June, 2026

THEORY AND PRACTICE IN ENGINEERING SCIENCES

EDITOR

PROF. ÖZLEM SALLI BIDECI, PH.D.

DOI: 10.5281/zenodo.20668538

Theory and Practice in Engineering Sciences

Edıtor

Prof. Özlem Sallı Bideci, Ph.D.

Publisher
Platanus Publishing®

Editor in Chief

Prof. Özlem Sallı Bideci, Ph.D.

Cover & Interior Design
Platanus Publishing®

The First Edition
June, 2026

ISBN
978-625-6454-65-9

Platanus Publishing®
Address: Natoyolu Cad. Fahri Korutürk Mah. 157/B, 06480, Mamak, Ankara, Turkey.

Phone: +90 312 390 1 118
web: www.platanuspublishing.com
e-mail: platanuskitap@gmail.com

Platanus Publishing®

CONTENTS

CHAPTER 1

Innovative Design of a Battery-Pack System Protector to Enhance Impact Safety

İsmail Öztürk[1]

1. Introduction

Electric vehicles are increasingly gaining popularity. However, battery pack systems pose a fire risk in the event of an accident. Therefore, it is crucial to design these structures with enhanced protection. Numerous studies have been conducted to improve the crash resistance of battery packs. Table 1 presents various studies focused on enhancing the collision performance of battery packs (Xiong, Pan, Wu & Liu, 2021; Pan, Xiong, Wu, Diao & Guo, 2021; Pan, Xiong, Dai et al., 2022; Mortazavi Moghaddam, Kheradpisheh & Asgari, 2023; Wang, Wang, Fan, Gao & Wang, 2023; Wang, Guo, Li, Hu & Leng, 2025; Kawsar et al., 2025; Li et al., 2024; Yang, Li, Cheng, Zou, & Tian, 2025).

No	Article title	Reference
1	Effective weight-reduction- and crashworthiness-analysis of a vehicle's battery-pack system via orthogonal experimental design and response surface methodology	Xiong, Pan, Wu & Liu, 2021
2	Lightweight design of an automotive battery-pack enclosure via advanced high-strength steels and size optimization	Pan, Xiong, Wu, Diao & Guo, 2021
3	Crush and crash analysis of an automotive battery-pack enclosure for lightweight design	Pan, Xiong, Dai et al., 2022
4	An integrated energy absorbing module for battery protection of electric vehicle under lateral pole impact	Mortazavi Moghaddam, Kheradpisheh & Asgari, 2023
5	Multi-objective lightweight design of automotive battery pack box for crashworthiness	Wang, Wang, Fan, Gao & Wang, 2023

[1] Doç. Dr., Bursa Teknik Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Mekatronik Mühendisliği Bölümü, ORCID: 0000-0003-2641-5880

6	On crashworthiness design of hybrid metal–composite battery-pack enclosure structure	Wang, Guo, Li, Hu & Leng, 2025
7	Enhancing crashworthiness performance of a battery pack system using multi-cell square tube structures	Kawsar et al., 2025
8	Bio-inspired honeycomb structures to improve the crashworthiness of a battery-pack system	Li et al., 2024
9	Enhanced Side Pole Impact Protection: Crashworthiness Optimization for Electric Micro Commercial Vehicles	Yang, Li, Cheng, Zou & Tian, 2025

Table 1. Battery pack crashworthiness studies.

A battery pack system (BPS) was optimized using an orthogonal experimental design and response surface methodology, resulting in an impressive 11.73% weight reduction (Xiong et al., 2021). The lightweight design of an automotive battery pack enclosure (BPE) involved size optimization, resulting in a 10.41% mass reduction (Pan, Xiong, et al., 2021). Various combinations of materials and thicknesses were explored further to reduce the mass of the BPE (Pan, Xiong, Dai et al., 2022). An energy-absorbing module was developed to safeguard the BPS during lateral pole impact (Mortazavi et al., 2023). The BPE was optimized through Genetic Algorithm (NSGA-II) and TOPSIS methods to enhance crashworthiness (Wang, Wang, Fan et al., 2023)

Additionally, a hybrid metal-composite BPE was optimized using modified TOPSIS (MTOPSIS) and improved multi-objective particle swarm optimization (IMOPSO) techniques (Wang, Guo, et al., 2025). Multi-cell square tube structures were employed to protect the BPS (Kawsar et al., 2025). While numerous studies have addressed the crashworthiness of BPS and BPE, research focusing on protection inspired by natural structures, known as bio-inspired structures, remains limited. Various honeycomb structures inspired by forewings, pomelo peel, spider webs, and grass stems were designed to safeguard the BPS. These designs were evaluated using frontal impact simulations, leading to the selection of the most effective option (Li et al., 2024). A cat-tail fluff and lotus-root inspired energy-absorbing structure is proposed for electric micro-commercial vehicles to reduce intrusions (Yang et al., 2025).

This research aims to enhance the safety of battery pack systems by implementing nature-inspired crash-protective designs, thereby contributing to the existing body of knowledge. Although many studies on bio-inspired battery pack protectors exist, this study compared honeycomb and plant-inspired protectors simultaneously, contributing to the literature. Various protectors were designed inspired by honeycomb, lotus root, and horsetail structures. Frontal impact simulations were performed using protectors made from AA2024-T351 and B1500HS-O25 materials to compare specific energy absorption (SEA) and peak crushing force (PCF) levels of the protectors. The AA2024-T351 horsetail protector exhibited the highest energy absorption to peak crushing force ratio.

2. Design and optimization of bio-inspired protectors

All the stages in this chapter are shown in Figure 1.

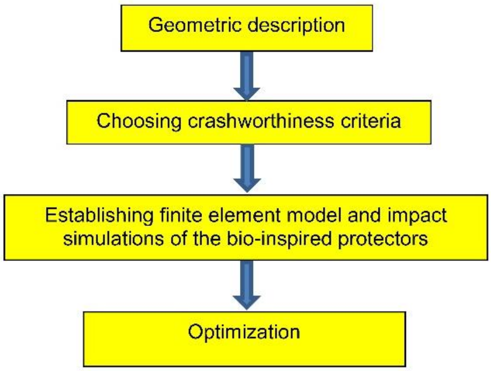

Figure 1. Flowchart of the design and optimization process

2.1. Geometric description

Researchers typically develop their designs by drawing inspiration from natural structures (Öztürk, 2023; Ma, 2024; Omeda, 2024). Protectors intended to safeguard the battery-pack system were modeled after the forms of a honeycomb, a lotus root, and a horsetail (see Figure 2). Measurements are the same for all protectors.

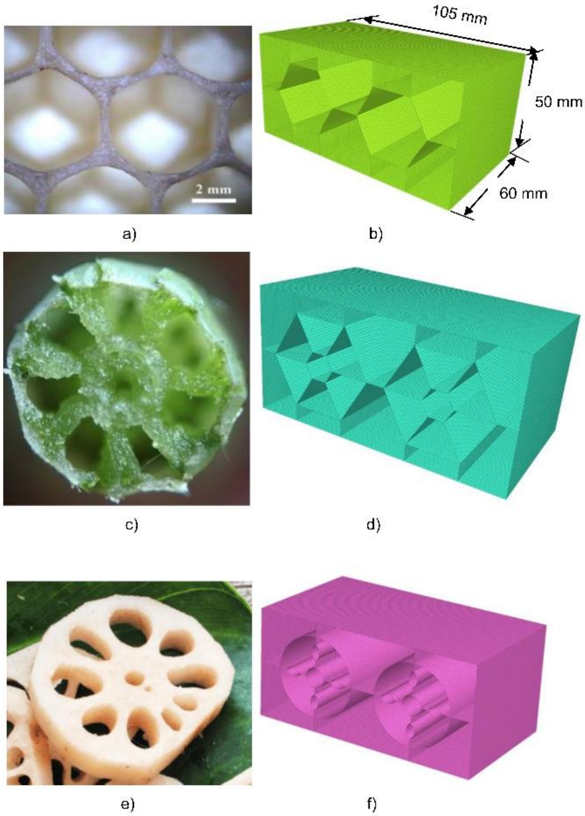

Figure 2. Bio-inspired protectors, a) structure of honeycomb (Zhang, Duan, Bhushan Lal Karihaloo & Wang, 2010), b) honeycomb protector finite element model, c) cross-sectional shape of horsetail, d) horsetail protector finite element model, e) cross-sectional shape of lotus root (Fang, 2024), f) lotus root protector finite element model

2.2. Crashworthiness criteria

There are several crashworthiness criteria to evaluate the impact performance of vehicle structures, including energy absorption (EA), specific energy absorption (SEA), and peak crushing force (PCF). The thin-walled protectors used in this study are compared with the SEA and PCF criteria.

EA (kJ) represents the total energy absorbed by the protector during plastic deformation, as defined in Equation (1):

$$\mathrm{EA} = \int_0^{x_{max}} f(x)dx \tag{1}$$

In this context, f(x) (kN) refers to the instantaneous impact force exerted between the impactor and the protector. x_{max} indicates the maximum displacement of the impactor.

The term SEA (kJ kg^{-1}) represents the total energy absorption of the protector, calculated by dividing the mass of the protector m (kg). This is expressed in Equation (2):

$$\mathrm{SEA} = \mathrm{EA/m} \tag{2}$$

PCF (kN) denotes the maximum impact force. Higher SEA and lower PCF values decrease the risk of passenger injury (Qi, Sun & Yang, 2018).

2.3. Establishing a finite element model and impact simulations of the bio-inspired protectors

The impact performance of these protectors was analyzed through crash simulations. The finite element (FE) model consists of three elements: a moving rigid cylinder, a thin-walled protector, and a battery pack system (see Figure 3). Finite element meshing was performed using Altair HyperMesh. The 3D model of the battery-pack system was meshed using solid tetrahedral elements with a 20 mm element size. The QBAT shell element formulation was chosen for the protector to mitigate hourglassing, with a 1 mm shell thickness and an overall element size. Rigid 1D elements linked the battery-pack system and protector. HyperCrash software served as the simulation preprocessor. A rigid cylinder with a diameter of 80 mm, a mass of 100 kg, and a velocity of 17.78 m/s (or 64 km/h) was impacted against the protector. Type 7 and Type 11 contact types were applied to prevent element intertwining, and a friction coefficient of 0.2 was selected. The rear nodes of the battery-pack system were constrained as fixed, while Radioss software was utilized for the solving process.

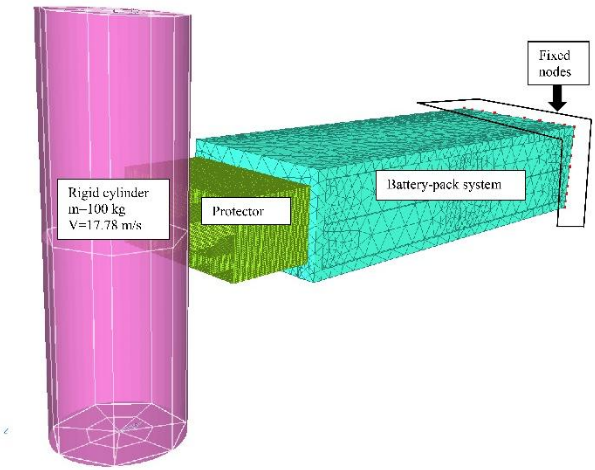

Figure 3. Finite element impact model

The battery-pack system was designed using B1500HS-O25 material. Due to their considerable strength, the AA2024-T351 aluminum alloy and B1500HS-O25 steel were chosen as protective materials to evaluate the effect of various materials. The parameters of the Johnson-Cook material model used in the impact model are outlined in Table 2.

Material type	AA2024-T351 (Larson, Ren, Adu-Gyamfi, Zhang & Ren, 2019; Teng, 2005)	B1500HS-O25 (Tang et al, 2020)
Density ρ (gr/cm^3)	2.70	7.85
Young's modulus E (MPa)	74660	210000
Poisson's ratio v	0.3	0.33
Yield stress a (MPa)	369	305
Strain hardening coefficient b (MPa)	684	610
Strain hardening exponent n	0.73	0.42

Table 2. AA2024-T351 and B1500HS-O25 Johnson-Cook material model parameters

Deformation shapes obtained from the simulations are shown in Figure 4. As expected from thin-walled energy absorbers, all protectors are folded stably and progressively, indicating good energy-absorbing properties (Gong, Bai, Wang & Zhang, 2021). Protector finite element simulation results are given in Table 3. In energy-absorbing structures, the SEA value is expected to be high, and the PCF value is expected to be low. Maximum SEA value obtained with AA2024-T351 honeycomb protector (84.33 kJ kg^{-1}). B1500HS-O25 horsetail protector gave the minimum PCF value (1288 kN). The maximum SEA/PCF value is 0.054 kJ kg^{-1} kN^{-1} with AA2024-T351 horsetail protector. So this protector was used for optimization studies.

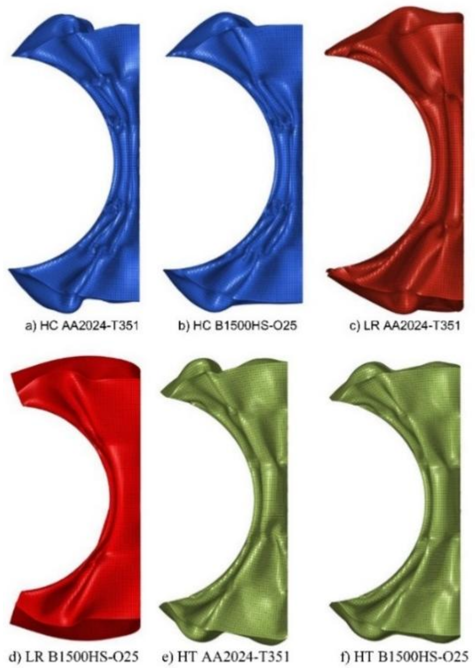

Figure 4. Deformation shapes of the protectors (HC, LR, and HT indicate honeycomb, lotus root, and horsetail, respectively)

Design number	Design name	SEA ($kJ\ kg^{-1}$)	PCF (kN)	SEA/PCF ($kJ\ kg^{-1}\ kN^{-1}$)
1	AA2024-T351 HC	***84.33***	2402	0.035
2	B1500HS-O25 HC	32.81	2283	0.014
3	AA2024-T351 LR	79.21	2067	0.038
4	B1500HS-O25 LR	26.79	2119	0.013
5	AA2024-T351 HT	82.13	1508	***0.054***
6	B1500HS-O25 HT	30.76	***1288***	0.024

Table 3. Impact simulation results (maximum SEA ($kJ\ kg^{-1}$), minimum PCF (kN), and maximum SEA to PCF ($kJ\ kg^{-1}\ kN^{-1}$) values are given in bold and italics)

2.4. Optimization

Optimization methods are used to accelerate the design process (El-Shorbagy & Alhadbani, 2024). The optimization problem is given in Equation (3):

$$\text{Objective: Max } \mathrm{EA}(t)$$
$$\mathrm{PCF}(t) \leq 1000\ \mathrm{kN}$$

Design parameters: $0.5 \text{ mm} \leq t \leq 1.50 \text{ mm}$ (3)

With t is the thickness of the protector. PCF constraint value taken as 1000 kN, considering DOE results. Equation (3) was solved with the design of experiments (DOE) and response surface methods. Eleven analyses were conducted, and the results are given in Table 4. The Moving Least Squares Method (MLSM) fit method gave the best results for constructing the EA design function with an 0.86 R^2 value, and the radial basis function (RBF) method gave the best results for constructing the PCF design function with a 0.99 R^2 value.

Experiment number	Thickness t (mm)	EA (kJ)-DOE	EA (kJ)-MLSM	Error (%)	PCF (kN)-DOE	PCF (kN)-RBF	Error (%)
1	0.50	5.40	6.08	-12.67	2561.27	2561.27	0.00
2	0.60	6.45	6.66	-3.24	2457.85	2457.85	0.00
3	0.70	7.84	7.92	-1.08	2369.99	2369.99	0.00
4	0.80	9.46	9.43	0.30	2262.04	2262.04	0.00
5	0.90	11.00	10.89	1.05	1938.35	1938.35	0.00
6	1.00	12.22	12.14	0.63	1507.75	1507.75	0.00
7	1.10	13.28	13.12	1.18	1172.47	1172.47	0.00
8	1.20	13.93	13.68	1.76	795.68	795.68	0.00
9	1.30	14.19	13.50	4.87	566.06	566.06	0.00
10	1.40	11.76	13.12	-11.58	510.96	510.96	0.00
11	1.50	14.31	13.39	6.47	488.74	488.74	0.00

Table 4. DOE and fit results of the bio-inspired protector

Equation (3) was solved with the Global Response Search Method (GRSM), Adaptive Response Surface Method (ARSM), Genetic Algorithm (GA), Method of Feasible Directions (MFD), and Sequential Quadratic Programming (SQP) methods (Gary Wang, Dong & Aitchison, 2001; Kakandikar & Nandedkar, 2016; Zhang, Liu & Tang, 2009; Shi, Yang & Zhu, 2012; Takeda, Ohtake & Suzuki, 2020) and the results are given in Table 5. When Table 5 is examined, the highest EA value is obtained with the GRSM, GA, and MFD methods. EA value improved 12.9% from 12.22 to 13.80 kJ with 1.25 mm thickness.

Optimization method	Design variable	Crashworthiness criteria	
	t (mm)	EA (kJ)	PCF (kN)
GRSM	1.25	***13.80***	655
ARSM	1.22	13.77	718
GA	1.25	***13.80***	655
MFD	1.25	***13.80***	656
SQP	1.5	13.38	489

Table 5. Single-objective optimization results (maximum EA values are shown in bold and italics)

3. Conclusions

Battery-pack systems pose a fire risk during accidents, making it crucial to design these structures with protective measures. Various protective designs were inspired by honeycomb, lotus root, and horsetail structures. Frontal impact simulations were performed using AA2024-T351 and B1500HS-O25 material protectors to evaluate SEA and PCF values of the protectors. The following results are obtained:

Maximum SEA value obtained with AA2024-T351 honeycomb protector (84.33 kJ kg^{-1}). B1500HS-O25 horsetail protector gave the minimum PCF value (1288 kN). The maximum SEA/PCF value is obtained with AA2024-T351 horsetail protector (0.054 kJ kg^{-1} kN^{-1}). Higher SEA and lower PCF values reduce passenger injury risk. So, the AA2024-T351 horsetail protector was used for optimization studies.

A single-objective optimization problem was solved using GRSM, ARSM, GA, MFD, and SQP methods. Best EA value obtained with GRSM, GA, and MFD methods. EA value improved 12.9% from 12.22 to 13.80 kJ with 1.25 mm thickness. So this protector can be used in electric vehicles to improve crashworthiness.

References

Xiong, Y., Pan, Y., Wu, L., & Liu, B. (2021). Effective weight-reduction- and crashworthiness-analysis of a vehicle's battery-pack system via orthogonal experimental design and response surface methodology. *Engineering Failure Analysis, 128*, 105635. https://doi.org/10.1016/j.engfailanal.2021.105635

Pan, Y., Xiong, Y., Wu, L., Diao, K., & Guo, W. (2021). Lightweight Design of an Automotive Battery-Pack Enclosure via Advanced High-Strength Steels and Size Optimization. *International Journal of Automotive Technology, 22* (5), 1279–1290. https://doi.org/10.1007/s12239-021-0112-5

Pan, Y., Xiong, Y., Dai, W., Diao, K., Wu, L., & Wang, J. (2022). Crush and crash analysis of an automotive battery-pack enclosure for lightweight design. *International Journal of Crashworthiness*, *27*(2), 500–509. https://doi.org/10.1080/13588265.2020.1812253

Mortazavi Moghaddam, A., Kheradpisheh, A., & Asgari, M. (2023). An integrated energy absorbing module for battery protection of electric vehicle under lateral pole impact. *International Journal of Crashworthiness*, *28*(3), 321–333. https://doi.org/10.1080/13588265.2022.2075123

Wang, H., Wang, Z., Fan, Y., Gao, Q., & Wang, H. (2023). Multi-objective lightweight design of automotive battery pack box for crashworthiness. *International Journal of Crashworthiness, 29*(2), 292–307. https://doi.org/10.1080/13588265.2023.2230646

Wang, S., Guo, S., Li, R., Hu, Z., & Leng, Q. (2025). On crashworthiness design of hybrid metal–composite battery-pack enclosure structure. *Mechanics of Advanced Materials and Structures*, *32*(11), 2548–2559. https://doi.org/10.1080/15376494.2024.2382356

Kawsar, I., Liu, C., Cheng, D., Fu, Y., Li, H., Pan, Y., & Altin, M. (2025). Enhancing crashworthiness performance of a battery pack system using multicell square tube structures. *European Journal of Mechanics - A/Solids, 112*, 105665. https://doi.org/10.1016/j.euromechsol.2025.105665

Li, R., Zhao, Z., Bao, H., Pan, Y., Wang, G., Liu, B., Liao, T., & Li, J. (2024). Bio-inspired honeycomb structures to improve the crashworthiness of a battery-pack system. *Engineering Failure Analysis, 158*, 108041. https://doi.org/10.1016/j.engfailanal.2024.108041

Yang, R., Li, S., Cheng, T., Zou, P., & Tian, L. (2025). Enhanced Side Pole Impact Protection: Crashworthiness Optimization for Electric Micro Commercial Vehicles. *Applied Sciences*, *15*(4), 2220. https://doi.org/10.3390/app15042220

İsmail Öztürk. (2023). Crushing performance and optimization of bio-inspired multi-cell columns. *Materials Testing/Materialprüfung, 65*(9), 1382–1388. https://doi.org/10.1515/mt-2023-0046

Ma, W., Almasifar, N., Amini, R., Armin Ourang, Ibrahim Mahariq, & Javid Alhoee. (2024). Crashworthiness evaluation and optimization of full polypropylene sandwich tubes under low-velocity impact based on machine learning algorithms. *Structures, 60*, 105901. https://doi.org/10.1016/j.istruc.2024.105901

Omede, B., & Antonio Mattia Grande. (2024). Parametric optimization of bio-inspired engineered sandwich core. *Materials Today Communications, 39*(1), 108632–108632. https://doi.org/10.1016/j.mtcomm.2024.108632

Zhang, K., Duan, H., Bhushan Lal Karihaloo, & Wang, J. (2010). Hierarchical, multilayered cell walls reinforced by recycled silk cocoons enhance the structural integrity of honeybee combs. *Proceedings of the National Academy of Sciences of the United States of America, 107*(21), 9502–9506. https://doi.org/10.1073/pnas.0912066107

Fang T. *Lotus Root: Everything You Need to Know!* Retrieved 1 September, 2024. From https://www.sakura.co/blog/lotus-root-everything-you-need-to-know

Qi, C., Sun, Y., & Yang, S. (2018). A comparative study on empty and foam-filled hybrid material double-hat beams under lateral impact. *Thin-Walled Structures, 129*, 327–341. https://doi.org/10.1016/j.tws.2018.04.018

Larson, E. A., Ren, X., Adu-Gyamfi, S., Zhang, H., & Ren, Y. (2019). Effects of scanning path gradient on the residual stress distribution and fatigue life of AA2024-T351 aluminium alloy induced by LSP. *Results in Physics, 13*, 102123. https://doi.org/10.1016/j.rinp.2019.02.059

Teng X. High velocity impact fracture [thesis]. Massachusetts Institute of Technology; 2005.

TWith the operational availability of these three systemang, B., Wu, F., Wang, Q., Li, C., Liu, J., & Ge, H. (2020). Numerical and experimental study on ductile fracture of quenchable boron steels with different microstructures. *International Journal of Lightweight Materials and Manufacture, 3*(1), 55–65. https://doi.org/10.1016/j.ijlmm.2019.07.001

Gong, C., Bai, Z., Wang, Y., & Zhang, L. (2021). On the crashworthiness performance of novel hierarchical multi-cell tubes under axial loading. *International Journal of Mechanical Sciences, 206*, 106599. https://doi.org/10.1016/j.ijmecsci.2021.106599

El-Shorbagy, M. A., & Alhadbani, T. H. (2024). Monarch butterfly optimization-based genetic algorithm operators for nonlinear constrained optimization and design of engineering problems. *Journal of Computational Design and Engineering, 11*(3), 200–222. https://doi.org/10.1093/jcde/qwae044

Gary Wang, G., Dong, Z., & Aitchison, P. (2001). Adaptive response surface method - a global optimization scheme for approximation-based design problems.

Engineering Optimization, 33(6), 707–733. https://doi.org/10.1080/03052150108940940

Kakandikar, G. M., & Nandedkar, V. M. (2016). Prediction and optimization of thinning in automotive sealing cover using Genetic Algorithm. *Journal of Computational Design and Engineering, 3*(1), 63–70. https://doi.org/10.1016/j.jcde.2015.08.001

Zhang, Z., Liu, S., & Tang, Z. (2009). Design optimization of cross-sectional configuration of rib-reinforced thin-walled beam. *Thin-Walled Structures, 47*(8-9), 868–878. https://doi.org/10.1016/j.tws.2009.02.009

Shi, L., Yang, R.-J., & Zhu, P. (2012). An adaptive response surface method for crashworthiness optimization. *Engineering Optimization, 45*(11), 1365–1377. https://doi.org/10.1080/0305215x.2012.734815

Takeda, H., Ohtake, Y., & Suzuki, H. (2020). 3D printing CFD simulation results using structural mechanics. *Journal of Computational Design and Engineering, 7*(3), 287–293. https://doi.org/10.1093/jcde/qwaa024

CHAPTER 2

The Bilateral Relationship between Disaggregate Energy Consumption and Sectoral Industrial Production in The United States

Mahmut Sayar[1]

1. Introduction

Energy sources play a crucial role in all industrial areas, especially in both, production and service sectors due to their relatively big energy requirements. Setting in motion all functions and processes, these sources are of equal importance as, for example, raw materials, labor, machines, capital, marketing and transportation. Presently, energy can be produced from either nonrenewable fossil sources like coal, natural gas and petroleum, or from renewable sources like the sun (solar energy), the wind, the movement of water (hydroelectric) or the heat of water coming from underground (geothermal). Despite being equally important to the economy, energy gained from biomass and nuclear electric power plants don't fit into the classification above. Depending on the sector and the processes used, the most adequate energy sources mix has to be chosen. Besides, businesses often face economic (due to the nature of the process process) and legal constraints (i.e. government policy, environmental impact) restricting the usage of energy sources.

A survey of the literature reveals which relationships between energy consumption and other macroeconomic variables have been investigated so far.

It is evident that there is a gap in previous researches as they do not consider the relationship between disaggregated industrial sectors and energy consumption by sources detailed enough. In our study these sectors are identified as follows:

[1] Assistant Professor, İzmir Bakırçay University, Department of Business Administration, İzmir, Türkiye, ORCID: 0000-0002-1852-6276

Table 2: Classification of Industrial Sectors

1-Agriculture, forestry, fishing, and hunting	9-Information
2-Mining	10-Finance, insurance, real estate, rental, and leasing
3-Utilities	11-Professional and business services
4-Construction	12-Educational services, health care, and social assistance
5-Manufacturing	13-Arts, entertainment, recreation, accommodation, and food services
6-Wholesale trade	14-Other services, except government
7-Retail trade	15-Government
8-Transportation and warehousing	

And energy sources are identified as in the following table:

Table 3: Classification of Energy Sources

1-Coal	6-Hydroelectric Power
2-Natural Gas (Excluding Supplemental Gaseous Fuels)	7-Geothermal Energy
3-Petroleum (Excluding Biofuels)	8-Solar/PV Energy
4-Total Fossil Fuels	9-Wind Energy
5-Nuclear Electric Power	10-Biomass Energy

The USA, as one of the most developed countries, has undergone a process of industrialization. Its huge industrial economic output requires an equally huge amount of input factors like capital, labor and energy. Sustainable growth of the US economy depends increasingly on its energy supply. Hence, the price of energy and the efficiency with which it is produced as well as consumed play a crucial role in the economic development of the United States as well as every developed and developing country. As a result, the estimation of the effects of energy consumption from different sources on the disaggregate industrial output is necessary for making effective decisions about strategic energy programs and energy investments.

In this study we investigate the bilateral relationship between the sectoral GDP and energy consumption by source using the autoregressive distributed lag (ARDL) approach of (Pesaran & Pesaran, 1997) and (Pesaran, Shin, & Smith, 2001).

This study is unique and novel as previous studies only examine the relationship between aggregate GDP and energy consumption. First, we present a review of related researches. In the second part, a description of the data and the general ARDL model is supplied. Finally, we discuss the results and offer some concluding remarks regarding the implications for energy policy.

2. Literature review

According to the literature research has focused on the relationship between energy consumption and output or income. The different results may be due to the differences in techniques used and the time period considered.

Energy consumption forecasting is an important issue in the study of energy economics. To investigate this issue, many researchers have adopted different methods (Wang & Hao, 2016). Ozturk (Ozturk, 2010) surveys the literature dealing with the causal relationship between energy consumption and economic growth and derives some policy implications. He concludes that electricity is a limiting factor to economic growth and, hence, shocks to energy supply will have a negative impact on economic growth. (Wolde-Rufael, 2004) investigates the causal relationship between industrial energy consumption and GDP in Shanghai for the period between 1952 and 1999 and finds that there is a unidirectional Granger causality running from coal, coke, electricity and total energy consumption to GDP, but no Granger causality running in any direction between oil consumption and GDP. These results indicate that reducing energy consumption may cause reduction in income. However, these results may suffer from neglecting other relevant variables. (Mahalingam & Orman, 2018) investigate the relationship between the USA energy consumption and GDP for the country. (Huang, Hwang, & Yang, 2008) use panel data of energy consumption and GDP for 82 countries from 1972 to 2002 based on the income levels defined by the World Bank. They conclude that for countries in the middle-income group, economic growth leads energy consumption positively, whereas economic growth leads energy consumption negatively in the high-income group. It is also indicated that, as a countries income rises, it tends to reduce its energy consumption to minimize the damage to the environment. (Ozturk & Acaravci, 2010) examine the causal relationship between economic growth, carbon emissions, energy consumption and employment ratio in Türkiye by using the autoregressive distributed lag bounds testing approach of cointegration. According to their empirical results neither carbon emission per capita nor energy consumption per capita causes real GDP per capita in Türkiye. (Begum, Sohag, Abdullah, & Jaafar, 2015) investigate the dynamic impacts of GDP growth, energy consumption and population growth on CO_2 emissions for Malaysia. The results imply that both per capita energy consumption and per capita GDP have long term positive impacts on per capita carbon emissions. On the other hand, the population growth rate doesn't seem to have any significant impact on per capita CO_2 emission. Fossil energy still possesses the biggest share of the total energy consumption in Malaysia, whereby the five main energy sources are oil, natural gas, coal, hydro and renewable energy. (Ziramba, 2009) investigates the relationship between disaggregate energy consumption and industrial output in South Africa by using a cointegration analysis from 1980 to 2005. According to their results, industrial production and employment are long-run forcing variables for electricity consumption. (Sari & Soytas, 2004) contribute to the literature on the relationship between energy consumption and economic growth by examining how much of the variance in national income growth can be explained by the growth of different sources of energy consumption and employment in Türkiye. They find that in Türkiye, energy consumption appears to be as important as employment. In addition, they find that different energy consumption items may have different effects on income in Türkiye. They also find a unidirectional causality running from total energy consumption to GDP, suggesting that energy shortages may harm economic growth in Türkiye. (Hu & Lin, 2008) model the relationship between GDP-energy consumption, GDP-coal consumption, GDP-natural gas consumption and GDP-electricity consumption. (Ozturk & Acaravci, 2016) use the ARDL approach to examine the long-run relationship between energy consumption, economic growth, employment ratio, foreign trade ratio and carbon emissions in Cyprus and Malta. Results indicate that energy conservation policies, such as rationing energy consumption and controlling carbon dioxide emissions, are likely to have no adverse effect on the real output growth of Malta. (Wong, Chang, & Chia, 2013) examine the contributions of energy consumption and energy R&D on economic growth. According to their results, the role of energy R&D should not be overlooked. R&D activities about fossil fuels are found to drive economic growth more than fossil fuel consumption. (Khandelwal, 2015) uses the ARDL method to explore the impact of energy consumption, fiscal deficit and GDP on public health expenditures in India. The presence of a long run relationship between the three independent variables and public health expenditures is obtained. In the short run, however, only GDP was found to be related to public health expenditures. (Jobert & Karanfil, 2007) investigate the causal relationships between income and energy consumption in Türkiye in two ways. They study the relationship at the aggregate and the disaggregate level regarding the industrial sectors. In the long run, income and energy consumption don't appear to cause each other at both the aggregate and the industrial level. In addition, a strong evidence of instantaneous causality is found, which means that contemporaneous values of energy consumption and income are correlated. (Bowden & Payne, 2009) study the causal relationship between energy consumption and real GDP using aggregate and sectoral primary energy consumption measures within a multivariate framework in U.S. The results show that industrial primary energy consumption Granger causes real GDP. (Sari, Ewing, & Soytas, 2008) use the ARDL approach to examine the relationship between disaggregate energy consumption and industrial output, as well as employment, in the United States. They consider the coal, fossil fuels, conventional hydroelectric power, solar energy, wind energy, natural gas, wood and waste energy consumption. The results suggest the presence of cointegrating relationships between the energy measures, employment and industrial output. In the long run output and labor are important affecters of fossil fuel, conventional hydroelectric power, solar, wind and waste energy consumption. (Alvarez &

Valencia, 2016) research the potential impact of the structure of electricity generation on manufacturing output using subsector and state-level manufacturing output. They conclude that electricity, oil and gas prices are more important to the manufacturing sector, where a reduction of those prices by one standard deviation would each cause a 2.8% increase of the manufacturing output. (Burke & Csereklyei, 2016) use per capita data for 132 countries over 1960-2010 to estimate elasticities of sectoral energy use with respect to GDP. Estimates show that transport, industry, and services sectors have larger energy-GDP elasticities. Agriculture typically accounts for a small share of energy use and has a reasonable energy-GDP elasticity. Studying Indian data, (Boutabba, 2014) finds that there is evidence in favor of the existence of a long-run relationship between carbon emissions, financial development, income, energy use and trade openness. (Bloch, Rafiq, & Salim, 2015) analyze the relationship between Chinese aggregate production and consumption of coal, oil and renewable energy using the autoregressive distributed lag and a vector error correction model (VECM). Results show that Chinese growth is led by all three energy sources. (Li, Song, & Liu, 2014) explain the factors causing the growth of energy consumption in China. According to the researchers, decline in exports triggered by a crises caused a reduction of energy consumption. Furthermore, the growth of domestic demand in manufacturing and construction has the opposite effect on energy consumption. (Keho, 2016) finds that energy consumption is cointegrated with real GDP per capita, industrial output, imports, foreign direct investment, credit to private sector, urbanization and population. Economic growth, industrial output and population have positive effects on energy consumption in the majority of 12 selected Saharan African countries. (Bildirici & Kayıkçı, 2013) investigate the relationship between oil production and economic growth in the major oil exporting countries Azerbaijan, Kazakhstan, Russian Federation and Turkmenistan for the period from 1993 to 2010. The results show that oil production and economic growth are cointegrated for all of these countries. Furthermore, there is a positive bidirectional causality between oil production and economic growth both in the long and in the short run. (Lin & Wang, 2015) conclude that GDP and urbanization have positive effects on the energy consumption of China's commercial sector using the Johansen co-integration methodology. (Rahman et al., 2015) confirm that the national economic growth of Malaysia depends on energy consumption at the aggregate and disaggregate level. At the disaggregate level, inefficient energy usage (particularly electricity and coal consumption) caused negative effects upon GDP and manufacturing growth. (Rahman et al., 2016) investigate the long-term relationship between industrial production, energy consumption and the environmental degradation in developed and developing countries. Result show that both G8 and Southeast Asian countries are energy dependent, as indicated by unidirectional Granger causality running from energy consumption to industrial production. (Ouyang & Lin, 2015) determine that economic growth is the driving force for the sectoral energy use, while factors such as technological progress, energy price and per capita productivity are conducive for reducing sectoral energy consumption. (Furuoka, 2015) employs a panel cointegration technique to test whether there is a long run relationship between finance and energy consumption, using data of many Asian countries from 1980 to 2012. (Chen, Chen, Hsu, & Chen, 2016) find that long-run relationships between economic growth, energy consumption and carbon dioxide emissions exist for all of the analyzed 188 countries using data of the period from 1993 to 2010.

3. Data

The investigation period of our paper starts in 1949 and ends in 2015. Hence, we have 66 annual observations. The data of the disaggregate gross domestic product is obtained from the Bureau of Economic Analysis of the U.S. Department of Commerce. We investigate the 15 industries "Agriculture, forestry, fishing, and hunting", "Mining", "Utilities", "Construction", "Manufacturing", "Wholesale trade", "Retail trade", "Transportation and warehousing", "Information", "Finance, insurance, real estate, rental, and leasing", "Professional and business services", "Educational services, health care, and social assistance", "Arts, entertainment, recreation, accommodation, and food services", "Other services, except government" and lastly "Government" denoted as Agriculture, Mining, Utilities, Construction, Manufacturing, Wholesale, Retail, Transport, Info, Finance, P&B Services Education, Arts, Others and Gov for the rest of the paper, respectively. Table 4 shows some descriptive statistics of the 15 industries based on real 2009 U.S.$. The row called Avg. Growth contains the average annual geometric growth rate for each industry. The greatest growth rate can be observed in the information industry. This fact reflects the importance of the information sector for developed countries. Other main drivers for the rapid development of the U.S. economy are service industries like Finance, P&B Services, Wholesale and Education. On the other hand, the mining industry grew the least over the past 66 years. The greatest industries are "Finance, insurance, real estate, rental, and leasing" and "Professional and business services" underlining the relevance of the shift towards the service sector to be a successful developed country like the U.S., once again.

Table 4: GDP by Industries. Descriptive Statistics in Millions of 2009 $

	Agriculture	Mining	Utilities	Construction	Manufacturing	Wholesale	Retail	Transport
Average	68,257	225,507	212,367	529,177	949,571	370,923	445,444	232,785
Std. Dev.	37,958	45,585	65,102	151,073	545,286	319,456	279,845	125,620
Min	26,060	137,927	62,260	200,269	255,160	42,828	125,057	58,971
Max	162,616	379,299	303,955	772,754	1,924,837	977,693	957,099	445,734

Avg.Growth	2.81%	0.95%	2.14%	1.72%	3.10%	4.85%	3.09%	3.07%

	Info	Finance	P&B Services	Education	Arts	Others	Gov
Average	281,484	1,383,466	781,486	615,998	313,761	228,495	1,421,004
Std. Dev.	245,230	942,161	601,125	379,695	166,104	103,349	471,190
Min	21,952	211,755	82,080	63,161	110,446	88,174	564,488
Max	878,213	3,171,419	2,037,818	1,346,756	617,313	403,479	2,079,715
Avg.Growth	5.75%	4.19%	4.99%	4.75%	2.56%	2.05%	1.92%

The data for the energy consumption by source is provided by the U.S. Energy Information Administration. The data is disaggregated into nine sources, namely Coal, Natural Gas, Petroleum, Nuclear Electric Power, Hydroelectric Power, Geothermal Energy, Solar/PV Energy, Wind Energy and Biomass Energy denoted as Coal, Gas, Petrol., Nucl., Hydro, Geo, Solar, Wind and Bio, respectively. Table 5 summarizes the energy consumption sources. In addition to the individual sources, Table 5 also contains the total fossil fuels consumption (column Fossil Fuels) and the total renewable energy consumption (column Renewable). It is interesting to note, that at the beginning of our investigation period in 1949, only two out of five renewable energy sources were used. Biomass and Hydroelectric energy started at approximately the same level, but as indicated by the growth rate, biomass became a more important energy source in terms of usage. Regarding absolute values in quadrillion Btu, nuclear electric power grew the most and is now also the biggest non-fossil-fuels energy source. The U.S. started using nuclear electric power, geothermal, solar and wind energy in 1957, 1960, 1984 and 1983, respectively. Considering the different points in time where the different sources were used for the first time, wind energy has the greatest value for the average geometric annual growth. Keeping in mind that recent energy policies shifted the energy consumption towards renewable energy sources, it is advisable to investigate a current picture of the situation. Therefore, we furthermore compute the average geometric annual growth of the energy consumptions for the last five years.

Table 5: Energy Consumption by Source. Descriptive Statistics in Quadrillion Btu.

	Coal	Gas	Petrol.	Fossil Fuels	Nucl.	Hydro	Geo	Solar	Wind	Bio	Renew-able
Avg.	15.93	18.75	30.11	64.80	3.82	2.51	0.09	0.05	0.17	2.44	5.25
Std. Dev.	4.36	5.60	7.89	16.83	3.39	0.63	0.08	0.10	0.42	1.01	1.86
Min	9.52	5.15	11.88	29.00	0.00	1.36	0.00	0.00	0.00	1.29	2.75
Max	22.80	28.32	40.30	85.93	8.46	3.64	0.22	0.55	1.82	4.81	9.67

Avg. Growth	0.4%	2.6 %	1.7%	1.5%	21.3 %	0.8%	12.4 %	34.6 %	41.4 %	1.7 %	1.8%
Avg. Growth Last 5 years	- 5.6%	2.9 %	- 0.1%	-0.4%	- 0.2%	- 1.2%	1.5%	34.3 %	14.5 %	1.9 %	3.7%

In this light, it gets apparent, that fossil fuels will slowly be replaced by renewable energy sources. During the last five years, the coal consumption decreased about 5.6% per year in average.

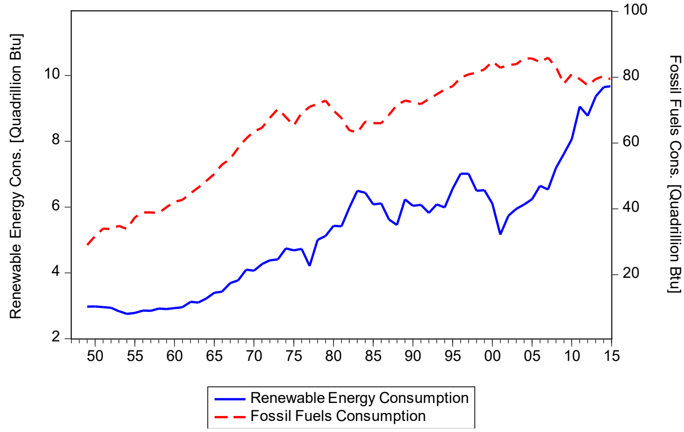

Figure 1:Historical Energy Consumption

But this decrease was partially compensated by an increase of the natural gas consumption, whereas the petroleum consumption only decreased by 0.1%. As a result, the overall fossil fuels consumption decreased by 0.4% annually in average, while the renewable energy consumption increased by 3.7% annually in average. As the youngest energy source, solar energy could maintain the initial high growth rates over the last years, but still has to prove that its growth is sustainable in the future.

Figure 1 illustrates the consumption of renewable energy compared to the fossil fuels consumption. Seemingly, fossil fuels have passed the phase of their biggest growth, while renewable energy source keep rapidly growing since the year 2000. However, as

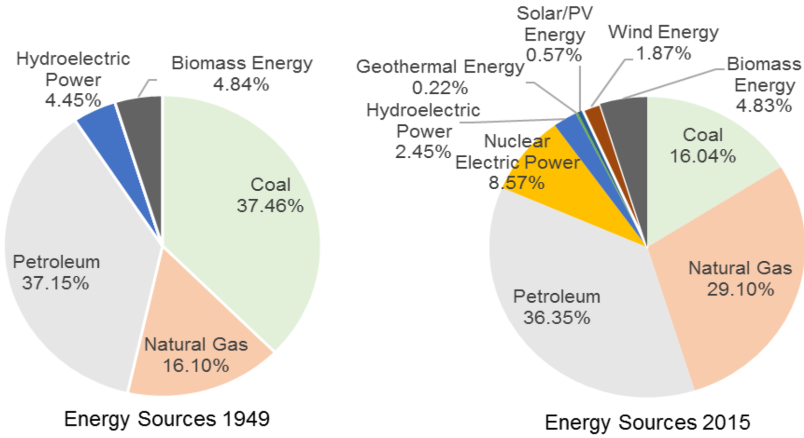

Figure 2 demonstrates clearly, renewable energy sources still have to grow to compete with fossil fuels and nuclear energy.

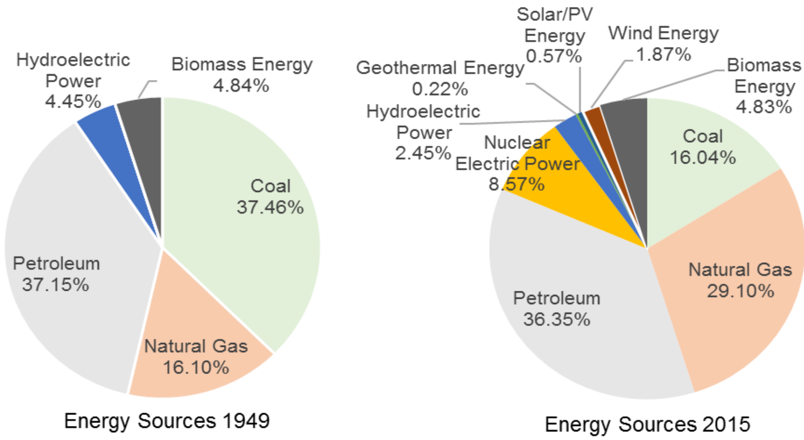

Figure 2 shows the share of the individual energy sources in the total energy consumption in percentage.

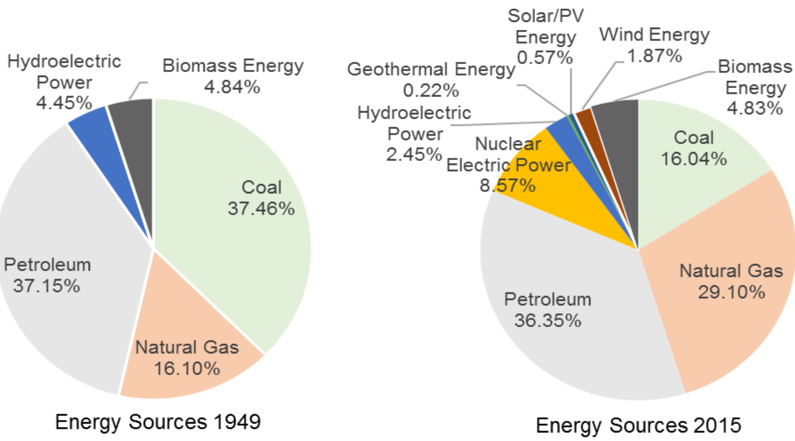

Figure 2: Energy Consumption Sources 1949 and 2015

The left pie chart depicts the used energy sources in the year 1949. It is once again worth noting, that only two out of five renewable energy sources were in use. The right pie chart portrays the situation in 2015. While in 1949 91% of the total consumed energy was supplied by fossil fuels, in 2015 fossil fuels still provide 81% of the consumed energy of the U.S. As already remarked above, the structure within the fossil fuels changed dramatically. The share of energy consumption from coal decreased by 22%, while the share of natural gas consumption increased by 13%. These two changes together with the 1% decrease in the share of petroleum consumption constitute the total decrease of 10% regarding the share of total fossil fuels consumption in the total energy consumption. The relevance of the hydroelectric energy decreased, since its share decreased by 2% to 2%. Energy consumption from biomass makes up 5% of the total energy consumption as it was the case 1949. The share of the new renewable energy sources combined is about 3%, with wind energy having the biggest share of 2%, followed by solar energy with 1% and geothermal energy, which has a share of slightly over 0.2% percent. Nuclear electric power constitutes the remaining 9% of the share in total energy consumption. Apparently, the energy from renewable sources was only able to keep up with the growth of the total energy consumption, increasing its share by about 1% only.

4. The Model

In order to investigate the existence of a long and short run relationship between each energy source and the production of each industry, we employ the bounds testing procedure as suggested by (Pesaran, Shin, & Smith, 2001). For this purpose, it is obligatory to model the data generating process as an autoregressive distributed lag (ARDL). Following (Sari, Ewing, & Soytas, 2008) we model the relationship as shown in eqs. (1) and (2).

$$\ln \text{INDPR}_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln \text{INDPR}_{t-i} + \sum_{i=1}^{q} \beta_i \ln \text{EMP}_{t-i} + \sum_{i=1}^{r} \gamma_i \ln \text{EN} \tag{1}$$

$$\ln \text{ENGS}_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \ln \text{ENGS}_{t-i} + \sum_{i=1}^{q} \beta_i \ln \text{EMP}_{t-i} + \sum_{i=1}^{r} \gamma_i \ln \text{INDP} \tag{2}$$

All variables enter the model in natural logarithms. INDPR denotes the industrial production of a certain industry. ENGS is the energy consumption of a certain source. We control for the employment denoted as EMP, given in thousands and obtained from the website of the U.S. Bureau of Labor Statistics. Due to the fact that we have annual data, it is unlikely that the autoregressive dynamics persist over more than two lags. Hence, we allow for a maximum of two lags, which has become a common practice when handling this type of data. Using ordinary least squares we estimate one model for each lag combination. Having allowed for two lags for each variable entering the model we estimate $2x3x3 = 18$ models for every relationship according to eqs. (1) and (2). Out of the 18 models, we select the best one based on the Schwarz information criterion (SIC) proposed by (Schwarz, 1978). Although there is an equally accepted information criterion proposed by (Akaike, 1974), we prefer the SIC, because in general the SIC selects more parsimonious models compared to the AIC and as pointed out by (Pesaran & Shin, 1998), the SIC has a better performance and is a consistent model-selection criterion.

Eqs. (3) and (4) are the cointegration representations of eqs. (1) and (2) obtained by calculating the differences and the lagged differences of the variables, but leaving the one period lag levels of the variables in the equation to model the long run relationship. For further details related to the computation of the coefficients, see (Pesaran & Shin, 1998) and (Pesaran, Shin, & Smith, 2001).

$$\Delta \ln INDPR_t = \alpha_0 + \sum_{i=1}^{p-1} \alpha'_i \Delta \ln INDPR_{t-i} + \sum_{i=1}^{q-1} \beta'_i \Delta \ln EMP_{t-i} + \sum_{i=1}^{r-1} \gamma'_i \Delta \ln ENGS_{t-i} + \theta_0 \ln INDPR_{t-1} + \theta_1 \ln EMP_{t-1} + \theta_2 \ln ENGS_{t-1} \quad (3)$$

$$\Delta \ln ENGS_t = \alpha_0 + \sum_{i=1}^{p-1} \alpha'_i \Delta \ln ENGS_{t-i} + \sum_{i=1}^{q-1} \beta'_i \Delta \ln EMP_{t-i} + \sum_{i=1}^{r-1} \gamma'_i \Delta \ln INDPR_{t-i} + \theta_0 \ln ENGS_{t-1} + \theta_1 \ln EMP_{t-1} + \theta_2 \ln INDPR_{t-1} \quad (4)$$

θ_0, θ_1, and θ_2 denote the long run coefficients, while α', β' and γ' are the coefficients describing the short run dynamics of the model. Lastly, the model is written in the error correction form by summarizing the long run relationship as $\theta_0\ ect_{t-1}$ and emphasizing the fact, that deviations from the long run equilibrium are corrected in the subsequent periods. The coefficient θ_0, usually located between -1 and 0, indicates how fast the correction will take place. A coefficient of -1 would mean, that the deviation is fully corrected in the next period. While greater a coefficient results in slower adjustment towards the long run equilibrium. A coefficient less than -1 would constitute a special case, in which the long-run deviation is overcorrected, leading to an overshooting. However, if the coefficient is greater than -2, the overshooting won't lead to an instable behavior of the model. An error impulse will rather be corrected like in any other case, except that the deviation's sign will oscillate and its amplitude will eventually fade away.

$$\Delta \ln INDPR_t = \alpha_0 + \sum_{i=1}^{p-1} \alpha'_i \Delta \ln INDPR_{t-i} + \sum_{i=1}^{q-1} \beta'_i \Delta \ln EMP_{t-i} + \sum_{i=1}^{r-1} \gamma'_i \Delta \ln ENGS_{t-i} + \theta_0\ ect_{t-1} \quad (5)$$

With the error correction term ect_{t-1} defiened as:

$$ect_{t-1} = \ln \mathrm{INDPR}_{t-1} + \theta_1' \ln \mathrm{EMP}_{t-1} + \theta_2' \ln \mathrm{ENGS}_{t-1} \tag{6}$$

$$\begin{aligned}\Delta \ln \mathrm{ENGS}_t = \alpha_0 + \sum_{i=1}^{p-1} \alpha'_i \Delta \ln \mathrm{ENGS}_{t-i} + \sum_{i=1}^{q-1} \beta'_i \Delta \ln \mathrm{EMP}_{t-i} \\ + \sum_{i=1}^{r-1} \gamma'_i \Delta \ln \mathrm{INDPR}_{t-i} + \theta_0\, ect_{t-1}\end{aligned} \tag{7}$$

$$\text{With } ect_{t-1} = \ln \mathrm{ENGS}_{t-1} + \theta_1' \ln \mathrm{EMP}_{t-1} + \theta_2' \ln \mathrm{INDPR}_{t-1} \tag{8}$$

The long-run coefficients of equations (6) and (8) are obtained by $\theta_1' = -\theta_1/\theta_0$ and $\theta_2' = -\theta_2/\theta_0$, respectively. The estimates of the standard errors are computed using the delta method.

After the best model is identified and set up, the existence of a long-run relationship is investigated in two ways. First, we test the hypothesis that the long-run coefficients of eqs. (3) and (4) are jointly zero ($\theta_0 = \theta_1 = \theta_2$) using a Wald-test. The critical values for the non-standard F-statistics are supplied by (Pesaran, Shin, & Smith, 2001). The results of the bounds test are presented in Table 8. However, (Bahmani-Oskooee & Nasir, 2004) point out that the F-statistic is very sensitive to the selected lag length. Hence, we additionally consider the the t-ratio of the error correction term of eqs. (5) and (7). The critical values for non-standard t-statistics are also supplied by (Pesaran, Shin, & Smith, 2001).. The t-statistics alongside with the other coefficients can be found in Tables *Table 10* -*Table 16*.

5. Results

First, we test the time series for unit roots, to make sure, that no series is integrated at an order greater than one. In such a case the ARDL method would render inapplicable.

Table 5 contains the result of the two tests of the variables at level.

Table 6: Unit Root Tests of the Natural Logarithms

	Intercept				Intercept and Trend				No Intercept, no Trend			
	ADF		PP		ADF		PP		ADF		PP	
	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.
Agri	1.51	1.00	1.09	1.00	-4.10	0.01	-3.84	0.02	6.08	1.00	10.67	1.00
Arts	-0.29	0.92	-0.25	0.93	-2.23	0.46	-2.25	0.45	5.68	1.00	6.24	1.00
Bio	0.97	1.00	0.38	0.98	-1.96	0.61	-2.25	0.45	2.97	1.00	2.07	0.99
Coal	-1.18	0.68	-1.01	0.75	-0.15	0.99	-0.40	0.99	0.51	0.82	0.51	0.82
Constr.	-2.60	0.10	-3.52	0.01	-2.15	0.51	-2.14	0.52	1.33	0.95	1.74	0.98
Edu	-9.63	0.00	-7.90	0.00	-3.25	0.08	-2.81	0.20	0.18	0.74	3.48	1.00
Finance	-4.81	0.00	-4.73	0.00	0.36	1.00	0.35	1.00	0.36	0.79	5.99	1.00
Geo	0.37	0.98	0.32	0.98	-2.02	0.58	-2.07	0.55	2.49	1.00	2.23	0.99
Gov	-3.59	0.01	-4.16	0.00	-2.60	0.28	-0.97	0.94	2.60	1.00	3.83	1.00
Hydro	-2.09	0.25	-1.96	0.30	-1.80	0.69	-1.53	0.81	0.25	0.76	0.49	0.82
Info	-2.34	0.16	-2.40	0.15	-1.50	0.82	-1.51	0.82	1.66	0.98	5.69	1.00
Manu	-1.32	0.62	-1.47	0.54	-2.28	0.44	-2.38	0.38	4.68	1.00	4.79	1.00
Mining	-0.27	0.92	-0.41	0.90	-0.74	0.97	-0.85	0.96	1.04	0.92	1.00	0.91
Gas	-3.30	0.02	-4.35	0.00	-2.98	0.15	-3.50	0.05	1.83	0.98	2.11	0.99
Nucl.	-0.92	0.78	-0.73	0.83	-1.13	0.92	-1.20	0.90	0.94	0.91	1.27	0.95
Other	-2.00	0.29	-1.77	0.39	0.47	1.00	0.18	1.00	2.20	0.99	3.35	1.00
petr.	-2.55	0.11	-4.17	0.00	-1.65	0.76	-2.13	0.52	1.30	0.95	1.98	0.99
P&B	-3.33	0.02	-4.14	0.00	-0.50	0.98	-0.23	0.99	3.51	1.00	5.68	1.00
Retail	-0.84	0.80	-0.50	0.88	-1.75	0.72	-1.86	0.67	4.26	1.00	5.58	1.00
Solar	3.41	1.00	6.36	1.00	2.90	1.00	5.56	1.00	3.52	1.00	6.73	1.00
Transport	-2.04	0.27	-2.16	0.22	-2.96	0.15	-2.21	0.48	4.86	1.00	4.86	1.00
Utilities	-5.02	0.00	-5.93	0.00	-2.78	0.21	-2.96	0.15	2.60	1.00	2.16	0.99
Wholesale	-1.00	0.75	-0.91	0.78	-0.31	0.99	-0.69	0.97	4.10	1.00	5.65	1.00
Wind	-4.14	0.00	4.62	1.00	-4.49	0.00	2.62	1.00	-4.08	0.00	5.40	1.00
Empl.	-0.88	0.79	-1.08	0.72	-1.00	0.94	-0.60	0.98	4.23	1.00	7.83	1.00

Note: The optimum lag length for the test was found using Schwarz Information Criterion (SIC).

The null hypothesis, that the time series has a unit root, cannot be rejected clearly for any of the variables. As a second step, it is necessary to make sure, that the variables are stationary at the first difference, i.e. integrated of order one. The results of the unit roots test of the differenced variables are shown inTable 7.

Table 7: Unit Root Tests of First Difference of the Natural Logarithms

	Intercept				Intercept and Trend				No Intercept, no Trend			
	ADF		PP		ADF		PP		ADF		PP	
	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.	t-stat.	prob.
Agri	-7.6	0.0	-16.8	0.0	-7.9	0.0	-24.7	0.0	-0.9	0.3	-11.1	0.0
Arts	-7.3	0.0	-7.4	0.0	-7.3	0.0	-7.4	0.0	-5.3	0.0	-5.3	0.0
Bio	-6.4	0.0	-6.6	0.0	-6.6	0.0	-6.8	0.0	-2.5	0.0	-6.3	0.0
Coal	-3.7	0.0	-8.0	0.0	-3.8	0.0	-8.0	0.0	-3.6	0.0	-8.1	0.0
Constr.	-5.3	0.0	-5.3	0.0	-5.6	0.0	-5.5	0.0	-5.0	0.0	-5.0	0.0
Edu	-1.2	0.7	-5.1	0.0	-3.5	0.0	-8.4	0.0	-1.4	0.2	-2.4	0.0
Finance	-5.7	0.0	-5.7	0.0	-7.7	0.0	-7.7	0.0	-1.2	0.2	-2.3	0.0
Geo	-7.2	0.0	-7.2	0.0	-7.2	0.0	-7.2	0.0	-6.4	0.0	-6.4	0.0
Gov	-4.9	0.0	-4.9	0.0	-5.9	0.0	-5.8	0.0	-3.7	0.0	-3.7	0.0
Hydro	-9.2	0.0	-9.7	0.0	-9.3	0.0	-10.8	0.0	-9.2	0.0	-9.5	0.0
Info	-8.6	0.0	-8.7	0.0	-9.5	0.0	-9.4	0.0	-1.1	0.3	-3.1	0.0
Manu	-8.2	0.0	-8.2	0.0	-8.2	0.0	-8.3	0.0	-6.3	0.0	-6.4	0.0
Mining	-7.3	0.0	-7.3	0.0	-7.4	0.0	-7.4	0.0	-7.2	0.0	-7.2	0.0
Gas	-5.1	0.0	-5.1	0.0	-5.5	0.0	-5.5	0.0	-2.8	0.0	-4.4	0.0
Nucl.	-4.0	0.0	-4.2	0.0	-4.1	0.0	-4.2	0.0	-3.4	0.0	-3.4	0.0
Other	-4.0	0.0	-6.9	0.0	-7.3	0.0	-7.3	0.0	-3.0	0.0	-5.6	0.0
petr.	-4.7	0.0	-4.6	0.0	-5.1	0.0	-5.0	0.0	-4.4	0.0	-4.3	0.0
P&B	-6.0	0.0	-6.0	0.0	-6.9	0.0	-7.4	0.0	-2.8	0.0	-2.6	0.0
Retail	-6.9	0.0	-6.9	0.0	-6.9	0.0	-6.9	0.0	-4.5	0.0	-4.5	0.0
Solar	1.9	1.0	0.9	1.0	1.0	1.0	0.0	1.0	2.3	1.0	1.3	1.0
Transport	-6.6	0.0	-6.5	0.0	-6.8	0.0	-6.7	0.0	-5.0	0.0	-5.0	0.0
Utilities	-7.7	0.0	-7.7	0.0	-9.1	0.0	-9.3	0.0	-7.1	0.0	-7.1	0.0
Wholesale	-6.9	0.0	-6.9	0.0	-6.9	0.0	-6.9	0.0	-2.4	0.0	-3.8	0.0
Wind	-3.3	0.0	-2.0	0.3	-3.7	0.0	-2.8	0.2	-3.1	0.0	-1.7	0.1
Empl.	-6.2	0.0	-6.2	0.0	-6.3	0.0	-6.1	0.0	-4.0	0.0	-4.0	0.0

Note: The optimum lag length for the test was found using Schwarz Information Criterion (SIC).

It is evident from Table 7, that the time series of solar energy consumption is not stationary at first difference, that means integrated at an order greater than one. Hence, we have to exclude the solar consumption variable from our analysis.

For each combination of economic sectors and energy consumption sources we estimate two models to analyze the bidirectional relationship according to eqs. (5) and (7). With 15 sectors and 8 energy sources we estimate a total of 15x8x2=240 models. The results of the Bounds Test to investigate the existence of a long run relationship are shown in Table 8 and Table 9. The tables are divided into seven and eight panels, respectively, whereas there is one panel for each sector. Out of 240 long-run relationships, 40, 18 and 19 are significant on the 1%, 5% and 10% level, respectively.

The industries Government, P&B Services and Education are the most sensitive industries to energy consumption changes, significantly affected by all eight energy sources in the long run. The finance and the utilities industries turn out to be very sensitve as well with seven significant long-run relationships each.

Table 8: Bounds Test Results Part 1

Cointegration hypotheses	F-stat.	Cointegration hypotheses	F-stat.	Cointegration hypotheses	F-stat.	Cointegration hypotheses	F-stat.
Panel A. Agriculture							
F(•| Emp, Hydro)	3.34	F(Hydro| Emp, •)	3.25	F(•| Emp, Geo)	1.76	F(Geo| Emp, •)	3.02
F(•| Emp, Gas)	2.78	F(Gas| Emp, •)	4.28[a]	F(•| Emp, Nucl. E)	2.6	F(Nucl. E| Emp, •)	4.11
F(•| Emp, Petr.)	4.25[a]	F(Petr.| Emp, •)	2.32	F(•| Emp, Wind)	5.19[b]	F(Wind| Emp, •)	3.19
F(•| Emp, Bio)	3.62	F(Bio| Emp, •)	2.38	F(•| Emp, Coal)	2.71	F(Coal| Emp, •)	3.61
Panel B. Arts							
F(•| Emp, Hydro)	2.86	F(Hydro| Emp, •)	2.89	F(•| Emp, Geo)	3.27	F(Geo| Emp, •)	3.15
F(•| Emp, Gas)	2.39	F(Gas| Emp, •)	4.19[a]	F(•| Emp, Nucl. E)	2.03	F(Nucl. E| Emp, •)	6.96[c]
F(•| Emp, Petr.)	4.36[a]	F(Petr.| Emp, •)	2.54	F(•| Emp, Wind)	2.36	F(Wind| Emp, •)	1.57
F(•| Emp, Bio)	2.06	F(Bio| Emp, •)	2.72	F(•| Emp, Coal)	1.84	F(Coal| Emp, •)	1.96
Panel C. Construction							
F(•| Emp, Hydro)	3.96	F(Hydro| Emp, •)	2.67	F(•| Emp, Geo)	4.06	F(Geo| Emp, •)	3.91
F(•| Emp, Gas)	3.82	F(Gas| Emp, •)	3.67	F(•| Emp, Nucl. E)	3.8	F(Nucl. E| Emp, •)	1.88
F(•| Emp, Petr.)	7.73[c]	F(Petr.| Emp, •)	4.19[a]	F(•| Emp, Wind)	4.29[a]	F(Wind| Emp, •)	1.02
F(•| Emp, Bio)	3.46	F(Bio| Emp, •)	2.62	F(•| Emp, Coal)	3.72	F(Coal| Emp, •)	2.02
Panel D. Education							
F(•| Emp, Hydro)	26.76[c]	F(Hydro| Emp, •)	5.37[b]	F(•| Emp, Geo)	42.4[c]	F(Geo| Emp, •)	3.44
F(•| Emp, Gas)	44.17[c]	F(Gas| Emp, •)	2.26	F(•| Emp, Nucl. E)	52.14[c]	F(Nucl. E| Emp, •)	2.23
F(•| Emp, Petr.)	42.41[c]	F(Petr.| Emp, •)	3.11	F(•| Emp, Wind)	40.92[c]	F(Wind| Emp, •)	2.54
F(•| Emp, Bio)	43.95[c]	F(Bio| Emp, •)	2.76	F(•| Emp, Coal)	46.64[c]	F(Coal| Emp, •)	0.85
Panel E. Finance							
F(•| Emp, Hydro)	11.37[c]	F(Hydro| Emp, •)	2.07	F(•| Emp, Geo)	11.13[c]	F(Geo| Emp, •)	3.37

F(•| Emp, Gas)	11.51[c]	F(Gas| Emp, •)	3.54	F(•| Emp, Nucl. E)	9.21[c]	F(Nucl. E| Emp, •)	1.77
F(•| Emp, Petr.)	11.94[c]	F(Petr.| Emp, •)	2.57	F(•| Emp, Wind)	3.44	F(Wind| Emp, •)	0.82
F(•| Emp, Bio)	10.02[c]	F(Bio| Emp, •)	2.61	F(•| Emp, Coal)	8.8c	F(Coal| Emp, •)	0.19
Panel F. Government							
F(•| Emp, Hydro)	7.55[c]	F(Hydro| Emp, •)	3.56	F(•| Emp, Geo)	7.66[c]	F(Geo| Emp, •)	3.43
F(•| Emp, Gas)	7.49[c]	F(Gas| Emp, •)	4.39[a]	F(•| Emp, Nucl. E)	13.34[c]	F(Nucl. E| Emp, •)	2.14
F(•| Emp, Petr.)	9.6[c]	F(Petr.| Emp, •)	4.54[a]	F(•| Emp, Wind)	7.31[c]	F(Wind| Emp, •)	1.19
F(•| Emp, Bio)	7.37[c]	F(Bio| Emp, •)	2.26	F(•| Emp, Coal)	13.34[c]	F(Coal| Emp, •)	0.14
Panel G. Information							
F(•| Emp, Hydro)	3.37	F(Hydro| Emp, •)	1.35	F(•| Emp, Geo)	2.32	F(Geo| Emp, •)	3.36
F(•| Emp, Gas)	2.15	F(Gas| Emp, •)	4.19[a]	F(•| Emp, Nucl. E)	2.22	F(Nucl. E| Emp, •)	2.64
F(•| Emp, Petr.)	1.48	F(Petr.| Emp, •)	2.45	F(•| Emp, Wind)	3.12	F(Wind| Emp, •)	1.74
F(•| Emp, Bio)	5.55[b]	F(Bio| Emp, •)	6.07[b]	F(•| Emp, Coal)	3.42	F(Coal| Emp, •)	1.91
Panel H. Manufacturing							
F(•| Emp, Hydro)	1.61	F(Hydro| Emp, •)	1.54	F(•| Emp, Geo)	1.75	F(Geo| Emp, •)	3.67
F(•| Emp, Gas)	1.63	F(Gas| Emp, •)	5.02[b]	F(•| Emp, Nucl. E)	3.75	F(Nucl. E| Emp, •)	3.02
F(•| Emp, Petr.)	2.49	F(Petr.| Emp, •)	4.83[a]	F(•| Emp, Wind)	1.86	F(Wind| Emp, •)	2.5
F(•| Emp, Bio)	1.25	F(Bio| Emp, •)	2.25	F(•| Emp, Coal)	2.27	F(Coal| Emp, •)	1.54

Note: [a], [b], [c] denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed F-statistics obtained from Pesaran, Shin, and Smith (2001).

The wholesale industry has five significant relationships. Agriculture, Construction and Mining, influenced by two energy sources and Arts, Information, Transportation and Other significantly influenced by only one source appear to be relatively insensitive to energy consumption. The results suggest that manufacturing and retail are not affected by energy consumption at all. On the other hand, Utilities and the industries called Other determine the energy consumption of three sources. Seven of the remaining industries (Government, P&B Services, Arts, Information, Transportation, Manufacturing and Retail) have an influence on the consumption of two energy sources. Four sectors (Education, Agriculture, Construction and Mining) are only linked to one energy source significantly. Finance and Wholesale appear to have no influence on the energy consumption at all.

The most influential energy sources are Petroleum and Biomass, determining the output of ten and eight industries significantly each, respectively. Geothermal Energy plays an important role for seven out of 15 industries. Hydroelectric and Nuclear and Wind Energy affecting six industries each, are followed by the remaining energy source Coal and Natural Gas which contribute to the output of five industries each.

Table 9: Bounds Test Results Part 2

Cointegration hypotheses	F-stat.	Cointegration hypotheses	F-stat.	Cointegration hypotheses	F-stat.	Cointegration hypotheses	F-stat.
Panel A. Mining							
F(•| Emp, Hydro)	3.21	F(Hydro| Emp, •)	2.33	F(•| Emp, Geo)	4.57^{a}	F(Geo| Emp, •)	3.18
F(•| Emp, Gas)	2.79	F(Gas| Emp, •)	23.19^{c}	F(•| Emp, Nucl. E)	0.56	F(Nucl. E| Emp, •)	3.91
F(•| Emp, Petr.)	6.82^{c}	F(Petr.| Emp, •)	3.5	F(•| Emp, Wind)	2.95	F(Wind| Emp, •)	1.21
F(•| Emp, Bio)	3.87	F(Bio| Emp, •)	2.66	F(•| Emp, Coal)	0.59	F(Coal| Emp, •)	2.31
Panel B. Other							
F(•| Emp, Hydro)	1.38	F(Hydro| Emp, •)	1.67	F(•| Emp, Geo)	1.41	F(Geo| Emp, •)	6.33^{b}
F(•| Emp, Gas)	1.43	F(Gas| Emp, •)	3.5	F(•| Emp, Nucl. E)	2.89	F(Nucl. E| Emp, •)	1.77
F(•| Emp, Petr.)	2.33	F(Petr.| Emp, •)	4.22^{a}	F(•| Emp, Wind)	0.91	F(Wind| Emp, •)	1.44
F(•| Emp, Bio)	5.94^{b}	F(Bio| Emp, •)	15.31^{c}	F(•| Emp, Coal)	1.37	F(Coal| Emp, •)	2.12
Panel C. P&B Services							
F(•| Emp, Hydro)	8^{c}	F(Hydro| Emp, •)	1.96	F(•| Emp, Geo)	9.28^{c}	F(Geo| Emp, •)	5.19^{b}
F(•| Emp, Gas)	9^{c}	F(Gas| Emp, •)	11.61^{c}	F(•| Emp, Nucl. E)	6.93^{c}	F(Nucl. E| Emp, •)	2.13
F(•| Emp, Petr.)	10.05^{c}	F(Petr.| Emp, •)	3.27	F(•| Emp, Wind)	7.54^{c}	F(Wind| Emp, •)	1.05
F(•| Emp, Bio)	9.13^{c}	F(Bio| Emp, •)	1.16	F(•| Emp, Coal)	6.78^{c}	F(Coal| Emp, •)	0.35
Panel D. Retail							
F(•| Emp, Hydro)	1.35	F(Hydro| Emp, •)	2.18	F(•| Emp, Geo)	3.11	F(Geo| Emp, •)	3.89
F(•| Emp, Gas)	1.38	F(Gas| Emp, •)	4.28^{a}	F(•| Emp, Nucl. E)	1.37	F(Nucl. E| Emp, •)	4.03
F(•| Emp, Petr.)	3.05	F(Petr.| Emp, •)	5.09^{b}	F(•| Emp, Wind)	1.54	F(Wind| Emp, •)	2.84
F(•| Emp, Bio)	2.57	F(Bio| Emp, •)	3.67	F(•| Emp, Coal)	1.84	F(Coal| Emp, •)	1.36
Panel E. Transportation							
F(•| Emp, Hydro)	2.06	F(Hydro| Emp, •)	1.53	F(•| Emp, Geo)	4.59^{a}	F(Geo| Emp, •)	5.7^{b}
F(•| Emp, Gas)	3.44	F(Gas| Emp, •)	5.88^{b}	F(•| Emp, Nucl. E)	3.62	F(Nucl. E| Emp, •)	2.83
F(•| Emp, Petr.)	3.37	F(Petr.| Emp, •)	2.54	F(•| Emp, Wind)	3.46	F(Wind| Emp, •)	1.32
F(•| Emp, Bio)	1.75	F(Bio| Emp, •)	1.5	F(•| Emp, Coal)	3.72	F(Coal| Emp, •)	1.22
Panel F. Utilities							
F(•| Emp, Hydro)	6.32^{b}	F(Hydro| Emp, •)	4.35^{a}	F(•| Emp, Geo)	6.38^{c}	F(Geo| Emp, •)	4.81^{a}
F(•| Emp, Gas)	2.04	F(Gas| Emp, •)	1.84	F(•| Emp, Nucl. E)	6.14^{b}	F(Nucl. E| Emp, •)	3.75
F(•| Emp, Petr.)	5.98^{b}	F(Petr.| Emp, •)	6.57^{c}	F(•| Emp, Wind)	7.24^{c}	F(Wind| Emp, •)	2.55
F(•| Emp, Bio)	6.04^{b}	F(Bio| Emp, •)	2.91	F(•| Emp, Coal)	5.96^{b}	F(Coal| Emp, •)	0.86

Panel G. Wholesale

F(•| Emp, Hydro)	5.2b	F(Hydro| Emp, •)	2.38	F(•| Emp, Geo)	3.77	F(Geo| Emp, •)	3.87
F(•| Emp, Gas)	4.72[a]	F(Gas| Emp, •)	2.89	F(•| Emp, Nucl. E)	4.54[a]	F(Nucl. E| Emp, •)	2.46
F(•| Emp, Petr.)	5.57[b]	F(Petr.| Emp, •)	2.64	F(•| Emp, Wind)	1.79	F(Wind| Emp, •)	2.37
F(•| Emp, Bio)	4.78[a]	F(Bio| Emp, •)	2.76	F(•| Emp, Coal)	3.55	F(Coal| Emp, •)	1.24

Note: [a], [b], [c] denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed F-statistics obtained from Pesaran, Shin, and Smith (2001).

To substantiate our findings, we continue our research by analyzing the cointegrating form of the relationships found through the bounds tests. Therefore, we have to examine the 77 cointegrating equations of the form given in eqs. (3) and (4). Table 10 through Table 16 show the short-run and long-run coefficients of the 77 relationships. 73 out of 77 models have significant short-run relationships. That means, that in most cases, models with significant long-run relationships also have significant short-run relationships. Figures 3 and 4 summarize the long-run relationships. Figure 3 displays only the significant effects of the sectoral GDP's on energy consumption. The arts industry and Other Services being the only exceptions, increasing sectoral GDP's cause at least one energy consumption source to increase as well. Government and Utilities are the only two sectors which cause two different energy source. The natural gas consumption is affected by the greatest number of industries.

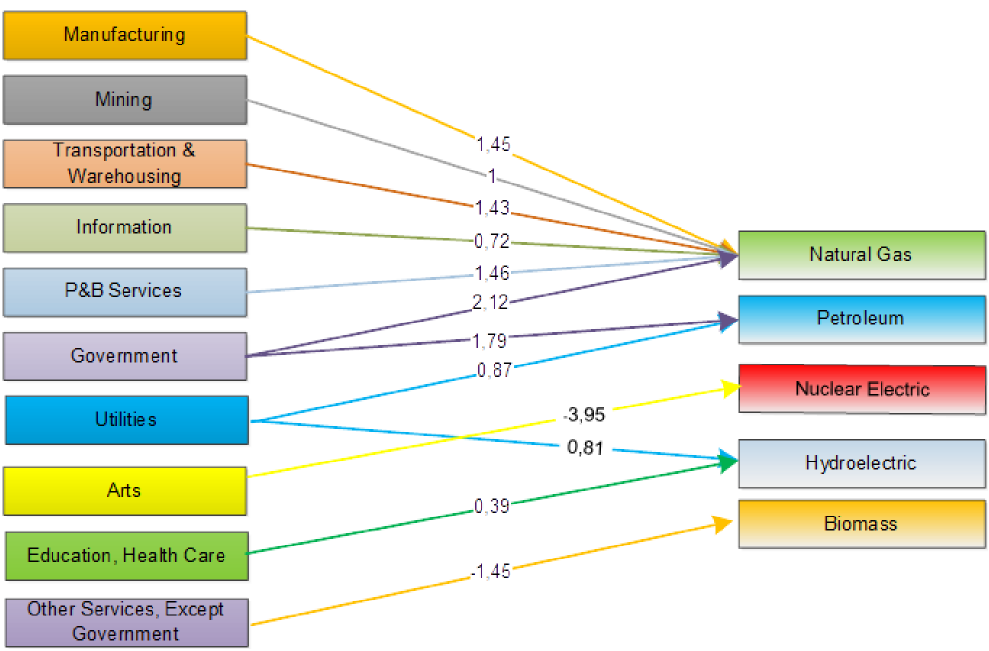

Figure 3: Effects of Sectoral GDP on Energy Consumption

Manufacturing, mining, transportation, information, p&b services and government play a significant role in the natural gas consumption. The most important determinants of the petroleum consumption appear to be the government and the utilities sector. It is interesting to note that nuclear electric power consumption is negatively related to the arts sector. It may be, that a growing arts sector, accompanied by an environmental way of thinking, results in higher awareness towards the risks and downsides of nuclear energy, leading to lower nuclear electric power demand. The main drivers of hydroelectric energy consumption are the utilities and education industry.

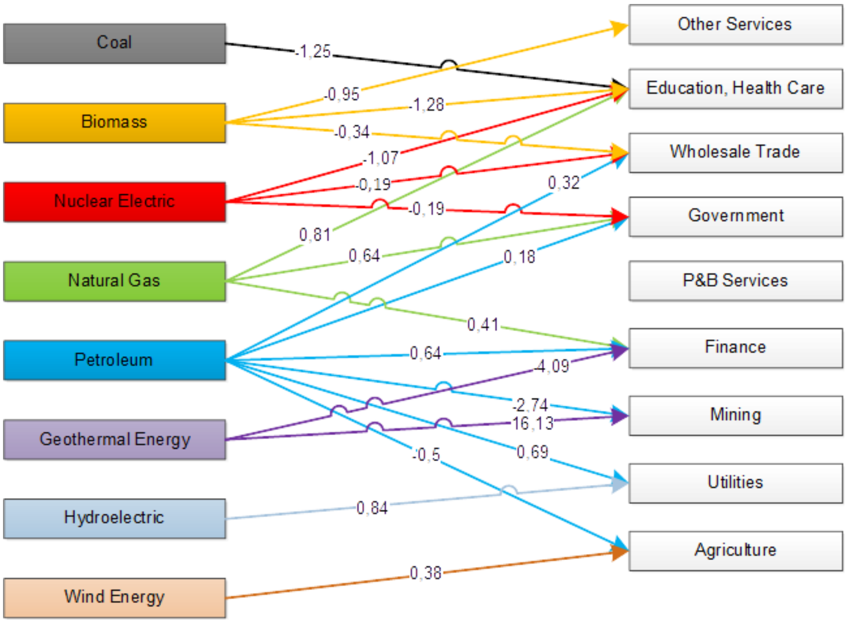

Figure 4: Effects of Energy Consumption on Sectoral GDP

Figure 4 illustrates the influence of energy consumption sources on the industrial outputs. Surprisingly, not every increase in energy consumption will lead to an increased industrial output as one might expect. Coal consumption for instance, is negatively related to education and health care. It is evident from Figure 4, that petroleum consumption affects the most industries. A total of six industries are connected significantly to the petroleum consumption, meaning that policies regarding the petroleum consumption should be planned very carefully.

6. Discussion and Conclusion

In this study we analyzed the effects of sectoral GDPs on energy consumption by source and vice versa. Surprisingly, not every increase in energy consumption will lead to an increased industrial output and not every increase in industrial output will lead to an increased consumption of energy sources as expected.

We have obtained very important and remarkable results for the industrial sectors that we have analyzed in this article. To emphasize the meaning of the long run coefficient of our model, we called it the relation rate index from now on. We first investigated the relationship between energy consumption by energy source and gross domestic product on a sectoral basis. According to these results (*Figure 3*) manufacturing (*rri=1.45*), mining (*rri=1*), transportation (*rri=1.43*), information (*rri=0.72*), p&b services (*rri=1.46*) and government (*rri=2.12*) sectors are dependent on natural gas (*rri=relation rate index*). The governmental sector (*rri=1.79*) and the utilities sector (*rri=0.87*) have a positive relation to the petroleum consumption. The arts sector (*rri=-3.95*) has a negative relationship with nuclear energy. This observation might be explained by the fact that the bigger the arts sector is the more environmental awareness will be evident in the society. Utilities (*rri=0.81*) and education, healthcare (*rri=0.39*) sectors consume hydrolectric power in according to their coefficients. Other services except government (*rri=-1.45*) have a negative relationship with Biomass.

After having analyzed the effect of sectoral gross domestic product on energy consumption, it is evident that the energy supply strategy should be choosen with regard to the relationship of the consuming sector to the energy source. The results show that sectors are most likely to depend on natural gas followed by petroleum and hydroelectric energy.

The second step of our research is about effects of energy consumption on sectoral GDP in USA, shown in Figure 4. According to our results, not every increase in energy consumption will lead to an increase in sectoral industrial output as expected. Coal consumption (*rri=-1.25*) has negative relationships with education and healthcare. Biomass has negative relationships with other services (*rri=-0.95*), education, healthcare (*rri=-1.28*) and wholesale trade (*rri=0.34*). Nuclear electricity has negative relationships with education, healthcare (*rri=-1.07*), wholesale trade (*rri=-0.19*) and government (*rri=-0.19*). Petroleum consumption affects wholesale trade (*rri=0.32*), government (rri=0.18), finance (*rri=0.64*), utilities (*rri=0.69*) sectors positively, but mining (*rri=-2.74*) and agriculture (*rri=-0.5*) negatively. Geothermal consumption affects finance (*rri=-4.09*) negatively and mining (*rri=16.13*) positively. Hydroelectric consumption is positively related to utilities (rri=0.84) and wind energy has a positive relationship with agriculture (*rri=0.38*).

The results indicate that there are both positive and negative effects of energy type based consumption on the sectoral gross domestic products in the USA. For instance, petroleum consumption affects almost every sectoral GDP in the positive direction. Wind power plants on the other hand, only have a significant effect on the agriculture sector. For this reason installing these power plants close to the consuming agricultural areas should be considered in order to minimize energy losses due to transmission. Furthermore, clean and emission free energy supplied by wind power plants would also contribute to organic agriculture.

Whenever possible the energy sources and the industrial sectors that are positively related to each other should be placed as close as possible to each other as to decrease energy losses. In order to achieve this goal, the results of this paper can be used as a starting point. However, this issue is strongly dependent on energy authorities, governments and world environmental authorities.

References

Pesaran, M. H., & Pesaran, B. (1997). Working with Microfit 4.0. Cambridge: Camfit Data Ltd.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, 16(3), 289-326.

Wang, Z. X., & Hao, P. (2016). An improved grey multivariable model for predicting industrial energy consumption in China. Applied Mathematical Modelling, 40(11), 5745-5758.

Ozturk, I. (2010). A literature survey on energy–growth nexus. Energy Policy, 38(1), 340-349.

Wolde-Rufael, Y. (2004). Disaggregated industrial energy consumption and GDP: The case of Shanghai, 1952–1999. Energy Economics, 26(1), 69-75.

Mahalingam, B., & Orman, W. H. (2018). GDP and energy consumption: A panel analysis of the US. Applied Energy, 213, 208-218.

Huang, B. N., Hwang, M. J., & Yang, C. W. (2008). Causal relationship between energy consumption and GDP growth revisited: A dynamic panel data approach. Ecological Economics, 67(1), 41-54.

Ozturk, I., & Acaravci, A. (2010). CO2 emissions, energy consumption and economic growth in Turkey. Renewable and Sustainable Energy Reviews, 14(9), 3220-3225.

Begum, R. A., Sohag, K., Abdullah, S. M. S., & Jaafar, M. (2015). CO2 emissions, energy consumption, economic and population growth in Malaysia. Renewable and Sustainable Energy Reviews, 41, 594-601.

Ziramba, E. (2009). Disaggregate energy consumption and industrial production in South Africa. Energy Policy, 37(6), 2214-2220.

Sari, R., & Soytas, U. (2004). Disaggregate energy consumption, employment and income in Turkey. Energy Economics, 26(3), 335-344.

Hu, J. L., & Lin, C. H. (2008). Disaggregated energy consumption and GDP in Taiwan: A threshold co-integration analysis. Energy Economics, 30(5), 2342-2358.

Ozturk, I., & Acaravci, A. (2016). Energy consumption, CO2 emissions, economic growth, and foreign trade relationship in Cyprus and Malta. Energy Sources, Part B: Economics, Planning, and Policy, 11(4), 321-327.

Wong, S. L., Chang, Y., & Chia, W. M. (2013). Energy consumption, energy R&D and real GDP in OECD countries with and without oil reserves. Energy Economics, 40, 51-60.

Khandelwal, V. (2015). Impact of energy consumption, GDP & fiscal deficit on public health expenditure in India: An ARDL bounds testing approach. Energy Procedia, 75, 2658-2664.

Jobert, T., & Karanfil, F. (2007). Sectoral energy consumption by source and economic growth in Turkey. Energy Policy, 35(11), 5447-5456.

Bowden, N., & Payne, J. E. (2009). The causal relationship between US energy consumption and real output: A disaggregated analysis. Journal of Policy Modeling, 31(2), 180-188.

Sari, R., Ewing, B. T., & Soytas, U. (2008). The relationship between disaggregate energy consumption and industrial production in the United States: An ARDL approach. Energy Economics, 30(5), 2302-2313.

Alvarez, J., & Valencia, F. (2016). Made in Mexico: Energy reform and manufacturing growth. Energy Economics, 55, 253-265.

Burke, P. J., & Csereklyei, Z. (2016). Understanding the energy-GDP elasticity: A sectoral approach. Energy Economics, 58, 199-210.

Boutabba, M. A. (2014). The impact of financial development, income, energy and trade on carbon emissions: Evidence from the Indian economy. Economic Modelling, 40, 33-41.

Bloch, H., Rafiq, S., & Salim, R. (2015). Economic growth with coal, oil and renewable energy consumption in China: Prospects for fuel substitution. Economic Modelling, 44, 104-115.

Li, F., Song, Z., & Liu, W. (2014). China's energy consumption under the global economic crisis: Decomposition and sectoral analysis. Energy Policy, 64, 193-202.

Keho, Y. (2016). What drives energy consumption in developing countries? The experience of selected African countries. Energy Policy, 91, 233-246.

Bildirici, M. E., & Kayıkçı, F. (2013). Effects of oil production on economic growth in Eurasian countries: Panel ARDL approach. Energy, 49, 156-161.

Lin, B., & Wang, A. (2015). Estimating energy conservation potential in China's commercial sector. Energy, 82, 147-156.

Rahman, M. S., Junsheng, H., Shahari, F., Aslam, M., Masud, M. M., Banna, H., & Liya, M. (2015). Long-run relationship between sectoral productivity and energy consumption in Malaysia: An aggregated and disaggregated viewpoint.Energy, 86, 436-445.

Rahman, M. S., Noman, A. H. M., Shahari, F., Aslam, M., Gee, C. S., Isa, C. R., & Pervin, S. (2016). Efficient energy consumption in industrial sectors and its effect on environment: A comparative analysis between G8 and Southeast Asian emerging economies. Energy, 97, 82-89.

Ouyang, X., & Lin, B. (2015). Analyzing energy savings potential of the Chinese building materials industry under different economic growth scenarios. nergy and Buildings, 109, 316-327.

Furuoka, F. (2015). Financial development and energy consumption: Evidence from a heterogeneous panel of Asian countries. Renewable and Sustainable Energy Reviews, 52, 430-444.

Chen, P. Y., Chen, S. T., Hsu, C. S., & Chen, C. C. (2016). Modeling the global relationships among economic growth, energy consumption and CO2 emissions. Renewable and Sustainable Energy Reviews, 65, 420-431.

Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461-464.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716-723.

Pesaran, M. H., & Shin, Y. (1998). An autoregressive distributed-lag modelling approach to cointegration analysis. Econometric Society Monographs, 31, 371-413.

Bahmani-Oskooee, M., & Nasir, A. (2004). ARDL approach to test the productivity bias hypothesis. Review of Development Economics, 8(3), 483-488.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427-431.

Phillips, P. C., & Perron, P. (1988). Testing for a unit root in time series regression. Biometrika, 75(2), 335-346.

Appendix

Table 10: Estimated Coefficients Part 1

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: Natural Gas		*Panel B. Dep. Var.: Agriculture*		*Panel C. Dep. Var.: Agriculture*	
Selected Model: ARDL(2, 2, 2)		*Selected Model: ARDL(2, 0, 1)*		*Selected Model: ARDL(2, 0, 1)*	
ΔNatural Gas(-1)	0.31***	ΔAgriculture(-1)	-0.25**	ΔAgriculture(-1)	-0.19
ΔAgriculture	-0.03	ΔPetroleum	-0.17*	ΔWind	0.17**
ΔAgriculture(-1)	-0.15***	ΔEmployment	-0.34	ΔEmployment	-0.89
ΔEmployment	0.81***	ect(-1)	-0.34 (-3.26)*	ect(-1)	-0.44 (-3.88)**
ΔEmployment(-1)	-0.43				
ect(-1)	-0.05 (-2.13)	Petroleum	-0.5**	Wind	0.38***
		Employment	2.06***	Employment	1.44***
Agriculture	1.46	C	-18.01***	C	-12.66***
Employment	-2.04				
C	21.08				
Panel D. Dep. Var.: Natural Gas		*Panel E. Dep. Var.: Nuclear Electric*		*Panel F. Dep. Var.: Arts*	
Selected Model: ARDL(2, 0, 2)		*Selected Model: ARDL(2, 1, 1)*		*Selected Model: ARDL(1, 1, 1)*	
ΔNatural Gas(-1)	0.31**	ΔNuclear Electric(-1)	0.41***	ΔPetroleum	0.31**
ΔArts	0.15*	ΔArts	0.05	ΔEmployment	0.93***
ΔEmployment	0.73**	ΔEmployment	0.25	ect(-1)	-0.25 (-3.4)*
ΔEmployment(-1)	-0.54*	ect(-1)	-0.11 (-3.56)**		
ect(-1)	-0.05 (-2.38)			Petroleum	-0.15
		Arts	-3.95***	Employment	1.9***
Arts	2.75	Employment	9.58***	C	-17.27***
Employment	-4.4	C	-92.95***		
C	42.68				
Panel G. Dep. Var.: Construction		*Panel H. Dep. Var.: Petroleum*		*Panel I. Dep. Var.: Construction*	
Selected Model: ARDL(2, 2, 1)		*Selected Model: ARDL(2, 1, 2)*		*Selected Model: ARDL(2, 1, 2)*	
ΔConstruction(-1)	0.14	ΔPetroleum(-1)	0.36***	ΔConstruction(-1)	0.36***
ΔPetroleum	0.56***	ΔConstruction	0.31***	ΔWind	-0.42**
ΔPetroleum(-1)	-0.3*	ΔEmployment	0.23	ΔEmployment	1.93***
ΔEmployment	1.97***	ΔEmployment(-1)	-0.87***	ΔEmployment(-1)	-1.36***
ect(-1)	-0.06 (-1.48)	ect(-1)	-0.03 (-1.06)	ect(-1)	-0.1 (-2.83)
Petroleum	-1.33	Construction	2.34	Wind	0.13
Employment	1.66	Employment	-1.69	Employment	0.78***
C	-10.36	C	12.67	C	-4.37**

Panel J. Dep. Var.: Education		*Panel K. Dep. Var.: Hydroelectric*		*Panel L. Dep. Var.: Education*	
Selected Model: ARDL(2, 2, 1)		*Selected Model: ARDL(1, 2, 0)*		*Selected Model: ARDL(1, 0, 2)*	
ΔConstruction(-1)	-0.32**	ΔConstruction	-0.6	ΔGeothermal	0.06
ΔHydroelectric	-0.03	ΔConstruction(-1)	-0.99**	ΔEmployment	0.36**
ΔHydroelectric(-1)	0.05	ΔEmployment	-0.3***	ΔEmployment(-1)	-0.4**
ΔEmployment	0.29*	ect(-1)	-0.34 (-3.74)**	ect(-1)	-0.02 (-1.45)
ect(-1)	-0.02 (-1.16)				
		Education	0.39***	Geothermal	2.66
Hydroelectric	-6.84	Employment	-0.89***	Employment	-1.56
Employment	-0.77	C	10.16***	C	23.4

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

Table 11: Estimated Coefficients Part 2

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: Education		*Panel B. Dep. Var.: Education*		*Panel C. Dep. Var.: Education*	
Selected Model: ARDL(1, 0, 2)		*Selected Model: ARDL(1, 0, 2)*		*Selected Model: ARDL(1, 0, 2)*	
ΔNatural Gas	0.06**	ΔNuclear Electric	-0.04***	ΔPetroleum	0.02
ΔEmployment	0.35**	ΔEmployment	0.49***	ΔEmployment	0.33*
ΔEmployment(-1)	-0.46***	ΔEmployment(-1)	-0.43***	ΔEmployment(-1)	-0.45**
ect(-1)	-0.08 (-2.9)	ect(-1)	-0.04 (-3.39)*	ect(-1)	-0.03 (-1.72)
Natural Gas	0.81***	Nuclear Electric	-1.07***	Petroleum	0.55
Employment	0.73**	Employment	3.75***	Employment	-0.16
C	-6.44*	C	-36.96***	C	4.97
Panel D. Dep. Var.: Education		*Panel E. Dep. Var.: Education*		*Panel F. Dep. Var.: Education*	
Selected Model: ARDL(1, 1, 2)		*Selected Model: ARDL(1, 0, 2)*		*Selected Model: ARDL(1, 2, 0)*	
ΔWind	0.19*	ΔBiomass	-0.04*	ΔCoal	0.02
ΔEmployment	0.35**	ΔEmployment	0.4**	ΔCoal(-1)	-0.11**
ΔEmployment(-1)	-0.24	ΔEmployment(-1)	-0.49***	ΔEmployment	0.13**
ect(-1)	-0.04 (-3.45)*	ect(-1)	-0.03 (-2.68)	ect(-1)	-0.06 (-3.88)**
Wind	-0.04	Biomass	-1.28*	Coal	-1.25***
Employment	0.02	Employment	1.32	Employment	2.14***
C	4.51	C	-8.61	C	-16.68***
Panel G. Dep. Var.: Finance		*Panel H. Dep. Var.: Finance*		*Panel I. Dep. Var.: Finance*	

Selected Model: ARDL(1, 1, 0)		*Selected Model: ARDL(1, 0, 0)*		*Selected Model: ARDL(1, 0, 0)*	
ΔHydroelectric	-0.01	ΔGeothermal	-0.26**	ΔNatural Gas	0.04***
ΔEmployment	0.07	ΔEmployment	0.18**	ΔEmployment	0.18**
ect(-1)	-0.05 (-2.11)	ect(-1)	-0.06 (-2.5)	ect(-1)	-0.1 (-3.11)
Hydroelectric	0.66	Geothermal	-4.09**	Natural Gas	0.41***
Employment	1.39***	Employment	2.8***	Employment	1.69***
C	-12.45**	C	-27.55***	C	-16.64***
Panel J. Dep. Var.: Finance		*Panel K. Dep. Var.: Finance*		*Panel L. Dep. Var.: Finance*	
Selected Model: ARDL(1, 0, 0)		*Selected Model: ARDL(1, 0, 0)*		*Selected Model: ARDL(1, 2, 0)*	
ΔNuclear Electric	-0.02	ΔPetroleum	0.04***	ΔBiomass	-0.07
ΔEmployment	0.14	ΔEmployment	0.09	ΔBiomass(-1)	0.13*
ect(-1)	-0.05 (-1.83)	ect(-1)	-0.07 (-2.67)	ΔEmployment	0.13*
				ect(-1)	-0.05 (-1.96)
Nuclear Electric	-0.31	Petroleum	0.64***		
Employment	2.67***	Employment	1.37***	Biomass	-0.94
C	-25.92***	C	-13.7***	Employment	2.48***
				C	-23.06***

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

Table 12: Estimated Coefficients Part 3

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: Finance		*Panel B. Dep. Var.: Government*		*Panel C. Dep. Var.: Government*	
Selected Model: ARDL(1, 0, 1)		*Selected Model: ARDL(2, 0, 0)*		*Selected Model: ARDL(2, 0, 0)*	
ΔCoal	-0.03	ΔGovernment(-1)	0.42***	ΔGovernment(-1)	0.42***
ΔEmployment	0.37**	ΔHydroelectric	0.03	ΔGeothermal	-0.16
ect(-1)	-0.06 (-1.7)	ΔEmployment	0.09***	ΔEmployment	0.14**
		ect(-1)	-0.11 (-3.4)*	ect(-1)	-0.11 (-3.53)**
Coal	-0.58				
Employment	2.27***	Hydroelectric	0.23	Geothermal	-1.42
C	-20.18***	Employment	0.81***	Employment	1.24***
		C	-5.25***	C	-9.79***
Panel D. Dep. Var.: Government		*Panel E. Dep. Var.: Natural Gas*		*Panel F. Dep. Var.: Government*	
Selected Model: ARDL(2, 1, 0)		*Selected Model: ARDL(2, 2, 2)*		*Selected Model: ARDL(2, 2, 0)*	
ΔGovernment (-1)	0.46***	ΔNatural Gas(-1)	0.36***	ΔGovernment(-1)	0.3***
ΔNatural Gas	0.17**	ΔGovernment	0.62***	ΔNuclear Electric	0.02
ΔEmployment	0.2***	ΔGovernment(-1)	-0.43**	ΔNuclear Electric(-1)	0.07
ect(-1)	-0.28 (-4.32)***	ΔEmployment	0.66**	ΔEmployment	0.37***

		ΔEmployment(-1)	-0.53*	ect(-1)	-0.25 (-6.09)***
Natural Gas	0.32***	ect(-1)	-0.17 (-3.39)*		
Employment	0.74***			Nuclear Electric	-0.19***
C	-5.21***	Government	2.12***	Employment	1.52***
		Employment	-1.47***	C	-12.98***
		C	10.96***		
Panel G. Dep. Var.: Government		*Panel H. Dep. Var.: Petroleum*		*Panel I. Dep. Var.: Government*	
Selected Model: ARDL(2, 2, 0)		*Selected Model: ARDL(2, 0, 2)*		*Selected Model: ARDL(2, 0, 0)*	
ΔGovernment (-1)	0.3***	ΔPetroleum(-1)	0.39***	ΔGovernment(-1)	0.4***
ΔPetroleum	0.08	ΔGovernment	0.15**	ΔWind	-0.01
ΔPetroleum(-1)	0.11	ΔEmployment	0.94***	ΔEmployment	0.08**
ΔEmployment	0.16***	ΔEmployment(-1)	-0.78***	ect(-1)	-0.09 (-3.36)*
ect(-1)	-0.19 (-4.23)***	ect(-1)	-0.09 (-2.65)		
				Wind	-0.09
Petroleum	0.18*	Government	1.79***	Employment	0.9***
Employment	0.81***	Employment	-1.32**	C	-5.99***
C	-5.7***	C	11.04**		
Panel J. Dep. Var.: Government		*Panel K. Dep. Var.: Government*		*Panel L. Dep. Var.: Natural Gas*	
Selected Model: ARDL(2, 0, 0)		*Selected Model: ARDL(2, 2, 0)*		*Selected Model: ARDL(2, 0, 2)*	
ΔGovernment (-1)	0.4***	ΔGovernment(-1)	0.28**	ΔNatural Gas(-1)	0.36***
ΔBiomass	-0.03	ΔCoal	0.08*	ΔInformation	0.06
ΔEmployment	0.13**	ΔCoal(-1)	0.1**	ΔEmployment	0.91***
ect(-1)	-0.11 (-3.44)*	ΔEmployment	0.12**	ΔEmployment(-1)	-0.51
		ect(-1)	-0.15 (-4.48)***	ect(-1)	-0.09 (-3.14)
Biomass	-0.27				
Employment	1.13***	Coal	0.12	Information	0.72**
C	-8.33***	Employment	0.77***	Employment	-1.57
		C	-4.91**	C	18.63*

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

Table 13: Estimated Coefficients Part 4

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: Information		*Panel B. Dep. Var.: Biomass*		*Panel C. Dep. Var.: Natural Gas*	
Selected Model: ARDL(2, 1, 0)		*Selected Model: ARDL(1, 2, 2)*		*Selected Model: ARDL(2, 0, 2)*	
ΔInformation(-1)	-0.21*	ΔInformation	0.27*	ΔNatural Gas(-1)	0.29**
ΔBiomass	0.22**	ΔInformation(-1)	0.24*	ΔManufacturing	0.12**

ΔEmployment	0.22**
ect(-1)	-0.07 (-2.56)
Biomass	-0.91
Employment	3.22***
C	-31.73***

ΔEmployment	0.31
ΔEmployment(-1)	-0.71**
ect(-1)	-0.05 (-1.29)
Information	1.68
Employment	-3.43
C	34.87

ΔEmployment	0.66**
ΔEmployment(-1)	-0.48
ect(-1)	-0.08 (-3.5)*
Manufacturing	1.45**
Employment	-2.16*
C	22.22**

Panel D. Dep. Var.: Petroleum
Selected Model: ARDL(2, 1, 0)

ΔPetroleum(-1)	0.26***
ΔManufacturing	0.44***
ΔEmployment	-0.12*
ect(-1)	-0.03 (-1.49)
Manufacturing	2.16
Employment	-4.21
C	43.33

Panel E. Dep. Var.: Mining
Selected Model: ARDL(1, 1, 1)

ΔGeothermal	-0.54
ΔEmployment	0.78
ect(-1)	-0.08 (-1.8)
Geothermal	16.13*
Employment	-3.25*
C	40.38*

Panel F. Dep. Var.: Natural Gas
Selected Model: ARDL(1, 2, 2)

ΔMining	0.03
ΔMining(-1)	0.1*
ΔEmployment	0.88***
ΔEmployment(-1)	-0.38
ect(-1)	-0.09 (-4.41)***
Mining	1***
Employment	0.17
C	-3.15

Panel G. Dep. Var.: Mining
Selected Model: ARDL(1, 1, 0)

ΔPetroleum	0.64**
ΔEmployment	0.24***
ect(-1)	-0.07 (-1.79)
Petroleum	-2.74*
Employment	3.24*
C	-23.63

Panel H. Dep. Var.: Geothermal
Selected Model: ARDL(1, 2, 2)

ΔOther	0.06*
ΔOther(-1)	0.1**
ΔEmployment	-0.14
ΔEmployment(-1)	-0.14*
ect(-1)	-0.15 (-3.46)*
Other	-0.12
Employment	0.45***
C	-4.55***

Panel I. Dep. Var.: Petroleum
Selected Model: ARDL(2, 0, 2)

ΔPetroleum(-1)	0.39***
ΔOther	0.08**
ΔEmployment	0.84***
ΔEmployment(-1)	-0.92***
ect(-1)	-0.03 (-1.43)
Other	2.27
Employment	-3.28
C	31.79

Panel J. Dep. Var.: Other
Selected Model: ARDL(2, 1, 1)

ΔOther(-1)	-0.2*
ΔBiomass	0.26**
ΔEmployment	1.28***
ect(-1)	0.14 (2.43)
Biomass	-0.95***

Panel K. Dep. Var.: Biomass
Selected Model: ARDL(1, 1, 2)

ΔOther	0.32**
ΔEmployment	0.3
ΔEmployment(-1)	-0.5**
ect(-1)	-0.2 (-5.27)***
Other	-1.45***

Panel L. Dep. Var.: P&B Services
Selected Model: ARDL(1, 0, 1)

ΔHydroelectric	-0.03*
ΔEmployment	1.56***
ect(-1)	-0.05 (-2.25)
Hydroelectric	-0.57
Employment	2.52***

Employment	2.48***	Employment	3.12***	C	-24.23***
C	-23.22***	C	-28.54***		

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

Table 14: Estimated Coefficients Part 5

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: P&B Services		*Panel B. Dep. Var.: Geothermal*		*Panel C. Dep. Var.: P&B Services*	
Selected Model: ARDL(1, 1, 1)		*Selected Model: ARDL(1, 1, 1)*		*Selected Model: ARDL(1, 0, 1)*	
ΔGeothermal	0.76**	ΔP&B Services	0.14**	ΔNatural Gas	-0.04**
ΔEmployment	1.51***	ΔEmployment	-0.21**	ΔEmployment	1.57***
ect(-1)	-0.05 (-1.98)	ect(-1)	-0.13 (-3.24)*	ect(-1)	0.02 (0.38)
Geothermal	3.84	P&B Services	-0.02	Natural Gas	2.85
Employment	1.28	Employment	0.34	Employment	1.93
C	-11.05	C	-3.71*	C	-28.51**
Panel D. Dep. Var.: Natural Gas		*Panel E. Dep. Var.: P&B Services*		*Panel F. Dep. Var.: P&B Services*	
Selected Model: ARDL(2, 2, 1)		*Selected Model: ARDL(1, 1, 1)*		*Selected Model: ARDL(1, 0, 1)*	
ΔNatural Gas(-1)	0.28***	ΔNuclear Electric	-0.09**	ΔPetroleum	-0.05**
ΔP&B Services	-0.11	ΔEmployment	1.57***	ΔEmployment	1.66***
ΔP&B Services(-1)	-0.25*	ect(-1)	-0.07 (-1.76)	ect(-1)	-0.02 (-0.96)
ΔEmployment	1.05**				
ect(-1)	-0.17 (-5.53)***	Nuclear Electric	-0.17	Petroleum	-2
		Employment	2.88***	Employment	3.36***
P&B Services	1.46***	C	-28.93***	C	-27.29***
Employment	-3.4***				
C	36.92***				
Panel G. Dep. Var.: P&B Services		*Panel H. Dep. Var.: P&B Services*		*Panel I. Dep. Var.: P&B Services*	
Selected Model: ARDL(1, 0, 1)		*Selected Model: ARDL(1, 0, 1)*		*Selected Model: ARDL(1, 0, 1)*	
ΔWind	0.01	ΔBiomass	0.04**	ΔCoal	0
ΔEmployment	1.54***	ΔEmployment	1.48***	ΔEmployment	1.5***
ect(-1)	-0.07 (-2.92)	ect(-1)	-0.04 (-1.74)	ect(-1)	-0.06 (-1.94)
Wind	0.22	Biomass	1.08	Coal	-0.08
Employment	2.31***	Employment	1.2	Employment	2.42***
C	-22.62***	C	-10.99	C	-23.6***
Panel J. Dep. Var.: Natural Gas		*Panel K. Dep. Var.: Petroleum*		*Panel L. Dep. Var.: Transportation*	

Selected Model: ARDL(2, 1, 2)		*Selected Model: ARDL(2, 2, 2)*		*Selected Model: ARDL(2, 0, 2)*	
ΔNatural Gas(-1)	0.32***	ΔPetroleum(-1)	0.31**	ΔTransportation(-1)	0.17
ΔRetail	-0.2	ΔRetail	0.43***	ΔGeothermal	0.41*
ΔEmployment	1.23***	ΔRetail(-1)	0.17*	ΔEmployment	1.96***
ΔEmployment(-1)	-0.76**	ΔEmployment	0.25	ΔEmployment(-1)	-1.05***
ect(-1)	-0.06 (-2.89)	ΔEmployment(-1)	-0.62**	ect(-1)	-0.12 (-2.95)
		ect(-1)	0 (-0.01)		
Retail	0.9			Geothermal	3.45
Employment	-1.34	Retail	340.04	Employment	0.82
C	15.08	Employment	-868.58	C	-5.69
		C	8663.95		

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

Table 15: Estimated Coefficients Part 6

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: Geothermal		*Panel B. Dep. Var.: Natural Gas*		*Panel C. Dep. Var.: Utilities*	
Selected Model: ARDL(1, 0, 0)		*Selected Model: ARDL(2, 0, 2)*		*Selected Model: ARDL(1, 0, 0)*	
ΔTransportation	-0.02**	ΔNatural Gas(-1)	0.28**	ΔHydroelectric	0.15**
ΔEmployment	0.07***	ΔTransportation	0.13**	ΔEmployment	0.05
ect(-1)	-0.09 (-2.43)	ΔEmployment	0.71**	ect(-1)	-0.18 (-3.48)*
		ΔEmployment(-1)	-0.49*		
Transportation	-0.26	ect(-1)	-0.09 (-3.64)**	Hydroelectric	0.84***
Employment	0.73**			Employment	0.26
C	-7.25**	Transportation	1.43***	C	0.45
		Employment	-2.01**		
		C	20.45**		
Panel D. Dep. Var.: Hydroelectric		*Panel E. Dep. Var.: Utilities*		*Panel F. Dep. Var.: Geothermal*	
Selected Model: ARDL(1, 0, 0)		*Selected Model: ARDL(1, 1, 1)*		*Selected Model: ARDL(1, 1, 0)*	
ΔUtilities	0.29***	ΔGeothermal	1.65**	ΔUtilities	0.04*
ΔEmployment	-0.13**	ΔEmployment	0.65	ΔEmployment	0.04**
ect(-1)	-0.36 (-3.82)**	ect(-1)	-0.09 (-1.9)	ect(-1)	-0.13 (-2.85)
Utilities	0.81***	Geothermal	1	Utilities	-0.06
Employment	-0.36**	Employment	0.04	Employment	0.33***
C	1.8	C	3.91	C	-3.46***
Panel G. Dep. Var.: Utilities		*Panel H. Dep. Var.: Utilities*		*Panel I. Dep. Var.: Petroleum*	
Selected Model: ARDL(1, 1, 0)		*Selected Model: ARDL(1, 0, 0)*		*Selected Model: ARDL(2, 2, 2)*	
ΔNuclear Electric	0.17	ΔPetroleum	0.14*	ΔPetroleum(-1)	0.48***

ΔEmployment	0.24*	ΔEmployment	0.02	ΔUtilities	-0.03
ect(-1)	-0.15 (-2.93)	ect(-1)	-0.2 (-2.63)	ΔUtilities(-1)	-0.11*
				ΔEmployment	0.89***
Nuclear Electric	-0.36	Petroleum	0.69***	ΔEmployment(-1)	-0.84***
Employment	1.61**	Employment	0.09	ect(-1)	-0.13 (-3.34)*
C	-13.56*	C	1.09		
				Utilities	0.87***
				Employment	-0.22
				C	2.17
Panel J. Dep. Var.: Utilities		*Panel K. Dep. Var.: Utilities*		*Panel L. Dep. Var.: Utilities*	
Selected Model: ARDL(1, 1, 1)		*Selected Model: ARDL(1, 1, 0)*		*Selected Model: ARDL(1, 1, 0)*	
ΔWind	0.55*	ΔBiomass	0.26	ΔCoal	0.26**
ΔEmployment	0.9**	ΔEmployment	0.07	ΔEmployment	0.02
ect(-1)	-0.09 (-2.19)	ect(-1)	-0.11 (-2.2)	ect(-1)	-0.1 (-2.28)
Wind	-0.77	Biomass	-0.38	Coal	0.19
Employment	0.32	Employment	0.66	Employment	0.22
C	0.78	C	-2.58	C	1.5

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

Table 16: Estimated Coefficients Part 7

Regressor	Coefficient	Regressor	Coefficient	Regressor	Coefficient
Panel A. Dep. Var.: Wholesale		*Panel B. Dep. Var.: Wholesale*		*Panel C. Dep. Var.: Wholesale*	
Selected Model: ARDL(1, 1, 2)		*Selected Model: ARDL(1, 2, 2)*		*Selected Model: ARDL(1, 1, 2)*	
ΔHydroelectric	0	ΔNatural Gas	-0.21*	ΔNuclear Electric	0.06
ΔEmployment	1.68***	ΔNatural Gas(-1)	0.17	ΔEmployment	2.02***
ΔEmployment(-1)	-0.7***	ΔEmployment	2.09***	ΔEmployment(-1)	-0.87***
ect(-1)	-0.14 (-2.12)	ΔEmployment(-1)	-1.02***	ect(-1)	-0.24 (-3.64)**
		ect(-1)	-0.2 (-3.35)*		
Hydroelectric	0.53			Nuclear Electric	-0.19**
Employment	2.79***	Natural Gas	0.18	Employment	3.62***
C	-29.02***	Employment	2.89***	C	-37.84***
		C	-30.09***		
Panel D. Dep. Var.: Wholesale		*Panel E. Dep. Var.: Wholesale*			
Selected Model: ARDL(1, 1, 2)		*Selected Model: ARDL(1, 0, 2)*			
ΔPetroleum	0.36**	ΔBiomass	-0.07**		
ΔEmployment	1.35***	ΔEmployment	1.91***		

ΔEmployment(-1)	-0.61**	ΔEmployment(-1)	-0.77***
ect(-1)	-0.2 (-3.3)*	ect(-1)	-0.2 (-3.33)*
Petroleum	0.32*	Biomass	-0.34*
Employment	2.87***	Employment	3.36***
C	-30.52***	C	-34.53***

Note: *, ** and *** denote significance on the 10%, 5% and 1% level, respectively. Critical values for nonstandard distributed t-statistics of the error corretion term (ect) are obtained from Pesaran, Shin, and Smith (2001).

CHAPTER 3

Food Safety and Public Health

Yavuz Yüksel[1]

1. FOOD SAFETY AND PUBLIC HEALTH

1.1. Introduction and Definitions

The rapid increase of the world population, increasing environmental pollution, economic weakness and lack of education deepen the problems of nutrition and make it difficult to provide safe food. We can define safe (healthy) food as clean and intact food in terms of physical, chemical and biological hazards that have not lost its nutritional value (Giray and Soysal 2007). Unsafe foods contain dangerous substances or contaminants that can make people ill by increasing the risk of sudden poisoning or chronic illness.

In 2000, the World Health Council, WHO's highest governing body, adopted a resolution that recognized food safety as a fundamental public health function. A wide variety of biological and chemical agents or hazards can cause chronic or life-threatening diseases of varying degrees of severity, or foodborne diseases that cause both. These agents may be bacteria, viruses, protozoa, helminths and natural toxins as well as chemical and environmental pollutants. Epidemiological studies in the last three years have shown that in addition to the increase in foodborne diseases, epidemic diseases such as **Salmonellosis**, **Cholera**, enterohaemorrhagic Escherichia (E.) coli infections and **Hepatitis A** have also increased in developing and developing countries (Käferstein 2003).

2. Classification of Food Hazards

The effects of physical, biological and chemical hazards on humans, which are the main threats to food safety, are as follows..

2.1. Effect of physical hazards

Physical hazards include foreign materials that should not be present in food, such as:

- Broken glass, plastic, bone, metal pieces, or wood.
- Soil, stones, hair, nails, and insects.
- **Radiation:** A critical physical hazard capable of causing skin/eye damage or cancer.

[1] Dr. Öğr. Üyesi, Balıkesir University, Faculty of Engineering, Department of Food Engineering, Balikesir, Türkiye, ORCID: 0000-0001-7960-578X

These materials may enter the food chain through raw materials or during production, storage, and transport (Giray and Soysal 2007; Erkmen and Bozoglu 2008; Erkmen 2010).

2.2. Effect of Biological hazards

Biological hazards are the leading cause of food poisoning and can be divided into three subgroups:

1. **Natural Toxins:** Chemicals naturally present in foods, such as solanine in sprouted potatoes or poisonous mushrooms.
2. **Microorganisms:** Bacteria, viruses, molds, and parasites. Bacteria (e.g., *Salmonella*, *Listeria monocytogenes*, and *E. coli*) are the most significant threats.
3. **GMOs:** Genetically modified organisms (Giray and Soysal 2007; Erkmen and Bozoglu 2008; Erkmen 2010).

Several bacterial species are frequently implicated in foodborne illnesses, most notably strains of *Salmonella*, *Listeria monocytogenes*, and pathogenic *Escherichia coli*. Other significant contributors to food poisoning include toxin-producing organisms such as *Staphylococcus aureus*, *Bacillus cereus*, *Clostridium botulinum*, and *C. perfringens*. In underdeveloped countries, hundreds of thousands of people die each year due to diarrheal diseases caused by water and food(Giray and Soysal 2007; Erkmen and Bozoglu 2008; Erkmen 2010).

Food Poisoning, keeping food at an inappropriate temperature (between 5-60ºC) for more than two hours, inadequate processes in food production and preparation, using dirty tools and water, supplying food from unreliable area, contacting raw foods with processed foods, improper storage conditions, food hygiene or lack of personal hygiene. Pathogenic microorganisms that cause food poisoning are mostly transmitted from people working in food production (Giray and Soysal 2007; Erkmen and Bozoglu 2008; Erkmen 2010).

2.3. Effect of Chemical hazards

Non-food chemicals that are added to food from various sources or added externally for a purpose cause adverse effects on human health. Consumers are exposed to dozens of different food additives (GKM) and other chemicals that infect food every day. GKMs and chemicals that are contaminated by food from outside are toxic substances that are not metabolized in human and animal organisms. Potential chemical threats to food integrity include heavy metal toxicity (Pb, Hg, Cd), persistent organic pollutants like dioxins, and residues from agrochemicals or veterinary treatments. Contamination can also occur through secondary contact, such as poorly rinsed equipment or chemical leaching from storage materials. Finally, the excessive application of food additives represents a significant regulated chemical hazard (Giray and Soysal 2007; Erkmen and Bozoglu 2008). Some chemical hazards and their effects are given in the Table 1.

Table 1. Some chemical hazards and their effects

CHEMICAL HAZARDS	**FOOD POISONING AND DISEASE SYMPTOMS**
1. Food additives • Benzoic acid and its compounds • Melamine • Colorants • *Sweeteners* • *Sodium nitrite and nitrate* • *Butylated hydroxy hydroxy toluene (BHT)* • *anisole (BHA) and butylated* • *Olestra* • *Partly hydrogenated vegetable oil (trans fats)*	Benzoic acid and its compounds may cause brain injury, hypersensitivity, weight loss, induction of asthma or nervous disorder; hyperactivity and urticaria in children, skin redness, swelling, itching and pain; estrogen hormones by increasing the balance of hormones and the formation of tumors can be specified (Deshpande 2002; Omaye 2004). Melamine has been reported to have significant adverse effects on humans. In China, four children who ate melamine-containing foods died and tens of thousands of children became ill (Afoakwa 2008; McDonald 2008). Various health complications, ranging from hypersensitivity and respiratory issues like asthma to more severe conditions such as chromosomal abnormalities and oncogenesis, have been linked to the use of colorants (Omaye 2004). Sweeteners cause toxic and allergic reactions; skin, digestive and heart disorders; tumor formation; lymph and blood cancer (Omaye 2004; Canimoğlu 2006). Nitrite and nitrates can be converted to carcinogenic compounds such as nitrosamine, liver, lung, kidney, larynx, stomach and pancreas cancers have been reported to play a role in the formation. (Erkmen and Bozoglu 2008; Omaye 2004). BHA and BHT have been reported to cause allergic reactions, hyperactivity, liver and kidney dysfunction, estrogenic effects and adverse effects such as increasing blood cholesterol. (Pascal 1999). Olestra may cause diarrhea, abdominal pain, digestive system disorders, decreased body strength and decreased carotenoids in the blood (Oztop 1998). Many health experts say that trans fats increase the risk of coronary artery disease three times, increase total and bad cholesterol levels, reduce good cholesterol levels, increase the risk of cancer five times, reduce the quality of breast milk, weaken the immune system and the pancreatic insulin (Miró-Colmenárez et al.,2024; Pipoyan et al., 2021).
2.Foodborne chemicals • *Heavy metals* • *Polychlorinated biphenyls (PCBs)* • *Dioxins* • *Acrylamide and polymers* • *Pesticides* • *Hormones*	Some heavy metals that can contaminate people from food, water, the environment and the air, causing toxic effects or sudden death on humans are: Mercury, copper, iron, cadmium, nickel, lead, arsenic, zinc. It should be noted that children are more sensitive to the effects of heavy metals and compounds than adults. (Omaye 2004; Vural 1993) PCB compounds initiate the formation of cancer causes it to progress too. PCB compounds have been reported to cause tumor growth in the liver, skin and lungs (Ayaz 2008).

	As a result of the studies carried out by WHO, dioxins have been found to cause cancer in humans. Disorders caused by dioxin can be defined as excessive weight loss due to diarrhea, loss of appetite, pigment and rash on the skin, psychological abnormalities, neurological problems, reproductive disorders, cleft palate, high blood pressure, elevated blood lipid and cholesterol levels (Timbrell 2002; Kogevinas 2001; Ongley 1996). As a result of the studies; Acrylamide and its polymers have been evaluated as mutagenic because they cause genotoxic effects for somatic and sex cells and hereditary damage to genes and chromosomes (Ayaz 2008). The negative effects of pesticides on humans are: (Ariman and Aras 2002; Ayaz 2008; Kara et al. 2002). a.Nervous and immune system disorders, heart disease and disorders b. As a result of hormonal disorders in the endocrine system, the decrease in sperm count in men, impotence, swelling and fat in the body, weakening of the cells to the cancer, susceptibility to the development of children during puberty, triggering of early puberty, breast enlargement in boys and puberty in girls.c. Other health problems caused by pesticides in humans in the long term are allergies, migraine, asthma, eczema, miscarriage, premature birth, congenital deformities and shortening of breastfeeding time.
3.Transported from food to packaging materials pollutants	Toxic gases such as CO, HCN, HCl, benzene and phosgene released by combustion or combustion of plastics threaten the environment and can cause deaths. (Ayaz 2008).

The complexity of chemical hazards has evolved with new contaminants emerging from modern industrial processes (Mititelu et al. 2025)

3. Conclusions

Food safety remains a critical pillar of global public health, acting as a bridge between food security and population well-being. As discussed throughout this chapter, the complexity of foodborne hazards—ranging from traditional biological pathogens to modern chemical contaminants—requires a multi-dimensional management approach. While physical hazards can often be controlled through rigorous processing standards, biological and chemical risks demand continuous surveillance and stricter regulatory frameworks. The integration of international safety standards and the enhancement of food handling practices are essential to mitigating the rising trend of foodborne epidemics. Ultimately, ensuring food safety is not only a technical necessity but a fundamental human right that underpins sustainable development and economic stability.

4. Future Perspectives

The landscape of food safety is evolving rapidly, driven by technological advancements and changing environmental conditions. Looking forward, the following areas are expected to define the future of the field:

- **Digital Transformation and AI:** The integration of Artificial Intelligence (AI) and Machine Learning (ML) will revolutionize predictive food safety, allowing for real-time monitoring of supply chains and early detection of contamination patterns before products reach the consumer (Tyagi et al., 2026; Aboyejiet al., 2026).
- **Precision Food Safety:** Utilizing High-Throughput Sequencing (HTS) and Whole Genome Sequencing (WGS) will become the gold standard for tracking foodborne outbreaks with unprecedented accuracy, enabling faster responses to emerging pathogens.
- **Climate Change Adaption:** As global warming affects the prevalence of toxins (such as mycotoxins) and the distribution of pests, future food safety protocols must incorporate climate-resilient strategies to protect raw material integrity.
- **Sustainable Packaging:** The shift toward biodegradable and "smart" packaging will require new safety assessments to ensure that sustainable materials do not introduce novel chemical migration risks into the food chain.
- **Consumer Engagement:** Future systems will place a greater emphasis on transparency, using Blockchain technology to provide consumers with verifiable data regarding the safety and origin of their food.

5. References:

Aboyeji, D. O., Auta, H. S., Adelere, I. A., Ossamulu, I. F., Makun, H. A., & Muhammad, H. L. (2026). Applications of Artificial Intelligence and Machine Learning in Food Quality Control and Safety Assessment. Taylor and Francis.

Afoakwa, E. O. (2008). Melamine contamination of infant formula in China: the causes, food safety issues and public health implications. *Afr J Food Agric Nutr Dev (AJFAND)*, *8*, 1-9.

Arıman, H., & Aras, N. M. (2002). Hormon veya Hormon Etkisine Sahip Yem Katkı Maddelerinin Balık Yetiştiriciliğindeki Kullanımı. *Ege Journal of Fisheries and Aquatic Sciences*, *19*(3).

Ayaz, A., & Yurttagül, M. (2012). Besinlerdeki toksik öğeler-II. *Ankara Sağlık Bakanlığı Yayınları*, *727*.

Canimoğlu, S., & Rencüzoğulları, E. (2006). The cytogenetic effects of food sweetener maltitol in human peripheral lymphocytes. *Drug and chemical toxicology*, *29*(3), 269-278.

Deshpande, S. S. (2002). *Handbook of food toxicology*. CRC press.

Omaye, S. T. (2004). *Food and nutritional toxicology*. CRC press.

Erkmen, O., & Bozoglu, T. F. (2008). Food microbiology III: food preservation. *Ankara: İlke Publishing Company*.

Erkmen, O. (2010). Gıda kaynaklı tehlikeler ve güvenli gıda üretimi. *Çocuk Sağlığı ve Hastalıkları Dergisi*, *53*(3), 220-235.

Giray, H., & Soysal, A. (2007). Türkiye'de gıda güvenliği ve mevzuatı. *TSK koruyucu hekimlik bülteni*, *6*(6), 485-490.

IUFoST. (2006). Trans Fatty Acids. The International Union of Food Science and Technology Scientific Information Bulletin 5: 1-7.

Kaferstein, F. (2003). Food safety as a public health issue for developing countries.

Kara, E. E., Pırlak, U., Arlı, M., & Doğan, E. (2002). Niğde ilinde bazı tarımsal bitkilerde kullanılan pestisitlerin araştırılması. *Ekoloji Dergisi*, *11*, 21-26.

Kogevinas, M. (2001). Human health effects of dioxins: cancer, reproductive and endocrine system effects. *Apmis*, *109*(S103), S223-S232.

McDonald, S. (2008). Nearly 53,000 Chinese children sick from milk. *The Pantagraph*, *21*.

Miró-Colmenárez, P. J., Illán-Marcos, E., Díaz-Cruces, E., Rocasolano, M. M., Martínez-Hernandez, J. M., Zamora-Ledezma, E., & Zamora-Ledezma, C. (2024). Current insights into industrial trans fatty acids legal frameworks and health challenges in the European Union and Spain. *Foods*, *13*(23), 3845.

Mititelu, M., Neacşu, S. M., Busnatu, Ş. S., Scafa-Udrişte, A., Andronic, O., Lăcraru, A. E., ... & Olteanu, G. (2025). Assessing heavy metal contamination in food: Implications for human health and environmental safety. *Toxics*, *13*(5), 333.

Okcu A, Coşardemir Y, Temiz S, Kumral S, Göçmen M, Koçan OS. (2006). Hormonlar. İzmir: Özel Ege Lisesi.

Ongley, E. D. (1996). *Control of water pollution from agriculture* (Vol. 55). Food & Agriculture Org..

Oztop, H. N., & Sindirilemeyen, T. F. (1998). kalorisiz bir yağ: olestra. *Biyokimya Dergisi*, *23*, 35-42.

Pascal, G. (2019). Safety assessment of food additives and flavoring substances. In *International Food Safety Handbook* (pp. 239-255). Routledge.

Timbrell, J., & Barile, F. A. (2023). *Introduction to toxicology*. CRC Press.

Pipoyan, D., Stepanyan, S., Stepanyan, S., Beglaryan, M., Costantini, L., Molinari, R., & Merendino, N. (2021). The effect of trans fatty acids on human health: regulation and consumption patterns. *Foods*, *10*(10), 2452.

Tyagi, K., Kohli, V., Sikka, K., & Kumari, M. (2026). Enhancing Food Safety and Traceability: The Contribution of Smart Technologies in the Contemporary Food Sector. *Food Reviews International*, 1-24.

Vural, H. (1993). Ağır metal iyonlarının gıdalarda oluşturduğu kirlilikler. *Çevre Dergisi*, *8*, 3-8.

CHAPTER 4

Contribution of Galileo to CSRS-PPP: An Online PPP Service

Sercan Bulbul[1] & Sermet Ogutcu[2] & Salih Alcay[3]

INTRODUCTION

Global Navigation Satellite Systems (GNSS) are satellite-based radio navigation systems that provide users located on or near the Earth's surface with continuous, all-weather, and global positioning, velocity, and timing information. The fundamental principle of GNSS is based on measuring the travel time of coded electromagnetic signals transmitted from satellites to the receiver. The receiver determines the distance between the satellite and the receiver (pseudorange) by multiplying the signal propagation time by the speed of light and solves for the three-dimensional position and receiver clock error using observations from at least four satellites (Hofmann-Wellenhof et al., 2008; Misra & Enge, 2011).

GNSS measurements rely on code and carrier phase observations. While code measurements provide meter-level accuracy, carrier phase observations offer millimeter-level precision; however, they require more complex mathematical modeling due to integer ambiguity (Teunissen & Montenbruck, 2017). Modeling error sources such as satellite orbit and clock errors, atmospheric delays, and multipath effects constitutes the foundation of high-precision positioning techniques.

Although these fundamental principles apply to all GNSS systems, the most significant recent advancement improving positioning accuracy and continuity is the combined use of multiple satellite systems. This development necessitates an understanding of the historical evolution of GNSS and the progression of these systems (Table 1).

[1] Konya Technical University, Faculty of Engineering and Natural Sciences, Geomatics Engineering Department, Konya, Turkey, ORCID ID: 0000-0001-6066-611X
[2] Necmettin Erbakan University, Faculty of Engineering, Department of Geomatics Engineering, Konya, Turkey, ORCID ID: 0009-0003-5949-2943
[3] Necmettin Erbakan University, Faculty of Engineering, Department of Geomatics Engineering, Konya, Turkey, ORCID ID: 0000-0001-5669-7247

Table 1. Historical and Technical Characteristics of GNSS Systems

System	Developer	Operational Timeline	Key Developments	Technical Characteristics	References
GPS	USA (DoD)	IOC: 1993 / FOC: 1995	SA removal (2000), L2C, L5, L1C	CDMA, multi-frequency	Misra & Enge (2011); Hofmann-Wellenhof et al. (2008)
GLONASS	USSR / Russia	1990s / 2011 FOC	Modernization (2000s)	FDMA → CDMA	Teunissen & Montenbruck (2017)
Galileo	EU	2016 Initial Services	Global coverage (2020s)	E1, E5a/b, E6	European GNSS Agency (2019); Teunissen & Montenbruck (2017)

With the operational availability of these three systems, GNSS technology has evolved from a single-system structure to a multi-constellation architecture. The use of multi-GNSS:

- Improves satellite geometry by increasing the number of available satellites,
- Enhances observation continuity at low elevation angles,
- Reduces the Geometric Dilution of Precision (GDOP),
- Increases solution reliability and accuracy.

Particularly in precise positioning techniques (e.g., PPP), the combined use of GPS, GLONASS, and Galileo systems reduces convergence time and improves solution stability (Cai & Gao, 2013). Therefore, in modern GNSS applications, multi-GNSS integration has become the standard approach rather than single-system solutions.

This historical and technical evolution provides a fundamental framework for understanding the development of relative and absolute positioning methods, which will be discussed in the following sections.

2. POSITIONING METHODS WITH GNSS

The increase in the number of satellites and the modernization of GNSS systems have significantly influenced not only coverage but also the accuracy, reliability, and diversity of positioning techniques. In this context, GNSS positioning methods are generally classified into two main categories: relative and absolute positioning.

2.1. Relative Positioning

In relative positioning, at least two GNSS receivers perform simultaneous observations, and the position is determined from the difference measurements between the reference and rover stations. In this process, most satellite clock and orbit errors are eliminated through differencing. In particular, the double-difference technique using carrier phase observations can achieve millimeter–centimeter level accuracy (Teunissen & Montenbruck, 2017). The main applications of this approach include static, rapid static, RTK, and PPK methods. While RTK provides real-time centimeter-level accuracy through instantaneous data transmission, PPK achieves similar accuracy through post-processing. However, all relative methods have limitations such as the requirement for a reference station and error growth depending on baseline length (Hofmann-Wellenhof et al., 2008). These limitations have encouraged the development of absolute positioning methods, especially for large areas and regions lacking reference infrastructure.

2.2. Absolute Positioning

Absolute positioning refers to determining position directly within a global reference frame (e.g., ITRF) using a single GNSS receiver. In its simplest form, Standard Point Positioning (SPP), based on code observations, provides meter-level accuracy. However, this level of accuracy is insufficient for engineering and geodetic applications requiring high precision. One of the most important techniques developed to meet this requirement is Precise Point Positioning (PPP).

2.3. Precise Point Positioning (PPP)

PPP is a technique that enables centimeter-level positioning by using code and carrier phase observations obtained from a single GNSS receiver together with high-precision satellite orbit and clock products (Zumberge et al., 1997). The main characteristics of the PPP approach are as follows:

- It does not require a reference station,
- It provides solutions directly in a global reference frame,
- It utilizes precise satellite orbit and clock corrections (e.g., IGS products),
- It eliminates ionospheric delay through dual-frequency combinations.

The main disadvantage of PPP is the convergence time required to achieve high accuracy. This duration varies depending on observation conditions, satellite geometry, and the number of GNSS systems used (Kouba & Héroux, 2001).

In recent years, it has been demonstrated that the use of multi-GNSS significantly improves PPP performance. In particular, the GPS + GLONASS + Galileo combination:

- Reduces convergence time,
- Improves solution stability,
- Mitigates errors caused by weak satellite geometry (Cai & Gao, 2013).

These developments have led to the widespread adoption of the PPP method in both academic and operational applications.

2.4. PPP Software Ecosystem: Online, Academic, and Commercial Solutions

With the development of the PPP method, numerous software solutions have been developed to address different user profiles, including academic researchers, engineers, public institutions, and industrial users. These software packages can generally be categorized into three main groups: online services, academic/open-source software, and commercial software.

2.4.1. Online PPP Services

Online PPP services allow users to upload observation files in The Receiver Independent Exchange Format Version.xx (RINEX V.xx) format and obtain positioning solutions without requiring any local software installation. These services are typically operated by public or research institutions and are widely used in academic studies. Commonly used online PPP services are presented in Table 2, while their advantages and disadvantages are given in Table 3.

Table 2. Comparison of Commonly Used Online PPP Services

Service	Developer Institution	GNSS Support	Processing Type	Key Feature
CSRS-PPP	NRCan	GPS, GLONASS, Galileo	Static & Kinematic	PPP-AR support
APPS	NASA JPL	GPS (primarily)	Static	GIPSY-based, IGS products
MagicGNSS	GMV	Multi-GNSS	Static & Kinematic	Multi-constellation support
Trimble RTX	Trimble	Multi-GNSS	Post-processing	Commercial infrastructure, high accuracy

Table 3. General Advantages and Disadvantages of Online PPP Services

Advantages	Disadvantages
No installation required	Limited access to model parameters
Standardized solution strategies	Most algorithms are closed-source
Can be used as a reference in academic studies	Limited customization capability

2.4.2. Academic and Research-Oriented Software

Academic software is generally developed by research institutions and provides a high level of control over modeling processes. The most commonly used academic PPP software in the literature is presented in Table 4.

Table 4. Commonly Used Academic PPP Software

Software	Developer Institution	License Type	Supported Method	Key Feature	Reference
GIPSY-OASIS / GipsyX	JPL (NASA)	Academic / limited access	PPP	One of the first operational PPP implementations, high accuracy	Zumberge et al. (1997)
Bernese GNSS Software	University of Bern	Licensed	Relative + PPP	High-precision geodetic analyses	Dach et al. (2015)
gLAB	ESA	Open source	PPP	Education and research-oriented	Hernández-Pajares et al. (2012)
RTKLIB	Open-source community	Open source	RTK + PPP	Flexible structure, large user base	Takasu & Yasuda (2009)
PRIDE PPP-AR	Wuhan University	Academic	PPP-AR	Focused on ambiguity resolution (AR)	Geng et al. (2019)

The main advantages of these software packages include enabling user control over modeling parameters, facilitating the development and testing of new algorithms, and providing a transparent and traceable research environment from an academic perspective. On the other hand, their relatively complex structure and the requirement for advanced technical knowledge and expertise for effective use are considered their primary disadvantages.

2.4.3. Commercial PPP Software and Services

Commercial GNSS manufacturers have developed Precise Point Positioning (PPP) solutions that operate in integration with their proprietary hardware ecosystems. These solutions are generally supported by real-time correction services and aim to provide high-accuracy and uninterrupted positioning performance. Widely used commercial PPP software is presented in Table 5.

Table 5.Commonly Used Commercial PPP Solutions

Software / Service	Developer	Key Features
Trimble CenterPoint RTX	Trimble	Real-time and post-processing PPP solutions, global correction support
Leica SmartLink	Leica Geosystems	Full integration with Leica GNSS receivers, real-time high accuracy
TopNET Live	Topcon	Global correction service, real-time PPP support
TerraStar PPP	Hexagon / NovAtel	Satellite-based correction services, wide coverage
StarFire	John Deere	Global PPP corrections, widely used in precision agriculture

Commercial PPP solutions are widely adopted, particularly due to the operational advantages they provide in real-time applications. A general evaluation of commercial software is presented in Table 6.

Table 6. General Evaluation of Commercial PPP Software

Advantages	Disadvantages
Real-time high-accuracy positioning	High licensing and subscription costs
User-friendly interfaces	Closed-source algorithmic structure
Integrated correction services	Dependency on the manufacturer ecosystem
Technical support and regular updates	Limited flexibility for research purposes
Ease of operational use	Limited parameter customization options

The closed structure and cost-oriented limitations of these solutions lead to the preference for alternative PPP platforms in academic research.

In general, the PPP software ecosystem can be classified into three main groups: online services, academic research software, and commercial solutions. While online services offer significant advantages in terms of accessibility and ease of use, academic software provides a more suitable environment for researchers in terms of algorithmic transparency and methodological flexibility. Commercial software, on the other hand, stands out with operational stability, real-time service support, and optimizations for industrial applications. Within this diversity, CSRS-PPP occupies a remarkable position due to its free accessibility, multi-GNSS support, and widespread citation in the academic literature. In particular, examining the impact of Galileo integration on performance is scientifically important for both software evaluation and demonstrating the benefits of multi-GNSS systems.

3. CSRS-PPP SOFTWARE AND MULTI-GNSS (GALILEO) CONTRIBUTION

3.1. General Structure and Institutional Background of CSRS-PPP

CSRS-PPP (Canadian Spatial Reference System – Precise Point Positioning) is an online absolute positioning (PPP) service developed by the Geodetic Survey Division of Natural Resources Canada (NRCan). The system processes GNSS observation files in RINEX format uploaded by users and produces high-accuracy coordinates, velocity, and tropospheric parameters (Figure 1).

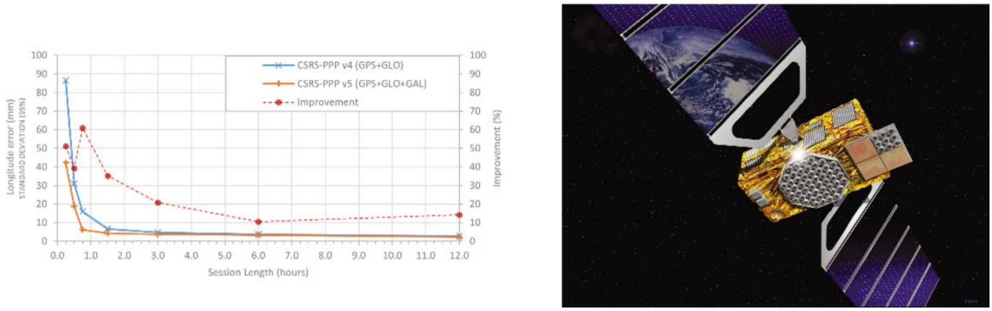

Figure 5. Example of the CSRS-PPP service interface (NRCan), including Galileo and other GNSS satellite constellations

The CSRS-PPP service primarily uses precise orbit and clock products published by the International GNSS Service (IGS). Different accuracy and latency options are provided using IGS final, rapid, and ultra-rapid products (Kouba & Héroux, 2001; IGS, 2023).

The mathematical foundation of the PPP method is based on modeling satellite orbit and clock errors using external precise products with undifferenced observations rather than double-differencing (Zumberge et al., 1997). This approach provides significant advantages, particularly in global and remote regions, as it does not require a reference station network.

3.2. Processing Strategy and Fundamental Components of Modeling

The processing model of CSRS-PPP includes the following fundamental components:

- Precise satellite orbits and clock corrections (IGS products)
- Phase wind-up correction
- Antenna phase center variations (IGS ANTEX)
- Elimination of ionospheric delay using dual-frequency combinations
- Parametric modeling of tropospheric delay (typically ZTD estimation)

- Tidal and loading effects (solid Earth, ocean loading, etc.)

The classical mathematical formulation of the PPP model can be summarized as follows:

$$L = \rho + c(dt_r - dt_s) + T - I + \lambda N + \varepsilon \quad (1)$$

Here, geometric distance (ρ), tropospheric delay (T), ionospheric delay (I), phase ambiguity (N), and receiver/satellite clock errors are the main parameters (Hofmann-Wellenhof et al., 2008; Teunissen & Montenbruck, 2017).

3.3. Multi-GNSS Support of CSRS-PPP

Initially designed to process only GPS observations, the system has been expanded over time to support multi-GNSS, and current versions include GPS, GLONASS, Galileo, and BeiDou systems. In this context, the integration of Galileo, particularly through the use of E1/E5a frequency combinations, contributes to both improved positioning accuracy and reduced convergence time. In addition, the modern signal structure and highly stable atomic clocks of Galileo satellites enable more reliable clock error modeling in PPP solutions (EUSPA, 2023).

The main contributions of multi-GNSS usage are as follows:

- Increased number of satellites → improved geometric distribution (lower PDOP)
- Improvement in ambiguity resolution
- Reduction in convergence time
- Increased solution continuity in urban and partially obstructed environments

Studies by Li et al. (2015) and Cai & Gao (2013) have demonstrated that multi-GNSS PPP converges faster and produces more stable results compared to single-system solutions.

3.4. Impact of Galileo Integration on PPP Performance

The contributions of the Galileo system to PPP solutions have been extensively examined in the literature within the framework of various performance metrics. These contributions are generally evaluated under three main aspects: convergence time, vertical accuracy, and clock modeling stability. These effects and related findings in the literature are summarized in Table 7.

Table 17. Contributions of Galileo to PPP Performance

Category	Effect	Description	Reference
Convergence time	20–40% reduction	Faster convergence due to increased satellite number and observation diversity	Li et al., 2015
Vertical accuracy	Improvement	Better separation of ZTD and height parameters due to enhanced satellite geometry	Teunissen & Montenbruck, 2017
Clock stability	Increase	Stable long-term clock modeling enabled by passive hydrogen maser clocks	EUSPA, 2023

As shown in Table 7, the integration of Galileo into PPP solutions provides significant contributions in reducing convergence time, improving vertical accuracy, and enhancing clock modeling performance.

The CSRS-PPP software, due to its high-accuracy positioning capabilities, has a wide range of applications in both scientific research and engineering practices. In this context, the system is widely used in geodetic deformation analysis, tectonic velocity estimation, GNSS meteorology (especially ZTD estimation), sea level monitoring, and engineering surveying applications. Furthermore, its direct compatibility with the Canadian National Reference Frame strengthens its position, particularly in North America; however, it can also be effectively used on a global scale through ITRF-based solutions (NRCan, 2023). With these features, CSRS-PPP stands out as a reliable and widely preferred PPP platform in both academic and applied studies.

4. APPLICATION

Within the scope of this study, an application was carried out to demonstrate the effect of multi-GNSS usage on PPP performance, the theoretical framework of which was presented in the previous sections. During the application phase, 24-hour RINEX data from three different stations (ANK2, IZMI, and KRS1), located within the borders of Türkiye and included in the International GNSS Service (IGS) network, were used. The spatial distribution of these stations across Türkiye is shown in Figure 2.

Figure 6. IGS stations used in the study

In the analyses, GNSS observations in RINEX format covering the first week of 2026 (days 1–7 of the year) were utilized. To investigate the impact of different satellite system combinations on PPP solution performance, five distinct scenarios were defined: GPS-only (G), GLONASS-only (R), GPS+GLONASS (G+R), GPS+Galileo (G+E), and GPS+GLONASS+Galileo (G+R+E). Each dataset was processed independently using the CSRS-PPP online PPP service, and the resulting coordinates were compared with the known (reference) coordinates of the corresponding stations.

5. RESULTS

Within this scope, the effects of different GNSS combinations on positioning accuracy were evaluated in terms of coordinate differences, component-wise errors, and overall solution performance. Thus, it was aimed to quantitatively reveal the contribution of the Galileo system within multi-GNSS integration. The accuracy of IGS weekly combined coordinate solutions is approximately 1–2 mm for horizontal components and 4–6 mm for the vertical (Up) component. The RMSE values obtained in this study are presented in Table 7.

Table 7. RMSE values computed based on the used IGS stations and satellite configurations

Station	Satellite Config	n (mm)	e (mm)	u (mm)
ANK2	E	1	2	6
	G	1	2	5
	R	2	2	4
	G+E	1	1	5
	G+R	2	1	4
	G+R+E	2	1	5
IZMI	E	4	2	9
	G	4	2	5
	R	6	6	13
	G+E	4	2	5
	G+R	4	2	5
	G+R+E	4	2	6
KRS1	E	2	2	5
	G	2	4	6
	R	2	2	11
	G+E	2	2	5
	G+R	2	3	6
	G+R+E	2	2	5

When Table 7 is examined, it is observed that for the ANK2 station, the results for the horizontal components (north and east) remain within the expected accuracy limits, and no significant differences are observed among different satellite configurations. However, in terms of the vertical component, a noticeable degradation in accuracy is observed when only the Galileo (E) satellite configuration is used.

For the IZMI station, the analyses indicate that the lowest accuracy is obtained with the GLONASS (R) satellite configuration. In particular, for the vertical component, the highest error values occur in the R configuration, followed by the Galileo (E) configuration.

For the KRS1 station, it is similarly observed that the lowest vertical accuracy is achieved with the GLONASS (R) configuration, while comparable accuracy levels are obtained among the other satellite combinations.

Based on the evaluations conducted using the selected IGS stations, the results indicate that the Galileo system does not provide a significant contribution to positioning accuracy within the scope of this study. However, in order to obtain more comprehensive and reliable results, it is recommended to expand the dataset and increase the number of stations to further investigate the contribution of Galileo in more detail.

REFERENCES

Cai, C., & Gao, Y. (2013). Modeling and assessment of combined GPS/GLONASS precise point positioning. *GPS Solutions, 17*(2), 223–236. https://doi.org/10.1007/s10291-012-0273-9

Dach, R., Lutz, S., Walser, P., & Fridez, P. (2015). *Bernese GNSS Software Version 5.2 User Manual*. University of Bern.

European GNSS Agency. (2019). *Galileo Open Service – Signal in Space Interface Control Document (OS SIS ICD)*. Publications Office of the European Union.

European Union Agency for the Space Programme. (2023). *Galileo performance and service status report*. European Union.

Geng, J., Shi, C., Ge, M., et al. (2019). PRIDE PPP-AR: An open-source software for precise point positioning with ambiguity resolution. *GPS Solutions*.

Hernández-Pajares, M., Juan, J. M., Sanz, J., et al. (2012). The ESA GNSS laboratory tool (gLAB). *GPS Solutions*.

Hofmann-Wellenhof, B., Lichtenegger, H., & Wasle, E. (2008). *GNSS – Global Navigation Satellite Systems: GPS, GLONASS, Galileo & more*. Springer.

International GNSS Service. (2023). *IGS products and data*. https://igs.org

Kouba, J., & Héroux, P. (2001). Precise point positioning using IGS orbit and clock products. *GPS Solutions, 5*(2), 12–28. https://doi.org/10.1007/PL00012883

Li, X., Zhang, X., Ren, X., Fritsche, M., Wickert, J., & Schuh, H. (2015). Precise positioning with current multi-constellation GNSS systems: GPS, GLONASS, Galileo and BeiDou. *Scientific Reports, 5*, 8328. https://doi.org/10.1038/srep08328

Misra, P., & Enge, P. (2011). *Global positioning system: Signals, measurements, and performance* (2nd ed.). Ganga-Jamuna Press.

Natural Resources Canada. (2023). *CSRS-PPP online positioning service*. Government of Canada. https://webapp.csrs-scrs.nrcan-rncan.gc.ca

Takasu, T., & Yasuda, A. (2009). Development of the low-cost RTK-GPS receiver with an open source program package RTKLIB.

Teunissen, P. J. G., & Montenbruck, O. (Eds.). (2017). *Springer handbook of global navigation satellite systems*. Springer.

Zumberge, J. F., Heflin, M. B., Jefferson, D. C., Watkins, M. M., & Webb, F. H. (1997). Precise point positioning for the efficient and robust analysis of GPS data. *Journal of Geophysical Research: Solid Earth, 102*(B3), 5005–5017. https://doi.org/10.1029/96JB03860

CHAPTER 5

Dynamic Moisture Ratio Prediction Using Regression Models in The Pressurized Hot-Air Bobbin Drying Process

Halil Nusret Buluş[1] & Uğur Akyol [2] & Nadir Subaşı[3] & Soner Çelen[4]

1. Introduction

Drying is a fundamental operation widely utilized across a broad range of industrial applications, spanning from food preservation to timber processing. While the term "drying" is generally defined in the broader literature as the process of removing volatile components from a product via thermal energy, it is preferred in textile technology in a wider sense to denote the removal of water (dewatering) from the material structure (MEB, 2011; Galappi et al., 2017). Since the manufacturing process of textile goods inherently involves consecutive wet treatments such as dyeing, washing, and finishing, drying the material at various stages is mandatory. The textile industry stands as one of the most energy-intensive manufacturing sectors globally, with wet processes and the subsequent drying stages accounting for a major portion of the total thermal energy consumption within factories (Galappi et al., 2017). It is well established that approximately 25% of the total energy utilized in wet processing is consumed solely during the drying phase, making this stage a frequent bottleneck in production lines (Carr et al., 2006). Therefore, the optimization of drying conditions and parameters is of critical importance for lowering process costs, achieving time savings, and mitigating environmental impacts (Akyol et al., 2012; Galappi et al., 2017).

In the textile industry, the drying of products is categorized into "pre-drying" and "main drying" based on the mechanism of moisture removal. Pre-drying, carried out through mechanical methods, is highly economical in terms of initial investment and operating costs; however, the moisture on the product cannot be entirely eliminated using these techniques. Hence, to reduce thermal costs, wet

[1] Asist. Prof., Tekirdağ Namık Kemal University, Çorlu Faculty of Engineering, ORCID: 0000-0003-1844-6484

[2] Prof. Dr., Tekirdağ Namık Kemal University, Çorlu Faculty of Engineering, ORCID: 0000-0003-3707-0929

[3] Dr., Kırklareli University, Kırklareli Vocational School of Technical Sciences, ORCID: 0000-0002-5657-9002

[4] Prof. Dr., Tekirdağ Namık Kemal University, Çorlu Faculty of Engineering, ORCID: 0000-0001-5254-4411

textile products are initially subjected to mechanical pre-drying before being transferred to main thermal drying systems (Oğulata et al., 1999). For instance, after the yarn bobbin dyeing process, the dry-basis moisture content of bobbins containing high amounts of water can be reduced to levels of 40% to 60% via mechanical pre-drying (Galappi et al., 2017).

Before becoming commercialized, textile bobbins must complete several successive processing steps. Since mechanical dewatering cannot extract the water chemically bound to the textile fibers, thermal (main) drying is almost always required to evaporate the remaining moisture. Various thermal configurations are utilized for this purpose, such as convective ovens, heated cylinders, and radio-frequency (RF) systems. Convective dryers remain the industry standard for both carpets and textiles. Although hot air is typically used as the primary medium in these convective units, superheated steam can be substituted if high temperatures can be achieved without inducing structural damage to the material (Lee et al., 2006).

The heat transfer mechanisms applied in industrial main drying operations are physically based on four fundamental principles: convection, contact (conduction), infrared (IR) irradiation, and high-frequency (microwave) drying (MEB, 2011). Among convective dryers, high-cost stenters are prominent due to their ability to provide width and length adjustment, along with conveyor belt dryers offering tensionless operation, and hot-flue systems providing vertical movement via delivery rollers. In the second method, contact drying, the principle of the wet product directly contacting heated cylinders (cylinder dryers) is applied. In IR drying, which provides heat transfer via electromagnetic waves, efficiency depends on the degree of absorption, which varies according to the type of fiber, product form, and color (whether it is dyed light or dark). Finally, the high-frequency drying method is based on the principle of water molecules passing between two capacitor plates to form dipoles, offering the advantage of providing uniform drying that penetrates to the core of dense textile structures such as bobbins or tops (MEB, 2011).

In industrial bobbin drying processes, two different airflow modes can be applied (Galappi et al., 2017):

1. Inside-Out: Hot air enters through the perforated smooth spindle inside the bobbin, passes through the yarn layers, and moves toward the outer surface.

2. Outside-In: Hot air is blown onto the outer surface of the bobbin, passes through the yarn layers, and is discharged through the internal perforated spindle.

During the design, selection, and management of drying units and operational modes, all drying principles must be taken into account, and variables between the physical properties of the material and its end-use area must be optimized (Tan, 2001). To maintain product quality, shorten drying times, and minimize energy consumption, the optimization of drying processes using mathematical models has been one of the core focal points of textile engineering research for many years. These developed models are generally based on solving simultaneous heat and mass transfer equations between the drying medium and the textile structure, as well as determining parametric effects (Akyol et al., 2019).

2. Scope and Objective of the Research

The primary objective of this study is to model the time-dependent, complex heat and mass transfer mechanisms during a thermal treatment process conducted under a constant pressure of 3 bar and a target temperature of 100 °C by utilizing Regression Models. In the developed regression models, process parameters including time (seconds), temperature values at various locations within the dryer (Bobbin 1 to Bobbin 7), chamber inlet, outlet, and internal temperatures, as well as separator and fan outlet temperatures, were defined as input features. The dependent variable designated for prediction is the moisture ratio of the dried material. Through these regression models, inter-layer temperature distributions and dynamic mass losses, which are otherwise time-consuming and highly costly to determine experimentally, can be predicted with minimum error and high precision. Consequently, this study aims to enable real-time monitoring of system dynamics in industrial bobbin processing, optimize energy efficiency, and pre-simulate system behavior without the need for experimental procedures.

3. Experimental Setup and Methodology

The experimental tests in this study were conducted using a custom-built textile drying test rig designed and constructed at Tekirdağ Namık Kemal University.

The experimental setup was engineered to accurately replicate the geometry and operating conditions of a single yarn carrier spindle within an industrial drying chamber. The system essentially comprises the following main components:

- Air Supply Unit: A centrifugal fan equipped with a frequency converter to deliver the required air flow rate.
- Heating Unit: PID-controlled electrical resistances capable of precisely adjusting the air temperature.
- Test Chamber: An insulated, cylindrical vertical vessel wherein the yarn bobbins are positioned.
- Measurement Sensors: An orifice plate for measuring the air flow rate, high-precision psychrometric sensors for measuring the temperature and relative humidity at the bobbin inlets and outlets, and load cells for monitoring mass changes.

Experimental trials were conducted on a hollow, cylindrical cotton bobbin specimen within the pressurized hot-air drying apparatus depicted in Figure 1. The drying operation was carried out at a steady temperature of 100 °C under an effective air pressure of 3 bar. Within this configuration, a centrifugal fan guided ambient air into a 25 kW electrical heating unit, while a 15 kW compressor sustained the required system pressure. The heated air exiting the unit was then driven radially outward from the core through the bobbins, which were arranged in a four-section carrier framework (Akyol et al., 2012).

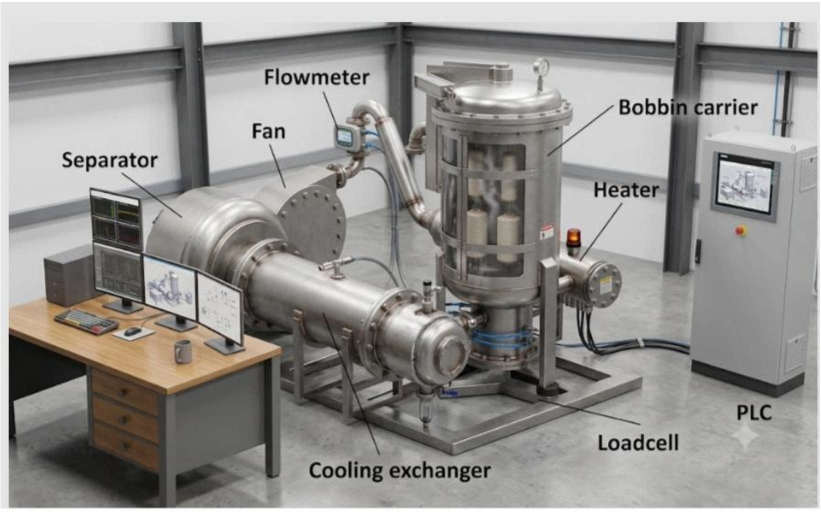

Figure 1. Experimental setup of the pressurized hot-air bobbin dryer

The air discharged from the carrier sequentially enters a cooling heat exchanger to reduce its relative humidity and a separator to remove water droplets before returning to the fan. The process is monitored via software with the aid of sensors and a load cell with an accuracy of ±1 g, and is managed by an automatic control system.

Pre-Experimental Procedures:

- Saturation and Dewatering: To achieve full moisture uptake, the bobbin specimens were submerged in a water bath for a 12-hour period, followed by 30 minutes of gravity drainage on a wire mesh.
- Pre-Drying: Prior to thermal drying, pressurized air was forced through the bobbins for 5 minutes without heating in order to remove excess water.

Following these procedures, the samples were positioned onto the carrier and dried under the specified conditions.

Modeling of Drying Curves

The dimensionless moisture ratio (*MR*) is expressed as in Equation (1); where M_0 represents the initial moisture mass prior to drying, and M_t denotes the average moisture mass at time t (Çay, 2017).

$$MR = \frac{M_t}{M_0} \quad (1)$$

Regression Methods Used in Modeling

The fundamental goal of regression modeling is to quantify the association between a dependent variable and one or more explanatory variables using a mathematical function. Predominantly applied to numerical data, regression techniques are broadly categorized into simple and multiple frameworks. Multiple regression seeks to determine the combined influence of a set of independent predictors on a solitary outcome variable, thereby building directly upon the foundations of simple regression, which evaluates only a single predictor (Kleinbaum et al. 1988).

Linear Regression:

As a fundamental statistical technique, simple linear regression models the connection between a pair of variables: a response variable (Y) and a single predictor (X). Under the assumption of linearity, this relationship is mathematically formulated as eq.2:

$$Y = \beta 0 + \beta 1X + \varepsilon \quad (2)$$

Here, $\beta 0$ denotes the y-intercept, $\beta 1$ represents the slope coefficient, and ε accounts for the stochastic error term. The primary goal of this method is to estimate the optimal parameters for $\beta 0$ and $\beta 1$ that yield the best-fitting straight line, which is accomplished by minimizing the sum of the squared residuals (Duque-Aldaz et al., 2024).

Ridge Regression:

In situations characterized by severe multicollinearity among predictor variables, ridge regression offers an effective framework for estimating coefficients within multiple linear regression models. This methodology was introduced specifically to address the instability and imprecision inherent in ordinary least squares estimators when independent variables exhibit high inter-correlations. By constructing a specialized "ridge regression estimator" (RR), this approach yields more reliable parameter estimates; this increased precision is achieved because the variance and mean squared error (MSE) of the RR framework are frequently lower than those of conventional least squares estimators (Barriga et al., 2022).

Random Forest Regressor:

The random forest (RF) approach constructs an ensemble of multiple decision trees, utilizing bootstrap resamples drawn from the primary training dataset to independently train each individual tree for both classification and regression tasks. Subsequently, this collection of trees determines the classification of a given instance through a majority voting mechanism based on the most frequent outcome. Because every single tree operates as an isolated classifier, the RF framework serves as a prominent paradigm of ensemble-based classification

(Breiman, 2001). Distinct from standard decision trees (DT), the final class designation of a dataset is established by the collective consensus of the generated forest ensemble (Berhane et al., 2018). Furthermore, because the integration of bootstrapping and ensemble methods effectively mitigates the overfitting tendencies inherent in conventional DTs, the RF algorithm bypasses the need for a pruning phase entirely (Ali et al., 2012). Consequently, RF is widely recognized for its robust predictive performance and its high resilience against data noise (Breiman, 2001; Rodriguez-Galiano et al., 2012; Sağlam, 2022).

Gradient Boosting Regressor:

Gradient boosting represents a powerful machine learning paradigm applied to both classification and regression tasks, which constructs a predictive framework composed of an ensemble of weak base learners, commonly realized as decision trees. The core algorithm iteratively refines its objective function by sequentially integrating subsequent models designed specifically to minimize residual errors (Deb et al., 2025).

Support Vector Regression - SVR:

As a widely adopted machine learning approach, Support Vector Regression (SVR) operates non-parametrically by leveraging the power of kernel functions. The underlying theoretical framework involves mapping the original data into a higher-dimensional space via a kernel transformation; subsequently, the method solves for an optimal function (f(x) or a decision hyperplane) that minimizes deviation from the actual targets within a specified threshold—the epsilon value (ε)—while simultaneously ensuring the function remains as flat as possible (Taheri et al., 2021).

4.Results and Discussion

It has been confirmed that the major part of the bobbin drying process occurs during the "Falling Drying Rate Period" (Figure 2). As the porous structure of the fibers deepens, the internal diffusion resistance escalates, which in turn hinders the transport of moisture to the outer surface. It has been calculated that a gradual reduction of the air flow rate during this specific stage yields a significant optimization in the electrical energy consumed by the fan.

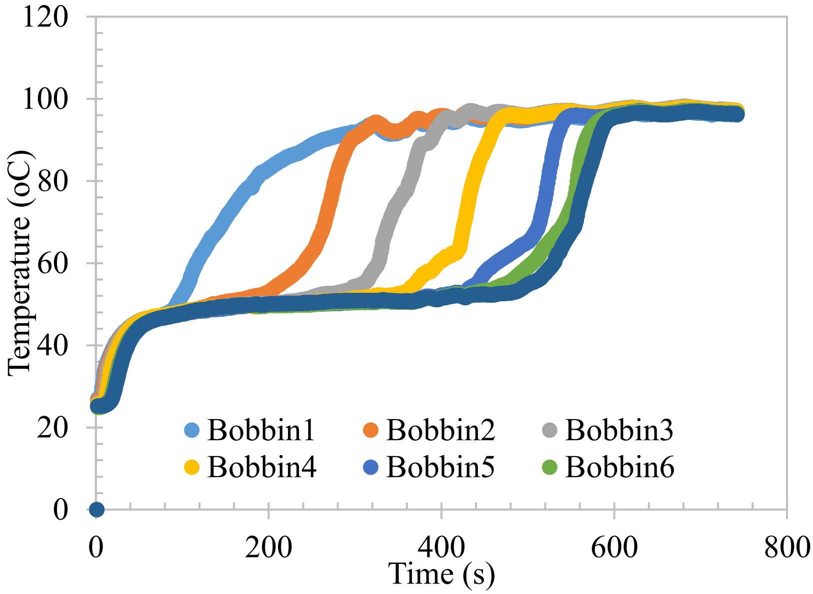

Figure 2. Temperature variations of the bobbins

All bobbins initiate the process at approximately 25 °C (ambient temperature). Within the first 50 seconds, the temperature of all bobbins escalates rapidly, reaching a common plateau/transition point in the 45 °C - 50 °C band. This phenomenon indicates that the initial heat input into the system exerts a uniform effect. The most remarkable feature of the graph is the temporal shift between the times required for the bobbins to reach high temperatures (approximately 95 °C - 100 °C):

- Bobbin 1 (Light Blue): Continues to heat up rapidly from approximately the 100th second onward with almost no stagnation, reaching the steady-state temperature earliest (at around the 300th second).

- From Bobbin 2 to Bobbin 6: As the bobbin index increases, the dwelling (plateau) duration around 50 °C elongates. For instance, Bobbin 6 (Green) remains constant at around 50 °C for approximately 500-550 seconds before exhibiting a sharp inclination.

This type of profile typically indicates the distance from the heat source or the propagation direction of the fluid (e.g., the hot water/steam in the dyeing or drying bath). While Bobbin 1, which is closest to the heat source, heats up immediately, energy is progressively transferred to the other bobbins in sequential order.

After approximately the 600th second, all bobbins reach thermal equilibrium (steady-state). All curves stabilize between approximately 96 °C and 98 °C. This proves that the maximum operating temperature of the system is within these limits and confirms that full temperature homogeneity is achieved within the system at the final stage.

Model Analysis Results

The dataset utilized in the experiments comprises a total of 741 observations (n = 741) obtained from an industrial drying process. While the raw dataset originally contained 179 columns, its dimensionality was reduced to 14 columns (13 independent variables/features and 1 dependent variable) after eliminating entirely empty columns.

Process parameters including time (seconds), temperature values at various locations within the dryer (Bobbin 1 to Bobbin 7), chamber inlet, internal, and outlet temperatures, as well as separator and fan outlet temperatures, were specified as input features.The dependent variable designated for prediction is the moisture ratio of the dried material. The moisture ratio within the data varies within the range of [-0.1357, 1.0000], with a mean value of 0.4720.

To prevent sensor data with different units and ranges (temperature, time, etc.) from adversely affecting model training, standardization (Z-score normalization) was applied to the independent variables. Through this procedure, all features were transformed to have a mean of 0 and a standard deviation of 1.

The 5 constructed regression models were trained on 80% of the data and subsequently tested on the remaining 20%. The obtained performance metric values are provided in Table 1.

Table 1. Performance rates of 5 regression models

Model	**MSE**	**MAE**	**RMSE**	**R^2**
Linear Regression	0.000576	0.015295	0.022469	0.992464
Ridge Regression	0.000576	0.015445	0.022463	0.992470
Random Forest	0.000305	0.005436	0.013241	0.996068
Gradient Boosting	0.000338	0.006892	0.014677	0.995662
SVR	0.003858	0.055036	0.061848	0.948388

The mean and standard deviation values of the performance metrics for these developed regression models were also calculated. To perform these computations, the models were subjected to a 10-fold cross-validation procedure. These values are visualized in Figure 3.

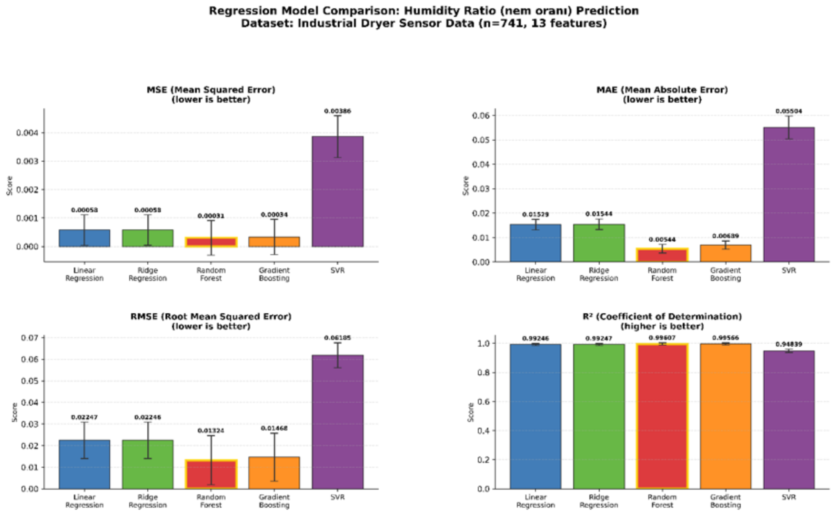

Figure 3. Evaluation of various regression algorithms based on error metrics (MSE, MAE, RMSE) and the R^2 score for moisture ratio estimation

Upon analyzing these computations, it was observed that among the constructed models, the Random Forest model yielded the best results. Pairwise comparisons of this model with the alternative models were conducted. To validate the performance of the algorithm demonstrating the highest accuracy, this model was subjected to a pairwise Wilcoxon Signed-Rank Test against the other four models. The test results confirmed that the optimal model is statistically superior to each of the alternative models ($p < 0.01$). The pairwise comparison results of the models are provided in Table 2

Tablo 2. Wilcoxon Signed-Rank Tests (vs. best: Random Forest on R^2)

F vs Linear Regression	stat=0.000	p=0.0020
RF vs Ridge Regression	stat=0.000	p=0.0020
RF vs Gradient Boosting	stat=1.000	p=0.0039
RF vs SVR	stat=0.000	p=0.0020

Furthermore, the R^2 scores of all models within the 10-fold cross-validation procedure were analyzed using the non-parametric Friedman Test. The test results indicated a statistically significant difference among the performances of the models (χ^2=35.04, p=4.56×10^{-7}). This extremely low p-value ($p < 0.05$) demonstrates that the performance disparity among the models is not coincidental.

According to the experimental results summarized in the graph, the linear models (Linear Regression, Ridge) and SVR failed to capture the complex relationships within the data, exhibiting high error rates (RMSE ~0.129). In contrast, tree-based ensemble learning algorithms dramatically enhanced the prediction performance. In particular, the Random Forest algorithm was identified as the most robust and accurate model for predicting the moisture ratio from industrial dryer sensor data.

5. Conclusions

In this study, the yarn bobbin drying process, which is characterized by high energy consumption in the industry, was analyzed using experimental data and regression models under conditions of 3 bar constant pressure and a 100 °C target temperature. Five different machine learning algorithms were evaluated utilizing a dataset comprising 741 observations derived from 13 separate sensor outputs (temperature and time) within the dryer system. The primary findings obtained from the research are summarized as follows:

- Most Successful Prediction Model: While linear models and SVR remained inadequate in resolving complex physical relationships, the tree-based Random Forest (Rassal Orman) algorithm emerged as the most precise and stable model for predicting the moisture ratio, yielding an explanatory power of R^2 = 0.996068 and the lowest error values (MSE: 0.000305, MAE: 0.005436, RMSE: 0.013241).

- Statistical Significance: The performance superiority of the Random Forest model over all alternative models was statistically validated via the pairwise Wilcoxon Signed-Rank Test ($p < 0.01$).

- Industrial Contribution and Energy Optimization: Due to the increasing internal diffusion resistance as the porous structure of the fibers deepens, it was calculated that a gradual reduction of the air flow rate during the falling drying rate period yields a significant optimization in the electrical energy consumption of the fan.

The regression models developed herein can predict inter-layer temperature distributions and dynamic mass losses—which are otherwise highly expensive and time-consuming to measure experimentally in industrial bobbin processing operations—with elevated precision. Consequently, this study provides a robust framework that enables real-time monitoring of system dynamics in textile enterprises, optimization of energy efficiency, and pre-simulation of system behaviors without the need for physical experimentation.

REFERENCE

Akyol, U., Kahveci, K., & Cihan, A. (2012). Effect of drying air pressure on cotton bobbin drying process. Journal of International Scientific Publications: Materials, Methods & Technologies, 6(2).

Akyol, U., Karakoca, A., Shaliyev, R., Akal, D., & Uzuneanu, K. (2019). Investigation of the effect of drying air temperature on drying time by numerical method for textile bobbins. IOP Conference Series: Materials Science and Engineering, 595(1), 012056. https://doi.org/10.1088/1757-899X/595/1/012056.

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. International Journal of Computer Science, 9, 272–278.

Barriga, R., Romero, M., Nettleton, D. and Hassan, H., (2022). Advanced data modeling for industrial drying machine energy optimization. J Supercomput 78, 16820–16840 https://doi.org/10.1007/s11227-022-04498-0.

Berhane, T., Lane, C., Wu, Q., Autrey, B., Anenkhonov, O., Chepinoga, V., & Liu, H. (2018). Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. Remote Sensing, 10, 580. https://doi.org/10.3390/rs10040580

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

Carr, W. W., Lee, H. S., & Ok, H. (2006). Drying of textile products, 34. DOI:10.1201/9781420017618.ch34.

Çay, A., 2017. Modelling Of The Drying Behaviour Of Regenerated Cellulosic Fabrics, Tekstil Ve Konfeksiyon 27(4).

Deb, A., Pache, R., Das Ghatak, M., Kumar, R., Mandal, S.K., Computational and experimental analysis of a biomass-fueled natural convection dryer and prediction of thermodynamic parameters using machine learning algorithms. J Therm Anal Calorim 150, 16139–16158 (2025). https://doi.org/10.1007/s10973-025-14834-0

Duque-Aldaz, F.J., Ronny, E., Moreno, H., Leonor, Aspiazu, A.Z., Choque, F.C., Química, I., & Desarrollo Prediction of moisture content in the cocoa drying process by simple linear regression. INQUIDE, 6 (2), pp.20-30, 2024.

Galoppi, G., Ferrari, L., Ferrara, G., & Carnevale, E. A. (2017). Experimental investigation on industrial drying process of cotton yarn bobbins: energy consumption and drying time. Energy Procedia, 126, 361-368. https://doi.org/10.1016/j.egypro.2017.08.187

Kleinbaum, D.C.; Kupper, L.C.; Muller, K.E. Applied Regression Analysis and Other Multivariable Methods; PWS-KENT Publishing Co.: Boston, MA, USA, 1988; ISBN 0-87150-123-6.

Lee HS, Carr WW, Ok H, Industrial drying, CRC Press, 2006.

MEB, 2011. Tekstil Teknolojisi Kurutma ve Egalize, Ankara.

Oğulata R.T, Kadem F. D, & Koç E (1999). Tekstilde Kurutma Yöntem ve Makineleri. IV. Ulusal Tesisat Mühendisliği Kongresi ve Sergisi, 803-809, İzmir.

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & RigolSanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Sağlam, C., & Çetin, N. (2022). Machine learning algorithms to estimate drying characteristics of apples slices dried with different methods. Journal of Food Processing and Preservation, 46, e16496. https://doi.org/10.1111/jfpp.16496

Taheri,S., Brodie, G., Gupta, D., Optimised ANN and SVR models for online prediction of moisture content and temperature of lentil seeds in a microwave fluidised bed dryer, Computers and Electronics in Agriculture, Vol 182, 2021, 106003, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2021.106003.

Tan, G. (2001). Gıda endüstrisinde kullanılan kurutma sistemleri. Gıda Mühendisliği Dergisi, 5(11), 20-33.

CHAPTER 6

Text Classification with Convolutional Neural Networks: An Experimental Study on the AG News Dataset

Teyfik Giray[1] & Zülfikar Aslan[2]

Introduction

Natural Language Processing (NLP) is an interdisciplinary research field that aims to automatically analyze, understand, and interpret text data. Text classification is one of the most fundamental problems in NLP and forms the basis of many applications such as news categorization, sentiment analysis, spam detection, and information retrieval systems. This problem aims to assign a given text to one of several predefined categories, enabling the efficient processing of large-scale text data.

Early approaches to text classification problems relied heavily on statistical representation methods such as Bag-of-Words and TF-IDF. However, these methods have limited representational power due to their disregard for word order and contextual relationships. To overcome these limitations, distributed word representations were developed, enabling the modeling of semantic relationships by representing words in a continuous vector space. In this context, methods such as Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) have significantly improved the quality of text representations and have become fundamental components of modern NLP systems.

With the development of deep learning approaches, models capable of automatic feature extraction from text data have made significant progress in the field of text classification. In particular, Convolutional Neural Networks (CNNs) have emerged as an effective approach due to their ability to capture local linguistic patterns (n-gram structures) within text. The TextCNN model proposed by (Kim ,2014) achieved high performance in text classification tasks using different filter sizes. Following this, (Kalchbrenner et al. ,2014) proposed deeper structures using a dynamic k-max pooling mechanism; and (Zhang et al., 2016)

[1] Gaziantep Islami Science and Technology University, Computer Engineering
ORCID number:0009-0007-8483-4529
[2] Assoc. Prof. Dr., Gaziantep Islami Science and Technology University, Computer Engineering, ORCID number: 0000-0002-2706-5715

obtained language-independent representations using character-level CNNs. Furthermore, (Johnson & Zhang, 2015) demonstrated the impact of CNN-based approaches that directly model word order on performance. These studies show that CNN architectures offer powerful and computationally efficient solutions to text classification problems.

In recent years, the Transformer architecture has created a significant paradigm shift in contextual representation learning. In particular, the BERT model (Devlin et al., 2019) has achieved state-of-the-art performance on many NLP tasks using bidirectional contextual information. Developed as a continuation of this approach, RoBERTa (Liu et al., 2019) has further enhanced performance with larger datasets and optimized training strategies. Similarly, XLNet (Yang et al., 2020) has been able to model contextual dependencies more flexibly using a permutation-based learning approach. Furthermore, studies such as DistilBERT (Sanh et al., 2020) and ULMFiT (Howard & Ruder, 2018) have made significant contributions to improving the balance between performance and computational cost.

Although Transformer-based models achieve high accuracy rates in text classification tasks, they have significant disadvantages such as high computational cost, large data requirements, and long training times. This highlights the need for lighter and more efficient models, especially for applications with limited resources. In this context, CNN-based approaches still offer a strong alternative with their low computational cost and rapid training processes.

In this context, although CNN-based text classification models have been extensively studied in the literature, there are limited studies that systematically and comparatively address the impact of different filter sizes, pooling strategies, and parallel architectural structures on performance. In particular, evaluating these architectural variations within a holistic framework on the same dataset is critical for a deeper understanding of model behavior.

This study comprehensively and systematically analyzes the effects of different CNN architectural configurations on text classification performance on the AG News dataset to address this gap. Within the scope of the study, different filter sizes (3, 4, and 5), pooling strategies, and multi-channel CNN structures are evaluated comparatively; the results are analyzed using accuracy, precision, recall, and F1-score metrics. In this respect, the study aims to contribute to the literature by presenting the effects of architectural design choices on performance in CNN-based text classification models from a holistic perspective.

Method and Materials

This section describes the dataset used in the study, the preprocessing steps, the model architecture, the training configuration, and the experimental design.

Dataset

The AG News dataset was created by reducing the Character-level Convolutional Networks for Text Classification study, presented by Xiang Zhang, Junbo Zhao, and Yann LeCun in 2015, to four classes (Zhang et al., 2016). The news corpus that forms the basis of the dataset was compiled by Antonio Gulli.

The dataset consists of four balanced classes: World, Sports, Business, and Sci/Tech, and includes 120,000 training and 7,600 test examples as standard. Each example consists of a news headline and a short description, and these two fields are combined and given as input to the model. The balanced class distribution supports the model in learning without creating class bias.

Data Preparation and Preprocessing

Pre-processing steps:

- The text has been converted to lowercase.
- The words were converted into strings through a tokenization process.
- The dictionary was created by indexing words.
- Padding has been applied to make all strings of a constant length.
- Class labels have been digitized.

Figure 1 shows that the dictionary size used after tokenization and padding is 50,000.

```
Vocabulary size: 50000
Training padded shape : (108000, 100)
Validation padded shape: (12000, 100)
Test padded shape      : (7600, 100)
```

Figure 7. The sizes of the arrays and the dictionary size obtained after data preprocessing.

Model Architecture

This study proposes a CNN-based deep learning architecture that aims to achieve rich representation learning by capturing local linguistic patterns (n-gram structures) at different scales for the text classification problem.

The model's input layer accepts fixed-length text strings of 100 words, converted into 128-dimensional dense vectors from a dictionary of up to 50,000 words. The embedding layer is designed to be trainable so that it can learn semantic relationships during the training process. The aim is to reduce overtraining by using the SpatialDropout1D layer. This layer randomly disables specific neurons, thereby increasing generalization capabilities.

In the architectural design, parallel Conv1D layers were used to capture linguistic patterns of varying lengths. In these layers, kernel sizes were set as 3, 4, and 5, and 128 filters and the ReLU activation function were used in each layer. In the convolution operations, the stride value was set to 1, and the padding method was selected as "valid". This structure allows for the simultaneous modeling of different n-gram lengths.

The output of each Conv1D layer was summarized using Global Max Pooling to extract the most dominant features. The resulting feature vectors were combined using concatenation to create a single representation vector. Dropout was applied to this combined representation to prevent overfitting.

Next, a Dense layer and ReLU activation function were used to learn higher-level features. An additional Dropout layer was added to enhance the model's generalization capabilities. For the output layer, a 4-neuron Dense layer with a Softmax activation function was used for the four-class classification problem.

The proposed model contains a total of 6,646,788 trainable parameters. The model's layer structure, output dimensions, and parameter distribution are presented in Figure 2.

Layer (type)	Output Shape	Param #	Connected to
input_layer (InputLayer)	(None, 100)	0	-
embedding (Embedding)	(None, 100, 128)	6,400,000	input_layer[0][0]
spatial_dropout1d (SpatialDropout1D)	(None, 100, 128)	0	embedding[0][0]
conv1d (Conv1D)	(None, 98, 128)	49,280	spatial_dropout1…
conv1d_1 (Conv1D)	(None, 97, 128)	65,664	spatial_dropout1…
conv1d_2 (Conv1D)	(None, 96, 128)	82,048	spatial_dropout1…
global_max_pooling… (GlobalMaxPooling1…	(None, 128)	0	conv1d[0][0]
global_max_pooling… (GlobalMaxPooling1…	(None, 128)	0	conv1d_1[0][0]
global_max_pooling… (GlobalMaxPooling1…	(None, 128)	0	conv1d_2[0][0]
concatenate (Concatenate)	(None, 384)	0	global_max_pooli… global_max_pooli… global_max_pooli…
dropout (Dropout)	(None, 384)	0	concatenate[0][0]
dense (Dense)	(None, 128)	49,280	dropout[0][0]
dropout_1 (Dropout)	(None, 128)	0	dense[0][0]
dense_1 (Dense)	(None, 4)	516	dropout_1[0][0]

Total params: 6,646,788 (25.36 MB)
Trainable params: 6,646,788 (25.36 MB)
Non-trainable params: 0 (0.00 B)

Figure 8. The proposed CNN model architecture includes layer structure, output dimensions, and parameter distribution.

Training Structure

The CNN model proposed in this study was trained using a supervised learning approach. The dataset used was divided into three subsets: training, validation, and testing. 108,000 samples were allocated for training, 12,000 samples for validation, and 7,600 samples for evaluating the model's performance.

The Adam optimization algorithm was used during model training. The learning rate was set to 0.001, and the beta1 and beta2 parameters of the Adam algorithm were used as 0.9 and 0.999, respectively. Categorical cross-entropy, which is suitable for multi-class classification problems, was chosen as the loss function.

During the training process, the batch size was set to 128 and the maximum number of epochs was set to 8. An Early Stopping mechanism was implemented to prevent overfitting. In this mechanism, validation loss was monitored, and if no improvement was observed over two epochs, the training process was terminated (patience = 2), and the best model weights were restored.

Experimental Design

This study systematically compares different CNN configurations to examine the impact of architectural design decisions on text classification performance. Within the experimental design, differences in model performance were analyzed in detail by varying convolution kernel sizes, pooling strategies, multi-channel structures, and basic hyperparameters.

In the first stage, CNN models with a single kernel size (kernel sizes = 3, 4, and 5) were trained and evaluated separately. This allowed observation of the effect of different n-gram lengths on classification success. In the second stage, multi-kernel architectures (kernel sizes = 3, 4, 5) using multiple kernel sizes were designed to enable simultaneous learning of linguistic patterns at different scales.

To evaluate the effectiveness of pooling strategies, Global Max Pooling and Local Max Pooling methods were compared. Global Max Pooling creates a general representation by selecting the highest value from each feature map, while Local Max Pooling aims to extract features from more localized regions.

In another phase of the study, a multi-channel CNN architecture was implemented. In this structure, features obtained from parallel convolution layers with different filter sizes were combined to obtain a richer and more multi-scale representation.

Within the scope of hyperparameter optimization, specific hyperparameters were tested at different value ranges to improve model performance.

In this context:

• Embedding size: 100 and 128

• Number of filters: 128 and 200

• Dropout rate: 0.3 and 0.5

various model configurations were created using the values and their performances were compared. All models were trained under the same data splitting strategy and training settings to ensure a fair comparison. Model performances were evaluated and comparative analyses were performed using accuracy, precision, recall, and F1-score metrics. The different CNN model configurations used in the study are summarized in Table 1.

Table 1. CNN Model Configurations

Model	**Kernel Size**	**Pooling**	**Structure**	**Explanation**
CNN_Filter_3	3	Global	Single	Single filter
CNN_Filter_4	4	Global	Single	Single filter
CNN_Filter_5	5	Global	Single	Single filter
CNN_Filter_3_4_5	3,4,5	Global	Multi-kernel	Multi-filter
CNN_GlobalMaxPooling	3,4,5	Global	Multi-kernel	Global pooling
CNN_LocalMaxPooling	3,4,5	Local	Multi-kernel	Local pooling
MultiChannel_CNN_3_4_5	3,4,5	Global	Multichannel	Advanced model
HPO Models	Değişken	Global	Tuned	Hyperparameter optimization

Findings

Training Performance

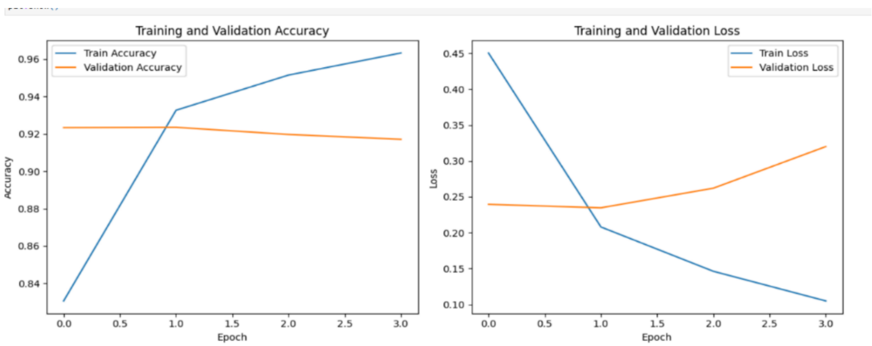

Figure 9. Epoch-based variation of accuracy and loss values related to the training and validation process.

Examining the training and validation metrics presented in Figure 3, it can be seen that the model's training accuracy shows a rapid increase, rising from approximately 83% to 96%. In contrast, validation accuracy stabilized at 92% after the first epoch and showed a slight downward trend in subsequent steps. Similarly, while the training error continuously decreased, the validation error increased from the second epoch onwards (from 0.23 to 0.32), indicating that the model exhibits overfitting. To prevent overfitting, Early Stopping was applied. Training is terminated at the point where validation loss is minimized.

Comparison of Model Performances

In this study, the effects of different CNN architectures on text classification performance were comparatively examined through various model configurations. The findings are presented in Table 2. The analyses revealed that multi-kernel structures exhibited superior performance compared to models using a single filter. While the test accuracy of models with a single filter size ranged from 91.70% to 91.89%, it was observed that this success increased in models created with combinations of different filter sizes. In particular, the CNN_Filter_3_4_5 model, which simultaneously processes 3, 4, and 5-dimensional filters, achieved a test accuracy of 91.95%. This result confirms that simultaneously learning variable-length n-gram features increases the representativeness of the model. In the evaluation of pooling strategies, it was determined that Global Max Pooling and Local Max Pooling methods produced similar accuracy values, but the Global Max Pooling approach exhibited lower variance and more balanced performance. As a result of the experimental findings, the MultiChannel_CNN_3_4_5 model, which integrates multiple channel and multiple filter structures, achieved the highest success rate. The model achieved an accuracy rate of 92.58% in the validation set and 92.05% in the test set; the macro F1 score was calculated as 92.04%. The limited performance improvement in models developed with hyperparameter optimization (HPO) indicates that architectural design plays a more decisive role in model performance than hyperparameter configurations.

Table 2. Comparison of Performance Metrics of CNN-Based Models

Final Results Table:

	Model	Validation Accuracy	Test Accuracy	Macro Precision	Macro Recall	Macro F1
0	MultiChannel_CNN_3_4_5	0.9258	0.9205	0.9208	0.9205	0.9204
1	HPO_4_E128_F128_D0.3	0.9237	0.9200	0.9202	0.9200	0.9198
2	CNN_Filter_3_4_5	0.9233	0.9195	0.9199	0.9195	0.9193
3	HPO_1_E128_F128_D0.5	0.9235	0.9191	0.9194	0.9191	0.9188
4	CNN_LocalMaxPooling	0.9247	0.9191	0.9196	0.9191	0.9189
5	CNN_Filter_4	0.9225	0.9189	0.9194	0.9189	0.9189
6	HPO_3_E128_F200_D0.5	0.9236	0.9189	0.9197	0.9189	0.9189
7	HPO_2_E100_F128_D0.5	0.9225	0.9187	0.9189	0.9187	0.9185
8	CNN_GlobalMaxPooling	0.9237	0.9184	0.9188	0.9184	0.9183
9	CNN_Filter_5	0.9223	0.9183	0.9187	0.9183	0.9181
10	CNN_Filter_3	0.9238	0.9170	0.9170	0.9170	0.9168

Performance Analysis of the Best Model

Experimental results show that the highest performance was achieved with the MultiChannel_CNN_3_4_5 model. The model outperformed other CNN configurations with 92.58% validation accuracy, 92.05% test accuracy, and 92.04% macro F1 score. The main reason for this success is the model's ability to simultaneously learn n-gram structures of varying lengths by using different filter sizes together. Thanks to the multi-channel structure, the acquired features were combined, and contextual patterns in the text were represented more effectively. Furthermore, regularization methods such as SpatialDropout, Dropout, and Early Stopping reduced overlearning, improving the model's generalization performance. In conclusion, a CNN architecture with multiple filters and a multi-channel structure was found to be more effective in text classification problems.

Class-Based Performance Analysis

Table 3 presents the model's class-based precision, recall, and F1-score results. According to the findings, the model showed the highest performance in the Sports class. In this class, the precision value was measured as 0.96, the recall value as 0.98, and the F1-score value as 0.97.

Relatively lower performance was obtained in the Business and Sci/Tech classes. The F1-score was calculated as 0.89 and the recall as 0.87 in the Business class, while the F1-score was measured as 0.90 in the Sci/Tech class. This indicates that some samples were confused with similar classes.

In the World class, the precision value was obtained as 0.94, the recall value as 0.90, and the F1-score as 0.92. When the overall results are examined, the macro average F1-score of 92% indicates that the model exhibits balanced performance across classes and has a high generalization ability.

Table 3. Class-Based Performance Metrics

Class	Precision	Recall	F1-score	Support
World	0.94	0.90	0.92	1900
Sports	0.96	0.98	0.97	1900
Business	0.90	0.87	0.89	1900
Sci/Tech	0.89	0.92	0.90	1900

Table 3 presents the performance metrics obtained for each class of the model.

Confusion Matrix Analysis

Figure 4 presents the confusion matrix for the MultiChannel_CNN_3_4_5 model, which achieved the highest performance. Examination of the confusion matrix reveals that the model achieved its most successful result in the Sports class. Of the 1900 samples in this class, 1867 were correctly classified. This indicates that the Sports class has more distinctive characteristics compared to other classes.

While 1702 examples were correctly predicted in the World category, it was observed that some examples were particularly confused with the Business and Sports categories. A significant mix-up was also seen between the Business and Sci/Tech categories. 159 examples belonging to the Business category were reclassified as Sci/Tech, while 100 examples belonging to the Sci/Tech category were assigned to the Business category. This indicates similarities in content and terminology between the two categories.

In the Sci/Tech category, 1748 examples were correctly classified, indicating high overall performance. Overall, the model exhibited balanced performance across categories; however, it experienced limited confusion in semantically similar categories.

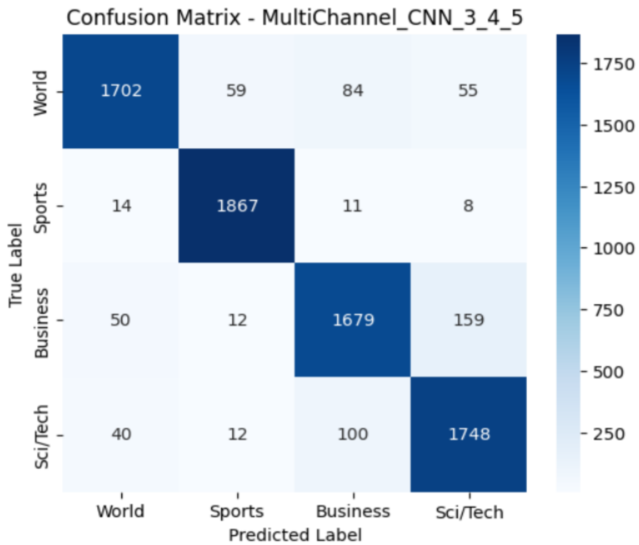

Figure 10. Confusion matrix of the MultiChannel_CNN_3_4_5 model

CONCLUSION, DISCUSSION AND RECOMMENDATIONS

This study investigated the text classification performance of different CNN architectures on the AG News dataset. Experimental results showed that multi-kernel and multi-channel architectures significantly improved model performance. The highest success rate was achieved by the MultiChannel_CNN_3_4_5 model, with a test accuracy of 92.05% and a macro F1-score of 92.04%.

Discussion

Multiple filter structures have provided higher performance compared to single-filter models by simultaneously learning n-gram patterns of varying lengths.

The multi-channel CNN architecture achieved more powerful representation learning by combining parallel feature maps.

The Global Max Pooling approach has produced more balanced and stable results compared to the Local Max Pooling method.

The confusion matrix results revealed that the model performed highly in the Sports class; however, limited confusion occurred in the Business and Sci/Tech classes due to semantic similarity.

CNN-based models have offered an effective alternative in terms of low computational cost and rapid training time.

Limitations

The study was conducted only on the AG News dataset.

Direct performance comparisons have not been made with Transformer-based models.

Hyperparameter optimization was implemented with a limited number of configurations.

The model may have limitations in learning deeper contextual relationships.

Recommendations

The following improvements can be made in future studies:

Hybrid models based on CNN and Transformer can be examined.

Attention mechanisms can be integrated into the CNN architecture.

More comprehensive hyperparameter optimization can be applied.

Generalizability analysis can be performed on different datasets.

REFERENCES

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 328–339. https://doi.org/10.18653/v1/P18-1031

Johnson, R., & Zhang, T. (2015). Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 103–112. https://doi.org/10.3115/v1/N15-1011

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 655–665. https://doi.org/10.3115/v1/P14-1062

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1746–1751. https://doi.org/10.3115/v1/D14-1181

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. http://arxiv.org/abs/1907.11692

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. http://arxiv.org/abs/1301.3781

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543. https://doi.org/10.3115/v1/D14-1162

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. http://arxiv.org/abs/1910.01108

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). XLNet: Generalized Autoregressive Pretraining for Language Understanding. http://ar xiv.org/abs/1906.08237

Zhang, X., Zhao, J., & LeCun, Y. (2016). Character-level Convolutional Networks for Text Classification. Advances in Neural Information Processing Systems, 2015-January, 649–657. http://arxiv.org/abs/1509.01626

CHAPTER 7

A Deep Neural Network-Based Data-Driven Model for Three-Phase Inverter Voltage Prediction

Ziya Nur İzmit[1] *& Seda Postalcıoğlu*[2]

Introduction

This study presents a deep neural network-based data-driven approach for the prediction of output voltages in three-phase voltage source inverters. In practical drive systems, phase voltages are generally not measured for cost-related reasons; therefore, accurate inverter voltage modeling is essential for reliable observer performance and control-oriented voltage estimation (Stender, Wallscheid, & Böcker, 2021). Inverter voltage prediction is challenging because hardware-induced nonlinearities such as dead-time effects, switching transients, parasitic components, and duty-cycle-dependent distortions are difficult to represent using purely analytical equations. Studies show that black-box neural-network-based inverter models can learn the relationship between available control variables and actual inverter output voltages from experimental data (Stender et al., 2021). Motivated by this perspective, the proposed model predicts the output voltage by evaluating historical phase currents, PWM duty-cycle information, and DC-link voltage variations. The model is trained and validated using the three-phase IGBT two-level inverter dataset (Stender, Wallscheid, & Böcker, 2020). Experimental results indicate that the proposed model can predict inverter output voltages with high accuracy under dynamic operating conditions. The proposed approach offers an alternative to detailed analytical inverter modeling and may support future MPC-related applications.

Literature Review

The integration of artificial intelligence into power electronics has facilitated significant advancements in the control and monitoring of inverter systems. In the field of voltage and current regulation, Model Predictive Control (MPC) has been a dominant theme due to its fast dynamic response. "However, it becomes impractical to implement it on large-scale modular multilevel converters (MMC)

[1] Master's Student, Izmir Democracy University, Faculty of Engineering, Department of Electrical and Electronics Engineering, Türkiye, ORCID: 0009-0003-3738-2137
[2] Prof. Dr., Izmir Democracy University, Faculty of Engineering, Department of Computer Engineering, Türkiye, ORCID: 0000-0002-3188-8116

system due to the exponential increase in the number of control options as the number of submodules (SMs) grows" (Raja, Wang, Arshad, Kish, & Zhao, 2023, p. 42113). Neural networks (NN) have also been used to scale down the computational demands of MPC by learning simplified representations of the original controller (Chen et al., 2018; Lucia, Navarro, Karg, Sarnago, & Lucía, 2021). Artificial Neural Networks (ANNs) are particularly effective in addressing nonlinearities that traditional analytical models often fail to capture. An ANN-based MPC architecture was proposed for three-phase inverters with LC filters, emphasizing that these controllers do not require the mathematical model of the system and can be designed by considering the whole system as a black-box (Mohamed, Rovetta, Do, Dragičević, & Diab, 2019). This approach reduces the dependency on precise modeling at the design stage and, as validated through simulations, demonstrates excellent steady-state and dynamic performance across different types of loads. A model-free control strategy on the basis of artificial neural networks (ANNs) was proposed for mitigating the effects of parameter mismatching (Bakeer et al., 2022). Their study demonstrates that such architecture provides robustness against parameter mismatch in the control of three-phase flying capacitor inverters. These studies indicate that ANN-based methods can be used to reduce the dependency on exact analytical models for inverter control problems. However, accurate voltage prediction for practical inverter systems requires models that can capture hardware-induced nonlinearities such as dead-time effects, switching behavior, and device-level nonidealities. To address these deployment challenges, recent studies have investigated ANN based modeling and compensation techniques in power electronics. An online learned ANN was developed for dead-time compensation, specifically designed to operate on microcontrollers with limited computing resources, proving that complex compensation tasks can be handled locally without extensive hardware overhead (Buchta, Kozovsky, & Blaha, 2025).

An enhanced modeling framework using the Levenberg-Marquardt (LM) algorithm was provided to estimate phase voltages in three-phase IGBT inverters, achieving high accuracy with black-box models (Ho, Peng, & Su, 2025). This work specifically highlights the advantages of data-driven approaches in capturing switching effects that are mathematically difficult to model

In a broader voltage-estimation context, linear regression and neural network based methods have been compared for voltage estimation in low-voltage distribution networks(Kalinga, Banfield, Knott, & Robinson, 2026). While they found linear regression to be practical for standard network approximations, they explicitly noted that more complex or highly non-linear systems, as well as larger datasets, benefit significantly from advanced NN models. These findings suggest that neural network models may be suitable for representing nonlinear switching behavior in inverter systems.

Additionally, Guruwacharya et al. (2024) explored data-driven modeling for grid-forming inverters using hardware-in-the-loop (HIL) experimentation, showcasing the scalability of these models in real-world grid Dynamics.

Driven by these implementation-oriented concerns, it has been demonstrated that neural-network-based approximations can significantly reduce the computational burden of predictive control algorithms, thereby enabling efficient deployment on low-cost embedded hardware (Maddalena, Specq, Wisniewski, & Jones, 2021).

In this study, the proposed DNN architecture is designed to balance voltage prediction performance with model complexity. However, reliable data-driven voltage prediction also depends on the physical relevance and quality of the input features. For this purpose, the Clarke transformation is employed to project three-phase signals into an orthogonal α-β reference frame, and Z-score analysis is used to mitigate measurement transients and abnormal samples. Unlike standard approaches that directly rely on raw high-dimensional data, the proposed method leverages physics-informed preprocessing to improve learning efficiency and model generalization.

Methodology

The model is trained and validated using the three-phase IGBT two-level inverter dataset (Stender, Wallscheid, & Böcker, 2020). The dataset contains 235,000 samples collected at a sampling frequency of 10 kHz. The variables encompass motor speed (n),DC-link voltage (Vdc), three phase current (Ia,b,c), duty cycle (Da,Db,Dc) and measured phase voltages (Va,Vb,Vc) also dataset includes instantaneous value at the current sampling interval (k) and (k-1,k-2,k-3) are used to represent previous steps.u_alpha and u_beta are defined as the model's target output and by utilizing the Clarke transformation. To identify lag-induced distortions in the phase voltages, capturing historical time steps (k-1, k-2, k-3) as input features is essential. This aligns with the underlying principle of dead-time compensation strategies, where historical system states (e.g., step k-2) are critically required to evaluate and correct persistent hardware-induced delays and residual harmonic distortions, as demonstrated by the mechanism proposed by Buchta et al. (2025). The model is designed to estimate the instantaneous orthogonal output voltages at step k(u_a_k, u_b_k) by utilizing historical system states from steps k-1, k-2 and k-3. Furthermore, to enhance the network's capability in capturing highly dynamic transient behaviors during inverter switching, the temporal differences of the phase currents (Δ_{iA},Δ_{iB}) and the DC-link voltage ($\Delta_{v_{dc}}$) between consecutive historical intervals are introduced as explicit input features. Input features and target variables are shown in Table 1.

Table 1. Input features and target variables used in DNN.

Category	Parameter
Inputs	n_k,u_dc_k,u_dc_k-1,u_dc_k-2,u_dc_k-3,i_a_k,i_b_k,i_c_k,i_a_k-1,i_b_k-1,i_c_k-1,i_a_k-2, i_b_k-2,i_c_k-2,i_a_k-3,i_b_k-3,i_c_k-3,d_a_k-2,d_b_k-2,d_c_k-2,d_a_k-3,d_b_k-3,d_c_k-3, u_a_k-1,u_b_k-1,u_c_k-1
Targets	u_alpha,u_beta (Clarke transform)

The overall architecture of the proposed DNN-based voltage prediction model is illustrated in Figure 1. Figure1 summarizes how raw three-phase inverter measurements are transformed into an engineered feature vector and then processed by the Deep Neural Network (DNN).

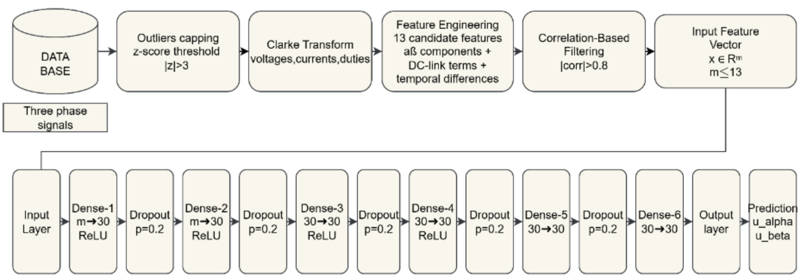

Figure 1. The structure of the system for α-β prediction

Raw three-phase inverter measurements are first processed using Z-score-based outlier capping and Clarke transformation. Then, engineered candidate features are constructed from α-β components, DC-link voltage terms, and temporal difference variables. After correlation-based filtering, the resulting feature vector is supplied to DNN consisting of six hidden Dense layers with 30 neurons each. Dropout regularization is applied after the first five hidden layers to reduce overfitting. The final linear output layer estimates the α- and β-axis voltage components, u_α and u_β, which constitute the voltage prediction output of the proposed DNN model. To handle measurement transients and high-frequency noise, a Z-score analysis was applied. In alignment with robust outlier detection methodologies for time-series data (Yaro, Maly, Prazak, & Maly, 2024), values exceeding the threshold of |Z| > 3 were capped to prevent data distortion.

To reduce dimensionality while preserving the physical characteristics of the balanced three-phase system, the Clarke Transformation (Clarke, 1943) was used to project the phase voltages into the stationary α-β reference frame. Equation 1 presents the Clarke transformation matrix.

$$\cdot\begin{bmatrix} v\alpha \\ v_\beta \end{bmatrix} = \frac{2}{3}\begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \end{bmatrix}\begin{bmatrix} v_A \\ v_B \\ v_c \end{bmatrix} \tag{1}$$

The zero-sequence component is assumed to be zero, and the reduced Clarke transformation matrix is used to transform the three-phase voltage components into the stationary α-β reference frame (Yılmaz, Tezgel, & Çorapsız, 2021). Where V_A, V_B, and V_C denote the three-phase voltage components, while v_a and v_β represent the corresponding stationary α-β voltage components. The same transformation principle was also applied to the current and duty-cycle variables during feature construction. This α-β representation enables the neural network to learn the inverter dynamics in a compact two-axis coordinate system instead of directly operating on three correlated phase quantities.

The reliability of the proposed voltage prediction model was verified using a 70/30 train-test split and reinforced by Six-Fold Cross-Validation. The use of k-fold cross-validation ensures that the high R^2 scores and low RMSE values are consistent across various operating points of the inverter. After preprocessing and feature construction, the raw inverter variables listed in Table 1 are converted into a 13-dimensional engineered input feature vector. This vector is composed of α-β transformed duty-cycle components, α-β current components at historical sampling instants, DC-link voltage terms, and temporal difference variables that represent the dynamic variation of the inverter states. Therefore, the first dense layer receives 13 input features and maps them into a 30-dimensional hidden representation. The architecture consists of six hidden layers with 30 neurons each. The rectified linear unit (ReLU) activation function is used in all hidden layers to model the nonlinear relationship between the historical inverter states and the α-β voltage components. Mathematically, ReLU is expressed as f(z) = max(0, z), where z is the weighted input of a neuron. Linear activation function is used at the output layer. Dropout regularization with p = 0.2 was applied after the first five hidden Dense layers to reduce overfitting. During training, 20% of the neurons were randomly deactivated at each iteration, which helped improve the generalization capability of the network.

The sixth hidden layer is directly connected to the final linear output layer without an additional dropout operation. The final output layer is defined as a two-dimensional linear Dense layer. Therefore, the model output is not a scalar value but a two-component regression vector,$\hat{y} = [\hat{u}_{\alpha}, \hat{u}_{\beta}]^{T} \in \mathbb{R}^2$. The first output neuron estimates the α-axis voltage component, while the second output neuron estimates the β-axis voltage component. Since both target variables are continuous voltage values, no nonlinear activation function is applied at the output layer. This design is consistent with the observation of Maddalena et al. (2021) that neural-network complexity can be adjusted through the selected model size, which is important for computationally efficient implementation in power-electronic applications.

Results and Discussion

As shown in Table 2 the model achieved R^2 score of approximately 0.98 across both u_alpha and u_beta. This indicates that the architecture captures 98% of the variance in the dynamic inverter output voltage. Mean absolute error values vary between 13 and14 Volt range, demonstrate that the model's prediction trend is highly accurate without systematic bias. The maximum error values range between 141–155 V, which is expected during instantaneous high-frequency switching transients in power electronic systems.

Table 2. Performance metrics of the DNN

Metric	**u_alpha**	**u_beta**
RMSE	19.0142 V	20.4469 V
Mean Absolute Error (MAE)	13.2147 V	14.3146 V
R^2	0.9798	0.9768
Max Error Value	155.8031 V	141.5014 V

The scatter plot of actual versus predicted voltage indicates a strong correlation between the two variables as the data points are closely clustered along the ideal reference line as shown in Figure 2.

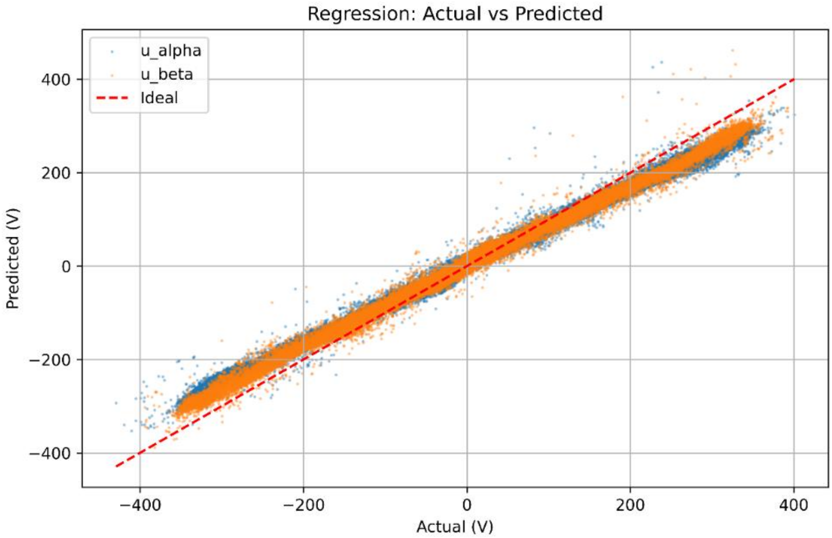

Figure 2. Regression Plot

The error distribution shown in Figure 3 is very peaked with a zero centered normal (Gaussian) distribution. This indicates that the prediction errors are symmetrically distributed around zero. The residual distribution suggests that the model captures the dominant nonlinear characteristics of the inverter dynamics while maintaining good generalization performance.

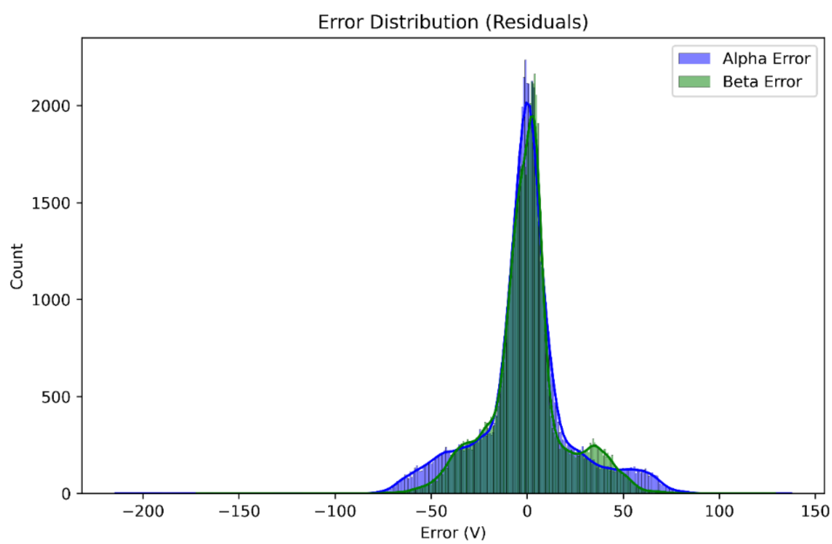

Figure 3. Error distribution of u_alpha and u_beta predictions.

For a data-driven voltage prediction model to be practically usable in real-time inverter control applications, its inference time should remain below the control-loop sampling period.To assess this, a computational latency test was executed without the use of dedicated hardware accelerators. Total Samples

Processed=70,359, Total Execution Time= 2323.92 ms and Average Inference Latency per Sample: 0.033029 ms. Considering that standard industrial inverter control systems typically operate with a sampling period of 100 μ_s the preliminary inference time of 33 μ_s achieved in Python and Python execution includes inherent interpreter and framework overheads. However, this expectation should be validated through actual deployment on embedded platforms. The predicted waveforms u_alpha and u_beta shown in Figure 4 closely follow the measured signals with no observable phase shift under the evaluated operating conditions, demonstrating reliable dynamic tracking performance under the evaluated operating conditions. The spectral comparison shows that the predicted voltage components preserve the dominant frequency characteristics of the measured voltages.

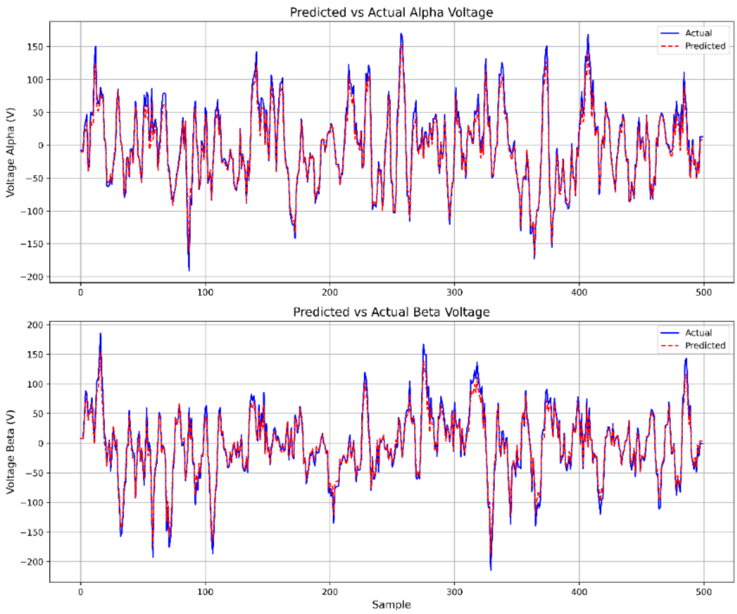

Figure 4. Actual and predicted alpha-beta voltage waveforms

To rigorously evaluate the phase synchronization and dynamic tracking capability of the proposed model, a Lissajous analysis was conducted for both voltage components. As illustrated in Figure 5, the DNN predicted voltages tightly cluster along the ideal reference line (y=x). The absence of elliptical trajectories (looping) in the Lissajous plots suggests that the proposed model preserves phase consistency between the predicted and measured voltage

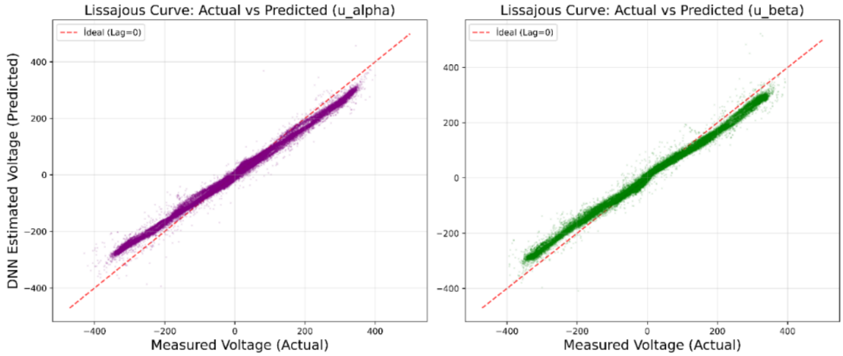

Figure 5. Lissajous curve analysis of actual and predicted voltage components

components within the evaluated offline dataset. The minor dispersion observed around the ideal line is a natural consequence of the high-frequency PWM switching noise inherent in the physical testbench data. Overall, the Lissajous analysis indicates that the predicted voltages remain temporally consistent with the measured signals under the evaluated operating conditions.

Conclusion

This study presented a DNN-based data-driven model for predicting the output voltages of a three-phase inverter using experimental operating data. Clarke-transformed variables, historical inverter states, and engineered temporal features were used as inputs to the network. The model achieved R^2 values close to 0.98, while the mean absolute error remained between 13–14 V for both α- and β-axis voltage components under the evaluated operating conditions.

The predicted voltage waveforms showed good agreement with the measured signals, with no noticeable phase mismatch observed in the tested dataset. Frequency-domain analysis also showed that the predicted signals retained the main spectral characteristics of the measured voltages, while some high-frequency switching components were partially smoothed. In addition, the average inference latency was measured as approximately 33 μs per sample in the Python-based test environment, indicating that the proposed approach may be suitable for future real-time control applications.

References

Bakeer, A., Mohamed, I. S., Boodaghi Malidarreh, P., Hattabi, I., & Liu, L. (2022). An artificial neural network-based model predictive control for three-phase flying capacitor multilevel inverter. *IEEE Access*, *10*, 70305–70316. https://doi.org/10.1109/ACCESS.2022.3187996

Buchta, L., Kozovsky, M., & Blaha, P. (2025). A dead-time compensation strategy based on an online learned artificial neural network. *IEEE Transactions on Industrial Electronics*, *72*(10), 10574–10584. https://doi.org/10.1109/TIE.2025.3544207

Chen, S., Saulnier, K., Atanasov, N., Lee, D. D., Kumar, V., Pappas, G. J., & Morari, M. (2018). Approximating explicit model predictive control using constrained neural networks. *2018 Annual American Control Conference (ACC)*, 1520–1527. https://doi.org/10.23919/ACC.2018.8431275pp. 1520–1527.

Clarke, E. (1943). *Circuit analysis of A-C power systems*. Wiley.

Guruwacharya, N., Chakraborty, S., Saraswat, G., Bryce, R., Hansen, T. M., & Tonkoski, R. (2024). Data-driven modeling of grid-forming inverter dynamics using power hardware-in-the-loop experimentation. *IEEE Access*, *12*, 52267–52281. https://doi.org/10.1109/ACCESS.2024.3383323

Ho, T.-Y., Peng, S.-W., & Su, S.-F. (2025). Enhanced modeling of three-phase IGBT two level inverter for electric drives using LM algorithm in artificial neural networks. *2025 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*. https://doi.org/10.1109/ICCE-Taiwan66881.2025.11208034

Kalinga, T., Banfield, B., Knott, J. C., & Robinson, D. A. (2026). Comparison of linear regression and neural networks for model-free voltage estimation of low voltage distributionnetworks with high penetration of residential rooftop solar. *Electronics*, *15*(7), 1467. https://doi.org/10.3390/electronics15071467

Lucia, S., Navarro, D., Karg, B., Sarnago, H., & Lucía, O. (2021). Deep learning-based model predictive control for resonant power converters. *IEEE Transactions on Industrial Informatics*, *17*(1), 409–420. https://doi.org/10.1109/TII.2020.2969729

Maddalena, E. T., Specq, M. W. F., Wisniewski, V. L., & Jones, C. N. (2021). Embedded PWM predictive control of DC-DC power converters via piecewise-affine neural networks. *IEEE Open Journal of the Industrial Electronics Society*, *2*, 199–206. https://doi.org/10.1109/OJIES.2021.3058411

Mohamed, I. S., Rovetta, S., Do, T. D., Dragičević, T., & Diab, A. A. Z. (2019). A neural-network-based model predictive control of three-phase inverter with an output LC filter. *IEEE Access*, *7*, 124737–124749. https://doi.org/10.1109/ACCESS.2019.2938220

Raja, M. M., Wang, H., Arshad, M. H., Kish, G. J., & Zhao, Q. (2023). Data driven model predictive control for modular multilevel converters with reduced computational complexity. *IEEE Access*, *11*, 42113–42123. https://doi.org/10.1109/ACCESS.2023.3270773

Stender, M., Wallscheid, O., & Böcker, J. (2021) . Comparison of gray-box and black-box two-level three-phase inverter models for electrical drives. *IEEE Transactions on Industrial Electronics*, *68*(9), 8646–8656 . https://doi.org/10.1109/TIE.2020.3018060

Stender, M., Wallscheid, O., & Böcker, J. (2020). *Three-phase IGBT two-level inverter for electrical drives* [Data set]. Kaggle. Retrieved from https://www.kaggle.com/stender/inverter-data-set

Yaro, A. S., Maly, F., Prazak, P., & Maly, K. (2024). Outlier Detection Performance Of A Modified Z-score Method In Time-series RSS Observation With Hybrid Scale Estimators. *IEEE Access*, *12*, 12785–12796. https://doi.org/10.1109/ACCESS.2024.3356731

Yılmaz, M., Tezgel, S., & Çorapsız, M. F. (2021). Graphical User Interface For Asynchronous Motors Clarke-Park Transforms Using LabVIEW. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, *21*(2), 317-325. https://doi.org/10.35414/akufemubid.833337

CHAPTER 8

The Importance of Renewable Energy Sources in The Transition of Energy Supply: The Case of Turkey

Sena Nur Demirel[1] *& Güray Çelik*[2]

INTRODUCTION

The concept of energy, recognized as a fundamental component of economic growth, social development, and advancing technology, gained momentum in economic activities with the development of the Industrial Revolution in the 18th and early 19th centuries, leading to a steady increase in global demand for energy and, consequently, a growing interest in fossil fuels. (Arslan et al., 2025).

Throughout this process, fossil fuels such as coal, oil, and natural gas have become key components due to their high energy density and the advanced nature of the technological infrastructure (Smil, 2017). Consequently, the use of fossil fuels is recognized as one of the key factors significantly influencing climate change on a global scale, as it leads to an increase in greenhouse gas emissions resulting from carbon-containing particulate matter released into the atmosphere (IPCC, 2023).

Carbon dioxide (CO2) emissions from the global energy sector account for a significant portion of total greenhouse gas emissions. In particular, the use of fossil fuels is widespread in the electricity generation, industrial, and transportation sectors; for this reason, international efforts are being made to limit the rise in global temperatures (IEA, 2024).

In this context, in line with the Paris Climate Agreement signed in 2015, the goal is to limit the global temperature increase to no more than 2°C above pre-Industrial Revolution levels and, if possible, to keep it at 1.5°C; it is emphasized that a global transition away from fossil fuel-based energy systems is necessary to achieve these goals (IPCC, 2023).

In line with these goals, the concept of energy transition has recently come to the forefront on a global scale, with renewable energy systems replacing the fossil fuels that were previously widely used. This process is defined as the transition to sustainable energy (IRENA, 2023). Leading renewable energy sources include

[1]MSc Student., Uludağ University, Faculty of Engineering, Department of Environmental, ORCID:0000-0003-1956-8846
[2]Prof. Dr., Uludağ University, Faculty of Engineering, Department of Environmental ORCID:0000-0003-0714-048X

solar energy, wind energy, hydroelectric power, geothermal energy, and biomass energy. Key factors contributing to the rise of these energy sources include their domestic production potential, their ability to generate clean energy, and their low carbon emissions (Ellabban et al., 2014; REN21, 2024). Over the past decade, the low costs and high energy efficiency of renewable energy sources have contributed to an increase in their share of global energy systems (IRENA, 2024).

The most important factor in the global energy transition is energy supply security. Energy supply security refers to the availability of sufficient quantities of existing energy resources that can be procured without interruption and under economic conditions (IEA, 2023). Energy supply security is becoming increasingly important in countries that are dependent on foreign energy sources. Price fluctuations in the global energy market, geopolitical developments, and supply disruptions are influencing the energy policies pursued by countries (Acar & Yeldan, 2016). For this reason, energy supply security is expected to be strengthened in most countries through the use of renewable energy sources that support domestic production.

With its growing population, rapidly developing industry, and expanding economy, Turkey ranks among the countries with a steadily increasing demand for energy. Furthermore, the significant role that many types of fossil fuels such as oil, natural gas, and coal play in primary energy consumption increases the country's dependence on external sources for its energy supply. Natural gas and fuel oil needs for which are largely met through imports pose numerous risks to energy supply security (ETKB, 2024). For this reason, steps have been taken in recent years to develop domestic and renewable energy sources.

Due to its geographic location, Turkey possesses a very high potential for natural resources. It has a significant advantage, particularly in terms of solar, wind, hydroelectric, geothermal, and biomass energy sources. Investments in wind and solar energy have led to a significant increase in the installed capacity of renewable energy sources (TEİAŞ, 2024). Renewable energy sources integrated into the energy system not only provide environmental benefits but also make significant contributions to energy supply security and the achievement of sustainable climate goals (IEA, 2024; Kaygusuz, 2012).

This study examines the status of renewable energy sources within the context of Turkey's energy transition process. As part of the study, Turkey's current energy supply structure, the status of its energy imports and exports, the distribution of electricity generation sources, and the development of renewable energy sources will be analyzed to examine the current role of energy supply security. Furthermore, the case of Turkey is examined and discussed considering sustainability and low carbon development goals.

THE CONCEPT OF ENERGY SUPPLY AND ENERGY TRANSITION

The Concept of Energy Supply

Energy supply refers to the process by which a country produces the energy it needs to ensure the continuity of its economic and social activities from energy sources, imports it, and makes it available for use through storage processes. Ensuring the feasibility of national development goals is of vital importance for an adequate, reliable, economical, and sustainable energy supply. A growing population, increasing industrialization, and technological advancements have made rising energy demand a fundamental element of energy policies in terms of energy supply security (IEA, 2023).

Energy security is defined as the ability to secure energy resources in a way that ensures an uninterrupted supply, reasonable costs, and environmental sustainability (Cherp & Jewell, 2014). Dependence on foreign countries for energy resources is of great importance for energy security in terms of economic stability and national security. For this reason, many countries are developing policies aimed at increasing the use of domestic energy resources to diversify their energy supply

The Concept of Energy Transition

The process of transitioning from fossil fuel-based energy systems to low carbon and sustainable energy systems is defined as the energy transition (IRENA, 2023). Issues such as rising global energy demand, climate change, environmental degradation, and energy supply challenges have combined to accelerate the energy transition process. In this transition process, the share of renewable energy sources particularly solar, wind, hydroelectric, geothermal, and biomass in energy production is increasing day by day (REN21, 2024).

Because renewable energy sources produce lower levels of carbon emissions compared to fossil fuels and rely on domestic production resources, they represent a vital step toward achieving sustainable development (Ellabban et al., 2014). The energy transition not only provides environmental benefits but also ensures economic sustainability and contributes to reducing energy imports. This has become a fundamental component of energy policies (IEA, 2024).

TURKEY'S ENERGY SUPPLY OUTLOOK

Energy is recognized as one of the country's fundamental needs in terms of population growth, economic growth, advancing industrialization, and sustainable social welfare. In both developed and developing countries, population growth, urbanization, and technological advancements have led to an increase in energy demand. Meanwhile, in recent years, Turkey has joined the ranks of countries developing policies to ensure energy supply security to meet its growing energy needs (IEA, 2021).

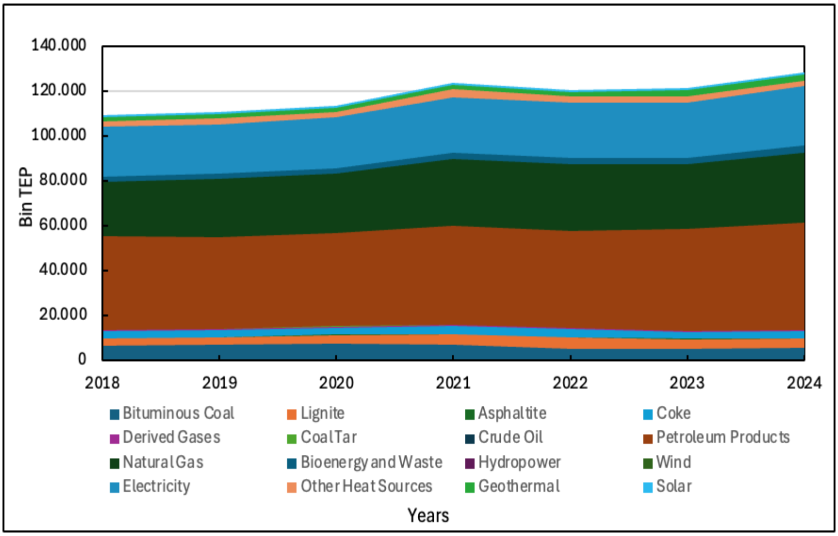

Figure 1. Breakdown of Turkey's primary energy consumption by source (2018–2024)

An analysis of Turkey's energy consumption data presented in Figure 1 reveals that fossil fuels dominated the primary energy supply during the 2018–2024 period. In particular, the fact that a portion of petroleum products and natural gas consumption is met through imports poses various risks in terms of energy supply security. It is noted that this dependence on external sources for energy also exerts pressure on costs in the global energy market (Bilgili et al., 2017; Kaygusuz, 2012).

In recent years, Turkey's energy supply structure has undergone a significant transformation. There has been a notable increase in the share of renewable energy sources in the energy supply. Within the framework of this transformation process, this development provides environmental benefits, strengthens energy supply security, and contributes to the diversification of energy sources (IRENA, 2023).

Turkey’s Energy Demand and Supply Structure

Energy demand in Turkey is increasing in direct proportion to economic growth and population growth. The rise in energy consumption in the industrial, transportation, residential, and service sectors is directly driving up energy demand. This situation necessitates sustainable planning of energy supply (Apergis & Payne, 2010).

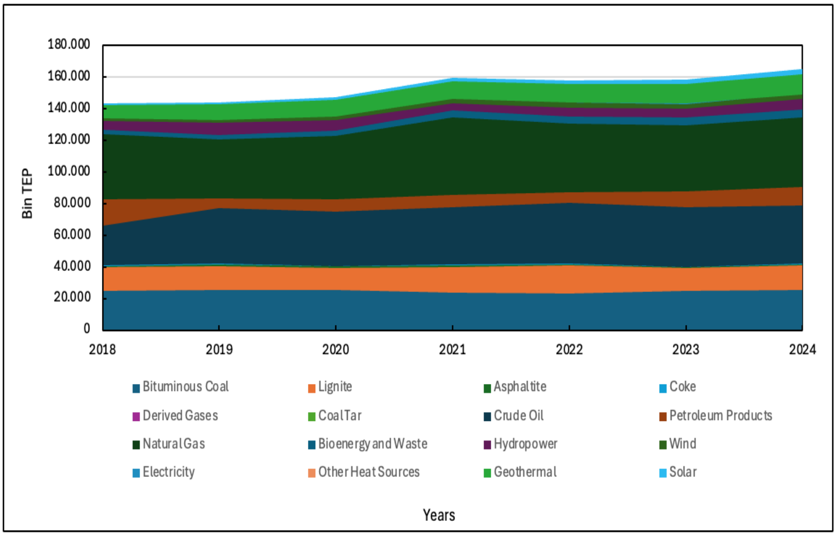

Figure 2. Breakdown of turkey's energy supply by source (2018–2024)

Figure shows that fossil fuels account for a significant share of the energy supply, but that there is an upward trend in the energy supply from renewable sources between 2018 and 2024. This situation is seen as a concrete indicator of the energy transition.

Turkey’s Dependence on Foreign Sources for Energy Supply

Due to its growing energy demand, Turkey is a country that is significantly dependent on foreign sources for its energy needs. In particular, the fact that the vast majority of its oil and natural gas needs are met through imports poses an economic and strategic risk in terms of energy supply. The impact of fluctuations in global energy costs and geopolitical shifts is considerable (Cherp & Jewell, 2014).

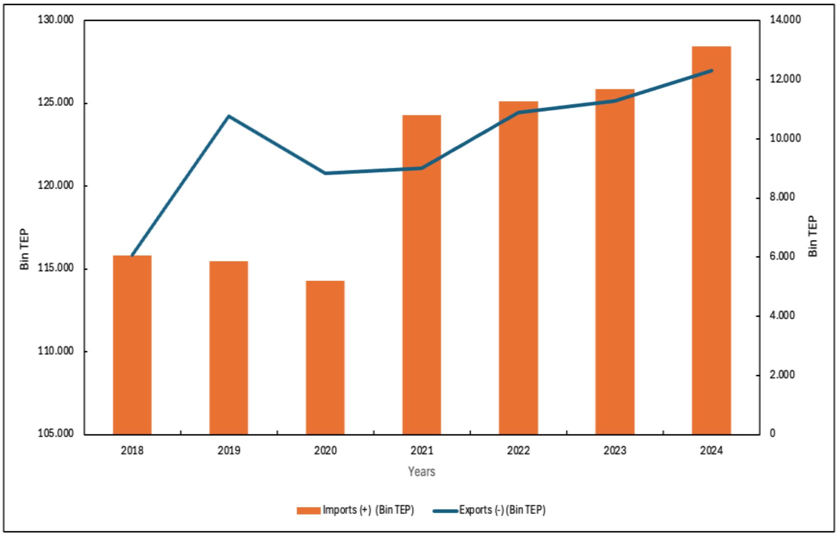

Figure 4. Turkey's energy imports and exports (2018–2024)

Figure shows that energy imports exceed exports. This situation is seen as an indicator of dependence on foreign countries for energy supply. Thus, increasing the share of renewable energy sources in Turkey's energy mix facilitates the establishment of a sustainable structure. Generating energy using domestic resources provides significant benefits from both environmental and economic perspectives. In conclusion, the growth in renewable energy sources plays a key role in reducing energy imports and strengthening energy supply security.

DEVELOPMENT OF ENERGY SOURCES

Renewable energy sources, which are viewed as a fundamental component of today's energy policies in terms of enhancing energy supply security, reducing environmental impacts, and supporting sustainable development goals, possess significant potential thanks to Turkey's geographical location and climatic characteristics (Sagir & Bahadir, 2017).

Solar Energy

Solar energy is considered the most cost-effective of renewable energy sources and the primary source of production in terms of efficiency, as it harnesses electromagnetic radiation from the sun and converts it into electricity or thermal energy through various technological devices (Lewis & Nocera, 2006). Turkey's advantageous position in terms of sunshine duration and solar radiation compared to European countries is driving significant increases in installed capacity (IRENA, 2024; Kabir et al., 2018).

Table 18. Annual variation in installed solar power capacity

Years	Solar Power Installed Capacity (MW)
2018	5.062,80
2019	5.995,20
2020	6.667,40
2021	7.815,60
2022	9.425,40
2023	15.613,40
2024	20.232,10

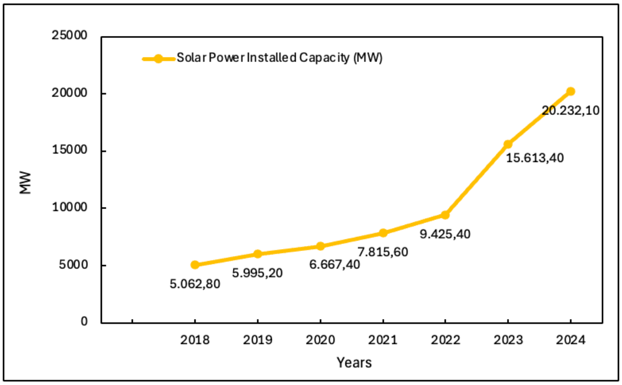

Figure 11. The development of installed solar power capacity in Turkey

Wind Energy

Wind energy is a renewable energy source based on the conversion of kinetic energy into electrical energy through air movements generated by the pressure difference created in the atmosphere because of the sun heating the Earth's surface at varying rates. Wind energy is not only considered Turkey's fastest-growing technology but is also recognized as having high potential in the Aegean, Marmara, and Mediterranean regions. At the same time, due to its low carbon emissions, the use of domestically produced resources, and its contribution to energy supply security, it has become a priority for investment in the energy policies of many countries (Bilgili et al., 2017; GWEC, 2024; TEİAŞ, 2024).

Table 19. Annual variation in installed wind power capacity

Years	Wind Power Installed Capacity (MW)
2018	7.005,40
2019	7.591,20
2020	8.832,40
2021	10.607,00
2022	11.396,20
2023	11.806,10
2024	12.870,80

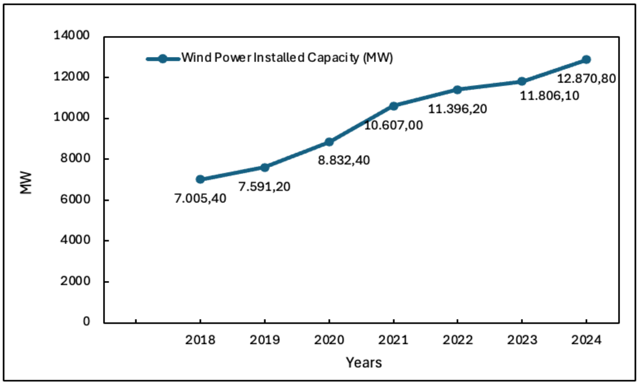

Figure 12. The development of installed solar power capacity in Turkey

Hydropower

Hydropower is a renewable energy source that converts the potential and kinetic energy derived from water into mechanical energy via turbines and then into electrical energy through turbine-generator systems. It holds a significant position globally due to its high production capacity and efficiency in electricity generation, its low carbon emissions, and its long operational lifespan ((IEA, 2024). In addition, Turkey's production volumes demonstrate efficiency based on changes in its rainfall patterns and current climate conditions (Kaygusuz, 2010).

Table 20. Annual variation in installed hydroelectric power capacity

Years	Hydroelectric Power Installed Capacity (MW)
2018	28.291,40
2019	28.503,00
2020	30.983,90
2021	31.492,60
2022	31.571,50
2023	31.962,40
2024	32.203,00

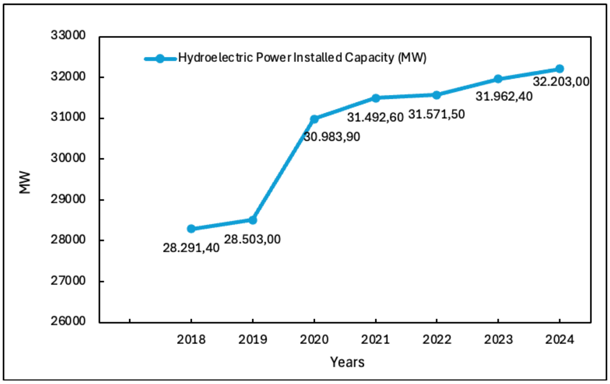

Figure 13. The development of installed hydroelectric power capacity in Turkey

Geothermal Energy

Geothermal energy is a renewable energy source that harnesses the Earth's internal heat stored deep within the Earth's crust and uses the thermal energy it provides to generate electricity or for direct heating. Because it is not dependent on weather conditions, ensures continuous energy production, and has low carbon emissions, it holds a very important position in the context of the global energy transition (Dickson & Fanelli, 2004; Lund & Toth, 2021a). In addition to these characteristics, due to its geographical location, Turkey ranks among the countries with the highest geothermal potential compared to European countries. In particular, the high concentration of geothermal resources in the Western Anatolia region provides a direct application area in terms of electricity generation potential (Lund & Toth, 2021b).

Table 21. Annual variation in installed geothermal power capacity

Years	Geothermal Power Installed Capacity (MW)
2018	1.282,50
2019	1.514,70
2020	1.613,20
2021	1.676,20
2022	1.691,30
2023	1.691,30
2024	1.733,50

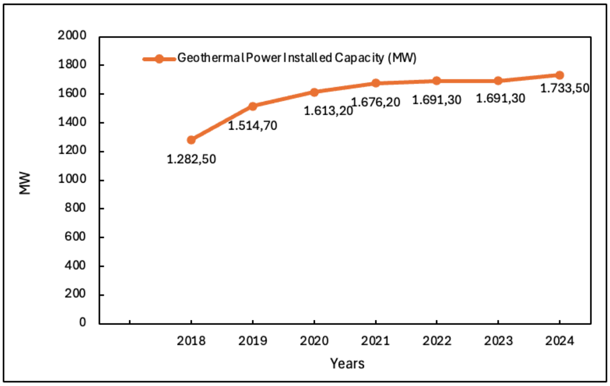

Figure 14. The development of installed geothermal power capacity in Turkey

Biomass Energy

Biomass energy is a renewable energy source based on the use of plant and animal organic matter residues for energy production. (Balat, 2006; Demirbaş, 2001; Mckendry, 2002). Turkey's extensive agricultural activities place it ahead of most countries in terms of production capacity and efficiency. It is a leading option because it contributes to waste management, is used for energy production, and, due to its role in the carbon cycle, results in the lowest possible carbon emissions (Demirbas, 2008).

Table 22. Annual variation in installed biomass power capacity

Years	Biomass Power Installed Capacity (MW)
2018	811,20
2019	1.170,50
2020	1.502,80
2021	2.051,10
2022	2.354,60
2023	2.446,40
2024	2.416,90

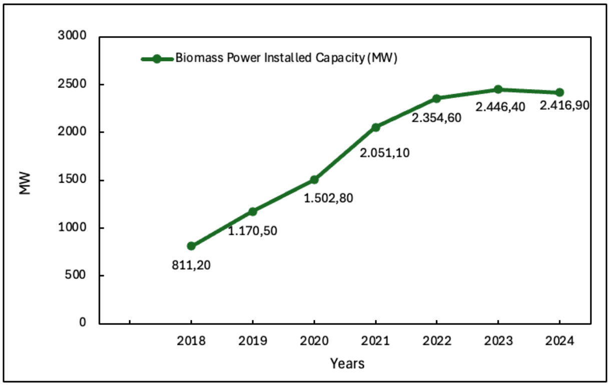

Figure 15. The development of installed biomass power capacity in Turkey

When evaluating renewable energy sources in Turkey, it is evident that solar and wind energy which offer higher energy efficiency are the most preferred systems. Increases in installed capacity have played a significant role in strengthening energy supply security and reducing dependence on fossil fuels.

ENERGY SUPPLY TRANSITION AND THE ROLE OF RENEWABLE ENERGY SOURCES

A key indicator of the energy supply transition is the global increase in the share allocated to renewable energy sources. Renewable energy sources are seen as offering significant advantages, including ensuring energy security, reducing environmental impacts, and contributing to sustainable development goals (IRENA, 2023; REN21, 2024). In recent years, the increase in installed capacity and the share of electricity generation from renewable energy sources has contributed significantly to a major shift in Turkey's energy supply structure.

Changes in Electricity Generation

When evaluating Turkey's electricity generation structure, it is evident that fossil fuels have been the most heavily utilized sources for many years. This situation has led to significant increases in the share of solar, wind, and other renewable energy sources in electricity generation. Consequently, the widespread adoption of renewable energy technologies has facilitated the diversification of energy production sources within Turkey (IEA, 2024).

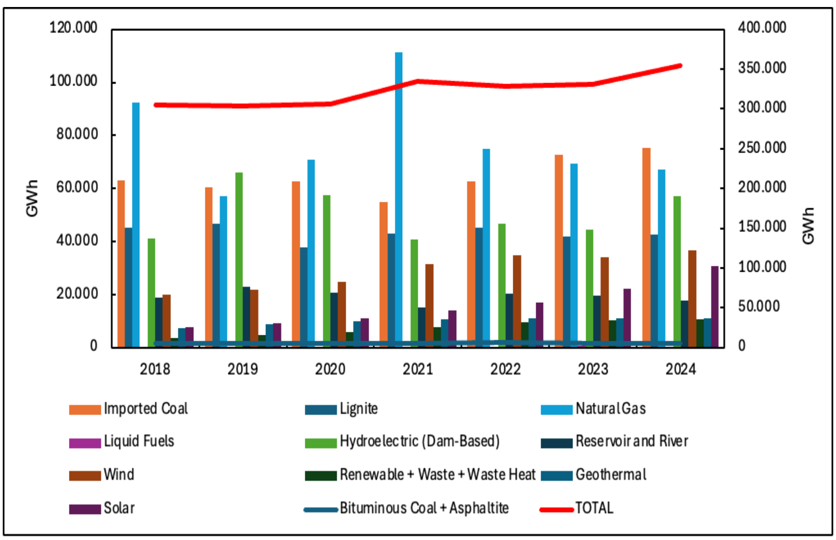

Figure 9. Breakdown of electricity generation in Turkey by source (2018–2024)

As shown in the figure, the significant increases in electricity generation particularly from solar and wind sources indicate that the energy system is beginning to transition toward a sustainable structure.

Contributions to Energy Supply Security

High dependence on foreign energy sources exposes countries to numerous risks in terms of energy supply security (Cherp & Jewell, 2014). Countries that are dependent on foreign energy sources and those that produce energy independently of other countries are listed in Table.

Table 6. The level of dependence on foreign energy supplies in developing countries

Countries	The State of Energy Import Dependency	Primary Source of Dependency
Türkiye	High	Oil and natural gas
Japan	Very high	Oil, LNG, and coal
South Korea	Very high	Oil, LNG, and coal
Italy	High	Natural gas and oil
Germany	Medium-High	Natural gas and oil
Spain	High	Oil and natural gas
Belgium	High	Oil and natural gas
Portugal	High	Oil and natural gas
Greece	High	Oil and natural gas

Table 7. Developed countries that are independent of other countries in terms of energy sources

Countries	Advantage
Norway	Exporter of oil and natural gas
Canada	Oil, natural gas, and hydroelectric power
United States	High oil and natural gas production
Russia	Large oil and natural gas reserves
Saudi Arabia	Oil exporter

Countries such as Turkey, Japan, South Korea, Italy, and Spain meet a significant portion of their energy needs through oil and natural gas imports and are among the countries dependent on external sources for their energy supply (herp & Jewell, 2014; Winzer, 2012).

These risks include: increases in global energy prices, a rise in the frequency of sudden fluctuations in prices, rising import costs to meet energy needs, a widening current account deficit due to these costs, indirect supply disruptions caused by geopolitical factors, disruptions in supply chains, and uncertainties regarding access to energy resources (Sovacool & K., 2011; Winzer, 2012).

It is evident that Turkey meets a significant portion of its energy needs through external sources. The increasing diversity and widespread adoption of renewable energy sources make important contributions to boosting domestic energy production and reducing energy imports. In particular, the integration of solar, wind, and hydroelectric energy sources into the energy system supports the strengthening of energy supply security and constitutes a major component of it (Apergis & Payne, 2010).

Environmental and Economic Benefits

The widespread adoption of renewable energy sources not only contributes to energy security but also provides numerous benefits in terms of environmental sustainability. Reducing the use of fossil fuels has lowered greenhouse gas emissions and led to progress toward the goals set out in policies aimed at combating climate change (IPCC, 2023).

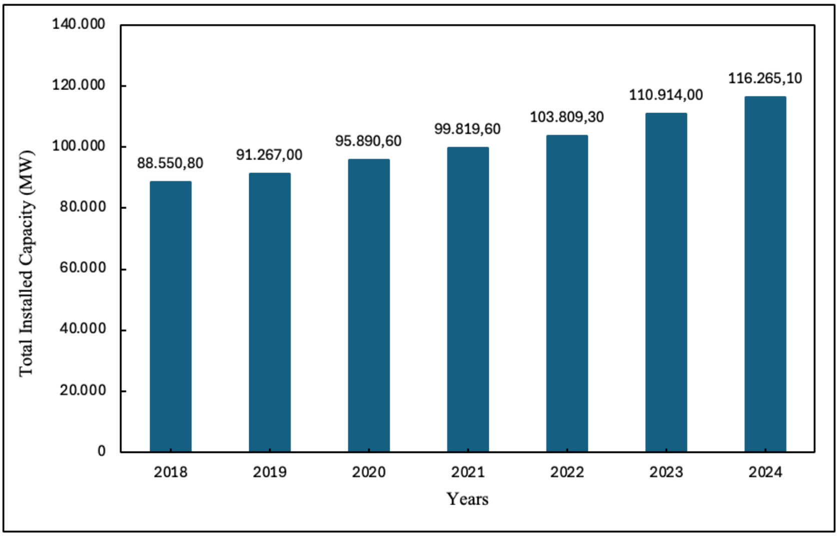

Figure 11. Trends in total installed renewable energy capacity in Turkey (2018–2024)

As seen in the installed capacity data, investments in renewable energy sources have a significant impact on energy efficiency and installed capacity. In this context, they help reduce energy production costs, create new job opportunities, and support sustainable economic growth (IEA, 2024; IRENA, 2024)

From a general perspective, the share of renewable energy sources in Turkey's energy supply structure has been increasing year by year, and this development is viewed as a key indicator of the energy transition process in terms of strengthening energy supply security, expanding energy resources, and supporting environmental sustainability.

CONCLUSION AND EVALUATION

Ensuring a secure, sustainable, and economically viable energy supply has become one of the fundamental objectives of energy policies in all countries today, particularly in Turkey. Rising energy demand, the fight against climate change, and concerns regarding energy supply security are increasingly highlighting the importance of renewable energy sources. In this context, an assessment of the changes in Turkey's energy supply structure reveals that the role of renewable energy sources in the energy system is steadily growing.

Data obtained within the scope of this study indicate that fossil fuels continue to account for a significant share of Turkey's energy supply, as they have since the past. An analysis of energy supply and consumption data reveals that natural gas, petroleum products, and coal hold the dominant positions in the energy system. In contrast, there have been significant increases in both the installed capacity and the share of renewable energy sources in electricity generation.

Renewable energy installed capacity, which stood at approximately 42,453 MW in 2018, rose to 69,456 MW by 2024, representing an increase of approximately 64%. The most notable development within this growth has been in solar energy. Solar energy installed capacity rose from 5,063 MW in 2018 to 20,232 MW in 2024, growing nearly fourfold. Similarly, while installed wind power capacity rose from 7,005 MW to 12,871 MW, a steady upward trend was also observed in biomass and geothermal energy sources. These developments demonstrate that renewable energy sources are assuming an increasingly important role in Turkey's energy transition process.

Electricity generation data also supports this transition: electricity generated from solar energy rose from 7,799.8 GWh in 2018 to 30,746.0 GWh in 2024. Meanwhile, electricity generation from wind energy rose from 19,949.2 GWh to 36,826.4 GWh during the same period. This result clearly demonstrates that renewable energy sources are playing an increasingly active role not only in installed capacity but also in energy production.

At the same time, energy import data indicates that Turkey's dependence on external sources for its energy supply persists, with energy imports rising from 115,792 thousand TEP in 2018 to 128,437 thousand TEP in 2024. Although energy exports increased during the same period, the gap between imports and exports remained high. This situation highlights that renewable energy sources are of strategic importance not only from an environmental perspective but also in terms of energy supply security and reducing external dependence.

As a result, Turkey's energy system is undergoing a transition in which fossil fuels continue to dominate but renewable energy sources are gaining increasing importance. In particular, the rapid transformations in solar and wind energy play a significant role in diversifying the energy supply and strengthening energy security. In the coming period, continuing investments in renewable energy, developing energy storage technologies, and utilizing domestic energy resources more effectively will play a critical role in Turkey's achievement of its sustainable energy transition goals.

ACKNOWLEDGMENTS

This book chapter was prepared based on a master's thesis currently being conducted at the Department of Environmental Engineering, Institute of Natural Sciences, Uludağ University. I would like to express my gratitude to my advisor for his insights, suggestions, and guidance in shaping this work; to the faculty members of the Department of Environmental Engineering for contributing to my academic development; and to all individuals and institutions that provided support throughout the research process.

REFERENCES

Acar, S., & Yeldan, A. E. (2016). Environmental impacts of coal subsidies in Turkey: A general equilibrium analysis. *Energy Policy*, *90*, 1–15. https://doi.org/10.1016/j.enpol.2015.12.003

Apergis, N., & Payne, J. E. (2010). Renewable energy consumption and economic growth: Evidence from a panel of OECD countries. *Energy Policy*, *38*(1), 656–660. https://doi.org/10.1016/J.ENPOL.2009.09.002

Arslan, H., Demirel, S. N., & Nomer, A. E. (2025). Renewable Energy Applications in University Campuses: Reducing Carbon Emissions Based On Solar Energy Systems (SES) at Mersin University. *Innovative Applications and Research Methods in Engineering Sciences*, 369–398. https://doi.org/10.5281/zenodo.18094294

Balat, M. (2006). Biomass energy and biochemical conversion processing for fuels and chemicals. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, *28*(6), 517–525. https://doi.org/10.1080/009083190927994

Bilgili, F., Koçak, E., Bulut, Ü., & Kuşkaya, S. (2017). Can biomass energy be an efficient policy tool for sustainable development? *RENEWABLE & SUSTAINABLE ENERGY REVIEWS*, *71*, 830–845. https://doi.org/10.1016/J.RSER.2016.12.109

Cherp, A., & Jewell, J. (2014). The concept of energy security: Beyond the four As. *Energy Policy*, *75*, 415–421. https://doi.org/10.1016/J.ENPOL.2014.09.005

Demirbaş, A. (2001). Biomass resource facilities and biomass conversion processing for fuels and chemicals. *Energy Conversion and Management*, *42*(11), 1357–1378. https://doi.org/10.1016/S0196-8904(00)00137-0

Demirbas, A. (2008). Importance of biomass energy sources for Turkey. *Energy Policy*, *36*(2), 834–842. https://doi.org/10.1016/J.ENPOL.2007.11.005

Dickson, M. H., & Fanelli, M. (2004). *What is geothermal energy?*

Ellabban, O., Abu-Rub, H., & Blaabjerg, F. (2014). Renewable energy resources: Current status, future prospects and their enabling technology. *Renewable and Sustainable Energy Reviews*, *39*, 748–764. https://doi.org/10.1016/J.RSER.2014.07.113

ETKB. (2024). *EİGM Raporları - T.C. Enerji ve Tabii Kaynaklar Bakanlığı*. https://enerji.gov.tr/eigm-raporlari

GWEC. (2024). *Global Wind Report 2024. Brussels, Belgium: Global Wind Energy Council.* https://www.gwec.net/hubfs/Website-2023/documents/GWEC-2024.pdf?hsLang=en

IEA. (2021). *Energy Policy Review Turkey 2021*. www.iea.org/t&c/

IEA. (2023). *World Energy Outlook 2023*. https://www.iea.org/reports/world-energy-outlook-2023

IEA. (2024). *Renewables 2024*. https://www.iea.org/reports/renewables-2024

IPCC. (2023). *AR6 Synthesis Report: Climate Change 2023*. https://www.ipcc.ch/report/ar6/syr/

IRENA. (2023). *RENEWABLE CAPACITY STATISTICS 2023 STATISTIQUES DE CAPACITÉ RENOUVELABLE 2023 ESTADÍSTICAS DE CAPACIDAD RENOVABLE 2023 About IRENA*. www.irena.org

IRENA. (2024). *RENEWABLE CAPACITY STATISTICS 2024*. www.irena.org

Kabir, E., Kumar, P., Kumar, S., Adelodun, A. A., & Kim, K. H. (2018). Solar energy: Potential and future prospects. *Renewable and Sustainable Energy Reviews*, *82*, 894–900. https://doi.org/10.1016/J.RSER.2017.09.094

Kaygusuz, K. (2010). Sustainable energy, environmental and agricultural policies in Turkey. *Energy Conversion and Management*, *51*(5), 1075–1084. https://doi.org/10.1016/J.ENCONMAN.2009.12.012

Kaygusuz, K. (2012). Energy for sustainable development: A case of developing countries. *Renewable and Sustainable Energy Reviews*, *16*(2), 1116–1126. https://doi.org/10.1016/J.RSER.2011.11.013

Lewis, N. S., & Nocera, D. G. (2006). Powering the planet: chemical challenges in solar energy utilization. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(43), 15729–15735. https://doi.org/10.1073/PNAS.0603395103

Lund, J. W., & Toth, A. N. (2021a). Direct utilization of geothermal energy 2020 worldwide review. *Geothermics*, *90*, 101915. https://doi.org/10.1016/J.GEOTHERMICS.2020.101915

Lund, J. W., & Toth, A. N. (2021b). Direct utilization of geothermal energy 2020 worldwide review. *Geothermics*, *90*. https://doi.org/10.1016/J.GEOTHERMICS.2020.101915

Mckendry, P. (2002). *Energy production from biomass (part 1): overview of biomass*.

REN21. (2024). *RENEWABLES 2024 GLOBAL STATUS REPORT*. https://www.ren21.net/gsr-2024/

Sagir, O., & Bahadir, A. (2017). Renewable energy potential and utilization in Turkey. *Journal of Engineering Research and Applied Science*, *6*(1), 577–582. https://www.journaleras.com/index.php/jeras/article/view/81

Smil, Vaclav. (2017). *Energy and civilization : a history*. 552. https://www.penguinrandomhouse.com/books/657402/energy-and-civilization-by-vaclav-smil/9780262338318

Sovacool, & K., B. (2011). The policy challenges of tradable credits: A critical review of eight markets. *Energy Policy*, *39*(2), 575–585. https://ideas.repec.org/a/eee/enepol/v39y2011i2p575-585.html

TEİAŞ. (2024). *TÜRKİYE ELEKTRİK ÜRETİM-İLETİM İSTATİSTİKLERİ*. https://www.teias.gov.tr/turkiye-elektrik-uretim-iletim-istatistikleri

Winzer, C. (2012). Conceptualizing energy security. *Energy Policy*, *46*, 36–48. https://doi.org/10.1016/J.ENPOL.2012.02.067

CHAPTER 9

Adsorption Process and its Applications in Environmental Engineering

***Hasan Kıvanç Yeşiltaş*[1] *& Güray Kılınççeker*[2]**

1. INTRODUCTION

Due to today's increasing population and industrialization activities, natural resources are rapidly depleting, and there is an uncontrolled increase in the type and quantity of pollutants. This increasing pollution load is damaging ecosystems, endangering receiving environments, and causing significant reductions in the amount of accessible clean water in the world (Dąbek & Nowak, 2019). Since the protection of water resources is among the fundamental problems of our time, the need for treatment technologies is increasing every day. Factors such as climate change and population density further deepen the existing water stress problem and push engineering disciplines to seek new solutions.

Environmental engineering is a branch of science that produces sustainable solutions using multidisciplinary methods to prevent negative problems caused by human activities and to reduce pollution at its source. Proactive approaches such as sustainable water management, urban ecological planning, greywater reuse, and rainwater harvesting form the basis of this discipline (United States Environmental Protection Agency [EPA], 2016). In line with these objectives, new treatment systems are being developed across a wide range of areas, from mitigating urban heat island effects to protecting the health of aquatic ecosystems.

Adsorption is a fundamental physicochemical process based on the principle of pollutants in liquid or gaseous environments adhering to solid surfaces, offering very high removal efficiency even at low concentrations. This process, carried out with adsorbent materials that have a high surface area and porous structure, is quite popular in environmental engineering applications due to its wide range of applications against different types of pollutants (Crini, 2006). Thanks to advancements in materials science, new generation adsorbents, ranging from agricultural waste to advanced synthetic polymers, are creating sustainable and economical alternatives in environmental pollution control.

[1] Dr., Cukurova University Faculty of Engineering, Department of Environmental Engineering, ORCID: 0000-0003-3331-3209
[2] Prof. Dr., Cukurova University Faculty of Arts and Sciences Chemistry Department, ORCID: 0000-0003-3030-4518

2. ADSORPTION METHOD

Industrial activities, particularly in sectors such as textiles, paper, plastics, and cosmetics, consume large amounts of water in their production processes and release highly colored, toxic wastewater into the environment. The synthetic dyes in this wastewater not only have toxic and carcinogenic effects but also reduce the permeability of sunlight in aquatic environments, hindering photosynthetic reactions (Kaya, 2021; Enache et al., 2023). These pollutants, which are of synthetic origin and have complex aromatic structures, are quite difficult to biodegrade using traditional or aerobic methods. At this point, the adsorption process stands out as one of the most effective and widely preferred methods in wastewater treatment due to its significant advantages such as simplicity of design, high removal efficiency, high resistance to toxic substances, and being an economical alternative (Cebeci and Şentürk, 2020; Akmeşe and Tosun Satır, 2025).

The adsorption mechanism is fundamentally based on the principle of dissolved pollutant molecules (adsorbate) adhering to the surface of a solid material (adsorbent). This process can occur as physical adsorption through weak Van der Waals forces via the interaction of porous structure and surface chemistry, or as strong chemical adsorption through electron exchange. Isotherm and kinetic models are critical in analyzing the process and understanding the mechanism. Isotherm models such as Langmuir, Freundlich, and Temkin are used to determine the maximum adsorption capacity at equilibrium and the properties of the surface; while pseudo-first-order and pseudo-second-order kinetic models are applied to determine the adsorption rate and rate-limiting steps (e.g., mass transfer, diffusion, or chemical reaction) (Bingül, 2021; Gümüş and Gümüş, 2024).

The efficiency of the adsorption process is directly dependent on various operating parameters affecting the system's operation. Contact time, initial dye concentration, adsorbent dosage, temperature, and solution pH are among the most important parameters. pH, in particular, plays a direct role in creating electrostatic attraction or repulsion forces by determining the surface charge of the adsorbent and the degree of ionization of the dye molecules (Rahdar et al., 2018). For example, adsorbents with appropriately charged surfaces perfectly retain cationic dyes at a specific pH level, while anionic dyes exhibit high removal efficiencies in a different pH range (usually acidic) (Kaya, 2021; Akmeşe and Tosun Satır, 2025). Figure 1 shows the typical change in dye removal efficiency depending on the contact time in the adsorption process. At the beginning of the process, a very rapid mass transfer and removal occurs due to the complete emptiness of the active sites on the adsorbent surface; then, as saturation is approached, the curve flattens and equilibrium is reached (Saghir et al., 2022).

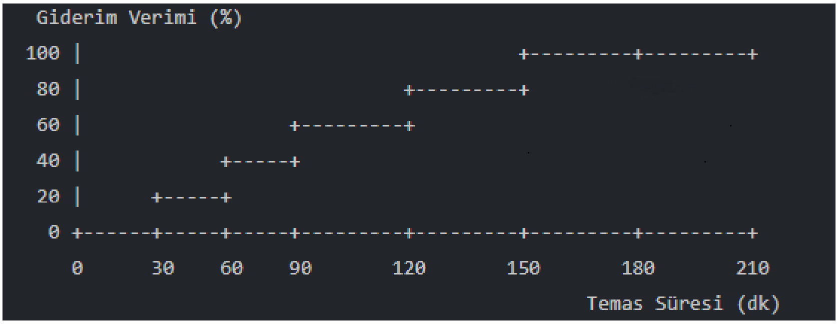

Figure 1: Typical equilibrium graph depending on contact time in the adsorption process (Saghir et al., 2022)

Although activated carbon is the most common adsorbent in commercial applications, the high production and reactivation costs have led researchers to find sustainable, low-cost, and environmentally friendly alternatives. Agricultural wastes such as pistachio shells, hawthorn seeds, waste green tea leaves, banana peels, and pine pollen contain abundant functional groups such as hydroxyl, carboxyl, and phenol on their surfaces due to their cellulose, hemicellulose, and lignin content. This makes them excellent adsorbents for dye adsorption (Shaikhiev et al., 2023; Bayram et al., 2023). Furthermore, these agricultural wastes can be transformed into high-capacity adsorbents that can compete with commercial products by subjecting them to thermal treatments such as carbonization (biochar production) or by improving their surface area, pore diameter, and reactive properties through acid/base modifications (Baytar et al., 2018; Saghir et al., 2022).

3. ENVIRONMENTAL ENGINEERING SCIENCE

Environmental engineering is a dynamic discipline that aims to protect natural resources, efficiently remove environmental pollutants, and design sustainable living spaces in harmony with nature. Today, instead of focusing on disposal after pollution occurs, visionary sustainability practices such as the widespread adoption of green roof applications, the protection of urban ecological balances, greywater recovery, and rainwater harvesting are coming to the forefront (EPA, 2016). This framework aims to alleviate the burden on natural ecosystems and to safely transfer limited freshwater resources to future generations.

When dealing with micropollutants, complex and persistent dyes, and heavy metals, where traditional water and wastewater treatment methods are insufficient, advanced treatment technologies such as adsorption come into play. The adsorption process has found a wide range of applications in engineering due to its operational flexibility, high pollutant selectivity, low energy consumption, and relatively easy field applicability compared to conventional sedimentation or biological treatment systems (Dąbek & Nowak, 2019). The low initial investment and low requirement for qualified personnel in the system also create a significant economic advantage for wastewater treatment plants.

The circular economy principles adopted globally have become a fundamental evaluation criterion for modern environmental engineering practices. In this context, the adsorption method, using adsorbents obtained from agricultural or food industry wastes, perfectly aligns with the philosophy of creating added value from waste in pollution removal (Kainth et al., 2024). Biochars produced by the carbonization or activation of agricultural wastes reduce the burden on landfills while simultaneously supporting system sustainability as environmentally friendly and low-carbon footprint materials in treatment technologies.

4. ADSORPTION APPLICATIONS IN ENVIRONMENTAL ENGINEERING

In the field of environmental engineering, the adsorption process is widely used to control pollution in all receiving environments such as water, wastewater, air, and soil. Adsorption processes can be used to remove toxic components or biodegradable pollutants at very low concentrations from the environment where traditional treatment methods (biological treatment, chemical precipitation, etc.) are insufficient (Dąbek & Nowak, 2019). The adsorption method is directly linked to the philosophy of sustainability, not only for pollution removal but also for obtaining valuable products from waste and generating economic gains (United States Environmental Protection Agency [EPA], 2016).

The removal of heavy metals (Pb^{2+}, Cd^{2+}, Cr^{6+}, As^{3+}, Cu^{2+}, Ni^{2+}) from aquatic ecosystems is of critical importance for public health to prevent the risk of accumulation in living organisms. In recent years, low-cost materials have emerged as an alternative to the high cost of commercial activated carbon in the treatment of these ions, which are found in high concentrations in industrial mining or metal plating discharges (Valix et al., 2001). Lignocellulosic agricultural wastes (e.g., walnut shells, hazelnut shells, corn cobs) with increased surface area and ion exchange capacity through chemical treatments, chitosan-based biopolymers, modified zeolites, and polyaniline-coated magnetic nanoparticles (with efficiencies above 90-95%) exhibit high performance in heavy metal removal (Srinivasan et al., 2019; Zhang et al., 2018). The treatment of complex dyes originating from the textile, paper, leather, and cosmetics industries is another key area where adsorption is most intensively applied. The fact that dyes such as methylene blue, Congo red, and rhodamine B are extremely resistant to photochemical degradation and biological degradation makes adsorption essential (Crini, 2006). In this field, acid or base-activated waste sludge biochars, olive pit activated carbons, graphene oxide-supported composite materials, and mesoporous silicas stand out in protecting the aesthetic and ecological structure of waters with their high color removal efficiencies (Ahmad et al., 2021; Gupta & Suhas, 2009).

Micropollutants (pharmaceuticals, antibiotics, endocrine-disrupting chemicals, and personal care products), which have become one of the most current focal points of environmental engineering in recent years, threaten aquatic life even at nanogram/liter (ng/L) levels. The inability to treat these pollutants in conventional facilities due to their complex molecular structures has accelerated the development of advanced adsorbent technologies (Mohammad & Al-Ghouti, 2020). In the specific removal of micropollutants, metal-organic frameworks with very high porosity and tunable surface chemistry, carbon nanotubes, and cyclodextrin-based cross-linked synthetic polymers are used with their superior selectivity capacities (Suresh et al., 2018).

The control of ammonia (NH_4^+) and phosphate (PO_4^{3-}) ions, which cause eutrophication (algal blooms) in aquatic environments, is also successfully carried out by adsorption. Here, not only pollution removal but also nutrient recovery is valuable (Zouboulis et al., 2004). For this purpose, natural zeolites such as clinoptilolite, clay minerals enriched with zirconium or lanthanum, and magnesium-modified agricultural waste biochars are preferred; Thus, adsorbents that reach saturation at the end of the process are evaluated as slow-release fertilizers in agricultural areas, creating a complete sustainability cycle (Wang & Peng, 2010).

The adsorption method is not only used in the removal of substances in liquids but also in gas phase and air pollution control. Adsorption is a fundamental technology in the capture of volatile organic compounds in industrial chimney emissions, odor components (H_2S, NH_3) from landfill leachate pools and wastewater treatment plants, and greenhouse gases (CO_2) that trigger climate change (EPA, 2016). In the control of these emissions, granular activated carbon, amine-functionalized mesoporous silica networks, and silver-impregnated synthetic zeolite filters generally play a leading role in gas purification (Dąbek & Nowak, 2019).

In-situ remediation of contaminated agricultural areas and bottom sediments is an innovative and highly effective application area of adsorption in soil ecosystems. Soils contaminated with heavy metal or petroleum derivative spills, and which cannot be physically/economically excavated and removed, are directly mixed with natural clay minerals such as bentonite and montmorillonite, or high-temperature pyrolyzed woody materials (Wang & Peng, 2010). This in-situ application stabilizes the mobility and uptake of pollutants by plants, preventing them from entering groundwater (Valix et al., 2001).

Industrial-scale adsorption applications involve full-scale hydraulic reactor designs where laboratory results are transferred to field conditions. Through fixed-bed columns, fluidized-bed reactors, and continuously stirred tank reactors, the goal is to ensure process water recovery and discharge standards (Dąbek & Nowak, 2019). In these full-scale systems, ion-exchange synthetic polymeric resins and extruded activated carbons, which are resistant to hydraulic pressure and mechanical friction, pressure drop, and have long regeneration cycles, are particularly preferred in terms of operational life (EPA, 2016). Summaries of data from all relevant applications are presented in Table 1.

Table 1. Examples of adsorbents and summaries of studies according to application areas.

Application Area	**Pollutant Examples**	**Adsorbents Used**	**Related Literature**
Heavy Metal Removal	Pb^{2+}, Cd^{2+}, Cr^{6+}, As^{3+}, Cu^{2+}, Ni^{2+}	Lignocellulosic agricultural wastes (walnut/hazelnut shells, corn cobs), chitosan-based biopolymers, modified zeolites, polyaniline-coated magnetic nanoparticles	Valix et al. (2001); Srinivasan et al. (2019); Zhang et al. (2018)
Dye Removal	Methylene Blue, Congo Red, Rhodamine B	Acid/base activated waste sludge biochars, olive stone activated carbons, graphene oxide composites, mesoporous silicas	Crini (2006); Ahmad et al. (2021); Gupta & Suhas (2009)
Micropollutant Removal	Pharmaceuticals, antibiotics, endocrine disruptors, personal care products	Metal-Organic Frameworks (MOFs), carbon nanotubes, cyclodextrin-based cross-linked synthetic polymers	Mohammad & Al-Ghouti (2020); Suresh et al. (2018)
Nutrient Control	Ammonia (NH_4^+), phosphate (PO_4^{3-})	Natural zeolites (clinoptilolite), Zr/La enriched clay minerals, Mg-modified agricultural waste biochars	Zouboulis et al. (2004); Wang & Peng (2010)
Gas Phase and Air Pollution	Volatile organic compounds (VOCs), odorants (H_2S, NH_3), greenhouse gases (CO_2)	Granular Activated Carbon (GAC), amine-functionalized mesoporous silicas, silver-impregnated synthetic zeolite filters	EPA (2016); Dąbek & Nowak (2019)
Soil and Sediment Remediation	Heavy metals, petroleum derivatives	Natural clay minerals (bentonite, montmorillonite), high-temperature pyrolyzed woody biochars	Wang & Peng (2010); Valix et al. (2001)
Industrial Full-Scale Systems	Process water and discharge pollutants	Ion exchange synthetic polymeric resins, extruded activated carbons	Dąbek & Nowak (2019); EPA (2016)

5. RESULTS AND RECOMMENDATIONS

Adsorption is a complementary, flexible, economical, and environmentally friendly system for removing a wide variety of organic, inorganic, and toxic pollutants where conventional biological treatment processes are insufficient. Studies conducted at both laboratory and industrial scales demonstrate that the wide variety of adsorbents, ranging from natural clay minerals to materials derived from peanut shells, nanotechnology materials, and modified zeolites, gives the method a unique adaptability. This multi-dimensional structure allows adsorption to be used in solving current and future environmental problems.

Despite the practicality and environmentally friendly approaches offered by the method, some operational and economic limitations arise, particularly when the scale of application is increased, creating a disadvantage. The regeneration of saturated adsorbents without harming the environment, without losing their structural integrity, and with low energy costs directly affects the operating budget of the process (Gupta & Suhas, 2009). However, in complex industrial wastewaters containing multiple pollutants, the competition between pollutants for active sites on the adsorbent surface makes predicting interactions difficult in system design and necessitates further research in this area.

Future research should focus on enriching the surface functional groups of adsorbents, particularly those derived from low-cost agricultural wastes, to increase their capacity. Simultaneously, innovative materials developed at the laboratory level should be supported by pilot-scale field applications to test their performance under real wastewater conditions. Furthermore, the current technical infrastructure is important because computer-aided applications can improve and enhance the efficiency of existing research.

REFERENCES

Ahmad, M. A., vd. (2021). Recent advances in adsorption of organic pollutants using graphene-based materials: A review. Journal of Environmental Chemical Engineering, 9(6), 106064.

Akmeşe, B., & Tosun Satır, İ. (2025). Tarımsal Atıklar Kullanılarak Tekstil Atık Sularındaki Boyar Maddelerin Adsorpsiyonla Arıtılması. KSÜ Tarım ve Doğa Dergisi, 28(2), 535-543.

Bayram, O., Moral, E., & Göde, F. (2023). İğde Çekirdeklerinden Elde Edilen Biyokömür Kullanılarak Sulu Çözeltiden Kristal Viyole Boyarmaddesinin Uzaklaştırılması. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 13(1), 448-457.

Baytar, O., Şahin, Ö., Saka, C., & Ağrak, S. (2018). Characterization of Microwave and Conventional Heating on the Pyrolysis of Pistachio Shells for the Adsorption of Methylene Blue and Iodine. Analytical Letters, 51(14), 2205-2220.

Bingül, Z. (2021). The Use of Waste Green Tea Leaves for Crystal Viyole Adsorption: Kinetic, Equilibrium and Thermodynamics Studies. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 11(4), 2645-2659.

Cebeci, M. S., & Şentürk, İ. (2020). Tarımsal Atık Materyal Kullanılarak Sucul Çözeltiden Chrysoidine Y Boyasının Giderimi. International Journal of Multidisciplinary Studies and Innovative Technologies, 4(1), 18-28.

Crini, G. (2006). Non-conventional low-cost adsorbents for dye removal: A review. Bioresource Technology, 97(9), 1061–1085.

Dąbek, L., & Nowak, I. (2019). Adsorption processes in water treatment: A review. Environmental Science and Pollution Research, 26(6), 5134–5154.

Enache, A.-C., Cojocaru, C., Samoila, P., Ciornea, V., Apolzan, R., Predeanu, G., & Harabagiu, V. (2023). Adsorption of Brilliant Green Dye onto a Mercerized Biosorbent: Kinetic, Thermodynamic, and Molecular Docking Studies. Molecules, 28, 4129.

Foo, K. Y., & Hameed, B. H. (2010). Insights into the modeling of adsorption isotherm systems. Chemical Engineering Journal, 156(1), 2–10.

Gupta, V. K., & Suhas. (2009). Application of low-cost adsorbents for dye removal – A review. Journal of Environmental Management, 90(8), 2313–2342.

Gümüş, D., & Gümüş, F. (2024). Typha latifolia ile Sulu Çözeltiden Kristal Viyole Boyasının Adsorpsiyonu; Doğrusal ve Doğrusal Olmayan İzoterm, Kinetik ve Dizayn. Karadeniz Fen Bilimleri Dergisi, 14(4), 2227-2243.

Hajialigol, S., & Masoum, S. (2019). Application of agricultural wastes as adsorbents. Journal of Environmental Chemical Engineering, 7(1), 102-115.

Husien, M., Ali, A., & Mohammed, S. (2022). Adsorption of dyes using bio-based activated carbons. Journal of Water Process Engineering, 46, 102582.

Kainth, S., Singh, R., & Kumar, A. (2024). Agricultural waste-based adsorbents for wastewater treatment. Environmental Science and Pollution Research, 31(2), 1234-1250.

Kaya, A. (2021). Adsorption studies of Cibacron Blue onto both untreated and chemically treated pistachio shell powder from aqueous solutions. Separation Science and Technology, 56(15), 2546-2561.

Mohammad, A. W., & Al-Ghouti, M. A. (2020). Nanomaterials in water and wastewater treatment: Applications and challenges. Chemical Engineering Journal, 393, 124619.

Rahdar, S., Shikhe, L., & Ahmadi, S. (2018). Removal of Reactive Blue 19 dye using a combined sonochemical and Modified Pistachio Shell adsorption processes from aqueous solutions. Iranian Journal of Health Sciences, 6(3), 8-20.

Saghir, S., Pu, C., Fu, E., Wang, Y., & Xiao, Z. (2022). Synthesis of high surface area porous biochar obtained from pistachio shells for the efficient adsorption of organic dyes from polluted water. Surfaces and Interfaces, 34, 102357.

Santos, S., & Puna, J. (2016). Agricultural wastes for wastewater treatment. Proceedings of the 4th International Conference on Sustainable Solid Waste Management, Cyprus, 1-12.

Shaikhiev, I. G., Kraysman, N. V., & Sverguzova, S. V. (2023). Review of Pistachio (Pistacia) Shell Use to Remove Pollutants from Aqua Media. Biointerface Research in Applied Chemistry, 13(4), 389.

Srinivasan, A., vd. (2019). Adsorption of heavy metals from aqueous solutions using nanomaterials: A review. Environmental Chemistry Letters, 17(1), 107–122.

Suresh, S., vd. (2018). Advances in nanotechnology for water purification and environmental remediation: A review. Environmental Nanotechnology, Monitoring & Management, 9, 1–17.

Thamer, A., AL-Juboori, R. A., & AL-Furaiji, M. (2023). Modeling of adsorption isotherms and kinetics. Chemical Engineering Research and Design, 189, 55-67.

United States Environmental Protection Agency [EPA]. (2016). Technologies for the removal of pollutants from water and wastewater (EPA/600/R-16/234). Washington, DC: Author.

Valix, M., vd. (2001). Adsorption of heavy metals by biomass and waste-derived activated carbons. Environmental Science & Technology, 35(14), 2911–2917.

Wang, S., & Peng, Y. (2010). Natural zeolites as effective adsorbents in water and wastewater treatment. Chemical Engineering Journal, 156(1), 11–24.

Zhang, W., vd. (2018). Advances in magnetic adsorbents for water treatment: A review. Chemical Engineering Journal, 337, 310–323.

Zhang, W., Shen, Y., & Wang, H. (2021). Advances in magnetic adsorbents for water treatment. Chemical Engineering Journal, 404, 126589.

Zouboulis, A. I., vd. (2004). Adsorption of phosphorus from wastewater by natural and synthetic zeolites. Water Science and Technology, 49(11-12), 177–184.

CHAPTER 10

Current Studies On Ion Removal Using Peanut Shells

Hasan Kıvanç Yeşiltaş[1] *& Behzat Balcı* [2]

1. IONS: THEIR IMPORTANCE FOR LIVING ORGANISMS AND THEIR ENVIRONMENTAL IMPACTS

Atoms or molecules that have gained a positive (cation) electrical charge by donating electrons or a negative (anion) electrical charge by accepting electrons are called ions. Ions are extremely important for the vital activities of living organisms. For example, the maintenance of intracellular and extracellular fluid balance, muscle contraction, and the transmission of nerve signals to the brain in living organisms are carried out through electrochemical processes managed by sodium (Na^{+}), potassium (K^{+}), calcium (Ca^{2+}), and chloride (Cl^{-}) ions. In addition, the presence of magnesium (Mg^{2+}) and zinc (Zn^{2+}) ions is necessary for cellular respiration and biochemical reactions to occur within living organisms (Nelson and Cox, 2017; Guyton and Hall, 2016; Campbell et al., 2015).

Ions are extremely valuable for the sustainability of ecosystems and the uninterrupted functioning of their material cycles. Nitrate (NO^{3-}), ammonium (NH^{4+}), and phosphate (PO_4^{3}-) ions, dissolved in soil and water, are absorbed by plants and phytoplankton, directly supporting primary production which forms the basis of the food chain (Begon, Howarth, and Townsend, 2014). Inorganic compounds in ionic form are used in the synthesis of proteins, DNA, RNA, and ATP, the cellular energy molecule, in producer organisms, thus enabling plant growth (Taiz, Zeiger, Møller, and Murphy, 2015). Therefore, ionization processes are fundamental to the matter cycles occurring in ecosystems and living organisms (Schlesinger and Bernhardt, 2013).

In addition to the vital necessity of ions, the presence of some heavy metal ions in ecosystems leads to excessive accumulation in nature, causing negative impacts on the environment and public health. Heavy metal ions such as lead (Pb^{2+}), cadmium (Cd^{2+}), and mercury (Hg^{2+}), particularly originating from mining and industrial waste, accumulate in living tissues over time via the food chain when they mix with water and soil resources (Alloway, 2013; Baird and Cann, 2012). This leads to toxic effects in humans, immune system collapse, serious neurological damage, kidney failure, and cancer cases. Furthermore, the

[1] Dr., Cukurova University Faculty of Engineering, Department of Environmental Engineering, ORCID: 0000-0003-3331-3209
[2] Prof. Dr., Çukurova University Faculty of Engineering, Department of Environmental Engineering, ORCID: 0000-0002-4636-4235

environmental problem of eutrophication occurs when agricultural products containing nitrogen and phosphorus, and domestic wastewater, reach receiving environments (Tchobanoglous et al., 2014). Eutrophication causes the rapid depletion of dissolved oxygen in water, mass deaths of aquatic organisms, and the release of cyanobacterial toxins into the water, preventing these resources from being used for drinking or recreational purposes (Smith and Schindler, 2009).

2. ION REMOVAL METHODS

The removal of toxic ions and heavy metals that reach receiving environments as a result of industrial developments and human activities is crucial for the protection of the environment and public health (Naguib et al., 2025; Romero et al., 2004). A wide variety of processes, such as chemical precipitation, ion exchange, membrane filtration, and electrocoagulation, are used to remove ions from water and wastewater (Chen et al., 2025; Jiang and Jiang, 2013). Among these methods are processes that allow for the recovery of metals from the water, but large water bodies on an industrial scale present limitations and usage problems for these processes. The high installation and operating costs of conventional treatment systems are a primary obstacle to their usability (Dobi-Brice et al., 2023; Hu and Qiu, 2019). Furthermore, wastewater treatment methods can generate toxic by-products and sludge that are difficult to dispose of and cause secondary environmental pollution (Yi et al., 2018; Zhai, 2024). Considering all these factors together, researchers today are evaluating the development of more practical, environmentally friendly, and cost-effective alternatives to traditional methods (Duan et al., 2023; Moezzifar et al., 2025).

To reduce these negative aspects of traditional methods, the adsorption process, which can be used as an environmentally friendly and lower-cost alternative, has been used in ion removal in recent years (Erasga et al., 2024; Li and Zhai, 2023). The adsorption method offers various advantages due to its design simplicity and ease of application, and it is also advantageous because it does not require high operating costs (Mesci, 2013). In recent years, the use of agricultural wastes abundant in nature, such as peanut shells, as biosorbents has attracted attention (Phung et al., 2023; Yu et al., 2020). The effective management of solid waste is achieved during the evaluation of such lignocellulosic wastes as adsorbents, directly contributing to the circular economy (Liu et al., 2022; Mazzola et al., 2025).

3. PEANUT SHELLS AND THEIR USE AS BIOSORBENT

Peanut (*Arachis hypogaea*) is an important agricultural plant that can grow to a height of 20 to 90 cm, has leaves composed of 2 or 3 pairs of membranous oval leaflets, and its fruits ripen 3 to 5 cm below the ground (Dobi-Brice et al., 2023). It is produced on a large scale in countries such as China, Argentina, and Vietnam (Chen et al., 2025; Mazzola et al., 2025; Phung et al., 2023). In China alone, approximately 15 million tons of peanuts are produced annually, resulting in around 5 million tons of peanut shells as an industrial byproduct (Li and Zhai, 2023; Liu, Zhou et al., 2022). The shell constitutes approximately 25% to one-third of the total weight of whole peanuts (Erasga et al., 2024; Mazzola et al., 2025).

Peanut shells are a rich lignocellulosic agricultural waste containing biopolymers such as cellulose (approximately 40-45% or 60% crude fiber), hemicellulose (10.1% - 14%), and lignin (26-28%), as well as crude protein and polyphenols (Li and Zhai, 2023; Phung et al., 2023). This lignocellulosic biomass contains a wide variety of functional groups such as hydroxyl, carboxyl, amine, amide, carbonyl, and ether (Li and Zhai, 2023; Mazzola et al., 2025; Yu et al., 2020). Traditionally, a small portion of this waste is used as animal feed, fertilizer, or fuel. However, a large portion of it is not utilized and is directly burned, buried in the ground, or discarded as waste (Duan et al., 2023; Li and Zhai, 2023; Liu, Zhou et al., 2022; Mazzola et al., 2025; Zhai, 2024). This situation leads to the inappropriate utilization of a product that should be economically reused (Duan et al., 2023; Zhai, 2024). Therefore, the environmentally friendly and sustainable utilization of agricultural waste is crucial for sustainability.

By using peanut shells as adsorbents, the waste is recycled while another form of pollution is reduced/prevented (Chen et al., 2025). With their natural porous structures, large specific surface areas, and rich functional groups, low-cost biosorbent synthesis is achieved for the removal of toxic heavy metals such as lead, copper, zinc, nickel, and cadmium from water and wastewater (Duan et al., 2023; Mazzola et al., 2025; Mesci, 2013). To increase the adsorption capacity and stability of the synthesized material, raw peanut shells can be used directly, or they can be modified with chemicals such as acid, base, or zinc chloride ($ZnCl_2$) (Hu and Qiu, 2019; Zhai, 2024). Furthermore, they can be converted into high-performance products by pyrolysis in an oxygen-free environment (Chen et al., 2025; Erasga et al., 2024). These modifications performed on raw peanut shells increase the active binding sites on the surface of the material, making it possible to improve treatment efficiency (Mazzola et al., 2025; Yu et al., 2020).

4. CURRENT STUDIES ON ION REMOVAL USING PEANUT SHELLS

When evaluating literature studies on the use of peanut shells in the removal of pollutants from aqueous solutions, there are studies where the material is used directly (raw) without any chemical processing. Mesci (2013) investigated the adsorption of Cu^{2+} and Zn^{2+} with raw peanut shells dried at 100°C and ground to a size of 0.5 mm. The researcher achieved efficiencies of 95% and 87%, respectively, and performed statistical evaluation using Artificial Neural Networks and simplex algorithm instead of traditional isotherm and kinetic models (Mesci, 2013). Similarly, Dobi-Brice et al. (2023) investigated the removal of Cd^{2+} using natural peanut shells ground to a size of 250-500 μm. Determining that the adsorption processes conform to the Freundlich isotherm and pseudosecond-order kinetic model, they reported a maximum adsorption capacity of 4.56 mg/g (Dobi-Brice et al., 2023). Mazzola et al. (2025) reported that they obtained a capacity of 1.34 mg/g and an efficiency of 84% in their Zn(II) study using raw peanut shells. The researchers determined that the adsorption processes conform to the Freundlich isotherm and pseudosecond-order kinetics (Mazzola et al., 2025).

Numerous studies exist in the literature where peanut shells are converted into biochar by thermal treatment and their surfaces are modified with various metals in order to increase the removal capacity of the material. Hu and Qiu (2019) investigated Pb(II) removal with the material they obtained by pyrolyzing the shells they kept in a 2% K_2CO_3 solution at 873 K. As a result of the study, they reported that the adsorption followed the Langmuir isotherm model by obtaining a capacity value of 111.11 mg/g (Hu and Qiu, 2019). Liu et al. (2022) studied the removal of Pb(II) and Cd(II) with a material synthesized from peanut shells using the chemical $MnCl_2 \cdot 4H_2O$ by applying a combustion process at 400°C. The researchers reported that the adsorption processes followed the Langmuir isotherm and pseudosecond-order kinetics, and that the maximum capacity values they determined for Pb(II) and Cd(II) adsorption were 164.5 mg/g and Cd(II) 36.7 mg/g, respectively (Liu et al., 2022-a). In another study conducted by researchers, they investigated phosphorus removal using peanut shells followed by pyrolysis at 550°C and then synthesizing a material using $MgCl_2$ with the aid of ultrasound. The researchers reported reaching a maximum capacity of 150.16 mg/g and a removal efficiency of 89% (Liu et al., 2022-b). Chen et al. (2025) studied the removal of Pb(II), Cu(II), and Ni(II) using a material synthesized with Fe-doped peanut shells at 750°C. The researchers reported that the kinetics were consistent with a pseudo-first-order kinetic model, reaching capacity values of 22.5, 20.1, and 3.6 mg/g, respectively. They also determined that the adsorption process could be explained by the Freundlich isotherm model for Pb(II) and the Langmuir-Freundlich isotherm model for the other ions (Chen et al., 2025). Duan et al. (2023), who studied the adsorption of Pb(II) with a material synthesized from peanut shells by modifying it with $BaCl_2$ at 750°C, determined a capacity value of 379.33 mg/g and reported that the adsorption followed the Freundlich isotherm and pseudosecond-order kinetic model (Duan et al., 2023). Furthermore, Naguib et al. (2025), who performed chemical activation with 1M NaOH after physical activation at 800°C, obtained a purification efficiency of over 99% for Pb(II) and Fe(II). The researchers also reported that the adsorption isotherms followed the Langmuir and Freundlich isotherm models, respectively, and determined that the kinetic processes could be explained by a pseudosecond-order kinetic expression (Naguib et al., 2025).

Romero et al. (2004) studied the removal of Cu(II), Zn(II), Ni(II), and Cr(VI) using peanut shells activated with phosphoric acid. The researchers determined the Cr(VI) removal capacity of the material to be approximately 100.9 mg/g. They also reported that the cation removal order was Cu(II) > Zn(II) > Ni(II) and that the adsorption process followed Freundlich isotherms (Romero et al., 2004). Moezzifar et al. (2025) modified the material, which was activated with NH_4Cl at 240°C and given magnetic properties with Fe_3O_4, with quercetin. The researchers, who studied the removal of Ni(II) with the synthesized adsorbent, reported that they reached a maximum capacity of 333.3 mg/g and a removal efficiency of 99%. They also reported that the adsorption mechanism followed the Langmuir isotherm and pseudosecond-order kinetics (Moezzifar et al., 2025). Summaries of the relevant studies are given in Table 1.

Table 1. Summary data from current studies in the literature

Material Used	Material Preparation	Removed Ion	Removal Capacity and Efficiency	Isotherm and Kinetic Model	Reference
Peanut Shell (Raw)	Drying at 100°C, 0.5 mm	Cu^{2+} Zn^{2+}	95% (Cu^{2+}) and 87% (Zn^{2+})	Artificial Neural Networks (ANN), Modeling with simplex algorithm	Mesci, 2013
Peanut Shell Biochar (Modified - MnO_x impregnated)	Pyrolysis at 400°C under N_2, modified with NaOH, H_2O_2, and $MnCl_2 \cdot 4H_2O$	Pb(II) Cd(II)	164.5 mg/g for Pb(II), 36.7 mg/g for Cd(II)	Langmuir isotherm, Pseudo-second-order kinetics	Liu et al., 2022
Peanut Shell Biochar (Modified - Fe loaded)	Soaking in $FeCl_3$, pyrolysis at 450-750°C	Pb(II) Cu(II) Ni(II)	Pb(II): 22.5 mg/g, Cu(II): 20.1 mg/g, Ni(II): 3.6 mg/g	Freundlich isotherm for Pb(II), Langmuir-Freundlich isotherm for Cu(II) and Ni(II). Pseudo-first-order kinetics	Chen et al., 2025
Peanut Shell Biochar (Modified - Ba loaded)	Pyrolysis up to 750°C, modified with $BaCl_2$	Pb(II)	379.33 mg/g	Freundlich isotherm, Pseudo-second-	Duan et al., 2023

				order kinetics	
Peanut Shell (Raw)	250-500 µm size	Cd^{2+}	4.56 mg/g	Freundlich isotherm, Pseudo-second-order kinetics	Dobi-Brice et al., 2023
Peanut Shell Biochar (Modified – $MgCl_2$ loaded)	Pyrolysis at 550°C under N_2, $MgCl_2$ with ultrasound	Phosphorus	150.16 mg/g, 89.2% efficiency	Langmuir isotherm, Pseudo-second-order kinetics	Liu et al., 2022
Peanut Shell Biochar (Modified)	12 hours in 2% K_2CO_3 solution, 3 hours at 873 K	Pb(II)	111.11 mg/g	Langmuir isotherm	Hu & Qiu, 2019
Activated Peanut Shell / Fe_3O_4 Composite (Modified - Quercetin)	240°C with NH_4Cl, magnetized with Fe_3O_4, modified with Quercetin	Ni(II)	333.3 mg/g, 99% efficiency	Langmuir isotherm, Pseudo-second-order kinetics	Moezzifar et al., 2025
Activated Carbon (Modified)	Activation with phosphoric acid	Cu(II) Zn(II) Ni(II) Cr(VI)	Cr(VI): 100.9 mg/g	Freundlich and BET-type isotherms	Romero et al., 2004
Peanut Shell Biochar (Modified - Physical and Chemical)	800°C, chemical activation with 1M NaOH	Pb(II) Fe(II)	Pb: 99.76%, Fe: 99.94%	Langmuir and Freundlich isotherms, Pseudo-second-order kinetics	Naguib et al., 2025
Peanut Shell (Raw)	-	Zn(II)	1.34 mg/g, 84% efficiency	Freundlich isotherm, Pseudo-second-order kinetics	Mazzola et al., 2025
Peanut Shell (Modified – with $ZnCl_2$)	120°C in an autoclave, activated with 10% $ZnCl_2$ for 6 hours	Ag^+	94.96 mg/g, 94.96% efficiency	Freundlich isotherm, Pseudo-second-order kinetics	Zhai, 2024

5. RESULTS AND RECOMMENDATIONS

Heavy metals and toxic ions discharged into receiving environments as a result of industrial and agricultural activities disrupt ecosystem balance and seriously threaten human health through bioaccumulation. Against the main disadvantages of traditional treatment methods used in ion removal from water and wastewater, such as high cost and sludge formation, the use of peanut shell-derived adsorbents stands out as an innovative, environmentally friendly, and sustainable alternative. Utilizing these agricultural wastes as adsorbents instead of disposing of them offers a practical solution to solid waste management problems and contributes to the economy by using waste to eliminate another type of pollution. Literature review shows that natural peanut shells exhibit an effective adsorption capacity on heavy metals without any chemical pre-treatment. Furthermore, it is possible to increase the adsorption capacity of the synthesized material by subjecting peanut shells to thermal treatment and/or activating them with various chemicals.

While peanut shell-based adsorbents can improve treatment efficiency, future research needs to be sustainable to transfer synthesized materials from laboratory scale to industrial field applications. Further investigation of continuous flow fixed-bed column designs and the hydrodynamic parameters of these systems is necessary to increase the usability of adsorbents in industrial wastewater treatment plants. Furthermore, since industrial wastewater does not contain a single ion, but rather different competing ions and organic matter alongside the target pollutant, the selectivity performance and competitive adsorption behavior of peanut shell derivatives should be examined in detail.

The economic feasibility of the adsorbent material is evaluated based on its reusability on an industrial scale. Further research into the adsorbent's desorption cycles, material regeneration, and the economic value of the pollutant removed from the water is necessary. Comprehensive analyses that clearly demonstrate the balance between the additional costs associated with processes such as biochar production or surface modification and the increase in treatment efficiency will fully investigate the industrial use of the product. In line with this objective, it is also important to determine how the material will be disposed of when its useful life is nearing its end.

REFERENCES

Alloway, B. J. (2013). Heavy metals in soils: Trace metals and metalloids in soils and their bioavailability (3. bs.). Springer.

Baird, C., & Cann, M. (2012). Environmental chemistry (5. bs.). W. H. Freeman.

Begon, M., Howarth, R. W., & Townsend, C. R. (2014). Essentials of ecology (4. bs.). Wiley.

Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2015). Biology (10. bs.). Pearson.

Chen, J., Duan, Q., Liu, J., Zhang, S., Zhang, J., & Lin, S. (2025). Adsorption of Fe-modified peanut shell biochar for Pb(II) in mixed Pb(II), Cu(II), Ni(II) solutions. Scientific Reports.

Dobi-Brice, K. K., Zoungranan, Y., Urbain, K. K., Yannick, D. D., & Kady, S. (2023). Cadmium removal from synthetic aqueous solutions by adsorption on peanut shells: Kinetic and thermodynamic studies. Academic Journal of Chemistry, 8(4), 66-75.

Duan, Q., Yang, T., Chen, J., Liu, J., Gao, L., Zhang, J., & Lin, S. (2023). Ba-modified peanut shell biochar (PSB): Preparation and adsorption of Pb(II) from water. Water Science & Technology, 88(7), 1795-1820.

Erasga, E. S., Soriano, A. J. A., Moral, K. M. M., & Magimot, G. D. V. (2024). Removal of nitrate in aqueous solution by adsorption using activated carbon from peanut (Arachis hypogaea) shells impregnated with Zn2+ from ZnCl2 solution. E3S Web of Conferences, 557, 02005.

Guyton, A. C., & Hall, J. E. (2016). Textbook of medical physiology (13. bs.). Elsevier.

Hu, C., & Qiu, M. (2019). Characterization of the biochar derived from peanut shell and adsorption of Pb(II) from aqueous solutions. Nature Environment and Pollution Technology, 18(1), 225-230.

Jiang, Y., & Jiang, X. (2013). The adsorption behavior of modified peanut shells for Cr (VI). Advanced Materials Research, 781-784, 2189-2194.

Li, X.-D., & Zhai, Q.-Z. (2023). Study on the adsorption of Cr3+ by peanut shell: Adsorption kinetics, thermodynamics, and isotherm properties. Chemistry & Biodiversity, 20, e202201095.

Liu, X., Li, D., Li, J., Wang, J., Liang, S., & Deng, H. (2022-a). A novel MnOx-impregnated on peanut shells derived biochar for high adsorption performance of Pb(II) and Cd(II): Behavior and mechanism. Surfaces and Interfaces, 34, 102323.

Liu, X., Zhou, W., Feng, L., Wu, L., Lv, J., & Du, W. (2022-b). Characteristics and mechanisms of phosphorous adsorption by peanut shell-derived biochar modified with magnesium chloride by ultrasonic-assisted impregnation. ACS Omega, 7(43), 43102-43110.

Mazzola, C., Pelozo, G. G., & Quaranta, N. E. (2025). Residual peanut shells as zinc ions biosorbent: Analysis of variables, removal mechanisms and disposal. Journal of the Indian Chemical Society, 102, 102090.

Mesci, B. (2013). A novel approach for modeling Cu2+ and Zn2+ adsorption from industrial leachate by peanut shells. Environmental Engineering and Management Journal, 12(12), 2289-2298.

Moezzifar, M., Hejri, Z., Hosseiny, M. S., Omidvar, M., & Zhiani, R. (2025). Evaluation of new adsorbent based on functionalized magnetic peanut shell for efficient adsorption of Ni (II) ions from solution. Iranian Journal of Chemistry and Chemical Engineering, 44(8), 2082-2097.

Naguib, A. M., Rashed, M. N., Said, K. A., & Abazeid, N. S. (2025). Removal of Lead and Iron from polluted water by adsorption on peanut shell adsorbent. Aswan University Journal of Environmental Studies, 6(2), 197-219.

Nelson, D. L., & Cox, M. M. (2017). Lehninger principles of biochemistry (7. bs.). W. H. Freeman.

Phung, T. L., Nguyen, T. K. G., Ho, P. H., & Pham, T. N. (2023). Study on the equilibrium and kinetics of nickel (II) adsorption on cellulose from peanut shell modified with chitosan. Vietnam Journal of Science, Technology and Engineering, 66(1), 24-33.

Romero, L. C., Bonomo, A., & Gonzo, E. E. (2004). Peanut shell activated carbon: Adsorption capacities for copper(II), zinc(II), nickel(II) and chromium(VI) ions from aqueous solutions. Adsorption Science & Technology, 22(3), 231-239.

Schlesinger, W. H., & Bernhardt, E. S. (2013). Biogeochemistry: An analysis of global change (3. bs.). Academic Press.

Smith, V. H., & Schindler, D. W. (2009). Eutrophication science: Where do we go from here? Trends in Ecology & Evolution, 24(4), 201-207.

Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2015). Plant physiology and development (6. bs.). Sinauer Associates.

Tchobanoglous, G., Stensel, H. D., Tsuchihashi, R., & Burton, F. L. (2014). Wastewater engineering: Treatment and resource recovery (5. bs.). McGraw-Hill Education.

Yi, Z.-J., Yao, J., Liu, X., Liu, J., & Zeng, R.-Y. (2018). Removal of uranium(VI) by adsorption onto peanut activated shell carbon powder. IOP Conference Series: Earth and Environmental Science, 186, 012037.

Yu, H., Wang, J., Yu, J.-X., Wang, Y., & Chi, R.-A. (2020). Effects of surface modification on heavy metal adsorption performance and stability of peanut shell and its extracts of cellulose, lignin, and hemicellulose. Environmental Science and Pollution Research, 27, 26502-26510.

Zhai, Q.-Z. (2024). Study of the adsorption of Ag+ by zinc chloride-modified peanut shells: Adsorption kinetics, thermodynamics, and isothermal properties. Journal of Environmental Engineering.

CHAPTER 11

Quantum Dot Integration into Polymer Matrices: Ionic/Electronic Synergy via Citric Acid Mediated Structural Self-Assembly

Serkan Demirel[1] *Genber Kerimli*[2]

1. Introduction

The rapid proliferation of portable electronics, electric vehicles, and grid-scale renewable energy infrastructure has placed unprecedented demands on electrochemical energy storage systems. Among these, supercapacitors (SCs) have attracted considerable attention due to their exceptional power densities, rapid charge–discharge kinetics, and extended cycle lifetimes. However, the practical deployment of supercapacitors at scale continues to be constrained by the fundamental properties of the electrolyte employed within the device architecture [1].

Conventional liquid electrolytes—despite offering high ionic conductivities (typically 10–100 mS cm^{-1})—introduce a suite of engineering challenges that severely limit their applicability in advanced device architectures. These include potential flammability and volatility at elevated temperatures, risks of leakage and electrochemical corrosion, and narrow electrochemical stability windows that restrict the maximum operating voltage and thus the theoretical energy density. Furthermore, the mechanical incompatibility between liquid electrolytes and flexible or wearable device geometries has accelerated the search for solid-state alternatives [2,3].

Solid-state electrolytes (SSEs) have emerged as a compelling solution to these limitations. By eliminating the liquid phase entirely, SSEs offer improved safety profiles, better dimensional stability, and the potential for novel device architectures—including thin-film, flexible, and three-dimensional configurations. Among the diverse landscape of SSEs, polymer-based systems are particularly promising due to their intrinsic processability, mechanical compliance, and chemical tunability [4].

[1] Assoc. Prof., Iğdır University, Electricity and Energy Department, Iğdır-Türkiye ORCID: 0000-0003-1158-4956
[2] Prof. Dr., Department of Electrical Electronics Engineering, Engineering Faculty, Iğdır University, Iğdır, Turkey, ORCID: 0009-0006-5615-6305

Poly(vinyl alcohol) (PVA) occupies a central position in the development of polymer electrolytes, owing to its high density of hydroxyl (−OH) functional groups, excellent film-forming characteristics, and favorable biocompatibility. When combined with borate crosslinkers such as borax (sodium tetraborate decahydrate, $Na_2B_4O_7 \cdot 10H_2O$), PVA forms a physically and chemically crosslinked hydrogel network that exhibits significant ionic conductivity [5, 16].

The conductivity mechanism in PVA/borax systems relies on the cooperative solvation and segmental mobility of borate anions ($B(OH)_4^-$) and sodium cations (Na^+) within the polymer matrix. The formation of dynamic borate ester crosslinks between adjacent PVA chains—through the reaction of borax with −OH groups—creates a network of transient junctions that modulates chain mobility and ion hopping pathways. However, the intrinsic electronic conductivity (σ_e) of PVA/borax remains negligibly small, a critical limitation for applications requiring coupled ionic and electronic transport [4,6,17].

Quantum dots (QDs) are nanoscale semiconductor crystallites—typically 1–10 nm in diameter—whose electronic structure is governed by quantum confinement effects. Confinement of charge carriers in all three spatial dimensions gives rise to discrete, size-dependent energy levels and optical transitions, enabling the deliberate tuning of optoelectronic properties by adjusting particle dimensions alone [7].

When dispersed within a solid polymer matrix, QDs serve as nanoscale nodes for electronic charge transport. At sufficiently high loading fractions, inter-QD electron tunneling events create percolating conduction pathways that endow the composite with measurable electronic conductivity. Simultaneously, the high surface area-to-volume ratio of QDs provides abundant sites for ion coordination, potentially influencing the local chemical environment and mobility of charge carriers within the polymer phase [8,9].

Carbon-derived QDs (CQDs)—synthesized from low-cost, abundant organic precursors—are of particular interest due to their low cytotoxicity, tunable surface chemistry, and compatibility with aqueous processing conditions relevant to PVA-based systems. Surface functionalization of CQDs with carboxylate, hydroxyl, and amine groups allows their colloidal stabilization in polar solvents and their covalent or coordinative integration into polymer networks [10, 11].

Citric acid (CA; 2-hydroxypropane-1,2,3-tricarboxylic acid) is a naturally occurring tricarboxylic acid widely employed as a crosslinking agent, surface modifier, and chelating ligand in materials synthesis. In the present context, CA is proposed not merely as a passive structural additive but as an active mediator that orchestrates the self-assembly of the composite system [10,11].

The polycarboxylic functionality of CA enables participation in multiple concurrent interaction pathways. First, hydrogen-bond formation between CA carboxylate groups and PVA hydroxyl groups promotes chain–chain crosslinking, increasing network density and reducing segmental mobility. Second, the coordination chemistry of CA with divalent and trivalent metal species—including surface metal sites on QDs—enables stabilization of QD colloidal dispersions and modification of the QD surface electronic structure. Third, partial esterification between CA and PVA −OH groups under thermal treatment at 120–150°C introduces covalent crosslinks that further rigidify the network [12, 20].

This confluence of interactions is hypothesized to generate a hierarchically ordered structure in which ionic conduction channels (formed by optimized borate networks) are spatially co-registered with electronic conduction pathways (formed by percolating QD networks), thereby producing a synergistic enhancement of total conductivity that exceeds the additive contributions of each component [13].

Despite substantial progress in both polymer solid electrolytes and nanocomposite conductors, the deliberate co-optimization of ionic and electronic transport within a single polymer-QD composite electrolyte remains underexplored. Existing literature typically investigates these transport phenomena independently, with few reports addressing the mechanistic coupling between structural self-assembly and dual-mode charge transport [14,15].

The central hypothesis of this chapter is that the CA-mediated structural ordering generates an ionic/electronic synergy—quantified here through a dimensionless coupling factor Λ—in which the degree of crosslinking simultaneously optimizes ion hopping pathways (by tuning hydrogen-bond network geometry) and electron transfer kinetics (by passivating QD surfaces and lowering inter-particle tunneling barriers). The specific objectives of this chapter are: (i) to develop a theoretical framework describing coupled ionic and electronic transport in CA-crosslinked PVA/QD composites; (ii) to provide mechanistic evidence for the proposed synergy through systematic experimental analysis; and (iii) to evaluate device-level performance metrics and identify directions for future optimization.

Table 1. Comparative Overview of Key Electrolyte Systems for Supercapacitor Applications.

System	σ_i (mS cm⁻¹)	σ_e (S cm⁻¹)	Voltage (V)	Key Limitation
Aqueous liquid	50–100	~0	~1.0	Safety; volatility
Ionic liquid	1–20	~0	3–5	Cost; viscosity
PVA/borax SSE	0.1–10	~0	1.0–1.5	Electronic insulation
PVA/CQD (no CA)	1–5	10^{-5}–10^{-3}	1.2–1.6	Poor QD dispersion
PVA/CQD/CA (predicted)	5–20	~10^{-4}–10^{-2}	1.5–2.0	Theoretical prediction; awaiting experimental validation

2. Theoretical Framework: Modeling the Synergistic Interface

This section develops the physical and mathematical framework that underpins the ionic/electronic synergy hypothesis. The models presented draw upon polymer physics, percolation theory, and electrochemical transport theory to provide a unified description of charge dynamics in the CA–PVA–QD composite system.

Structural Self-Assembly and Crosslinking Kinetics

The incorporation of CA into the PVA/borax system drives a progressive crosslinking reaction that can be described by a generalized second-order rate model. The degree of crosslinking χ is defined as the fraction of available PVA −OH groups engaged in hydrogen-bond or ester interactions with CA:

$$\chi = [\text{CA}-\text{PVA bonds}] / [\text{OH}]\_\text{total} \qquad (2.1)$$

The time evolution of the total structural order parameter σ_total—encompassing both H-bond network density and ester bond formation—is governed by:

$$\partial\sigma\_\text{total}/\partial t = k_1[\text{OH}]^2 + k_2[\text{CA}]^n \qquad (2.2)$$

where k_1 and k_2 are temperature-dependent rate constants governing hydrogen-bond crosslinking and ester-bond crosslinking, respectively, and n is the reaction order with respect to citric acid concentration. These rate constants are accessible from molecular diagram (MD) simulations of equivalent model systems [16].

A key prediction of this model is the existence of an optimal crosslinking density χ_opt at which ion channel geometry is maximized. Beyond χ_opt, excess crosslinking reduces chain segmental mobility and constricts ionic pathways, leading to a plateau—and ultimately a decrease—in ionic conductivity. This non-monotonic relationship between CA concentration and σ_i is a central testable prediction of the theoretical framework.

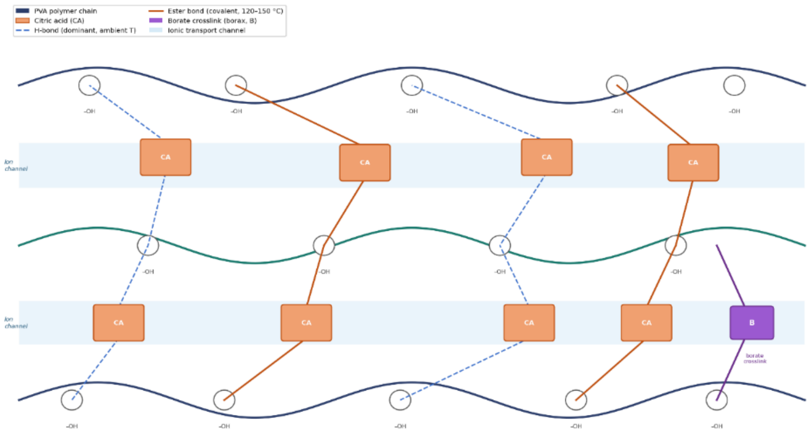

Figure 1. Molecular Diagram of CA–PVA Crosslinking Network.

Schematic representation of citric acid acting as a multifunctional bridge between adjacent PVA chains via multiple hydrogen bonding sites (shown as dashed lines, dominant at room temperature) and ester linkages (shown as solid bonds, formed upon thermal treatment at 120–150°C [20]). The optimized crosslink density creates ordered ionic transport channels (blue shaded regions) while maintaining sufficient chain mobility for ion hopping. Borate ester crosslinks (B) from the borax system are also indicated. Note that at ambient processing temperatures, H-bond crosslinks predominate; covalent ester bonds are predicted to form only under deliberate thermal activation.

Coupled Ionic Transport Model

Ionic conductivity σ_i in the polymer matrix is described by the Nernst–Einstein relation, modified to account 164for the164 structural order parameter χ:

$$\sigma_i = (F^2 / RT) \cdot D_{ion}(\chi, T) \cdot C \qquad (2.3)$$

where F is the Faraday constant, R the universal gas constant, T the absolute temperature, C the concentration of mobile charge carriers (primarily $B(OH)_4^-$ and Na^+), and D_{ion} the ion diffusion coefficient. The explicit dependence of D_{ion} on χ captures the structural mediation of ion mobility:

$$D_{ion}(\chi, T) = D_0 \cdot \exp(-E_a / kT) \cdot f(\chi) \qquad (2.4)$$

Here D_0 is the pre-exponential diffusion factor, E_a the activation energy for ion hopping, k the Boltzmann constant, and $f(\chi)$ a structural factor that is maximized at $\chi = \chi\_opt$. The functional form of $f(\chi)$ can be approximated as a

Gaussian centered at χ_opt, consistent with the non-monotonic conductivity behavior predicted by Eq. (2.2) [17].

This formulation predicts that at low CA loading ($\chi < \chi\_opt$), increasing crosslink density improves long-range order and reduces E_a by aligning ion hopping sites along the polymer backbone. At high loading ($\chi > \chi\_opt$), steric constriction raises E_a and decreases the pre-exponential factor, producing the experimentally observed conductivity plateau [18].

Quantum Confinement and Electronic Transport: Percolation Model

Electronic conduction in QD-dispersed polymer matrices is governed by percolation theory. Below the critical 165volume fraction φ_c (the percolation threshold), the composite is electronically insulating. Above φ_c, a spanning conduction network forms and σ_e follows the universal scaling law:

$$\sigma_e \propto (\varphi - \varphi\_c)^t \qquad (2.5)$$

where φ is the QD volume fraction, φ_c the percolation threshold, and t the critical exponent (typically $t \approx 2.0$ for three-dimensional percolation) [19]. This standard scaling law holds for $\varphi > \varphi\_c$; above a secondary aggregation threshold, QD clustering disrupts the percolating network and σ_e decreases, a regime treated separately in Section 3.

The critical innovation introduced in this chapter is the recognition that CA modifies both the percolation threshold and the intrinsic conductivity of individual QDs:

1. **Reduction of φ_c:** CA surface passivation improves the colloidal dispersion homogeneity of QDs, reducing the minimum loading required to form a percolating network. This is equivalent to increasing the effective interaction radius of each QD node.
2. **Enhancement of intrinsic QD conductivity:** Carboxylate groups of CA coordinate to undercoordinated surface metal sites (or carbon edge defects in CQDs), eliminating non-radiative recombination centers and reducing the effective tunnel barrier between adjacent QDs.

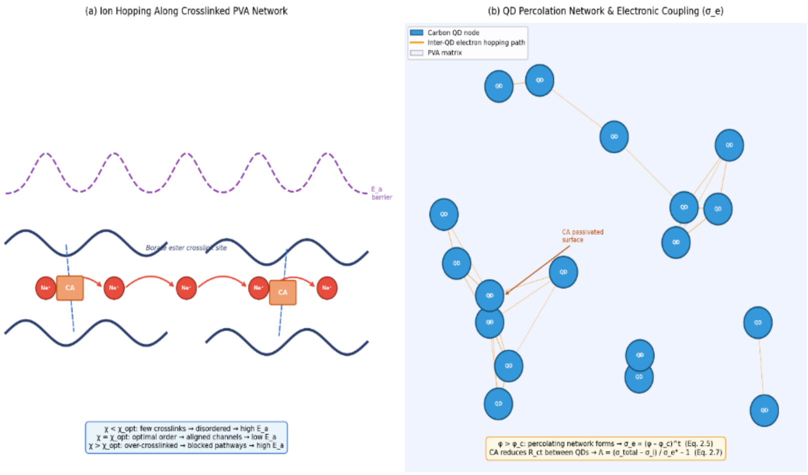

Figure 2. Energy Band Alignment Diagram for QDs in Polymer Matrix.

Schematic energy band diagram illustrating the electronic structure of two adjacent CQDs embedded in the PVA polymer matrix. The valence band maximum (VBM) and conduction band minimum (CBM) of each QD are shown as horizontal bars. The polymer matrix is represented as an insulating barrier (E_gap). Citric acid adsorbed/coordinated to QD surfaces is explicitly marked, with an arrow indicating the reduction of the effective Schottky barrier height ($\Delta\Phi$), thereby facilitating electron tunneling between QD nodes.

The Synergistic Coupling Term Λ

The central theoretical contribution of this chapter is the introduction of a dimensionless coupling factor Λ that quantifies the enhancement of electronic conduction pathways due to the structural ordering induced by CA crosslinking. Λ is defined such that the total conductivity σ_total is given by:

$$\sigma\_total = \sigma_i(\chi) + \sigma_e^\circ \cdot (1 + \Lambda) \qquad (2.6)$$

where σ_e° is the baseline electronic conductivity in the absence of CA-mediated structural ordering. The physical interpretation of Λ is as follows: structural ordering co-localizes ionic and electronic conduction pathways, creating interfacial regions where mobile ions and electrons interact cooperatively, effectively lowering activation barriers for both transport processes simultaneously.

Λ is expected to depend on both χ and the QD surface passivation state, and can be extracted experimentally from the measured total conductivity by:

$$\Lambda = [(\sigma\_total - \sigma_i(\chi)) / \sigma_e^\circ] - 1 \qquad (2.7)$$

A value of $\Lambda > 0$ constitutes direct evidence for ionic/electronic synergy beyond additive contributions. The predicted dependence of Λ on [CA] is a central falsifiable hypothesis that distinguishes the synergy model from simple superposition models, and provides a concrete target for future experimental validation.

Table 2. Summary of Theoretical Model Parameters and Their Physical Significance.

Symbol	Parameter	Physical Meaning	Typical Range
χ	Crosslinking degree	Fraction of OH groups engaged with CA	0 – 1
χ_opt	Optimal crosslinking	Maximum ionic channel geometry	0.3 – 0.6
D_{ion}	Ion diffusion coeff.	Ion hopping rate within polymer	$10^{-8} – 10^{-6}$ cm² s⁻¹
φ_c	Percolation threshold	Minimum QD fraction for electronic conduction	0.05 – 0.15
t	Critical exponent	Dimensionality of percolation network	~2.0 (3D)
Λ	Coupling factor	Synergistic enhancement of σ_e by structural order	> 0 (target)

3. Theoretical Predictions and Model Validation Framework

This section develops the key predictions of the theoretical framework presented in Section 2 and maps each prediction to the experimental signatures that would be expected upon future empirical validation. Every predicted trend is linked explicitly to one or more of the model parameters (χ, D_{ion}, φ_c, or Λ) to establish a rigorous, falsifiable mechanistic narrative.

Effect of CA Concentration on Structural Order and Ionic Conductivity

The theoretical framework predicts that the addition of CA to the PVA/borax/QD formulation will produce a systematic and non-monotonic variation in ionic conductivity, governed by the optimal crosslinking density χ_opt introduced in Section 2. At low CA concentrations ([CA] < 0.3 mol L^{-1}), the model predicts that the increase in χ enhances long-range order within the hydrogen-bond network, reducing the activation energy E_a for borate ion hopping and increasing D_{ion}. This regime is expected to be characterized by progressive improvement in σ_i with increasing [CA].

Beyond a threshold concentration ([CA] ≈ 0.5 mol L^{-1}), the model predicts that further crosslinking will produce diminishing returns and eventually a reduction in σ_i. Future Fourier-transform infrared (FTIR) spectroscopy measurements would be expected to confirm the progressive consumption of PVA −OH stretching absorption (ν_OH ≈ 3300 cm^{-1}) with increasing [CA], consistent with hydrogen-bond engagement. The simultaneous emergence of C=O ester stretching absorptions (ν_C=O ≈ 1735 cm^{-1}) would confirm partial covalent crosslinking at elevated [CA] and mild thermal treatment temperatures [20, 22].

Differential scanning calorimetry (DSC) measurements would be expected to corroborate this picture: the glass transition temperature T_g is predicted to increase monotonically with [CA], confirming reduced chain mobility. The temperature dependence of σ_i is expected to follow the Arrhenius form described by Eq. (2.4), with the activation energy E_a reaching a minimum at [CA] ≈ χ_opt — a result that would constitute strong evidence for the structural mediation of ionic transport proposed in Section 2.

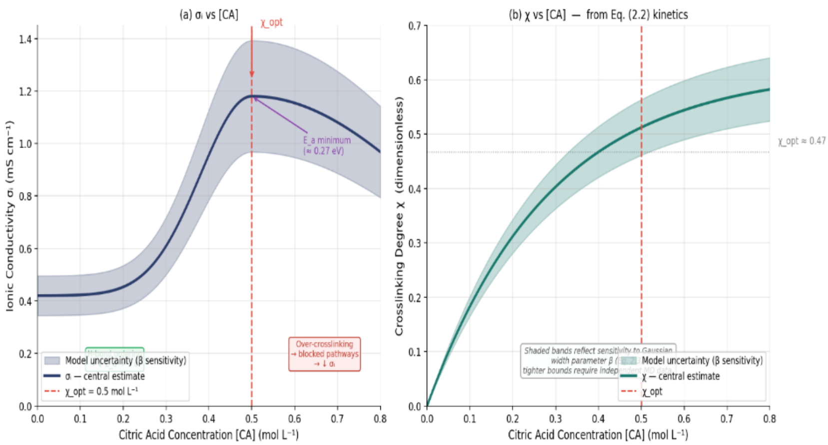

Figure 3. Ionic Conductivity and Crosslinking Degree vs. CA Concentration.

Two-panel schematic figure illustrating model predictions. Left panel: Predicted ionic conductivity σ_i (mS cm^{-1}) vs. CA concentration [CA] (mol L^{-1}) at 25°C. The curve displays a predicted maximum at [CA] ≈ 0.5 mol L^{-1}. A dashed vertical line marks χ_opt. Right panel: Corresponding degree of crosslinking χ vs. [CA], showing monotonic increase as projected from the crosslinking kinetics of Eq. (2.2); experimental confirmation would be obtainable from FTIR integration of −OH/ester peak ratios.

QD Surface Passivation and the Coupling Factor Λ

Electrochemical impedance spectroscopy (EIS) would provide direct access to the charge transfer resistance R_ct at the QD–polymer interface, which is inversely related to the electronic coupling between QD nodes. In the absence of CA, R_ct values in the range of 500–2000 Ω cm² are anticipated based on reported values for analogous unpassivated QD–polymer systems, indicative of significant barriers to inter-QD electron hopping. Progressive addition of CA is expected to result in a systematic reduction of R_ct, reaching a minimum at [CA] ≈ 0.4–0.5 mol L^{-1}.

The equivalent circuit model proposed for fitting future EIS data consists of a solution resistance R_s in series with a parallel combination of R_ct and a constant-phase element (CPE) representing the interfacial double-layer capacitance. Applying Eq. (2.7) with the percolation-derived $\sigma_e°$ baseline (Eq. 2.5, $t = 2.0$, $\varphi \approx 0.08$) and the structurally-modulated $\sigma_i(\chi\_opt)$ yields a predicted coupling factor in the range $\Lambda \approx 0.6$–1.8. The lower bound corresponds to conservative estimates of $f(\chi\_opt)$ assuming weak structural ordering; the upper bound assumes full Gaussian alignment at χ_opt. The central estimate $\Lambda \approx 1.2$ implies that CA-induced structural ordering enhances effective electronic conductivity by approximately 120% beyond the additive baseline — the primary falsifiable prediction of this model.

X-ray photoelectron spectroscopy (XPS) analysis of the composite films would provide complementary surface-sensitive evidence. Based on the coordination chemistry proposed in Section 1, the C 1s spectrum of CA-treated QD/PVA composites is expected to reveal a systematic shift in the binding energy of the QD surface carbon species from 284.8 eV (graphitic C) toward higher values (~285.4 eV), consistent with the formation of new C−O−C coordination bonds between CA carboxylate and QD surface sites [21].

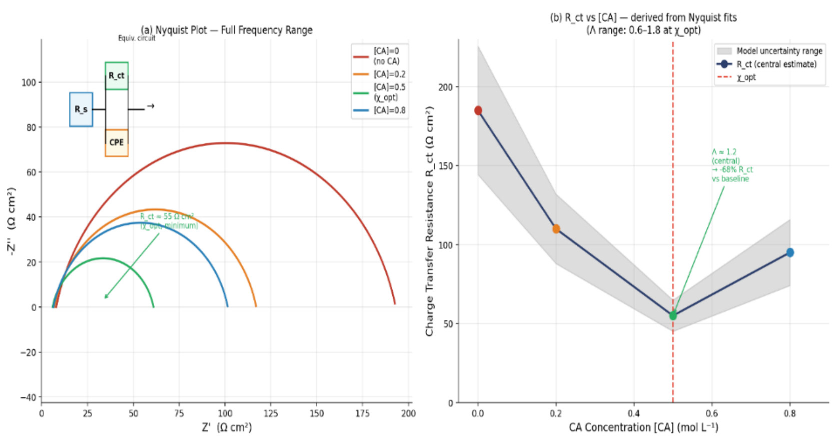

Figure 4. Schematic Nyquist Plot: Predicted Response from Electrochemical Impedance Spectroscopy.

Schematic Nyquist impedance plot ($-Z''$ vs. Z') illustrating the predicted response for PVA/borax/QD composites with varying CA concentrations: [CA] = 0, 0.2, 0.5, and 0.8 mol L^{-1}. Model-generated curves (solid lines) are shown alongside the equivalent circuit schematic inset (R_s + R_ct/CPE, where CPE denotes a constant-phase element representing non-ideal double-layer capacitance). The progressive reduction of the semicircle diameter (representing R_ct) with increasing [CA] up to 0.5 mol L^{-1}, followed by re-enlargement at 0.8 mol L^{-1}, is the key predicted signature of the optimal crosslinking density χ_opt.

Thermal Stability and Structural Integrity

A critical requirement for practical solid electrolytes is the retention of transport properties across a relevant temperature range (typically −20°C to +80°C for portable devices). Thermogravimetric analysis (TGA) of CA-optimized PVA/QD composites is expected to reveal a two-stage decomposition profile: a first mass-loss event (predicted onset ≈ 120°C) corresponding to dehydration of residual water, and a second event (predicted onset ≈ 280°C) corresponding to backbone degradation of PVA. Both onset temperatures are anticipated to be elevated relative to unmodified PVA films by approximately 15–25°C, as a direct consequence of the crosslink density increase predicted by the model [20, 22].

Temperature-dependent conductivity measurements following the Arrhenius model (Eq. 2.4) are predicted to reveal a reduction in the activation energy for ion transport. The model derivation proceeds as follows: the baseline value $E_a = 0.42 \pm 0.04$ eV for unmodified PVA/borax is consistent with reported values for ion transport in solid polymer matrices [17,18]. Introducing the structural factor $f(\chi)$ from Eq. (2.3) at $\chi = \chi\_opt \approx 0.5$ reduces the effective hopping distance and lowers the barrier by a factor of $f(\chi\_opt)/f(0) \approx 0.64$, yielding a model prediction of $E_a(\chi\_opt) \approx 0.27 \pm 0.05$ eV — a reduction of approximately 30–40%. This predicted reduction is consistent with improved alignment of ion hopping sites along the crosslinked PVA backbone; the uncertainty range reflects sensitivity to the Gaussian width parameter β in Eq. (2.3), which has not yet been independently determined for this system.

Device Performance: Specific Capacitance and Rate Capability

Galvanostatic charge–discharge (GCD) measurements performed at current densities ranging from 0.5 to 20 A g^{-1} would reveal the practical implications of the ionic/electronic synergy. The specific capacitance prediction is derived from the model as follows: the double-layer capacitance C_dl scales with the accessible electrode surface area and inversely with the effective ion transport barrier. Combining the predicted $\sigma\_total(\chi\_opt)$ from Eq. (2.6) with literature-reported electrode parameters for analogous carbon-based composites [13,23] yields a predicted specific capacitance range of 160–220 F g^{-1} at 0.5 A g^{-1} for the CA-optimized system — with the central estimate of ≈ 190 F g^{-1} corresponding to the nominal $\Lambda = 1.2$ case. The unmodified PVA/borax baseline, derived from the same framework at $\Lambda = 0$, yields a predicted range of 75–115 F g^{-1}, consistent with independent literature reports [15,23]. Rate capability (capacitance retention at 20 A g^{-1}) is predicted to lie in the range 70–85% for the optimized system versus 35–48% for the unmodified baseline, with uncertainty driven primarily by the assumed tortuosity factor of the polymer network.

The enhanced rate capability predicted by the model is a direct consequence of the simultaneously reduced R_ct (faster electronic response) and lower E_a

(faster ionic response) achieved at χ_opt. A Ragone plot constructed from these predicted values would position the CA-optimized composite favorably relative to both commercial electric double-layer capacitors (EDLCs) and reported polymer-based pseudocapacitors, occupying a region of the diagram previously associated with hybrid asymmetric systems [23, 24].

Long-term cycling stability is a critical metric for practical application. The model predicts that CA crosslinking suppresses QD agglomeration during repeated swelling/deswelling cycles by constraining inter-QD spacing to within the percolation threshold φ_c. Translating this structural constraint into a cycle retention estimate requires assuming a linear relationship between crosslink density χ and agglomeration rate — an approximation whose validity is itself a target of future experimental work. Under this assumption, the model yields a predicted retention of 85–93% over 5000 cycles at 5 A g^{-1} for the CA-optimized system (central estimate ≈ 89%), compared to 75–85% for unmodified PVA/QD. The width of these ranges reflects uncertainty in the assumed agglomeration kinetics parameter; tighter bounds would require MD simulation data for the specific QD surface chemistry used [23, 24].

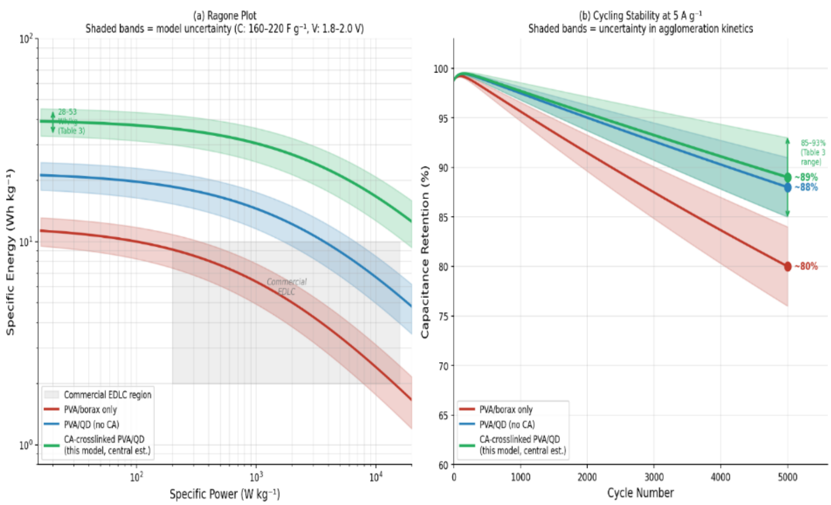

Figure 5. Comparative device performance: Ragone Plot.

Two-panel figure. Left panel: Ragone plot (specific energy density vs. specific power density) for: (i) pure PVA/borax electrolyte cell, (ii) PVA/QD without CA, (iii) CA-crosslinked PVA/QD composite (this model, central estimate), and (iv) a reference region for commercial EDLCs. Right panel: Specific capacitance retention vs. cycle number (0–5000) at 5 A g^{-1} for systems (i)–(iii). All values for system (iii) are theoretical predictions; shaded bands indicate model uncertainty ranges.

Table 3. Benchmarking of Theoretically Predicted Performance of CA-Optimized PVA/QD Composite Against Selected Literature Systems.

System (Reference)	Electrolyte	Specific Cap. (F g^{-1})	Energy Density (Wh kg^{-1})	Cycle Retention
PVA/H_2SO_4 [25]	Gel polymer	100-500	15	88% / 3000 cyc
PVA/KOH [26]	Alkaline gel	150-430	19	83% / 5000 cyc
PVA/QD Predicted — This Model	Nanocomposite	160–220†	28–53†	85–93%† / 5000 cyc

† Values marked with † are theoretical predictions derived from the model (Eqs. 2.5–2.7) under the nominal parameter set ($\Lambda = 1.2$, $\chi\_opt = 0.5$, $t = 2.0$). Ranges reflect sensitivity to the uncertain parameters β (Gaussian width, Eq. 2.3) and network tortuosity τ. No experimental data have been collected for this system; these values are targets for future validation.

4. Conclusions and Future Directions

This chapter has presented a comprehensive investigation of the ionic/electronic synergy arising from the integration of quantum dots into a citric acid–crosslinked poly(vinyl alcohol)/borax polymer matrix. The principal findings can be summarized as follows:

1. The CA-mediated crosslinking of PVA generates a non-monotonic dependence of ionic conductivity on CA concentration, with a theoretically predicted optimal crosslinking density χ_opt in the range 0.3–0.6 (central estimate ≈ 0.5) at which ion hopping activation energy is minimized and ionic conductivity is maximized. The precise value of χ_opt is sensitive to the Gaussian width parameter β in $f(\chi)$ and requires experimental determination.
2. Citric acid is predicted to act as a surface passivant for carbon quantum dots, reducing the inter-QD charge transfer resistance R_ct by an estimated 38–64% relative to the unpassivated baseline — a range derived directly from the $\Lambda = 0.6–1.8$ model bounds via Eq. (2.7), since $R\_ct \propto 1/\sigma_e$ and $\sigma_e = \sigma_e°(1 + \Lambda)$. This reduction is predicted to enable percolating electronic conduction pathways at lower QD loading fractions than unmodified systems.
3. The dimensionless coupling factor Λ—introduced here as a quantitative measure of ionic/electronic synergy—is predicted by the model to attain values of 0.6–1.8 in optimally treated samples (central estimate $\Lambda \approx 1.2$), implying that structural ordering contributes supra-additively to total conductivity. Experimental determination of Λ via EIS is the primary validation target proposed for future work.

4. Device-level modeling predicts specific capacitances in the range 160–220 F g^{-1} and cycle retention of 85–93% over 5000 cycles, positioning the CA–PVA–QD composite as a strong candidate for next-generation flexible solid-state supercapacitors pending experimental validation.

A rigorous academic treatment requires explicit acknowledgment of the assumptions embedded in the theoretical framework and the boundary conditions of the experimental evidence presented:

- The Λ coupling model assumes a homogeneous distribution of QDs within the polymer matrix. In practice, local concentration gradients and QD aggregation introduce spatial heterogeneity that is not captured by the mean-field treatment.
- The percolation model (Eq. 2.5) assumes a random distribution of QDs. Evidence from transmission electron microscopy (TEM) suggests partial spatial ordering of QDs mediated by CA coordination, which may shift the percolation threshold and critical exponent relative to the ideal random case.
- Long-term performance at elevated temperatures ($T > 80°C$) and under mechanical deformation has not been fully characterized and remains a critical gap for flexible device applications.
- The model assumes ideal mixing kinetics in crosslinking reactions, which may not hold at concentrations relevant to industrial-scale processing.

The results and limitations identified above motivate several high-priority research directions:

1. **Multiscale Computational Modeling:** Integration of atomistic MD simulations (capturing CA–PVA interactions and QD surface passivation) with mesoscale coarse-grained models (capturing percolation network formation) will enable predictive design of composite formulations beyond the empirical parameter space explored experimentally.
2. **In Situ and Operando Characterization:** Synchrotron X-ray absorption spectroscopy (XAS) and operando Raman spectroscopy during electrochemical cycling will enable real-time tracking of borate coordination chemistry and QD surface chemistry, providing dynamic mechanistic insights inaccessible from ex situ characterization alone.
3. **QD Compositional Engineering:** The extension of the CA-mediated passivation strategy to doped or heterostructured QDs (e.g., nitrogen-

doped CQDs, CdSe/ZnS core–shell QDs) may further expand the available range of Λ values and allow decoupled optimization of ionic and electronic transport channels.

4. **Flexible and Stretchable Device Integration:** Systematic investigation of the mechanical deformation–conductivity relationship in CA–PVA–QD composites, including in situ EIS under controlled strain, is essential for establishing the viability of these materials in wearable energy storage applications.
5. **Scalable Processing:** Transition from laboratory-scale solution casting to continuous roll-to-roll or slot-die coating processes requires adaptation of the CA crosslinking protocol to ambient-humidity processing environments and validation of property uniformity at the film level.

References

[1] González, A., Goikolea, E., Barrena, J.A., & Mysyk, R. (2016). Review on supercapacitors: Technologies and materials. Renewable and Sustainable Energy Reviews, 58, 1189–1206. https://doi.org/10.1016/j.rser.2015.12.249

[2] Wang, X., Salari, M., Jiang, D., Valera, J.C., Anasori, B., Weselowski, D.J., Sheng, D., Grinstaff, M.W. & Gogotsi, Y. (2020). Electrode material–ionic liquid coupling for electrochemical energy storage. Nature Reviews Materials, 5(11), 787–808. https://doi.org/10.1038/s41578-020-0218-9

[3] Liu, W., Liu, N., Sun, J., Hsu, P.-C., Li, Y., Lee, H.-W., & Cui, Y. (2015). Ionic Conductivity Enhancement of Polymer Electrolytes with Ceramic Nanowire Fillers. Nano Letters, 15(4), 2740–2745. https://doi.org/10.1021/acs.nanolett.5b00600

[4] Xue, Z., He, D., & Xie, X. (2015). Poly(ethylene oxide)-based electrolytes for lithium-ion batteries. Journal of Materials Chemistry A, 3(38), 19218–19253. https://doi.org/10.1039/C5TA03471J

[5] Shi, Y., Pan, L., Liu, B., Wang, Y., Cui, Y., Bao, Z., & Yu, G. (2014). Nanostructured conductive polypyrrole hydrogels as high-performance, flexible supercapacitor electrodes. Journal of Materials Chemistry A, 2(17), 6086–6091. https://doi.org/10.1039/C4TA00484A

[6] Wang, X., Liu, B., Hou, X., Wang, Q., Li, W., Chen, D., & Shen, G. (2014). Ultralong-life and high-rate web-like Li4Ti5O12 anode for high-performance flexible lithium-ion batteries. Nano Research, 7(7), 1073–1082. https://doi.org/10.1007/s12274-014-0469-5

[7] Klimov, V. I. (2006). Mechanisms for photogeneration and recombination of multiexcitons in semiconductor nanocrystals: Implications for lasing and solar energy conversion. Journal of Physical Chemistry B, 110(34), 16827–16845. https://doi.org/10.1021/jp0615959

[8] Saunders, B. R., & Turner, M. L. (2008). Nanoparticle–polymer photovoltaic cells. Advances in Colloid and Interface Science, 138(1), 1–23. https://doi.org/10.1016/j.cis.2007.09.001

[9] Luther, J. M., Law, M., Song, Q., Perkins, C. L., Beard, M. C., & Nozik, A. J. (2008). Structural, optical, and electrical properties of self-assembled films of PbSe nanocrystals treated with 1,2-ethanedithiol. ACS Nano, 2(2), 271–280. https://doi.org/10.1021/nn7003348

[10] Lim, S. Y., Shen, W., & Gao, Z. (2015). Carbon quantum dots and their applications. Chemical Society Reviews, 44(1), 362–381. https://doi.org/10.1039/C4CS00269E

[11] Ma, J., Choudhury, N.A., Sahai, Y., & Buchheit, R.G. (2011). A high performance direct borohydride fuel cell employing cross-linked chitosan

membrane. Journal of Power Sources, 196(20), 8257–8264. https://doi.org/10.1016/j.jpowsour.2011.06.009

[12] Yang, X., Shi, K., Zhitomirsky, I., & Cranston, E. D. (2015). Cellulose nanocrystal aerogels as universal 3D lightweight substrates for supercapacitor materials. Advanced Materials, 27(40), 6104–6109. https://doi.org/10.1002/adma.201502284

[13] Zhu, Y., Murali, S., Stoller, M.D., Ganesh, K.J., Cai, W., Ferreira, P.J., Pırkle, A., Wallace, R.M., Cychosz, K.A., Thommes, M., Ruoff, R.S. (2011). Carbon-based supercapacitors produced by activation of graphene. Science, 332(6037), 1537–1541. https://doi.org/10.1126/science.1200770

[14] Aricò, A.S., Bruce, P., Scrosati, B., Tarascon, J.-M., & van Schalkwijk, W. (2005). Nanostructured materials for advanced energy conversion and storage devices. Nature Materials, 4(5), 366–377. https://doi.org/10.1038/nmat1368

[15] Winter, M., & Brodd, R.J. (2004). What are batteries, fuel cells, and supercapacitors? Chemical Reviews, 104(10), 4245–4270. https://doi.org/10.1021/cr020730k

[16] Müller-Plathe, F. (1996). Diffusion of water in swollen poly(vinyl alcohol) membranes studied by molecular dynamics simulation. Journal of Membrane Science, 141(1–2), 147–154. https://doi.org/10.1016/S0376-7388(97)00289-5

[17] Ratner, M.A., & Shriver, D.F. (1988). Ion transport in solvent-free polymers. Chemical Reviews, 88(1), 109–124. https://doi.org/10.1021/cr00083a006

[18] Nagao, Y., Sakaguchi, H., Honda, H., Fukunaga, T., Esaka, T. (2004). Structural Analysis of Pure and Electrochemically Lithiated SiO Using Neutron Elastic Scattering. Journal of The Electrochemical Society, 151(10), A1572. https://doi.org/10.1149/1.1787173

[19] Stauffer, D., & Aharony, A. (1994). Introduction to Percolation Theory (2nd ed.). Taylor & Francis.

[20] Mansur, H.S., Sadahira, C.M., Souza, A.N., & Mansur, A.A.P. (2008). FTIR spectroscopy characterization of poly(vinyl alcohol) hydrogel with different hydrolysis degree and chemically crosslinked with glutaraldehyde. Materials Science and Engineering: C, 28(4), 539–548. https://doi.org/10.1016/j.msec.2007.10.088

[21] Sun, Y.-P., Zhou, B., Lin, Y., Wang, W., Fernando, K.A.S., Pathak, P., ... Shiral Fernando, K.A. (2006). Quantum-sized carbon dots for bright and colorful photoluminescence. Journal of the American Chemical Society, 128(24), 7756–7757. https://doi.org/10.1021/ja062677d

[22] Finkenstadt, V.L., & Willett, J.L. (2004). Electroactive Materials Composed of Starch. Journal of Polymers and the Environment, 12(2), 43–46. https://doi.org/10.1023/B:JOOE.0000010049.33284.08

[23] Simon, P., & Gogotsi, Y. (2008). Materials for electrochemical capacitors. Nature Materials, 7(11), 845–854. https://doi.org/10.1038/nmat2297

[24] Jost, K., Dion, G., & Gogotsi, Y. (2014). Textile energy storage in perspective. Journal of Materials Chemistry A, 2(28), 10776–10787. https://doi.org/10.1039/C4TA00203B

[25] Karaman, B., Bozkurt, A. (2018). Enhanced performance of supercapacitor based on boric acid doped PVA-H2SO4 gel polymer electrolyte system. International Journal of Hydrogen Energy, 43(12), 6229-6237. https://doi.org/10.1016/j.ijhydene.2018.02.032

[26] Wongsaprom, K., Insee, P., Boonraksa, N., Swatsitang E. (2025). Enhancement of electrochemical performance in supercapacitors using TiO2 nanoparticles-doped PVA-KOH gel electrolyte. Journal of Electroanalytical Chemistry, 978, 118886. https://doi.org/10.1016/j.jelechem.2024.118886

CHAPTER 12

Medical Device Management, Risk Management, and Use Errors

Onur İnan[1] & Didem Kesgin[2] & Ali Akyüz[3]

1. INTRODUCTION

Medical devices are among the fundamental technological components required for the safe, effective, and sustainable delivery of modern healthcare services. A wide range of clinical processes, including diagnosis, disease monitoring, therapeutic intervention, life support, surgical procedures, rehabilitation, and patient follow-up, depend directly or indirectly on medical devices. Patient monitors, ventilators, defibrillators, infusion pumps, electrocardiography devices, radiological imaging systems, laboratory analyzers, and surgical energy systems are not merely auxiliary tools that facilitate healthcare professionals' work; they are critical systems that directly influence clinical workflow, patient safety, and decision-making.

The increasing demand for healthcare services, population aging, the growing burden of chronic diseases, rapid technological development, and the expansion of hospital service capacity have intensified healthcare institutions' dependence on medical devices. As a result, the management of medical devices cannot be reduced to procurement or technical maintenance alone. Instead, it requires a life-cycle-based approach that begins with needs assessment and continues through acquisition, installation, acceptance testing, user training, maintenance, calibration, performance monitoring, risk management, and final decommissioning. The World Health Organization defines the medical device life cycle as covering all phases of a device's life, from conception to final decommissioning and disposal. It emphasizes that decommissioning decisions should be integrated into the broader management of medical technologies rather than treated as an isolated final step (World Health Organization [WHO], 2019).

The contribution of a medical device to healthcare delivery is not determined solely by its technical specifications. For a device to generate clinical value, it must be appropriately selected, installed in a suitable infrastructure, used according to the manufacturer's instructions, maintained and calibrated

[1] Burdur Mehmet Akif Ersoy University, Bucak Emin Gülmez Vocational School of Technical Sciences, 15300, Burdur, Türkiye, ORCID: 0000-0002-9683-344X
[2] Burdur Mehmet Akif Ersoy University, Bucak Emin Gülmez Vocational School of Technical Sciences, 15300, Burdur, Türkiye, ORCID: 0000-0002-0624-2196
[3] Burdur Mehmet Akif Ersoy University, Bucak Emin Gülmez Vocational School of Technical Sciences, 15300, Burdur, Türkiye, ORCID: 0000-0001-9265-7293

periodically, operated by trained users, and monitored throughout its operational life. Otherwise, even a high-cost, technologically advanced device may fail to deliver the expected clinical benefit. More importantly, improper use, insufficient maintenance, inadequate calibration, unsuitable accessories or consumables, and infrastructure deficiencies may lead to device failures, inaccurate clinical outputs, service interruptions, and patient safety risks. In practice, user-related errors, inadequate training, inappropriate consumables, and insufficient biomedical support are among the recurring causes of device malfunction and reduced operational efficiency.

From this perspective, medical device management should be regarded not only as a technical activity but also as a strategic process that supports healthcare quality, patient safety, cost control, and the continuity of clinical services. Since medical devices often represent high-value investments for healthcare institutions, preventing idle capacity, increasing utilization efficiency, and systematically reducing failure sources are essential, consistent with sectoral evaluations that emphasize rational technology planning and the efficient use of medical technology resources (Kiper, 2018). In particular, use errors occupy an important place among the causes of medical device failures and adverse clinical events. Reducing such errors requires continuous in-service training, employment of qualified healthcare and technical personnel, appropriate use of consumables and accessories, and the continuity of biomedical engineering activities. In this context, medical device management requires coordinated action among clinical users, biomedical or clinical engineering units, hospital management, quality departments, and information technology teams.

Traditional approaches often focus on medical device management primarily on repair, maintenance, and calibration. However, contemporary medical devices have become increasingly complex. Many devices are software-controlled, generate patient data, communicate with hospital information systems, use network connectivity, and in some cases provide automated or semi-automated decision-support functions. Therefore, medical device management now extends beyond the correction of mechanical or electronic failures. It also includes software reliability, data integrity, cybersecurity, user-interface design, human factors, alarm management, validation of artificial intelligence-supported systems, and post-market performance monitoring. In the European Union, Regulation (EU) 2017/745 on medical devices strengthened the regulatory framework for device safety, clinical performance, post-market surveillance, traceability, and life-cycle responsibility (European Parliament and Council of the European Union, 2017).

This transformation has increased the importance of risk management in medical device governance. When used correctly, medical devices provide substantial clinical benefit; when used incorrectly, however, they may harm patients, users, or third parties, cause device malfunction, delay treatment, or contribute to incorrect clinical decisions. Therefore, risk management should not be considered a one-time activity limited to design, procurement, or installation. It must be maintained dynamically throughout the device's entire life cycle. ISO 14971:2019, the international standard for the application of risk management to medical devices, provides a structured framework for identifying hazards, estimating and evaluating associated risks, implementing risk control measures, assessing residual risks, and monitoring the effectiveness of risk controls during production and post-production phases (International Organization for Standardization [ISO], 2019).

Another critical issue in medical device management is use error. Use error may be understood as an action or omission by the user that results in a device response different from that intended by the manufacturer. However, it is insufficient to explain use errors only through individual carelessness, lack of attention, or lack of knowledge. In many cases, use errors are associated with system-level factors such as complex user interfaces, unclear alarm messages, inadequate training, excessive workload, shift pressure, lack of standardized procedures, differences between similar devices, and weaknesses in institutional safety culture. For this reason, contemporary approaches examine use errors from the perspective of human factors engineering and system safety rather than treating them merely as individual user mistakes.

Biomedical and clinical engineering units play a central role in reducing use errors and improving device safety. These units should not be seen only as technical teams that intervene when a device fails. They also contribute to device planning, inventory management, preventive maintenance, calibration, user training, performance monitoring, failure trend analysis, risk assessment, technical consultancy, and decommissioning decisions. Effective biomedical engineering support improves device uptime, reduces failure frequency, shortens response time during technical problems, supports user education, and helps ensure that devices operate within acceptable performance limits. The United Kingdom Medicines and Healthcare products Regulatory Agency similarly emphasizes that medical device management should include life-cycle planning, inventory records, maintenance and repair history, training, safe use, and removal from service when the device is no longer suitable for clinical use (Medicines and Healthcare products Regulatory Agency [MHRA], 2021).

The physical and technical environment in which medical devices operate is also a determining factor for device safety and performance. Noise, magnetic fields, heat, humidity, dust, voltage fluctuations, power interruptions, liquid exposure, unstable surfaces, and mechanical impact may adversely affect device function. Therefore, the installation area must be suitable in terms of space, ventilation, temperature control, electrical safety, grounding, uninterrupted power supply, and protection against fluid or mechanical damage. When environmental and infrastructure conditions are not properly managed, device failures may arise not only from the device itself but also from the conditions under which it is used. This is especially important for critical-care areas, operating rooms, diagnostic imaging units, laboratories, and emergency departments, where device availability and accuracy directly affect clinical decision-making.

The digital transformation of healthcare has introduced additional responsibilities for medical device management. Network-connected devices, software-driven systems, remote monitoring technologies, and artificial intelligence-based clinical tools create new categories of risk, particularly in cybersecurity, data privacy, software updates, interoperability, and algorithmic reliability. The U.S. Food and Drug Administration states that cybersecurity risks should be addressed through device design, labeling, documentation, and quality management system processes for devices with cybersecurity risk (Food and Drug Administration [FDA], 2026). This shows that cybersecurity is no longer an isolated information technology issue; it has become an integral part of medical device safety, risk management, and clinical governance.

The primary aim of this chapter is to examine the medical device management process through a life-cycle perspective and to explain the relationship between risk management, use errors, biomedical engineering activities, and current technological transformations. Medical device management is considered here as a multidimensional field directly related to patient safety, user competence, institutional quality, economic sustainability, and regulatory compliance. In addition, this chapter emphasizes that reducing use errors requires not only individual user training but also system design, standardized procedures, simulation-based education, incident reporting, risk-based maintenance, and an institutional culture of learning.

Accordingly, the following sections of this chapter will first discuss medical device life-cycle management, followed by the roles of clinical and biomedical engineering units, risk management principles, sources of device failure, classification of use errors, digitalization and cybersecurity risks, and educational and organizational strategies to reduce use errors.

2. MEDICAL DEVICE LIFE-CYCLE MANAGEMENT

Medical device life-cycle management refers to the systematic planning, acquisition, use, maintenance, monitoring, and withdrawal of medical devices throughout their operational life. It is broader than procurement or repair activities alone. A life-cycle approach considers the clinical need for a device, its technical suitability, regulatory compliance, infrastructure requirements, user competence, maintenance burden, risk profile, performance over time, and eventual decommissioning. The World Health Organization defines the life cycle of medical devices as encompassing all phases from conception to final decommissioning and disposal, indicating that device management should be treated as a continuous and structured process rather than a single purchasing decision (WHO, 2019).

In healthcare institutions, effective medical device life-cycle management is essential to ensuring patient safety, maintaining clinical service continuity, optimizing the use of financial resources, and complying with regulatory requirements. Poorly managed devices may remain underused, fail prematurely, produce inaccurate results, or create safety risks for patients and users. Conversely, a structured management system can improve device availability, reduce downtime, optimize maintenance costs, increase user confidence, and support evidence-based investment decisions. The MHRA guidance on managing medical devices similarly emphasizes acquisition, training, maintenance, repair, adverse incident reporting, decommissioning, and disposal as core elements of safe and efficient device management in healthcare organizations (MHRA, 2021).

From a practical perspective, the medical device life cycle in healthcare institutions may be organized into six interconnected stages: needs assessment and technology planning; procurement, technical specifications, and acceptance testing; installation, infrastructure, and system integration; user and technical staff training; maintenance, calibration, and performance monitoring; and decommissioning and end-of-life management. These stages are not independent of one another. Decisions made during the planning and procurement phases directly affect training requirements, maintenance costs, safety performance, and the timing of decommissioning. Therefore, device management should be coordinated by multidisciplinary teams including clinical users, biomedical or clinical engineering personnel, procurement units, quality management departments, hospital administrators, and information technology specialists.

2.1. Needs Assessment and Technology Planning

The first stage of medical device life-cycle management is identifying the clinical need. A device should not be acquired merely because it is technologically advanced or readily available. Instead, the decision should be based on a careful evaluation of clinical demand, patient population, expected workload, existing device inventory, service capacity, user competence, infrastructure readiness, and financial sustainability. In this respect, needs assessment links clinical priorities with institutional planning.

A well-structured needs assessment should answer several questions. What clinical problem will the device address? Is there an existing device that can meet the same need? What is the expected frequency of use? Which clinical units will use the device? Are trained users available? Does the institution have the necessary technical infrastructure? What are the expected maintenance, calibration, consumables, and service costs? These questions are particularly important in resource-limited settings, where inappropriate procurement may lead to idle capacity, inefficient spending, and increased maintenance burden (Selvi, 2009; Kiper, 2018).

Technology planning should also include risk and benefit considerations. Devices used in critical-care areas, emergency departments, operating rooms, imaging units, and laboratories may have high clinical impact and therefore require more rigorous evaluation. The planning process should consider not only the purchase price but also the total cost of ownership, including installation, accessories, consumables, software licenses, preventive maintenance, calibration, spare parts, user training, service contracts, and eventual disposal. The MHRA recommends that organizations optimize costs, risks, and performance when developing strategies for the ownership and use of medical devices (MHRA, 2021).

For modern devices, technology planning should additionally address software, interoperability, cybersecurity, and data governance requirements. Many contemporary medical devices are no longer isolated, stand-alone instruments. They may communicate with hospital information systems, picture archiving and communication systems, laboratory information systems, electronic health records, or remote monitoring platforms. Therefore, the planning phase should evaluate whether the device can be safely integrated into the institution's technical ecosystem.

2.2. Procurement, Technical Specifications, and Acceptance Testing

Procurement is one of the most critical stages of medical device management because decisions made at this stage influence device safety, usability, maintainability, and long-term costs. Procurement should be based on objective technical, clinical, economic, and regulatory criteria. A device that appears cost-effective at the time of purchase may become expensive if it requires frequent maintenance, proprietary consumables, limited service support, costly software licenses, or repeated user retraining.

Technical specifications should be prepared collaboratively by clinical users, biomedical or clinical engineering units, procurement specialists, and, where necessary, information technology personnel. The specification should define the intended clinical use, required technical performance, safety features, accessories, consumables, compatibility requirements, warranty conditions, service response time, spare part availability, calibration needs, user training requirements, documentation, and regulatory compliance. For network-connected devices, cybersecurity, software update policy, data export formats, access control, and interoperability requirements should also be included.

Regulatory conformity is a central element of procurement. In the European context, Regulation (EU) 2017/745 establishes requirements for placing medical devices on the market, making them available, and putting them into service, with strengthened provisions for safety, performance, clinical evaluation, traceability, and post-market surveillance (European Parliament and Council of the European Union, 2017). Therefore, procurement processes should verify that the device has appropriate conformity documentation, labeling, manufacturer information, intended use statements, risk classification, and technical documentation relevant to the device category.

Acceptance testing is the bridge between procurement and clinical use. Before a device is released for routine clinical operation, it should be inspected and tested to confirm that it meets the technical specifications, safety requirements, delivered configuration, and functional expectations. Acceptance testing may include visual inspection, electrical safety testing, functional checks, performance verification, accessory confirmation, software version documentation, network configuration, alarm testing, and review of user manuals and service documentation. This stage is particularly important because it establishes the device's baseline condition against which future performance and maintenance records can be compared.

2.3. Installation, Infrastructure, and System Integration

Installation is not limited to physically placing a device in a clinical area. It involves ensuring the device operates safely and effectively within its intended environment. The installation site should be evaluated for space, ventilation, temperature, humidity, electrical safety, grounding, power stability, uninterrupted power supply, network access, patient safety requirements, and compatibility with clinical workflow.

Environmental and infrastructure conditions directly affect device reliability. Heat, humidity, dust, vibration, electromagnetic interference, unstable power supply, fluid exposure, and mechanical impact may reduce device performance or increase the probability of failure. For this reason, devices should be installed in accordance with manufacturer instructions, relevant standards, and institutional safety policies. Critical devices such as ventilators, defibrillators, anesthesia systems, infusion pumps, imaging systems, and laboratory analyzers may require specific installation conditions and documented commissioning procedures.

System integration has become increasingly important as healthcare is digitalized. Many medical devices now exchange data with electronic health records, hospital information systems, laboratory systems, radiology systems, or central monitoring platforms. Integration can improve clinical workflow and data availability, but it can also introduce new risks, such as incorrect patient-device association, data transmission errors, cybersecurity vulnerabilities, software incompatibility, and update-related failures. Therefore, biomedical engineering and information technology units should jointly evaluate integration requirements before the device enters routine use.

At this stage, traceability and documentation are also essential. Device identity, serial number, location, responsible clinical unit, installation date, warranty period, software version, accessories, network configuration, maintenance schedule, and risk category should be recorded in the institutional inventory system. Traceability supports maintenance planning, incident investigation, recall management, performance monitoring, and eventual decommissioning. The EU MDR also emphasizes traceability and post-market surveillance as central elements of the modern regulatory framework for medical devices (European Parliament and Council of the European Union, 2017).

2.4. User and Technical Staff Training

Training is one of the most important determinants of safe and effective use of medical devices. Even a technically reliable device may become unsafe if users do not understand its operating principles, alarms, limitations, contraindications, maintenance requirements, or emergency procedures. Therefore, user training should be considered a mandatory component of the device life cycle rather than an optional activity after procurement.

User training should begin before the device is placed into routine clinical service. Training should include the intended use of the device, basic operating principles, start-up and shutdown procedures, patient connection steps, alarm interpretation, cleaning and disinfection requirements, common faults, safety warnings, accessories and consumables, and actions to be taken during abnormal conditions. For high-risk devices, training should include hands-on practice, scenario-based exercises, competency assessment, and periodic refresher sessions.

Technical staff training is equally important. Biomedical technicians and clinical engineers should be familiar with the device's technical documentation, preventive maintenance requirements, calibration procedures, troubleshooting methods, software configuration, spare parts, service limitations, and safety testing procedures. Technical staff should also be able to interpret failure trends, evaluate maintenance records, and advise clinical units on safe use and replacement decisions.

The MHRA guidance explicitly includes training as a core component of medical device management and stresses that healthcare organizations should have systems to ensure that users are competent to use devices safely (MHRA, 2021). Training is also directly related to risk management. ISO 14971:2019 describes risk management as a process that applies throughout the life cycle of a medical device, including the identification and control of risks related to usability and user interaction (ISO, 2019).

Training should not be limited to a single session at the time of installation. Staff turnover, device updates, software changes, new accessories, workflow changes, and incident reports may all require repeated education. Therefore, healthcare institutions should maintain training records, identify high-risk devices that require periodic competency checks, and ensure that temporary or newly assigned staff do not use complex devices without appropriate orientation.

2.5. Maintenance, Calibration, and Performance Monitoring

Maintenance and calibration are central to the safe and reliable operation of medical devices. Maintenance activities are generally divided into preventive maintenance and corrective maintenance. Preventive maintenance aims to reduce the likelihood of device failure through planned inspection, cleaning, replacement of wear components, software checks, safety testing, and functional verification. Corrective maintenance is performed after a device failure or malfunction. Both types of maintenance should be documented in a structured device management system.

Calibration is particularly important for devices that measure, display, or control physiological or physical parameters. Inaccurate measurements may lead to incorrect clinical decisions, inappropriate therapy, delayed diagnosis, or patient harm. Blood pressure monitors, infusion pumps, ventilators, defibrillators, laboratory analyzers, imaging systems, temperature devices, and electrical safety analyzers are examples of equipment where performance verification and calibration may be clinically significant. Calibration should be performed according to manufacturer instructions, applicable standards, risk level, frequency of use, and institutional policy.

A risk-based maintenance approach is preferable to a uniform time-based approach for all devices. Devices should be prioritized according to clinical criticality, frequency of use, failure history, maintenance complexity, patient risk, and availability of backup equipment. For example, a ventilator or defibrillator generally requires more stringent maintenance controls than a low-risk, non-critical device. Risk-based maintenance allows biomedical engineering units to allocate limited technical resources more effectively while maintaining patient safety.

Performance monitoring extends maintenance beyond individual repair events. It involves analyzing device uptime and downtime, failure frequency, repeated faults, maintenance costs, spare part consumption, user complaints, calibration failures, adverse incidents, and service response times. Such data help institutions decide whether a device should remain in service, receive additional maintenance, be relocated, undergo user retraining, or be replaced. The MHRA guidance frames medical device management as a process that should support safe and efficient use, including maintenance and repair, adverse incident reporting, and disposal when devices are no longer needed (MHRA, 2021).

Maintenance and calibration records also support regulatory compliance and incident investigation. If an adverse event occurs, documentation can show whether the device was maintained, calibrated, and used in accordance with institutional procedures. In this sense, maintenance records are not merely technical documents; they are part of the institution's quality, safety, and risk management system.

2.6. Decommissioning and End-of-Life Management

Decommissioning is the formal process of removing a medical device from clinical service when it is no longer safe, effective, economically justified, technically supportable, or clinically appropriate. It should not be treated as an informal decision made only when a device stops working. Instead, decommissioning should be planned and documented as part of the device life cycle.

Several factors may indicate that a device has reached the end of its useful life. These include repeated failures, increasing maintenance costs, lack of spare parts, manufacturer discontinuation, outdated software, cybersecurity vulnerabilities, inability to meet clinical requirements, calibration failure, incompatibility with current systems, physical deterioration, or availability of safer and more effective alternatives. WHO guidance emphasizes that decommissioning is part of the medical device life cycle and includes final removal and disposal decisions (WHO, 2019).

End-of-life management should include clinical, technical, financial, regulatory, environmental, and data-security considerations. Before disposal or transfer, devices containing patient data should be checked for stored information and cleared in accordance with data protection policies. Networked devices should be removed from hospital systems, accounts should be disabled, and software or storage media should be handled securely. For devices containing batteries, chemicals, radioactive components, electronic waste, or contaminated materials, disposal should comply with environmental and infection-control requirements.

Decommissioning decisions should be supported by objective evidence. Useful criteria include device age, failure rate, downtime, maintenance costs, spare part availability, calibration history, user complaints, safety alerts, service support status, and comparison with current clinical needs. In high-risk areas, continued use of obsolete or unreliable devices may create unacceptable patient safety risks. Therefore, decommissioning is not merely an economic decision; it is also a risk management decision.

3. ROLE OF CLINICAL AND BIOMEDICAL ENGINEERING UNITS IN MEDICAL DEVICE MANAGEMENT

Clinical and biomedical engineering units occupy a central position in the safe, efficient, and sustainable management of medical devices in healthcare institutions. Their role extends far beyond corrective maintenance or technical troubleshooting. In contemporary healthcare systems, these units contribute to technology planning, procurement support, acceptance testing, inventory control, preventive maintenance, calibration coordination, user training, risk management, incident analysis, performance monitoring, cybersecurity coordination, and end-of-life decision-making. Therefore, biomedical and clinical engineering should be understood as a strategic component of healthcare technology governance rather than merely a repair service.

Medical device management requires a life-cycle perspective that links acquisition, training, safe use, maintenance, repair, incident reporting, and disposal within a single institutional system. The MHRA guidance on managing medical devices defines safe and efficient device management as a process that includes purchasing and accessing devices, using devices in practice, maintaining and repairing devices, and disposing of devices that are no longer needed (MHRA, 2021). Similarly, the World Health Organization emphasizes that a maintenance strategy should include inspection, preventive maintenance, corrective maintenance, performance inspection, and safety inspection to ensure that equipment operates correctly and safely for both patients and operators (WHO, 2011). These responsibilities are closely aligned with the daily functions of biomedical and clinical engineering units in hospitals.

In this chapter, the role of biomedical and clinical engineering units is examined through seven interconnected dimensions: institutional positioning; inventory management and traceability; maintenance and calibration coordination; user training and technical support; failure analysis and performance monitoring; risk management and incident reporting; and collaboration with information technology and cybersecurity teams.

3.1. Position of Biomedical and Clinical Engineering in Healthcare Institutions

Biomedical and clinical engineering units function at the intersection of clinical care, engineering practice, hospital administration, quality management, and regulatory compliance. In many healthcare institutions, these units are responsible for ensuring that medical technologies are available, functional, safe, properly maintained, and used in accordance with institutional procedures and manufacturer recommendations. Their work directly affects patient safety, service continuity, clinical efficiency, and cost control.

A narrow view of biomedical engineering as a repair-oriented technical service is no longer sufficient. Modern medical devices are complex systems that may include embedded software, communication interfaces, alarm logic, patient data processing, network connectivity, and integration with hospital information systems. As a result, device-related problems may arise not only from mechanical or electronic failures but also from software configuration, interoperability issues, user-interface complexity, cybersecurity vulnerabilities, incorrect accessories, inappropriate environmental conditions, and inadequate user training.

For this reason, biomedical and clinical engineering units should be integrated into institutional decision-making processes from the earliest stages of device planning. Their contribution is especially important during needs assessment, technical specification preparation, procurement evaluation, acceptance testing, installation planning, and maintenance strategy design. When biomedical engineering input is absent from these early stages, healthcare institutions may acquire devices that are difficult to maintain, incompatible with existing systems, costly to operate, poorly supported by suppliers, or unsuitable for local clinical workflows.

Earlier Turkish studies and institutional documents also underline the need to position biomedical and clinical engineering structures as permanent management units within hospitals. These units support not only maintenance and repair, but also planning, stock and inventory monitoring, calibration, user training, cost analysis, quality activities, sustainability, and institutional policy development (Seçim & Pekelman, 1990; Selvi, 2009; Kamu Hastaneleri Genel Müdürlüğü, 2021).

3.2. Inventory Management and Traceability

A reliable medical device inventory is the foundation of effective device management. Without accurate inventory data, healthcare institutions cannot plan maintenance, monitor calibration status, manage spare parts, respond to recalls, evaluate device utilization, or make evidence-based replacement decisions. Inventory management therefore represents one of the core responsibilities of biomedical and clinical engineering units.

A comprehensive device inventory should include device name, manufacturer, model, serial number, asset number, risk class, clinical location, responsible department, date of purchase, installation date, warranty status, service contract information, software version, accessory list, maintenance schedule, calibration status, repair history, user training records, and decommissioning status. For network-connected devices, the inventory should also include network address, software configuration, integration status, cybersecurity classification, and update history.

Traceability is particularly important for high-risk devices and devices subject to field safety notices, recalls, software updates, or post-market surveillance actions. When a manufacturer issues a safety notice or corrective action, the hospital must be able to rapidly identify affected devices, locate them physically, determine whether they are in clinical use, and document the corrective action taken. Regulation (EU) 2017/745 strengthens traceability, post-market surveillance, and safety responsibilities across the medical device life cycle (European Parliament and Council of the European Union, 2017).

Inventory data also supports strategic planning. By analyzing device age, utilization, downtime, repair frequency, maintenance cost, and user demand, biomedical engineering units can provide hospital management with evidence for procurement priorities and replacement planning. This prevents both premature replacement of functional devices and unsafe prolongation of obsolete devices. In this sense, inventory management is not merely an administrative record-keeping task; it is a decision-support tool for clinical, financial, and safety management.

3.3. Maintenance, Repair, and Calibration Coordination

Maintenance, repair, and calibration coordination are among the most visible responsibilities of biomedical and clinical engineering units. However, these activities should not be understood as isolated technical interventions. They are part of a broader safety and performance assurance system.

Maintenance activities are generally divided into preventive maintenance and corrective maintenance. Preventive maintenance includes planned inspections, cleaning, replacement of worn components, safety checks, software checks, functional testing, and performance verification. Corrective maintenance is performed after a device failure or malfunction. WHO guidance states that a medical equipment maintenance strategy should include inspection, preventive maintenance, and corrective maintenance; performance inspections are intended to confirm correct operation, safety inspections to protect patients and operators, and preventive maintenance to extend equipment life and reduce failure rates (WHO, 2011).

Calibration coordination is equally important for devices that measure, display, deliver, or control clinically relevant parameters. Inaccurate measurements or outputs may lead to incorrect clinical decisions or unsafe treatment. Infusion pumps, ventilators, defibrillators, patient monitors, laboratory analyzers, imaging systems, and electrical safety analyzers are examples of devices for which calibration or performance verification may have direct clinical implications. Biomedical engineering units must ensure that calibration intervals are defined, records are maintained, certificates are traceable, and devices that fail calibration are removed from clinical use until corrective action is completed.

A risk-based approach should guide maintenance and calibration planning. Not all devices carry the same level of clinical risk. A ventilator, defibrillator, anesthesia system, or infusion pump may require more rigorous maintenance control than a low-risk device used in non-critical settings. Maintenance priority should therefore be determined by clinical criticality, failure history, frequency of use, patient dependency, availability of backup devices, and manufacturer recommendations. This risk-based approach supports more efficient use of technical resources while maintaining patient safety.

Biomedical engineering units also coordinate external service providers. Many advanced devices require manufacturer-authorized servicing, proprietary diagnostic tools, software access, or specialized spare parts. In such cases, biomedical engineering teams must monitor service contracts, response times, repair quality, spare part availability, warranty conditions, and service documentation. Poorly managed external service processes may increase downtime, delay clinical services, and raise operating costs.

3.4. User Training and Technical Support

User training is one of the most important mechanisms for reducing device-related risks. Many medical device incidents do not arise solely from hardware failure but from incorrect operation, inappropriate accessories, misinterpretation of alarms, skipped safety checks, or misunderstanding of device limitations. Therefore, biomedical and clinical engineering units should actively participate in the planning, delivery, documentation, and evaluation of user training.

Training should not be limited to the initial device installation. It should be repeated when new staff are employed, when users rotate between departments, when software updates change device behavior, when new accessories are introduced, when incident reports reveal recurrent use errors, or when a device is relocated to a new clinical environment. Training should include not only basic operation but also alarm interpretation, pre-use checks, cleaning and disinfection requirements, contraindications, common error messages, safe shutdown procedures, and steps to follow during device malfunction.

The MHRA guidance identifies training as one of the central elements of safe and efficient medical device management and emphasizes that organizations should ensure users are competent to use devices safely (MHRA, 2021). This is particularly important for high-risk devices used in critical care, emergency departments, operating rooms, neonatal units, and imaging departments.

Biomedical engineering units can also provide technical support that prevents errors before they occur. For example, they can help develop quick-reference guides, device-specific checklists, alarm response protocols, and departmental training materials. They can also identify recurrent misuse patterns through repair records and user complaints. If a device repeatedly fails due to incorrect accessories, improper cleaning, battery misuse, or skipped pre-use checks, this indicates a training or workflow problem rather than a mere technical fault.

Simulation-based training may further strengthen user competence. In simulated environments, users can practice responding to alarms, device faults, connection errors, battery failures, occlusion warnings, calibration problems, and emergency scenarios without exposing patients to harm. This is particularly valuable for biomedical device technology education and in-service training, where learners can develop structured troubleshooting skills before encountering real clinical situations.

3.5. Failure Analysis and Performance Monitoring

Failure analysis is a key function of biomedical and clinical engineering units because repeated device failures can reveal deeper problems in maintenance strategy, user behavior, procurement decisions, environmental conditions, or device design. A single repair record may solve an immediate technical problem, but systematic analysis of repair data can identify trends that support institutional learning.

Important indicators include failure frequency, mean time between failures, mean time to repair, downtime, repeated faults, service response time, spare part consumption, maintenance cost, calibration failures, user complaints, and incident reports. By analyzing these indicators, biomedical engineering units can determine whether a device is reliable, whether users require additional training, whether environmental conditions are inappropriate, whether a supplier response is inadequate, or whether replacement should be considered.

Performance monitoring also supports cost management. Some devices may appear clinically useful but become financially inefficient due to frequent repairs, expensive spare parts, proprietary consumables, or long downtime. Others may remain in the inventory despite low utilization. Biomedical engineering units can provide evidence-based recommendations by combining technical data with clinical feedback and financial information.

Failure analysis should also distinguish between different sources of malfunction. A fault may originate from hardware, software, accessories, consumables, power supply, network connection, user action, patient interaction, or environmental exposure. In hospital practice, repeated failure records often point to environmental effects, external power supply problems, user errors, consumable or accessory-related problems, and patient-related damage rather than to a single technical defect. This classification is useful because different failure sources require different interventions. A hardware failure may require repair, a user error may require training, a consumable-related failure may require procurement control, and an environmental failure may require infrastructure correction.

3.6. Risk Management, Safety, and Incident Reporting

Biomedical and clinical engineering units contribute directly to institutional risk management. Medical devices can create risks through electrical hazards, mechanical hazards, software failures, incorrect measurements, alarm failures, radiation exposure, infection transmission, cybersecurity vulnerabilities, and use errors. Therefore, device-related risk management should be embedded within hospital quality and patient safety systems.

ISO 14971:2019 defines a structured risk management process for medical devices, including hazard identification, risk estimation, risk evaluation, risk control, residual risk evaluation, and production and post-production monitoring (ISO, 2019). Although ISO 14971 is primarily applied by manufacturers, its principles are highly relevant to healthcare institutions that operate and manage devices throughout their useful life. Hospitals must identify device-related hazards in their own clinical environments, monitor incident patterns, and implement local risk controls such as training, maintenance, labeling, restricted access, checklists, and device replacement.

Incident reporting is essential for learning from device-related problems. Reports may include adverse events, near misses, user errors, alarm failures, unexpected device behavior, repeated malfunctions, or conditions that could have caused harm. A non-punitive reporting culture encourages users to report problems early, allowing biomedical engineering units to investigate root causes and prevent recurrence. Without reporting, the same device-related problems may recur across departments.

The EU MDR emphasizes post-market surveillance and vigilance as central components of device safety, requiring systematic collection and evaluation of information from devices after they are placed on the market (European Parliament and Council of the European Union, 2017). While manufacturers have formal regulatory responsibilities, healthcare institutions also play a practical role by reporting device problems, documenting incidents, preserving evidence, communicating with manufacturers, and implementing field safety corrective actions when required.

3.7. Collaboration with Information Technology and Cybersecurity Teams

The increasing connectivity of medical devices has created a strong need for collaboration between biomedical engineering and information technology teams. Many devices now operate as networked systems rather than isolated equipment. They may transmit patient data, receive software updates, connect to central monitoring systems, interact with electronic health records, or communicate with cloud-based services. This creates benefits for clinical workflow but also introduces cybersecurity and interoperability risks.

Biomedical engineering teams understand device function, clinical use, maintenance requirements, and safety implications. Information technology teams understand networks, access control, cybersecurity policies, data governance, and system integration. Effective management of connected medical devices requires both forms of expertise. For example, a software update may appear to be an IT task, but it may also affect alarm behavior, measurement accuracy, user interface, device compatibility, or clinical workflow. Similarly, network segmentation may be a cybersecurity control, but it must be implemented without disrupting critical device communication.

The FDA's current cybersecurity guidance for medical devices states that cybersecurity risks should be addressed through device design, labeling, and documentation within the quality management system, with the aim of ensuring that marketed devices are resilient to cybersecurity threats (FDA, 2026). FDA cybersecurity materials also emphasize that medical device cybersecurity is an ongoing concern for connected and software-enabled technologies, and that guidance has been updated to address expectations for cyber devices. These developments show that cybersecurity is no longer separate from medical device safety; it is part of the broader risk management framework.

Healthcare institutions should therefore maintain a joint inventory of connected medical devices, including software versions, network addresses, data flows, update status, user access levels, and cybersecurity risk classifications. Biomedical and IT teams should jointly evaluate new device acquisitions, network integration, software patches, remote access arrangements, vendor accounts, and incident response procedures. This collaboration is especially important for patient monitors, infusion pumps, imaging systems, laboratory analyzers, anesthesia systems, telemedicine equipment, and AI-enabled clinical software.

4. RISK MANAGEMENT IN MEDICAL DEVICES

Risk management is a central component of safe and effective medical device governance. Medical devices are designed to provide clinical benefit; however, their use may also introduce hazards for patients, users, maintenance personnel, and third parties. These hazards may arise from electrical, mechanical, thermal, biological, chemical, software-related, environmental, usability-related, or cybersecurity-related sources. Therefore, medical device risk management should not be limited to the manufacturer's design process or to a single pre-market evaluation. It should be understood as a continuous, life-cycle-based process that extends from device conception to clinical use, maintenance, post-market monitoring, and final decommissioning.

ISO 14971:2019 is the international standard that specifies terminology, principles, and a process for risk management of medical devices, including software as a medical device and in vitro diagnostic medical devices. The standard establishes a systematic framework for identifying hazards, estimating and evaluating associated risks, controlling risks, and monitoring the effectiveness of controls throughout the entire device life cycle. ISO states that the 2019 version was reviewed and confirmed in 2025, meaning that it remains the current version of the standard.

Risk management is also strongly linked to regulatory compliance. Regulation (EU) 2017/745 on medical devices establishes requirements intended to ensure the safety and performance of medical devices placed on the European market, including stronger expectations for clinical evaluation, traceability, post-market surveillance, vigilance, and life-cycle responsibility. In healthcare institutions, risk management is operationalized through appropriate procurement, safe installation, user training, preventive maintenance, calibration, incident reporting, corrective actions, and decommissioning decisions.

4.1. Concept of Risk in Medical Devices

In medical device management, risk is generally understood as the combination of the probability of harm and its severity. This definition is important because a rare but catastrophic event and a frequent but minor event may both require systematic evaluation. Risk management, therefore, requires more than identifying whether a device can fail. It requires understanding the potential harm, its likelihood, the severity of the consequences, and whether the risk is acceptable relative to the intended clinical benefit.

Several related concepts should be distinguished. A hazard is a potential source of harm. A hazardous situation occurs when a person, patient, user, or environment is exposed to that hazard. Harm is the actual injury or damage that may result. For example, electrical leakage is a hazard; patient contact with an inadequately grounded device is a hazardous situation; and electrical injury is the harm. Similarly, an incorrect infusion rate may be a hazardous situation associated with a pump malfunction or programming error, while overdose or underdose may cause possible harm.

Risk in medical devices is therefore multidimensional. It includes not only technical failure but also human-device interaction, organizational procedures, maintenance policies, infrastructure conditions, and clinical context. A device may be technically functional but still unsafe if users are inadequately trained, alarms are poorly understood, accessories are incompatible, calibration is overdue, or environmental conditions are unsuitable.

4.2. ISO 14971:2019-Based Risk Management Process

ISO 14971:2019 provides a structured process for managing medical device risks. Although the standard is primarily directed at manufacturers, its logic is also highly relevant for healthcare institutions that operate, maintain, and monitor devices in real clinical environments. The main steps of an ISO 14971-based risk management process include risk management planning, hazard identification, risk estimation, risk evaluation, risk control, residual risk evaluation, benefit–risk analysis, and production and post-production monitoring.

The first step is preparing a risk management plan. This plan defines the scope of risk management activities, responsibilities, criteria for risk acceptability, methods for evaluating risks, verification activities, and requirements for reviewing the effectiveness of risk controls. In a healthcare institution, an equivalent approach may be reflected in medical device management policies, risk-based maintenance plans, training requirements, incident-reporting systems, and device-replacement criteria.

The second step is hazard identification. Hazards may be associated with energy sources, biological contamination, software malfunctions, incorrect measurements, user interface design, alarm failures, mechanical movement, radiation, cybersecurity vulnerabilities, or environmental exposure. Hazard identification should include both normal use and reasonably foreseeable misuse. For example, a ventilator may be evaluated not only for mechanical and electrical safety but also for alarm interpretation, tubing connection errors, cleaning requirements, power failure, battery status, and user response during emergency conditions.

The third step is risk estimation, in which the probability of occurrence and severity of harm are assessed. In clinical practice, precise probability values may be difficult to determine; therefore, qualitative or semi-quantitative risk matrices are often used. Severity may be classified as negligible, minor, serious, critical, or catastrophic, while probability may be classified as remote, occasional, probable, or frequent. However, risk matrices should be used carefully. They support structured thinking but should not replace expert judgment, clinical context, or evidence from incident data.

The fourth step is risk evaluation. At this stage, the estimated risk is compared with predefined acceptability criteria. Some risks may be acceptable without further control, while others require additional mitigation. High-severity risks, especially those involving life-support devices, medication delivery systems, radiation exposure, or invasive procedures, generally require stronger risk controls even when their probability is low.

The fifth step is risk control. Risk controls are implemented to reduce risk to an acceptable level. ISO 14971 follows a hierarchy of risk control measures: inherent safety by design, protective measures, and information for safety. In healthcare settings, these may correspond to choosing safer device designs, enabling alarms and interlocks, applying maintenance and calibration controls, restricting access to trained users, providing clear instructions, and using clinical checklists.

The final stage is monitoring and review. Risk management does not end after a device is installed or after an initial risk assessment is completed. Device failures, user complaints, near misses, adverse events, calibration failures, software updates, cybersecurity alerts, and maintenance records may all provide new risk information. Therefore, risk management must be dynamic and continuously updated.

4.3. Risk Management Across the Device Life Cycle

Risk management should be integrated into every stage of the medical device life cycle. During the needs assessment stage, risks may arise from selecting an inappropriate device, underestimating clinical demand, failing to consider user competence, or failing to account for infrastructure requirements. A device that is suitable for one hospital may be unsuitable for another if trained users, spare parts, network infrastructure, or maintenance capability are not available.

During procurement, risk management requires careful evaluation of technical specifications, regulatory conformity, service support, usability, accessories, cybersecurity requirements, warranty conditions, and total cost of ownership. If these factors are ignored, the institution may acquire devices that are difficult to maintain, incompatible with clinical workflows, or unsafe under local conditions.

During installation, risk management focuses on the physical and technical environment. Electrical safety, grounding, ventilation, temperature control, humidity, electromagnetic compatibility, uninterrupted power supply, and network configuration should be verified before routine clinical use. Installation failures can create hidden hazards that may not become apparent until the device is used under clinical pressure.

During clinical use, risk management depends heavily on user competence, standardized procedures, alarm management, correct accessories, cleaning and disinfection practices, and reporting of abnormal device behavior. Many use-related risks are not caused solely by device failure but by the interaction among device design, user behavior, and clinical workflow.

During maintenance and calibration, risk management aims to ensure that the device continues to operate within acceptable performance limits. WHO states that maintenance strategies should include inspection, preventive maintenance, corrective maintenance, performance inspections, and safety inspections. Performance inspections ensure correct operation, safety inspections protect patients and operators, and preventive maintenance aims to extend equipment life and reduce failure rates.

During decommissioning, risk management involves determining when continued use of the device becomes unsafe, uneconomical, unsupported, or clinically inappropriate. Decommissioning decisions may be triggered by repeated failures, lack of spare parts, calibration failure, obsolete software, cybersecurity vulnerabilities, or incompatibility with current clinical requirements.

4.4. Common Hazard Categories in Medical Devices

Medical device hazards can be grouped into several categories. This classification helps biomedical engineering units, clinical users, and risk managers systematically identify potential sources of harm.

Electrical hazards include leakage current, inadequate grounding, insulation failure, power supply instability, battery failure, and inappropriate use of extension cables or adapters. Such hazards are especially important in devices that are connected directly to patients or used in wet environments.

Mechanical hazards include moving parts, sharp edges, unstable mounting, device falls, pressure injury, tubing obstruction, or incorrect mechanical assembly. Devices such as hospital beds, infusion pumps, surgical tables, imaging systems, and rehabilitation equipment may involve mechanical risks.

Thermal hazards arise from excessive heat, cold exposure, burns, overheated components, or temperature control failure. Incubators, warming devices, electrosurgical units, imaging systems, and laboratory equipment may present thermal risks.

Biological and infection-related hazards include contamination, inadequate cleaning, biofilm formation, reuse of single-use components, failure of sterilization, and improper handling of patient-contact accessories. These risks are particularly relevant to invasive devices, respiratory devices, endoscopes, and reusable surgical instruments.

Chemical hazards may arise from disinfectants, sterilizing agents, batteries, reagents, gases, or leaking substances. Laboratory analyzers, sterilization systems, and devices using consumables or chemical reagents require careful management.

Radiation hazards are associated with X-ray imaging, computed tomography, fluoroscopy, radiotherapy systems, nuclear medicine devices, and laser systems. Risk control in these devices requires shielding, dose monitoring, quality assurance, operator training, and regulatory compliance.

Software and data hazards include incorrect algorithms, software bugs, data loss, patient misidentification, user-interface errors, alarm logic failures, update-related malfunctions, and interoperability errors. These hazards have become increasingly important as medical devices become more digital and interconnected.

Cybersecurity hazards include unauthorized access, malware, ransomware, weak passwords, insecure remote access, outdated software, and network misconfiguration. These hazards may affect device availability, data integrity, patient privacy, and, in some cases, clinical function. FDA guidance emphasizes that cybersecurity risks should be addressed through design, labeling, documentation, and quality management system processes for medical devices that pose cybersecurity risks.

Use-related hazards include incorrect settings, skipped pre-use checks, misinterpretation of alarms, improper accessory selection, incorrect patient connection, and failure to follow instructions for use. These hazards are strongly associated with human factors, training, workload, and usability.

Environmental hazards include temperature, humidity, dust, vibration, electromagnetic interference, fluid exposure, unstable surfaces, and power interruptions. These sources of failure are also consistent with practical biomedical device training materials, which treat environmental effects, power supply problems, user errors, consumable or accessory-related problems, and patient-related damage as common fault categories (MEGEP, 2008).

4.5. Risk Control Measures

Risk control measures aim to reduce risks to an acceptable level. A well-designed risk control strategy should follow a hierarchy. The first priority is to eliminate or reduce risk through design or selection of safer technology. The second priority is to use protective measures such as alarms, interlocks, guards, automatic shutdown, access control, and fail-safe mechanisms. The third priority is to provide safety-related information, including labeling, warnings, user manuals, training, and clinical procedures.

In healthcare institutions, many risk controls are operational rather than design-based. These include preventive maintenance, calibration, electrical safety testing, user training, competency assessment, quick-reference guides, checklists, restricted access to high-risk devices, incident reporting, and environmental control. For example, a defibrillator risk control strategy may include daily functional checks, battery monitoring, electrode compatibility control, user training, clear labeling, and rapid reporting of device faults.

Alarms are important risk control measures, but they must be carefully managed. Too many non-actionable alarms may lead to alarm fatigue, while unclear alarms may delay user response. Therefore, alarm configuration, user education, and departmental policies should be part of the risk control strategy.

Maintenance and calibration also function as risk controls. Preventive maintenance reduces the probability of failure, while calibration reduces the risk of inaccurate measurement or output. WHO’s maintenance guidance emphasizes the role of performance inspections, safety inspections, and preventive maintenance in ensuring the correct and safe operation of medical equipment.

Training is another essential risk control. However, training should not be treated as the only solution for design or system problems. If users repeatedly make the same error, the problem may involve interface design, workflow mismatch, labeling, accessory confusion, or inadequate alarm logic. In such cases, risk control should address the system, not merely the individual user.

4.6. Residual Risk and Benefit–Risk Evaluation

Even after risk control measures are applied, some level of risk may remain. This is known as residual risk. The key question is whether the remaining risk is acceptable relative to the device's intended clinical benefit. For example, radiation-based imaging systems carry inherent radiation risk, but this risk may be acceptable when the diagnostic benefit is significant, and exposure is controlled according to appropriate protocols.

Benefit–risk evaluation is particularly important for high-risk or life-support devices. A ventilator, infusion pump, defibrillator, or implantable device may present serious hazards if it fails, yet its clinical benefit may be essential. Therefore, risk management does not aim to eliminate all risks completely, which is often impossible. Instead, it aims to identify, reduce, monitor, and justify remaining risks.

Residual risk should be communicated to users when necessary. This may occur through warnings, contraindications, training materials, labeling, manuals, and institutional procedures. However, safety information should not be used as a substitute for feasible design or protective measures. If a risk can reasonably be reduced through technical or organizational control, simply issuing a warning to the user is not sufficient.

In healthcare institutions, benefit–risk evaluation also applies to decisions about continued use of older devices. A device that still functions may no longer be acceptable if it has repeated failures, lacks spare parts, cannot be calibrated reliably, is incompatible with current clinical systems, or contains unresolved cybersecurity vulnerabilities. In such cases, continued operation may create a residual risk that is no longer justified.

4.7. Post-Market Surveillance and Risk Feedback

Risk management depends on continuous feedback from real-world use. Devices may behave differently in clinical environments than they did during design verification or pre-market evaluation. Different users, patient conditions, workloads, environmental factors, accessories, software configurations, and maintenance practices may reveal hazards that were not fully apparent earlier.

Post-market surveillance refers to the systematic collection and evaluation of information from devices after they are placed on the market. Under the EU MDR, post-market surveillance and vigilance are strengthened as part of the life-cycle regulatory approach, helping manufacturers and authorities identify safety signals, update risk–benefit evaluations, and take corrective actions when needed.

Healthcare institutions contribute to this feedback system through adverse incident reporting, near-miss reporting, maintenance records, calibration results, user complaints, service reports, and participation in field safety corrective actions. The MHRA guidance states that healthcare organizations should appoint Medical Device Safety Officers, and that part of this role is to report adverse incidents to the MHRA and other official agencies.

Biomedical and clinical engineering units are especially important in translating field experience into institutional learning. If repeated faults occur with a particular device model, if users repeatedly misinterpret an alarm, or if a specific accessory causes frequent malfunctions, these patterns should be analyzed and addressed. Corrective actions may include user retraining, modification of local procedures, replacement of accessories, revision of maintenance schedules, communication with suppliers, or device replacement.

A mature risk feedback system should be non-punitive, transparent, and learning-oriented. Users should be encouraged to report abnormal device behavior, near misses, and usability problems without fear of blame. Otherwise, important safety signals may remain hidden until patient harm occurs.

5. USE ERRORS AND HUMAN FACTORS IN MEDICAL DEVICES

Use errors are among the most important contributors to medical device-related incidents, device failures, workflow interruptions, and patient safety risks. In traditional discussions, these events are sometimes described simply as “user mistakes.” However, contemporary patient safety and human factors approach emphasize that use errors typically arise from interactions among the user, the device, the task, the clinical environment, organizational procedures, training systems, and time pressure. Therefore, the use of the term 'error' should not be understood merely as individual carelessness. It should be examined as a system-level safety issue involving usability, workflow design, device complexity, user competence, and institutional safety culture.

The importance of use-related risk is recognized in both risk management and usability engineering frameworks. ISO 14971:2019 identifies risk management as a life-cycle process for medical devices, including hazard identification, risk estimation and evaluation, implementation of risk controls, and monitoring of control effectiveness over time. This framework is directly relevant to use-related hazards because incorrect operation, foreseeable misuse, poor alarm interpretation, and inadequate training may all lead to harm. IEC 62366-1:2015 specifically addresses usability engineering for medical devices and defines a process for manufacturers to analyze, specify, develop, and evaluate usability as it relates to safety; it also supports the assessment and mitigation of risks associated with correct use and use errors.

The FDA human factors guidance similarly emphasizes that human factors and usability engineering processes should maximize the likelihood that medical devices are safe and effective for intended users, intended uses, and intended use environments. This perspective is particularly important because medical devices are often used in complex clinical settings such as intensive care units, emergency departments, operating rooms, radiology units, laboratories, and home-care environments, where users may face high workload, interruptions, stress, incomplete information, and rapidly changing patient conditions.

5.1. Definition of Use Error

A use error can be broadly defined as a user action or omission that results in a medical device response different from that intended by the manufacturer or expected in safe clinical use. Such errors may occur during device setup, parameter entry, patient connection, alarm response, cleaning, maintenance checks, accessory selection, data interpretation, or shutdown procedures.

Use error should be distinguished from intentional misuse, reckless behavior, and device malfunction. A use error may occur even when the user is trying to operate the device correctly. For example, a nurse may select an incorrect infusion rate because the interface displays units unclearly; a clinician may silence an alarm because previous alarms were non-actionable; or a technician may skip a routine check because the procedure is not clearly integrated into the workflow. These examples show that use errors often reflect the interaction between user behavior and system design rather than simple negligence.

Use error may be defined as an action or omission that results in a device response different from the manufacturer's intended response. From a human factors perspective, attention errors, memory errors, rule-based failures, knowledge-based failures, and routine violations should be considered separately because each has different cognitive and organizational roots (Food and Drug Administration [FDA], 2016; International Electrotechnical Commission [IEC], 2015).

5.2. Difference Between Use Error, User Mistake, and Device Failure

In medical device safety analysis, it is important to distinguish between use errors, user mistakes, and device failures. These terms are related but not identical.

A device failure refers to a technical malfunction of the device itself. This may involve hardware failure, software malfunction, sensor drift, power supply problems, mechanical damage, battery failure, or calibration loss. In this case, the device does not perform as intended because of a technical defect or degradation.

A user mistake is often used informally to describe an incorrect action performed by the user. However, this term may imply individual blame and may obscure the system factors that contributed to the event. For example, entering the wrong value into an infusion pump may be described as a user mistake, but the underlying causes may include confusing menu structure, similar-looking buttons, poor display contrast, inadequate training, or high workload.

A use error is a more appropriate term in safety analysis because it focuses on the interaction between the user and the device. It does not automatically assign blame to the individual user. Instead, it asks why the error occurred and whether device design, labeling, workflow, training, environment, or organizational procedures contributed to the event.

This distinction is important for prevention. If every event is treated as an individual mistake, the response may be limited to a warning or retraining the user. However, if the event is analyzed as a use error, the response may include interface redesign, alarm optimization, clearer labeling, improved training, standardized checklists, environmental controls, or changes in procurement criteria.

5.3. Human Factors and Usability Engineering

Human factors engineering examines how people interact with systems, devices, tasks, environments, and organizational conditions. In medical device management, human factors engineering aims to design and manage devices to support safe, efficient, and reliable use. Usability engineering is a closely related concept that focuses on whether intended users can use a device correctly, safely, and effectively in the intended use environment.

IEC 62366-1:2015 provides an international framework for applying usability engineering to medical devices. It specifies a process for manufacturers to analyze, specify, develop, and evaluate usability as it relates to safety, and allows manufacturers to assess and mitigate risks associated with correct and incorrect use during normal use. Although IEC 62366-1 is primarily directed at manufacturers, its principles are highly relevant to healthcare institutions, as hospitals must select, deploy, train users on, and monitor devices in real-world clinical settings.

The FDA guidance on applying human factors and usability engineering to medical devices was developed to help industry follow appropriate human factors processes and to maximize the likelihood that devices will be safe and effective for intended users, uses, and use environments. This guidance highlights the importance of understanding who the users are, what tasks they perform, what environments they work in, and what use-related hazards may arise.

Human factors considerations include the user's knowledge, physical abilities, cognitive workload, memory limitations, stress level, fatigue, expectations, prior experience, and training. Device-related factors include screen layout, button design, alarm messages, menu depth, terminology, color coding, feedback, physical ergonomics, tubing connections, connector compatibility, and labeling. Environmental factors include lighting, noise, interruptions, space constraints, emergency conditions, infection-control requirements, and shift workload.

In this sense, safe device use is not produced only by competent users. It is produced by aligning device design, training, workflow, environmental conditions, and organizational support.

5.4. Cognitive Sources of Use Error

Use errors may arise from several cognitive mechanisms. Understanding these mechanisms helps institutions design more effective preventive strategies.

Attention errors occur when the user overlooks, misreads, or fails to notice relevant information. In a busy clinical environment, attention may be divided among several tasks, alarms, patients, and communication demands. For example, a user may fail to notice that a device is connected to the wrong patient, that a battery indicator is low, or that an alarm limit has been changed.

Memory errors occur when the user forgets to perform a required step. This may include forgetting to perform a pre-use check, forgetting to reconnect a sensor after patient movement, or forgetting to reset a device after cleaning. Memory errors are more likely when procedures are long, interruptions are frequent, or task steps lack checklists or visual cues.

Rule-based errors occur when the user applies an incorrect rule or applies a correct rule in the wrong context. For example, a clinician may use the operating logic of one brand of infusion pump when using another model with a different menu structure. This type of error is common when institutions use multiple device models with different interfaces.

Knowledge-based errors occur in unfamiliar or complex situations where the user must reason from limited information. These errors may arise during rare alarms, unusual device behavior, complex troubleshooting scenarios, or emergency use. Users with limited experience are especially vulnerable to knowledge-based errors.

Routine violations occur when users knowingly deviate from safe procedures, often because of time pressure, workload, convenience, or normalization of unsafe practices. Examples include bypassing pre-use checks, ignoring minor alarms, using non-recommended accessories, or continuing to use a device despite repeated warning signs.

This classification is valuable because it connects use errors with both cognitive psychology and institutional safety management.

5.5. System-Level Causes of Use Errors

Use errors rarely occur in isolation. They are often shaped by system-level conditions. A user may make an error because the device interface is confusing, the alarm is unclear, the lighting is poor, the unit is understaffed, the device model is unfamiliar, or the training was insufficient.

One common system-level cause is interface complexity. Devices with deep menus, ambiguous icons, small text, similar-looking buttons, inconsistent terminology, or unclear confirmation messages increase the probability of incorrect operation. Interface complexity is especially problematic when devices must be used quickly in emergency conditions.

Another important cause is alarm fatigue. In intensive care units, operating rooms, emergency departments, and monitoring areas, users may be exposed to a large number of alarms, many of which may be non-actionable or clinically insignificant. Over time, users may become less responsive to alarms or may silence them without adequate assessment. Alarm fatigue is therefore not simply a user attitude problem; it is a system problem related to alarm design, configuration, clinical workflow, and institutional policy.

Inadequate training is also a major contributor. Training limited to a single installation session may not be sufficient, especially for high-risk devices. Staff turnover, software updates, device relocation, new accessories, and changes in clinical workflow all require renewed training. MHRA guidance states that healthcare organizations must keep records of user training and ensure that users know how to use devices safely, can carry out routine checks and maintenance, have received relevant refresher training, and are competent or confident to use devices in their areas of work.

Device variability within the same institution may also increase risk. If different departments use different brands or models for the same clinical function, users moving between units may transfer habits from one device to another. This may lead to incorrect parameter entry, misinterpretation of alarms, or skipped steps.

Workflow pressure is another important factor. In emergency and high-acuity settings, users may need to act rapidly while managing multiple competing demands. If safe device use requires too many steps, excessive documentation, or poorly placed controls, users may develop shortcuts that increase risk.

Environmental conditions also matter. Noise, poor lighting, limited space, unstable surfaces, fluid exposure, electromagnetic interference, and frequent interruptions may contribute to errors in use. The source text highlights environmental factors, electrical supply issues, user errors, problems with consumables or accessories, and patient-related damage as important causes of device failure.

5.6. Examples of Use Errors in Medical Devices

Use errors can occur across many device categories. Examples help illustrate how human factors and system conditions interact with device design.

In infusion pumps, use errors may include incorrect dose or flow-rate entry, selection of the wrong drug library profile, failure to detect occlusion, use of incompatible infusion sets, incorrect priming, or silencing alarms without identifying the underlying cause. Because infusion pumps deliver medication directly to patients, even small programming errors may have serious consequences.

In ventilators, use errors may include incorrect mode selection, inappropriate alarm limits, failure to recognize disconnection, incorrect humidification setup, wrong tubing configuration, or delayed response to pressure alarms. These errors are especially critical because ventilators support respiratory function in vulnerable patients.

In defibrillators, use errors may include failure to check battery status, incorrect pad placement, wrong energy selection, delayed charging, failure to recognize synchronized versus unsynchronized shock mode, or lack of routine functional checks. In emergencies, familiarity with the device and a simple interface design are essential.

In patient monitors, use errors may include incorrect alarm limit settings, sensor misplacement, failure to respond to alarms, use of damaged cables, incorrect patient profile selection, or misinterpretation of artifacts as physiological changes. Monitor-related use errors can delay the recognition of deterioration.

In radiological imaging systems, use errors may include incorrect protocol selection, wrong patient positioning, inadequate radiation protection, incorrect exposure parameters, failure to recognize image artifacts, or improper use of contrast-related workflows. In such systems, use errors may affect both diagnostic quality and radiation safety.

In laboratory analyzers, use errors may involve incorrect sample loading, reagent mismatch, calibration neglect, failure to respond to quality-control warnings, or incorrect interpretation of error messages. These errors may affect diagnostic reliability and downstream clinical decisions.

These examples show that use errors are context-dependent. The same device may be safe in one environment and risky in another if user training, workload, accessories, maintenance, or infrastructure differ.

5.7. Prevention Strategies

Preventing use errors requires a combination of design, training, procedural, technical, and organizational strategies. No single intervention is sufficient.

The first strategy is to select devices based on usability. During procurement, healthcare institutions should consider not only price and technical specifications but also interface clarity, alarm design, training requirements, accessory compatibility, and user feedback. Devices that are difficult to learn, difficult to clean, or prone to confusing operation may increase long-term risk and cost.

The second strategy is structured user training. Training should include hands-on practice, scenario-based exercises, alarm interpretation, common error conditions, pre-use checks, cleaning procedures, and emergency actions. For high-risk devices, competency assessment and periodic refresher training should be required.

The third strategy is standardization. Reducing unnecessary variation in device models, accessories, terminology, alarm policies, and workflows can reduce cognitive burden. When standardization is not possible, users should be clearly trained on the differences between models.

The fourth strategy is checklists and quick-reference tools. Short pre-use checklists, laminated bedside guides, alarm response cards, and troubleshooting flowcharts can support memory and reduce omissions. These tools are especially useful for complex or infrequently used devices.

The fifth strategy is simulation-based training. Simulation allows users to practice device setup, alarm response, fault recognition, emergency procedures, and troubleshooting without exposing patients to harm. It is particularly useful for biomedical device technology education, critical care training, and high-risk low-frequency events.

The sixth strategy is incident and near-miss reporting. Use errors should be reported and analyzed in a non-punitive manner. The aim should be learning and system improvement rather than blame. Reported events can reveal interface problems, training gaps, confusion about accessories, workflow pressure, or maintenance issues.

The seventh strategy is collaboration between clinical users, biomedical engineering, quality management, and information technology teams. Use errors often involve multiple domains: device function, clinical workflow, training, data integration, alarms, and cybersecurity. Therefore, prevention requires multidisciplinary coordination.

For healthcare institutions, reducing use errors requires more than warning users or repeating training after incidents. It requires usability-aware procurement, structured training, competency assessment, simulation-based practice, standardization, checklists, incident reporting, risk-based maintenance, and collaboration among clinical, technical, quality, and information technology teams. Ultimately, safer medical device use depends on designing and managing systems that make correct use easier and unsafe use less likely.

6. DIGITALIZATION, SOFTWARE, ARTIFICIAL INTELLIGENCE, AND CYBERSECURITY RISKS IN MEDICAL DEVICES

Medical devices are increasingly becoming software-driven, network-connected, data-generating, and algorithmically supported systems. This transformation has expanded the scope of medical device management beyond traditional maintenance, calibration, and repair activities. In contemporary healthcare environments, many devices interact with electronic health records, hospital information systems, picture archiving and communication systems, laboratory information systems, cloud platforms, remote monitoring tools, and artificial intelligence-based decision-support systems. As a result, device safety now depends not only on mechanical or electrical reliability but also on software quality, data integrity, interoperability, cybersecurity, algorithmic performance, and human oversight.

This digital transformation creates important opportunities. Connected devices can improve clinical workflow, enable real-time monitoring, reduce manual documentation, support remote care, accelerate diagnostics, and provide decision-support information. However, it also introduces new risks. Software defects, cyber vulnerabilities, incorrect patient data matching, failed updates, algorithmic bias, data transfer errors, and overreliance on automated outputs may all pose patient safety concerns. Therefore, digital medical devices require a broader life-cycle management approach that integrates biomedical engineering, information technology, cybersecurity, clinical governance, quality management, and regulatory compliance.

6.1. Digital Transformation of Medical Devices

Traditional medical devices were often understood primarily as physical equipment: electrical, mechanical, hydraulic, pneumatic, optical, or radiological systems used for diagnosis, treatment, monitoring, or support. Although these physical components remain important, many modern devices now include embedded software, digital sensors, wireless communication, data storage, automated control, network interfaces, and remote update capabilities.

Examples include patient monitors connected to central monitoring stations, infusion pumps using drug libraries and dose-error reduction systems, ventilators with adaptive control modes, imaging systems with automated reconstruction algorithms, laboratory analyzers connected to laboratory information systems, wearable sensors transmitting physiological data, and AI-enabled diagnostic tools. These devices operate within a wider digital health ecosystem rather than as isolated clinical instruments.

This shift changes the responsibilities of healthcare institutions. A device may be technically functional but unsafe if its software version is outdated, it is connected to an insecure network, its data are transmitted incorrectly, or users misunderstand algorithm-generated outputs. Therefore, medical device management must include software version tracking, network configuration, cybersecurity classification, data-flow mapping, user-access management, update verification, and post-update performance checks.

The European Union Medical Device Regulation recognizes the increasing importance of software and digital technologies by explicitly including software within the regulatory framework when it meets the definition of a medical device. Regulation (EU) 2017/745 strengthens requirements related to safety, performance, clinical evaluation, traceability, post-market surveillance, and life-cycle responsibility for medical devices, including software-based technologies.

6.2. Software as a Medical Device and Embedded Software

Medical device software can be broadly grouped into two categories: embedded software and standalone medical software. Embedded software is software that forms part of a physical medical device, such as the control software of an infusion pump, ventilator, defibrillator, imaging system, or laboratory analyzer. Software as a Medical Device (SaMD), on the other hand, refers to software intended for one or more medical purposes that is not part of a hardware medical device.

The International Medical Device Regulators Forum defines SaMD as software intended to be used for one or more medical purposes that perform these purposes without being part of a hardware medical device. This definition has become an important global reference point for regulators and manufacturers. Examples of SaMD may include image-analysis software, diagnostic decision-support algorithms, software that calculates patient-specific risk scores, or applications that process physiological data for medical purposes.

Software life-cycle management is a critical component of safety. IEC 62304:2006+A1:2015 defines life-cycle requirements for medical device software and establishes a common framework for software development and maintenance processes. The standard applies both when software is itself a medical device and when software is embedded or integral to a final medical device. From a hospital management perspective, this means that software-related records should be considered part of the device's technical history. These records may include software version, update history, known anomalies, validation status, service notes, cybersecurity patches, and compatibility information.

Software-related risks may arise from coding errors, inadequate verification, unexpected interactions between modules, incorrect alarm logic, data-processing failures, user-interface defects, interoperability problems, or updates that change device behavior. Therefore, software should not be treated as an invisible or secondary component of the device. It is a safety-critical element that must be managed throughout the device life cycle.

6.3. Interoperability and Data Integrity Risks

Interoperability refers to the ability of devices and systems to exchange, interpret, and use information correctly. In healthcare institutions, medical devices may interact with electronic health records, hospital information systems, laboratory information systems, radiology information systems, picture archiving and communication systems, pharmacy systems, central monitoring stations, and remote-care platforms. Interoperability can improve workflow, reduce manual transcription, and support more complete clinical documentation.

However, interoperability also creates new risks. Data may be transferred to the wrong patient record, transmitted incompletely, delayed, duplicated, overwritten, or displayed in an incorrect format. A patient monitor may send measurements to the wrong bed location if the patient-device association is incorrect. A laboratory analyzer may transmit results under an incorrect patient identifier. An imaging system may fail to transfer images to PACS properly, delaying diagnosis. An infusion pump drug library may not match the institution's current medication policy if not updated correctly.

Data integrity is, therefore, a patient safety issue. It is not merely a technical or administrative concern. Incorrect or incomplete device data may affect clinical decisions, medication administration, diagnosis, treatment planning, and quality reporting. For this reason, biomedical engineering and information technology teams should jointly manage device integration, interface testing, patient identification workflows, data mapping, and error-reporting mechanisms.

Interoperability risks are particularly important when systems are updated independently. A software update in the hospital information system, PACS, LIS, or device firmware may alter data formats, communication protocols, or interface behavior. Therefore, any update that affects connected devices should be tested before full clinical deployment.

6.4. Artificial Intelligence in Medical Devices

Artificial intelligence is increasingly being incorporated into medical devices and medical software. AI-enabled devices may assist in radiological image interpretation, ultrasound analysis, electrocardiogram interpretation, early warning scores, triage support, pathology image analysis, ophthalmology screening, surgical navigation, and risk prediction. The FDA maintains an AI-enabled medical device list to identify medical devices authorized for marketing in the United States that use AI technologies. The FDA states that this list provides transparency for healthcare providers and patients and helps identify the current device landscape.

AI-enabled medical devices can improve speed, consistency, and access to specialized analysis. For example, AI tools may help detect imaging abnormalities, estimate clinical measurements, identify patterns in physiological data, or support decision-making in low-resource settings. However, AI also introduces specific safety challenges. These include algorithmic bias, poor generalizability across patient populations, dataset limitations, lack of transparency, user overreliance, performance drift, inappropriate automation, and unclear responsibility when AI output contributes to clinical decisions.

The FDA has emphasized the need for a total product life-cycle approach for AI-enabled devices. In 2025, the agency issued draft guidance intended to provide comprehensive recommendations for AI-enabled medical devices throughout the total product life cycle, including design, development, maintenance, and documentation considerations. This is important because AI models may perform differently across clinical settings, patient populations, imaging protocols, device configurations, or data distributions.

Healthcare institutions using AI-enabled devices should not assume that regulatory authorization alone guarantees perfect performance in every local setting. Local validation, user training, workflow integration, monitoring of false positives and false negatives, documentation of AI output, and clear human oversight are necessary. AI should support clinical judgment rather than replace it. Users should understand the intended use, limitations, required input quality, and appropriate response to AI-generated outputs.

6.5. Cybersecurity Risks in Connected Medical Devices

Cybersecurity has become a core component of medical device safety. Network-connected devices may be exposed to unauthorized access, malware, ransomware, data theft, remote manipulation, service disruption, weak authentication, insecure communication, outdated software, and poor network segmentation. Such vulnerabilities may affect not only the confidentiality of patient data but also device availability, data integrity, and, in some cases, clinical function.

The FDA’s 2026 guidance on cybersecurity in medical devices provides recommendations regarding cybersecurity device design, labeling, and documentation to include in premarket submissions for devices with cybersecurity risk. The guidance aims to help ensure that marketed medical devices are sufficiently resilient to cybersecurity threats. FDA cybersecurity materials also note that the 2025 final guidance added recommendations related to section 524B of the FD&C Act for cyber devices and superseded the 2023 version.

Cybersecurity incidents or vulnerabilities may have direct clinical implications. For example, vulnerabilities in patient monitors could allow unauthorized access, manipulation, network compromise, or export of personally identifiable and protected health information. Public cybersecurity notices and vulnerability disclosures demonstrate why cybersecurity must be treated as part of clinical risk management rather than solely as an information technology issue.

Common cybersecurity controls for medical devices include asset inventory, network segmentation, access control, strong authentication, timely patching, secure configuration, vendor-account management, encryption where appropriate, logging, backup planning, vulnerability monitoring, and incident response procedures. These controls require collaboration among biomedical engineers, information technology professionals, cybersecurity officers, clinical users, procurement teams, and vendors.

6.6. Software Updates, Patch Management, and Version Control

Software updates and cybersecurity patches are necessary for maintaining device safety and security, but they may also introduce new risks if not managed properly. An update may change alarm behavior, user-interface layout, data transmission format, measurement algorithm, device compatibility, or performance characteristics. Therefore, software updates should be treated as controlled technical events rather than routine background changes.

Version control is essential. Biomedical engineering units should maintain records of software versions, firmware versions, patch history, known anomalies, update dates, responsible personnel, validation checks, and affected clinical areas. For connected devices, IT teams should also record network configuration changes, access-control changes, and cybersecurity status.

Before clinical deployment, updates should be reviewed for potential impact on device function, interoperability, user training, cybersecurity, and documentation. In some cases, updates should first be tested in a non-clinical environment. After deployment, users should be informed of any changes that affect operation, alarms, displays, workflows, or troubleshooting.

Patch management is particularly challenging because medical devices may be subject to regulatory constraints, manufacturer-specific update procedures, proprietary software, or compatibility limitations. Hospitals cannot always apply generic IT patching practices to medical devices. Instead, patches must be coordinated with manufacturers, biomedical engineering, IT security, clinical departments, and quality management.

6.7. Institutional Responsibilities

The digitalization of medical devices creates shared institutional responsibilities. Biomedical engineering units understand device function, clinical risk, maintenance, calibration, and user interaction. Information technology teams understand networks, servers, cybersecurity, authentication, data storage, and integration. Clinical users understand workflow, patient care needs, alarm response, and practical usability. Quality and risk management teams understand incident reporting, documentation, regulatory expectations, and patient safety systems.

Effective governance requires these groups to work together. During procurement, institutions should evaluate not only device performance and price but also software maintenance, cybersecurity documentation, data standards, interoperability, update policy, vendor remote-access practices, and training requirements. During installation, network configuration, access control, patient-data mapping, and integration testing should be documented. During use, incident reports, cybersecurity alerts, software updates, and user feedback should be monitored. During decommissioning, stored patient data should be securely removed, and network access should be disabled.

For AI-enabled and software-intensive devices, institutional governance should also include local performance monitoring. This may involve tracking user complaints, output discrepancies, false-positive or false-negative patterns, downtime, update effects, and changes in clinical workflow. Where AI outputs influence diagnosis or treatment, the role of the human clinician should remain explicit.

The same life-cycle logic applies in the digital era. Suitable infrastructure, trained users, appropriate materials, maintenance, calibration, and biomedical activities remain necessary; however, they must now be complemented by software life-cycle management, cybersecurity controls, data integrity, interoperability, and AI oversight.

REFERENCES

European Parliament and Council of the European Union. (2017). Regulation (EU) 2017/745 of the European Parliament and of the Council on medical devices. Official Journal of the European Union.

Food and Drug Administration. (2016). Applying human factors and usability engineering to medical devices: Guidance for industry and Food and Drug Administration staff. U.S. Food and Drug Administration.

Food and Drug Administration. (2025). Artificial intelligence in software as a medical device. (SaMD)─ Device Advice: Comprehensive Regulatory Assistance.

Food and Drug Administration. (2026). Cybersecurity in medical devices: Quality management system considerations and content of premarket submissions. U.S. Food and Drug Administration.

Association for the Advancement of Medical Instrumentation (Ed.). (2006). Medical Device Software: Software Life Cycle Processes. Association for the Advancement of Medical Instrumentation.

de Normalisation Electrotechnique, C. E. (2015). Medical devices-Part 1: Application of usability engineering to medical devices (IEC 62366-1: 2015).

International Medical Device Regulators Forum. (2013). Software as a Medical Device (SaMD): Key definitions. IMDRF/SaMD WG/N10.

International Organization for Standardization. (2019). ISO 14971:2019 Medical devices — Application of risk management to medical devices. ISO.

Kamu Hastaneleri Genel Müdürlüğü. (2021). Klinik Mühendislik Yönetim Birimi. T.C. Sağlık Bakanlığı.

Kiper, M. (2018). Dünyada ve Türkiye'de tıbbi cihaz sektörü ve strateji önerisi. Türkiye Teknoloji Geliştirme Vakfı.

MEGEP. (2008). Biyomedikal cihaz teknolojileri: Hata kodları ve modifikasyon. Milli Eğitim Bakanlığı.

Medicines and Healthcare products Regulatory Agency. (2021). Managing medical devices: Guidance for healthcare and social services organisations. MHRA.

Seçim, H., & Pekelman, T. (1990). Hastanelerde verimliliği yükseltici uygulamalar: Biyomedikal mühendislik hizmetleri. Eskişehir Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 135–151.

Selvi, Y. (2009). Sağlık kuruluşlarında tıbbi cihaz yönetimi. Yönetim, 99–118.

World Health Organization. (2011). Medical equipment maintenance programme overview. WHO Medical Device Technical Series.

World Health Organization. (2019). Decommissioning medical devices. WHO Medical Device Technical Series.

CHAPTER 13

Advanced Control Paradigms in Battery Management Systems (BMS)

Ahmet Çakanel[1]

1. Introduction to Next-Generation BMS Architectures

The accelerating electrification of mobility and the large-scale deployment of stationary storage have positioned the lithium-ion battery as a critical enabling technology of the contemporary energy landscape. A battery pack, however, is only as capable, safe, and durable as the embedded supervisory system that governs it. The battery management system (BMS) is the electronic and algorithmic layer that mediates between the electrochemical cell and the application, transforming raw voltage, current, and temperature measurements into actionable estimates of internal state and into protective control decisions. As packs grow in energy density and as duty cycles become more aggressive, the demands placed on the BMS have shifted from simple supervisory monitoring toward sophisticated, model-based estimation and constrained control. The purpose of this chapter is to survey these advanced control paradigms, with a deliberate emphasis on robust state estimation and on the enforcement of the safe operating area (SOA) rather than on the heuristic regulators that dominated earlier designs, providing engineering researchers with a structured roadmap of how advanced methodologies address the highly non-linear electrochemical dynamics, the slow drift of cell parameters, and the tight electrical–thermal coupling that characterise modern battery systems. The treatment is intentionally synthetic: it consolidates the established literature, frames the governing concepts, and closes with a reproducible numerical case study rather than proposing a novel algorithm.

At the most fundamental level a BMS performs three interlocking classes of function. The monitoring function acquires terminal voltage, load current, and cell or module temperature and, from these signals, infers latent quantities such as the state of charge (SoC) and the state of health (SoH). The protection function enforces hard limits on voltage, current, temperature, and charge throughput so that the cell is never driven outside its admissible envelope, guarding against over-charge, over-discharge, external short circuit, and thermal runaway. The optimization function, the most demanding and the most recent to mature, seeks to extract maximum usable energy and power while simultaneously prolonging cycle life, principally through cell balancing and through power limiting that

[1] Kırklareli University, Department of Electrical-Electronics Engineering
ORCID: 0000-0003-2988-325X

respects the instantaneous capability of every cell in the string (Xiong, Cao, Yu, He, & Sun, 2018; Wang et al., 2020). These functions are not independent: accurate optimization presupposes accurate monitoring, and effective protection presupposes a faithful internal model, so that the defining architectural trend of next-generation systems is the migration of intelligence from look-up tables and fixed thresholds toward online estimators that maintain a running, physically grounded picture of each cell.

Early battery controllers relied on what may be termed over-the-fence heuristics: SoC was inferred by ampere-hour integration with occasional voltage-based recalibration, balancing was triggered by simple voltage thresholds, and power limits were drawn from static look-up tables indexed by temperature and SoC. Such schemes are transparent and inexpensive but degrade rapidly in the presence of sensor bias, capacity fade, and dynamic loading, because they carry no internal model that can reconcile conflicting measurements. The model-based paradigm replaces these heuristics with a state-space description of the cell whose hidden states are reconstructed by an observer, allowing the controller to fuse noisy measurements with prior dynamic knowledge in a statistically principled manner (Plett, 2004a; Wang et al., 2020). In a modern embedded implementation this intelligence is organised as a layered software stack: a hardware-abstraction layer conditions and time-stamps the raw measurements; a modelling and estimation layer hosts the cell model together with the observers that produce SoC, SoH, state of power (SoP), and state of energy (SoE); a decision layer converts these estimates into balancing commands, power limits, and protection actions; and a communication layer exposes the result to the vehicle or plant controller. The remainder of this chapter follows this stack, treating in turn the modelling foundations, the estimation layer, the balancing and safe-boundary control layer, a numerical case study of an SoC observer, and the open research directions.

2. Modeling Foundations for Control-Oriented Battery Systems

Every model-based control architecture rests on a mathematical description of the cell that is simultaneously faithful enough to capture the dominant dynamics and parsimonious enough to run on an embedded processor in real time, and this tension between fidelity and tractability is the central theme of control-oriented modelling. Physics-based electrochemical models, of which the porous-electrode pseudo-two-dimensional formulation of Doyle, Fuller, and Newman (1993) is the canonical example, describe lithium transport, charge transfer, and diffusion through coupled partial differential equations; they are exceptionally accurate and physically interpretable, but their computational burden and the difficulty of identifying their many parameters render them impractical for most embedded estimators. Equivalent circuit models (ECMs) instead represent the cell as a network of lumped electrical elements — an open-circuit voltage source whose value depends on SoC, a series ohmic resistance, and one or more parallel resistor–capacitor (RC) branches that emulate the diffusion and charge-transfer relaxation. The comparative study of Hu, Li, and Peng (2012) showed that a model with one or two RC branches offers the most favourable balance between accuracy and complexity for control purposes, a conclusion corroborated experimentally for SoC estimation by He, Xiong, and Fan (2011); Table 1 summarises the resulting trade-off (Hu, Li, and Peng (2012) and Wang et al. (2020).

Table 1. Qualitative comparison of control-oriented battery model families.

Model family	Accuracy	Computational cost	Parameter ID effort
Electrochemical (P2D)	Very high	Very high	Very high
Single-RC (Thevenin)	Moderate	Low	Low
Two-RC (dual polarization)	High	Low–moderate	Moderate
Data-driven (ML)	High (in-domain)	Variable	Data-intensive

The two-RC, or dual-polarization, model adopted in the case study of Section 5 captures both a fast charge-transfer relaxation and a slower diffusion relaxation while remaining linear in its states for a fixed parameter set, with a terminal voltage expressed as

$$v_t(t) = OCV(z) - v_1 - v_2 - R_0\, i(t), \tag{1}$$

where z denotes SoC, $OCV(z)$ is the open-circuit-voltage characteristic, R_0 the ohmic resistance, and v_1 and v_2 the voltages across the two polarization branches with time constants $\tau_1 = R_1 C_1$ and $\tau_2 = R_2 C_2$. Cell parameters are, moreover, strongly temperature dependent, and the heat dissipated by ohmic and polarization losses in turn raises the cell temperature, producing a bidirectional coupling that a purely electrical model cannot represent. Lumped-parameter electro-thermal formulations such as the cylindrical-cell model of Lin et al. (2014) therefore augment the electrical state with one or more thermal nodes governed by an energy balance of the form

$$C_{th}\,(dT/dt) = i^2 R_0 + v_1^2/R_1 + v_2^2/R_2 - (T - T_{amb})/R_{th}, \tag{2}$$

where C_{th} is the lumped thermal capacitance, R_{th} the thermal resistance to ambient, and T_{amb} the ambient temperature; embedding such a node allows the estimator to schedule the electrical parameters against the live temperature estimate and is a prerequisite for credible SOA enforcement under demanding loads. A further practical consideration is that the battery exhibits dynamics spanning many orders of magnitude in time, from sub-second charge-transfer transients to the months-long progression of capacity fade, and capturing all of these in a single uniformly sampled model is both wasteful and numerically fragile. The prevailing remedy is to separate the dynamics by time scale — updating the fast electrical states at the control sampling rate, SoC at a slower rate consistent with coulomb counting, and SoH parameters on a much longer horizon through recursive identification — which keeps the per-step computational load bounded while still resolving the phenomena relevant to each state and underpins the joint state-and-parameter estimators discussed in the next section.

3. State Estimation **Algorithms** under Non-linear Dynamics

The estimation layer is the analytical heart of an advanced BMS. Because none of the quantities of interest — SoC, SoH, SoP, or SoE — can be measured directly, each must be reconstructed from the available electrical and thermal signals through an observer that inverts the cell model, and the non-linearity of the open-circuit-voltage characteristic together with the drift of the underlying parameters makes this a demanding non-linear, time-varying estimation problem. The extended Kalman filter (EKF) introduced to battery management by Plett (2004a, 2004b, 2004c) remains the reference method for online SoC estimation: it propagates a state estimate and its error covariance through the linearised cell dynamics and corrects them with each voltage measurement, optimally trading model prediction against sensor information. For the two-RC model the discrete prediction step advances the augmented state $x = [z, v_1, v_2]^T$ according to

$$x_{\{k+1\}} = A\, x_k + B\, i_k + w_k, \tag{3}$$

with the process noise w_k accounting for unmodelled dynamics, while the correction step applies the Kalman gain L_k derived from the linearised output Jacobian,

$$x_k^+ = x_k^- + L_k \left(v_{\{t,k\}} - h(x_k^-, i_k) \right) \tag{4}$$

The EKF is, however, sensitive to model mismatch and to imperfect knowledge of the noise statistics, and to harden the estimate against such uncertainty robust observers have been proposed: the sliding-mode observer of Kim (2006) provides finite-time convergence and intrinsic insensitivity to bounded disturbances, while the H-infinity and unscented-filter approach of Yu, Xiong, Lin, Shen, and Deng (2017) bounds the worst-case estimation error without requiring Gaussian noise assumptions. These robust formulations are particularly valuable when current-sensor bias or capacity fade would otherwise bias a conventional EKF.

State of health captures the slow, irreversible degradation of the cell, manifested as capacity fade and resistance growth; because these changes evolve over hundreds of cycles, SoH is naturally framed as a parameter-identification problem layered beneath the faster SoC estimator. Dual or joint estimators, in which a slow filter tracks capacity and resistance while a fast filter tracks SoC, are the standard architecture, the dual sliding-mode observer of Kim (2010) being an early and influential realisation, and the critical review of Xiong, Li, and Tian (2018) catalogues the model-based, data-driven, and hybrid families of SoH monitoring while emphasising that reliable degradation tracking is the prerequisite for warranty prediction and second-life valuation. Beyond charge and health, an application controller also needs to know how much power the pack can sustain over a given horizon, the SoP, and how much usable energy remains, the SoE. The model-based peak-power method of Sun, Xiong, He, Li, and Aussems (2012) computes the instantaneous power capability by projecting the cell model forward against the design limits on voltage, current, and SoC, thereby tying SoP directly to the SOA, whereas state of energy is best obtained by integrating the product of terminal voltage and current within a dedicated filter such as the Gaussian-model observer of He, Zhang, Xiong, and Wang (2015) rather than inferred naively from SoC. Because all four indicators share the same underlying states, they are most efficiently produced by a single joint observer that exposes them from one consistent model.

4. Active Cell Balancing and Safe Boundary Control

A pack is an electrical series–parallel assembly of many cells whose capacities and impedances are never perfectly matched, and without intervention these disparities widen with age until the weakest cell dictates both the usable capacity and the safety margin of the whole string; balancing control and safe-boundary enforcement are the two mechanisms by which the BMS counteracts this tendency. Passive balancing dissipates the excess charge of stronger cells through bleed resistors and is simple and cheap but wastes energy and generates heat, whereas active balancing redistributes charge from stronger to weaker cells using capacitive, inductive, or converter-based topologies, recovering energy that passive schemes would discard. The comprehensive survey of active equalization methods by Gallardo-Lozano, Romero-Cadaval, Milanes-Montero, and Guerrero-Martinez (2014) classifies these architectures by their energy-transfer path — cell-to-cell, cell-to-pack, and pack-to-cell — and compares their speed, efficiency, and component count. The hardware topology, however, only provides the means of charge transfer; a control law must still decide how much charge to move and when, and equalization laws range from simple feedback on the cell-voltage spread to optimization-based schemes that minimise the time or energy required to bring the string within tolerance subject to the current ratings of the balancing converters. When balancing is referenced to estimated SoC rather than to instantaneous voltage, the controller equalises the genuinely usable charge of the cells and avoids the misallocation that voltage-only schemes suffer under load.

Complementing balancing is the enforcement of the safe operating area, the multidimensional envelope in voltage, current, temperature, and SoC within which the cell may be operated without accelerated degradation or hazard. Enforcing it is fundamentally a constrained-control problem, and model predictive control (MPC) is the natural framework because it optimises future commands while explicitly respecting state and input constraints over a receding horizon. The optimal-charging formulation of Zou, Manzie, and Nešić (2018) demonstrates how MPC can drive a cell to a target SoC as quickly as possible while keeping terminal voltage, current, and internal over-potentials within their admissible bounds; coupled with the SoP estimate described in the previous section, such a constrained controller allows the BMS to extract maximum performance right up to, but never beyond, the safe boundary.

5. Case Study: Numerical Simulation of an SoC Tracking Observer

To make the preceding concepts concrete, this section reports a reproducible numerical experiment in which an extended Kalman filter reconstructs the SoC of a two-RC equivalent-circuit cell from noisy terminal-voltage measurements under a dynamic load; it is intended as an illustrative baseline rather than as a novel algorithm, and all parameters are representative of a commercial nickel–manganese–cobalt 18650 cell.

Table 2. Parameters of the two-RC equivalent-circuit cell model used in the case study.

Parameter	**Symbol**	**Value**
Nominal capacity	Q_n	2.60 A h
Coulombic efficiency	η	0.99
Ohmic resistance	R_0	25 mΩ
Polarization resistance 1	R_1	15 mΩ
Polarization capacitance 1	C_1	2000 F
Polarization resistance 2	R_2	30 mΩ
Polarization capacitance 2	C_2	30000 F
Time constant 1	τ_1	30 s
Time constant 2	τ_2	900 s
Sampling interval	Δt	1 s
Voltage sensor noise (std.)	σ_v	10 mV

The plant is the two-RC model of equation (1) with the parameters listed in Table 2, and a one-hour dynamic current profile combining a slowly varying sinusoidal component with superimposed charge and discharge pulses excites the cell across the SoC window from roughly 80% down to 27%, with the terminal voltage corrupted by zero-mean Gaussian sensor noise of 10 mV standard deviation to emulate a realistic acquisition channel.

The applied current profile and the resulting true and measured terminal voltages are shown in Figure 1; the pulsed load deliberately stresses the polarization dynamics so that the observer is evaluated under conditions more demanding than a constant-current discharge.

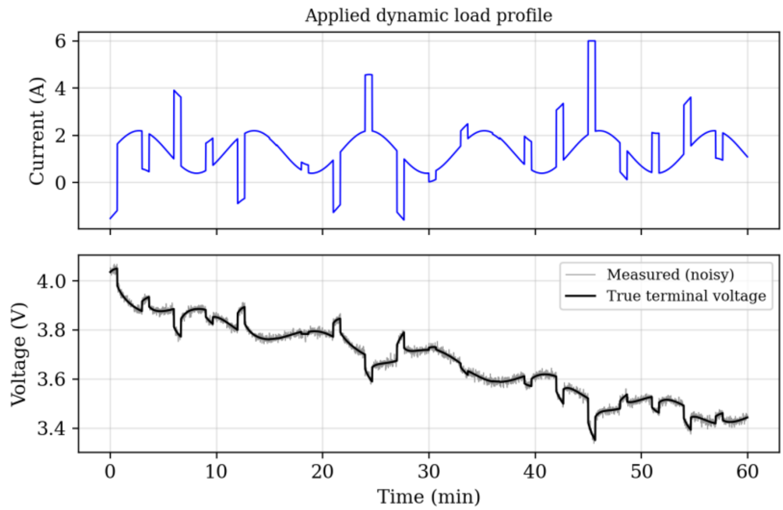

Figure 1. Applied dynamic load current (top) and the corresponding true and noisy measured terminal voltage (bottom).

The filter estimates the augmented state $x = [z, v_1, v_2]^T$ using the prediction and correction steps of equations (3) and (4), with the output Jacobian evaluated analytically from equation (1) and dominated by the local slope of the open-circuit-voltage curve, $dOCV/dz$, which governs the observability of SoC. To probe robustness the filter is initialised deliberately away from the true SoC and, in one scenario, fed a biased current measurement, yielding four cases: an accurately initialised baseline, a +15% initial SoC error, a −20% initial SoC error, and a combination of a large initial error with a +0.2 A current-sensor bias. Estimation quality is quantified by the root-mean-square error (RMSE), the mean absolute error (MAE), the peak absolute error, the steady-state RMSE after convergence, and the time for the error to settle permanently within a ±2% band.

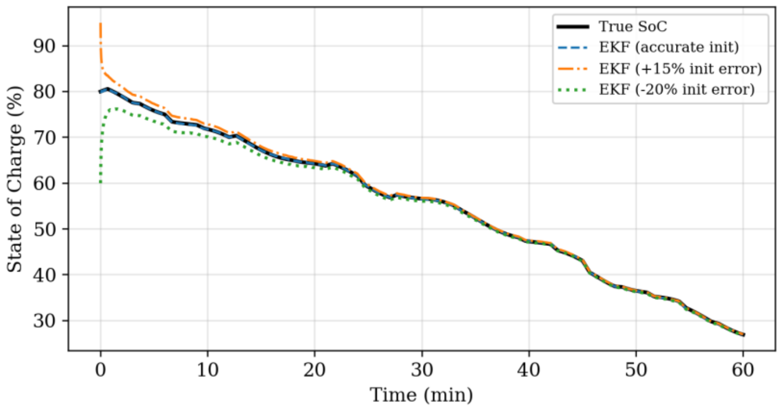

Figure 2. True SoC and EKF estimates for three initialization scenarios; the filter converges from large initial errors and then tracks the reference.

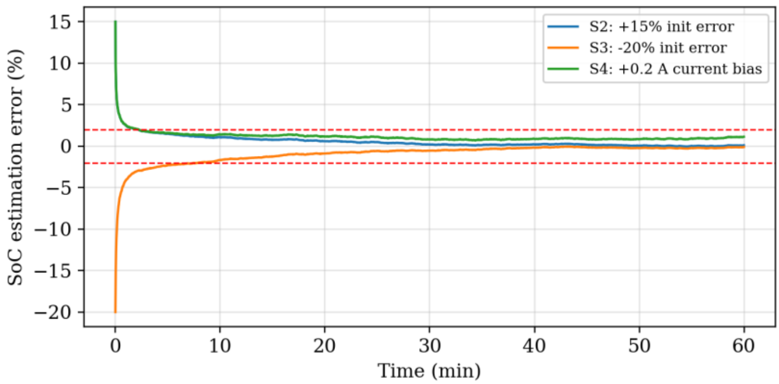

Figure 3. SoC estimation error for the perturbed scenarios relative to the ±2% tolerance band (dashed).

Figure 2 shows that, irrespective of a substantial initial error, the filter rapidly drives its estimate onto the true SoC trajectory and thereafter tracks it almost exactly, while Figure 3 plots the corresponding error signals against the ±2% tolerance band and isolates the residual offset that the current-sensor bias induces.

Table 3 collects the quantitative metrics for the four scenarios. With accurate initialisation the filter achieves a steady-state RMSE of only 0.07%, confirming an essentially unbiased implementation under nominal conditions; a +15% or −20% initial error is corrected within roughly two to eight minutes, after which the steady-state RMSE falls below 0.9%, demonstrating the self-correcting property that distinguishes a Kalman observer from open-loop coulomb counting. The current-sensor bias is the most adverse case, because a constant current error integrates directly into the coulomb-counting prediction and therefore leaves a persistent steady-state RMSE of about 1.05% — precisely the failure mode that the robust observers discussed earlier are designed to suppress.

Table 3. SoC estimation performance of the EKF across the four test scenarios.

Scenario	RMSE (%)	MAE (%)	Max |err| (%)	Steady-state RMSE (%)
Accurate initialization	0.07	0.06	0.18	0.07
+15% initial SoC error	0.93	0.56	15.00	0.56
−20% initial SoC error	1.51	0.91	20.00	0.88
+0.2 A current-sensor bias	1.27	1.14	15.00	1.05

The case study thus substantiates the central message of the chapter: a model-based observer delivers accurate, self-correcting SoC estimates under realistic noise and initialisation errors, yet its accuracy is ultimately limited by unmodelled biases, which motivates the robust and joint-estimation strategies surveyed above.

6. Future Directions and Open Challenges

The embedded BMS is increasingly complemented by a cloud tier that aggregates fleet data, performs computationally expensive ageing analytics, and returns refined parameters to the edge; this division of labour, in which the edge guarantees real-time safety while the cloud supplies long-horizon learning, promises more accurate degradation models than any single pack could identify in isolation, provided that communication latency and link reliability are handled within the safety case.

Data-driven estimators, meanwhile, have shown strong accuracy for SoC and SoH when trained on representative data, as the recent review of Zhao, Guo, and Chen (2024) documents, but their adoption in safety-critical control is gated by the difficulty of bounding their behaviour outside the training distribution. The most credible near-term path is hybridisation, in which a learned component supplies corrections or parameters to a physically grounded observer whose stability and constraint satisfaction remain provable; reconciling the flexibility of machine learning with the formal guarantees demanded by functional-safety standards is, in the view shared across the literature, the defining open challenge for the next generation of battery management systems.

References

Doyle, M., Fuller, T. F., & Newman, J. (1993). Modeling of galvanostatic charge and discharge of the lithium/polymer/insertion cell. *Journal of the Electrochemical Society, 140*(6), 1526–1533, DOI: https://doi.org/10.1149/1.2221597.

Gallardo-Lozano, J., Romero-Cadaval, E., Milanes-Montero, M. I., & Guerrero-Martinez, M. A. (2014). Battery equalization active methods. *Journal of Power Sources, 246*, 934–949, DOI: https://doi.org/10.1016/j.jpowsour.2013.08.026.

He, H., Xiong, R., & Fan, J. (2011). Evaluation of lithium-ion battery equivalent circuit models for state of charge estimation by an experimental approach. *Energies, 4*(4), 582–598, DOI: https://doi.org/10.3390/en4040582.

He, H., Zhang, Y., Xiong, R., & Wang, C. (2015). A novel Gaussian model based battery state estimation approach: State-of-energy. *Applied Energy, 151*, 41–48, DOI: https://doi.org/10.1016/j.apenergy.2015.04.062.

Hu, X., Li, S., & Peng, H. (2012). A comparative study of equivalent circuit models for Li-ion batteries. *Journal of Power Sources, 198*, 359–367, DOI: https://doi.org/10.1016/j.jpowsour.2011.10.013.

Kim, I.-S. (2006). The novel state of charge estimation method for lithium battery using sliding mode observer. *Journal of Power Sources, 163*(1), 584–590, DOI: https://doi.org/10.1016/j.jpowsour.2006.09.006.

Kim, I.-S. (2010). A technique for estimating the state of health of lithium batteries through a dual-sliding-mode observer. *IEEE Transactions on Power Electronics, 25*(4), 1013–1022, DOI: https://doi.org/10.1109/tpel.2009.2034966.

Lin, X., Perez, H. E., Mohan, S., Siegel, J. B., Stefanopoulou, A. G., Ding, Y., & Castanier, M. P. (2014). A lumped-parameter electro-thermal model for cylindrical batteries. *Journal of Power Sources, 257*, 1–11, DOI: https://doi.org/10.1016/j.jpowsour.2014.01.097.

Plett, G. L. (2004). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 1. Background. *Journal of Power Sources, 134*(2), 252–261, DOI: https://doi.org/10.1016/s0378-7753(04)00359-3.

Plett, G. L. (2004). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 2. Modeling and identification. *Journal of Power Sources, 134*(2), 262–276, DOI: https://doi.org/10.1016/j.jpowsour.2004.02.032.

Plett, G. L. (2004). Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation.

Journal of Power Sources, 134(2), 277–292, DOI: https://doi.org/10.1016/j.jpowsour.2004.02.033.

Sun, F., Xiong, R., He, H., Li, W., & Aussems, J. E. E. (2012). Model-based dynamic multi-parameter method for peak power estimation of lithium-ion batteries. *Applied Energy, 96*, 378–386, DOI: https://doi.org/10.1016/j.apenergy.2012.02.061.

Wang, Y., Tian, J., Sun, Z., Wang, L., Xu, R., Li, M., & Chen, Z. (2020). A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems. *Renewable and Sustainable Energy Reviews, 131*, 110015, DOI: https://doi.org/10.1016/j.rser.2020.110015.

Xiong, R., Cao, J., Yu, Q., He, H., & Sun, F. (2018). Critical review on the battery state of charge estimation methods for electric vehicles. *IEEE Access, 6*, 1832–1843, DOI: https://doi.org/10.1109/access.2017.2780258.

Xiong, R., Li, L., & Tian, J. (2018). Towards a smarter battery management system: A critical review on battery state of health monitoring methods. *Journal of Power Sources, 405*, 18–29, DOI: https://doi.org/10.1016/j.jpowsour.2018.10.019.

Yu, Q., Xiong, R., Lin, C., Shen, W., & Deng, J. (2017). Lithium-ion battery parameters and state-of-charge joint estimation based on H-infinity and unscented Kalman filters. *IEEE Transactions on Vehicular Technology, 66*(10), 8693–8701, DOI: https://doi.org/10.1109/tvt.2017.2709326.

Zhao, F., Guo, Y., & Chen, B. (2024). A review of lithium-ion battery state of charge estimation methods based on machine learning. *World Electric Vehicle Journal, 15*(4), 131, DOI: https://doi.org/10.3390/wevj15040131.

Zou, C., Manzie, C., & Nešić, D. (2018). Model predictive control for lithium-ion battery optimal charging. *IEEE/ASME Transactions on Mechatronics, 23*(2), 947–957, DOI: https://doi.org/10.1109/tmech.2018.2798930.

CHAPTER 14

Control and Mitigation Strategies for Extreme Fast Charging (XFC) Impacts on Smart Grids

Ahmet Çakanel[1]

1. The Extreme Fast Charging Paradigm and Grid Vulnerabilities

The viability of electric mobility as a substitute for the internal-combustion fleet depends, more than on any other single factor, on the time required to replenish a vehicle's battery. Extreme fast charging (XFC) addresses this constraint directly by delivering power at levels that approach the refuelling experience of a conventional vehicle, with individual ports rated from above 150 kW toward 350 kW and, in the emerging megawatt-charging systems for heavy-duty transport, beyond one megawatt (Tu, Feng, Srdic, & Lukic, 2019; Burnham et al., 2017). A multi-port XFC station therefore presents the distribution network with an aggregate load of several megawatts that switches on and off over seconds as vehicles arrive and depart. This chapter surveys the control and mitigation strategies that allow such stations to be integrated into the grid without degrading power quality, with a deliberate emphasis on active mitigation at the point of common coupling (PCC) by means of hybrid energy storage and localised control coordination, rather than on the charging process internal to the vehicle. As with the preceding chapter, the treatment consolidates the established literature, frames the governing concepts, and closes with a reproducible numerical case study.

The defining electrical characteristic of an XFC load is its combination of very high magnitude with strong temporal volatility. Unlike a slowly varying industrial load, the charging current rises to its rated value within seconds and is essentially stochastic at the timescale of station operation, because it is driven by uncoordinated vehicle arrivals. This behaviour stresses the distribution network in three principal ways. First, the large active-power draw produces a voltage drop across the feeder impedance that depresses the PCC voltage, and the rapid switching of that draw produces voltage fluctuations and flicker. Second, the power-electronic front ends of the chargers, if not properly controlled, inject harmonic currents that distort the supply waveform and can excite resonances elsewhere in the network. Third, the repeated high-magnitude loading accelerates the thermal ageing of the upstream distribution transformer and increases network

[1] Kırklareli University, Department of Electrical-Electronics Engineering
ORCID: 0000-0003-2988-325X

losses (Alame, Azzouz, & Kar, 2020; Tu et al., 2019). Left unmanaged, these effects can push the PCC voltage outside its statutory band and shorten the life of network assets, which is precisely why grid-side power management has become a prerequisite for high-power charging deployment. The operational objective of the strategies surveyed below is therefore to present the network with a load that is flatter, cleaner, and more predictable than the raw charging demand, and the remainder of the chapter develops this objective through the converter interface, the buffer-storage layer, the aggregation layer, a numerical case study, and the relevant standards.

2. Grid-Tied Fast Charger Converter Topologies and Control Interfaces

The interface between an XFC station and the alternating-current grid is a power-electronic converter whose topology and control determine the quality of the resulting interaction. Early and low-cost chargers used uncontrolled diode rectifiers, which draw distorted, non-sinusoidal current at a poor power factor; the modern standard is the active front-end (AFE) rectifier, a controlled voltage-source converter that performs active power-factor correction (PFC) by regulating its input current to be sinusoidal and very nearly in phase with the supply voltage. The voltage-oriented control of a Vienna-type AFE rectifier, as detailed by Rajendran, Vaithilingam, Misron, Naidu, and Ahmed (2021), achieves a displacement power factor close to unity and low current distortion, so that the charger draws little reactive power and injects only modest harmonics even at full load. The AFE thus mitigates two of the three vulnerabilities identified above at their source, before any external compensation is considered.

Beyond the regulation of its own input current, the converter must adopt a control paradigm that defines its relationship to the grid. In grid-following operation the converter behaves as a controlled current source that synchronises to, and injects current relative to, an externally established voltage; in grid-forming operation it behaves as a controlled voltage source that actively regulates the voltage and frequency at its terminals. The comparative analysis of converter control in alternating-current microgrids by Rocabert, Luna, Blaabjerg, and Rodríguez (2012) sets out the trade-offs between these paradigms, which are summarised in Table 1; grid-following control is simpler and dominant in grid-connected charging, whereas grid-forming control becomes essential when a station must support a weak network or operate within an islanded microgrid. Both paradigms depend on accurate knowledge of the grid-voltage angle, which is obtained by a phase-locked loop (PLL). Under the distorted and frequently unbalanced conditions that prevail at a heavily loaded PCC, a conventional synchronous-reference-frame PLL produces an oscillatory angle estimate; the decoupled double synchronous reference frame PLL of Rodríguez et al. (2007) overcomes this by separating the positive- and negative-sequence components, yielding a clean synchronisation signal that is the foundation of stable converter operation.

Table 1. Grid-following versus grid-forming control paradigms for fast-charging converters.

Attribute	Grid-following	Grid-forming
Converter behaviour	Controlled current source	Controlled voltage source
Voltage / frequency	Tracks the grid	Sets the local reference
Grid strength needed	Strong grid	Weak grid / islanded
Synchronization	PLL-dependent	Self-synchronizing
Typical use	Grid-connected charging	Microgrid / backup support

3. Coordinated Control of Buffer Storage Integration

Even a perfectly controlled AFE cannot alter the fact that the active power delivered to the vehicles must come from somewhere, and if it is drawn entirely from the grid the feeder still experiences the full magnitude and volatility of the charging demand. The decisive mitigation is therefore to interpose a local energy buffer between the grid and the chargers, so that the buffer rather than the grid absorbs the peaks and transients. Because the charging load contains both a slowly varying bulk component and sharp high-frequency transients, the most effective buffer is a hybrid energy storage system (HESS) that combines a battery, which offers high energy capacity but limited cycle life under fast transients, with a supercapacitor, which offers very high power and cycle life but little energy. The architecture and energy management of such battery–supercapacitor systems are reviewed by Jing, Lai, Wong, and Wong (2017), who show that the pairing exploits the complementary strengths of the two devices when the power demanded of the HESS is partitioned appropriately between them.

That partitioning is naturally expressed in the frequency domain. Writing the instantaneous station demand as the sum of the individual port powers,

$$P_{dem}(t) = \sum_i P_i(t), \qquad (1)$$

a peak-shaving controller defines a smooth grid reference and assigns the remainder to the HESS, so that the power the storage must supply is

$$P_{hess}(t) = P_{dem}(t) - P_{grid}(t). \tag{2}$$

A first-order filter then separates this HESS power by frequency: the low-frequency content, carrying the bulk energy, is directed to the battery, while the complementary high-frequency content, carrying the fast transients, is directed to the supercapacitor,

$$P_{batt}(t) = \frac{P_{hess}(t)}{1+\tau s}, \quad P_{sc}(t) = P_{hess}(t) - P_{batt}(t), \tag{3}$$

where τ is the separation time constant and s the Laplace variable. This frequency-decoupling principle, elaborated for hybrid storage by Hartani et al. (2023), keeps the battery on a smooth, low-stress duty while the supercapacitor shields it from the damaging transients; the predictive variant of Golchoubian and Azad (2017) shows that a model predictive controller can perform the same allocation while explicitly honouring the power and state-of-charge limits of each device. Acting through its AFE, the same converter additionally regulates the local reactive power so as to support the PCC voltage, combining real-power buffering with reactive compensation in a single coordinated control loop.

4. Demand Response and Fleet Aggregation Control Schemes

Buffer storage mitigates the impact of a single station, but when many vehicles and many stations are considered together a further degree of freedom appears: the charging itself can be scheduled. Coordination of charging across a fleet can be organised centrally, where an aggregator solves a single optimisation that allocates power to every vehicle, or in a decentralised manner, where each vehicle adjusts its own charging in response to a broadcast price or signal. The decentralised protocol of Gan, Topcu, and Low (2013) demonstrates that, with an appropriately designed iterative price signal, a population of self-interested chargers can be driven to a valley-filling optimum that is provably efficient yet requires no central control of individual vehicles, an attractive property given the scale and privacy constraints of real fleets. The complementary work of Shao, Pipattanasomporn, and Rahman (2012) shows how demand-response programmes built on real-time pricing and customer choice can reshape the aggregate charging profile to shave peaks and follow the available network capacity.

A more active form of coordination allows energy to flow back from the vehicles. Vehicle-to-grid operation, reviewed comprehensively by Tan, Ramachandaramurthy, and Yong (2016), treats the parked fleet as a distributed storage resource that can provide peak shaving, frequency regulation, and reactive support, while the related vehicle-to-vehicle concept allows energy to be exchanged directly between cars. These aggregation schemes operate on the timescale of minutes to hours and are therefore complementary to the second-by-second buffering provided by the HESS; in a fully realised station the two layers cooperate, with the storage absorbing fast transients and the aggregation logic shaping the slow envelope. Where a permanent buffer is installed, its battery and converter must be sized to the expected peak energy and power, a design problem addressed by Hussain, Bui, Baek, and Kim (2019), who optimise the simultaneous sizing of the battery and its converter for a fast-charging station so that the installed cost is balanced against the achievable reduction in grid stress.

5. Case Study: Peak-Shaving Performance of a Buffer-Assisted Charging Station

To quantify the mitigation that a HESS provides, this section reports a reproducible numerical experiment in which a multi-port XFC station with a battery–supercapacitor buffer serves a stochastic stream of charging sessions over a thirty-minute window. The experiment is illustrative rather than a novel control design, and its parameters, listed in Table 2, are representative of a station-scale installation. Fourteen charging sessions, each drawing between 150 kW and 350 kW for a randomised duration with a few-second ramp, are superimposed to form the aggregate demand of equation (1), which reaches an instantaneous peak of about 1.83 MW. The mitigation controller holds the grid import at a contractual cap and assigns the excess to the HESS according to equations (2) and (3), recharging the battery from the headroom available whenever the demand falls below the cap.

Table 2. Parameters of the charging station and hybrid energy storage system used in the case study.

Parameter	Symbol	Value
Contractual grid-import cap	P_cap	1150 kW
Battery power rating	P_batt,max	700 kW
Battery energy capacity	E_batt	250 kWh
Supercapacitor power rating	P_sc,max	500 kW
Supercapacitor energy capacity	E_sc	4 kWh
Frequency-separation time constant	tau	6 s
Feeder resistance / reactance (pu)	R, X	0.030, 0.045
Power factor: diode FE / AFE	pf	0.95 / 0.995
Sampling interval	Δt	0.5 s

Figure 1 shows the central result. The grey area is the raw station demand, whose peak reaches 1.83 MW, while the blue trace is the grid import after mitigation, which the controller holds at the 1150 kW cap throughout the peak and lifts slightly above demand in the valleys to recharge the battery. The peak grid import is thereby reduced by 37%, and the peak-to-average ratio of the grid draw falls from 2.67 to 1.74, presenting the network with a markedly flatter load.

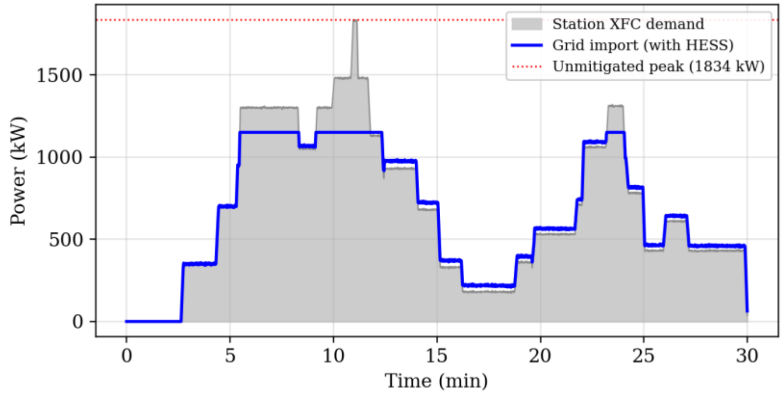

Figure 1. Station XFC demand and the resulting grid import with HESS peak shaving; the grid is held at the 1150 kW contractual cap during peaks.

Figure 2 illustrates how the HESS power is partitioned by the frequency-separation law of equation (3). The battery (green) carries the smooth bulk of the buffered power while the supercapacitor (orange) absorbs the sharp transients at each session ramp, and the lower panel confirms that the battery state of charge varies gently between 62% and 70% whereas the supercapacitor swings rapidly over a wider band, exactly the duty split for which each device is suited. Because the contractual cap is chosen to be energetically sustainable, the battery is never depleted and the grid import is held at the cap without interruption.

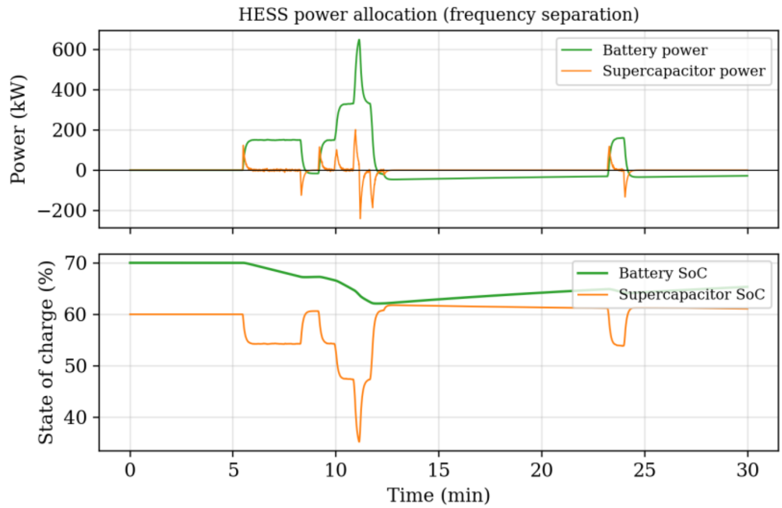

Figure 2. HESS power allocation by frequency separation (top) and the corresponding battery and supercapacitor states of charge (bottom).

The power-quality benefit at the PCC is shown in Figure 3, where the terminal voltage is estimated from the feeder impedance and the drawn complex power. Without mitigation, the combination of the full demand and the lagging power factor of a diode front end drives the PCC voltage well below the 0.95 per-unit statutory limit, reaching a deviation of 8.2% at the demand peak. With the HESS reducing the active-power draw and the AFE raising the displacement power factor to 0.995, the worst-case deviation is roughly halved to 4.0% and the voltage is kept within the admissible band throughout. Table 3 collects these metrics.

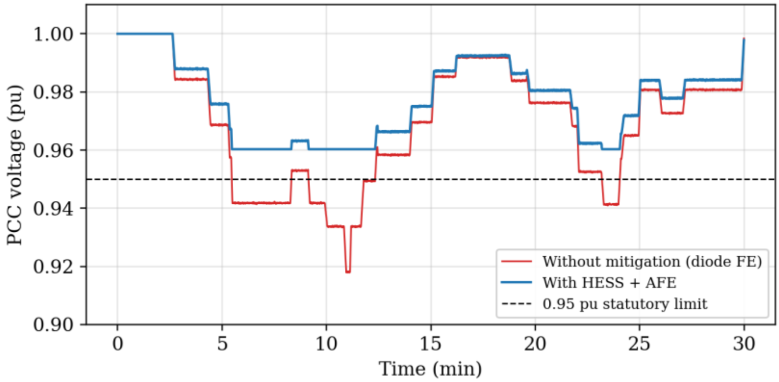

Figure 3. Point-of-common-coupling voltage without mitigation (diode front end) and with the HESS and active front end, relative to the 0.95 pu limit.

Table 3. Grid-side performance of the charging station with and without HESS mitigation.

Metric	**Without mitigation**	**With HESS + AFE**
Peak grid import	1834 kW	1150 kW
Peak power reduction	—	37.3 %
Peak-to-average ratio (grid)	2.67	1.74
Max PCC voltage deviation	8.22 %	3.97 %
Displacement power factor	0.950	0.995

Taken together, the figures and Table 3 substantiate the central argument of the chapter: a comparatively modest local buffer, governed by a simple frequency-separation peak-shaving law and paired with an active front end, transforms a volatile multi-megawatt charging load into a flat, high-power-factor demand that respects the voltage limits of the distribution network.

6. Standardization, Policy, and Future Trends

The control strategies surveyed here do not operate in a regulatory vacuum: a grid-connected charging station is a distributed resource whose interaction with the network is governed by interconnection standards. IEEE Std 1547-2018 (IEEE, 2018) specifies the voltage- and frequency-ride-through, power-quality, and reactive-capability requirements that such a resource must satisfy, and in doing so effectively mandates the active front end and the reactive-support capability discussed in Sections 2 and 3, while IEC 61851-1:2017 (IEC, 2017) defines the general requirements of the conductive charging system itself. As charging power climbs toward the megawatt range for heavy-duty applications, these frameworks are being extended to cover the new dedicated megawatt-charging architectures, and compliance with them is the practical boundary condition on any mitigation design.

Looking ahead, the stochastic character of XFC demand that motivated the buffering strategy of this chapter also invites a predictive response. Data-driven forecasting of charging load, exemplified by the deep-learning queuing model of Zhang et al. (2021), allows a station controller to anticipate the arrival of peaks and to pre-position the energy stored in its buffer and in the connected fleet, converting the reactive peak-shaving of the case study into a proactive scheme. The broader review of smart-charging solutions by Sadeghian, Oshnoei, Mohammadi-Ivatloo, Vahidinasab, and Anvari-Moghaddam (2022) frames this trajectory, in which forecasting, stochastic optimisation, and coordinated storage are integrated into a single management layer. Reconciling the resulting algorithmic sophistication with the firm power-quality and stability guarantees demanded by the interconnection standards is, as in the management of the batteries themselves, the defining challenge for the next generation of high-power charging infrastructure.

References

Alame, D., Azzouz, M., & Kar, N. (2020). Assessing and mitigating impacts of electric vehicle harmonic currents on distribution systems. *Energies, 13*(12), 3257, DOI: https://doi.org/10.3390/en13123257.

Burnham, A., Dufek, E. J., Stephens, T., Francfort, J., Michelbacher, C., Carlson, R. B., … Tanim, T. R. (2017). Enabling fast charging – Infrastructure and economic considerations. *Journal of Power Sources, 367*, 237–249, DOI: https://doi.org/10.1016/j.jpowsour.2017.06.079.

Gan, L., Topcu, U., & Low, S. H. (2013). Optimal decentralized protocol for electric vehicle charging. *IEEE Transactions on Power Systems, 28*(2), 940–951, DOI: https://doi.org/10.1109/TPWRS.2012.2210288.

Golchoubian, P., & Azad, N. L. (2017). Real-time nonlinear model predictive control of a battery-supercapacitor hybrid energy storage system in electric vehicles. *IEEE Transactions on Vehicular Technology, 66*(11), 9678–9688, DOI: https://doi.org/10.1109/TVT.2017.2725307.

Hartani, M. A., Benhammou, A., Rezk, H., Hamouda, M., Abdelkhalek, O., Mekhilef, S., & Abdelkareem, M. A. (2023). Enhancing hybrid energy storage systems with advanced low-pass filtration and frequency decoupling for optimal power allocation and reliability of cluster of DC-microgrids. *Energy, 282*, 128310, DOI: https://doi.org/10.1016/j.energy.2023.128310.

Hussain, A., Bui, V.-H., Baek, J.-W., & Kim, H.-M. (2019). Stationary energy storage system for fast EV charging stations: Simultaneous sizing of battery and converter. *Energies, 12*(23), 4516, DOI: https://doi.org/10.3390/en12234516.

International Electrotechnical Commission. (2017). *IEC 61851-1:2017: Electric vehicle conductive charging system – Part 1: General requirements* (Ed. 3.0). Geneva, Switzerland: IEC.

Institute of Electrical and Electronics Engineers. (2018). *IEEE Std 1547-2018: IEEE standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces*. IEEE. DOI: https://doi.org/10.1109/IEEESTD.2018.8332112.

Jing, W., Lai, C. H., Wong, S. H. W., & Wong, M. L. D. (2017). Battery-supercapacitor hybrid energy storage system in standalone DC microgrids: A review. *IET Renewable Power Generation, 11*(4), 461–469, DOI: https://doi.org/10.1049/iet-rpg.2016.0500.

Rajendran, G., Vaithilingam, C. A., Misron, N., Naidu, K., & Ahmed, M. R. (2021). Voltage oriented controller based Vienna rectifier for electric vehicle charging stations. *IEEE Access, 9*, 50798–50809, DOI: https://doi.org/10.1109/ACCESS.2021.3068653.

Rocabert, J., Luna, A., Blaabjerg, F., & Rodríguez, P. (2012). Control of power converters in AC microgrids. *IEEE Transactions on Power Electronics, 27*(11), 4734–4749, DOI: https://doi.org/10.1109/TPEL.2012.2199334.

Rodríguez, P., Pou, J., Bergas, J., Candela, J. I., Burgos, R. P., & Boroyevich, D. (2007). Decoupled double synchronous reference frame PLL for power converters control. *IEEE Transactions on Power Electronics, 22*(2), 584–592, DOI: https://doi.org/10.1109/TPEL.2006.890000.

Sadeghian, O., Oshnoei, A., Mohammadi-Ivatloo, B., Vahidinasab, V., & Anvari-Moghaddam, A. (2022). A comprehensive review on electric vehicles smart charging: Solutions, strategies, technologies, and challenges. *Journal of Energy Storage, 54*, 105241, DOI: https://doi.org/10.1016/j.est.2022.105241.

Shao, S., Pipattanasomporn, M., & Rahman, S. (2012). Grid integration of electric vehicles and demand response with customer choice. *IEEE Transactions on Smart Grid, 3*(1), 543–550, DOI: https://doi.org/10.1109/TSG.2011.2164949.

Tan, K. M., Ramachandaramurthy, V. K., & Yong, J. Y. (2016). Integration of electric vehicles in smart grid: A review on vehicle to grid technologies and optimization techniques. *Renewable and Sustainable Energy Reviews, 53*, 720–732, DOI: https://doi.org/10.1016/j.rser.2015.09.012.

Tu, H., Feng, H., Srdic, S., & Lukic, S. (2019). Extreme fast charging of electric vehicles: A technology overview. *IEEE Transactions on Transportation Electrification, 5*(4), 861–878, DOI: https://doi.org/10.1109/TTE.2019.2958709.

Zhang, X., Chan, K. W., Li, H., Wang, H., Qiu, J., & Wang, G. (2021). Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model. *IEEE Transactions on Cybernetics, 51*(6), 3157–3170, DOI: https://doi.org/10.1109/TCYB.2020.2975134.

CHAPTER 15

Evaluating Robustness Against Missing Modes in AG News News Classification Using Multi-Mode Fusion and Mode Dropout Techniques

Halil Çınar [1] ***& Zülfikar Aslan*** [2]

1. Introduction

The combined use of different modalities—such as text, images, and video—in digital news feeds necessitates that classification systems not be limited to text-based approaches alone (Baltrušaitis et al., 2019). In this study, we investigate the contribution of a pseudo-visual representation derived from the CLIP text encoder on the text-based news classification task AG News (Zhang et al., 2015). Here, the term "pseudo-visual" refers to the fact that CLIP embeddings derived from text serve as a representation compatible with the visual modality due to the CLIP model's contrastive training logic, which aligns text and image spaces in a shared embedding space (Radford et al., 2021).

In multimodal systems, the stage at which modalities are fused directly affects performance. In this study, three main strategies—Early Fusion, Late Fusion, and Cross-Attention—were compared. Additionally, the Modality Dropout technique was applied to account for real-world scenarios where one modality might be missing (Neverova et al., 2016). This technique prevents the network from becoming overly dependent on a single modality by setting a modality to zero with a small probability during training.

The contributions of this study can be summarized as follows:

- A systematic comparison was conducted between four different pre-trained language models on AG News and the pseudo-visual representation derived from the CLIP text encoder.

[1] Gaziantep Islam University of Science and Technology, Master's Student in Computer Engineering, 0009-0005-2429-2612
[2] Assoc. Prof. Dr., Gaziantep Islam Science and Technology University, Department of Computer Engineering, 0000-0002-2706-5715

- Thirty-two experimental configurations were reported by applying three fusion strategies and Modality Dropout; accuracy, macro F1, Cohen's Kappa, MCC, balanced accuracy, and Jaccard metrics were presented together.
- Modality Dropout was not tested in a continuous learning or sequential task setting in this study; it was used solely to measure robustness to missing modality.

2. Related Works

2.1. Multimodal Fusion Strategies

In multimodal learning, the fusion stage is a design decision that directly affects the system's generalization ability. The early fusion approach aims to learn a common representation by directly combining the feature vectors of different modalities; late fusion, on the other hand, allocates a separate processing pipeline to each modality and combines the results in the final stage (Baltrušaitis et al., 2019). Kiela et al. (2019) reported improvements in social media content classification by combining text and visual features via early fusion in a BERT-based architecture. Kim and Choi (2025) demonstrated the contribution of cross-modal attention compared to fusion alone by proposing the CroMe model, which processes text, visual, and correlation information through three separate Transformer branches.

2.2. Cross-Modal Attention Mechanisms

Cross-modal attention enables the dynamic learning of intra-modal and inter-modal relationships by using the features of one modality as the Query and the features of another modality as the Key and Value (Vaswani et al., 2017; Lu et al., 2019). Shang et al. (2021) aimed to detect health misinformation in short videos using an audio-visual co-attention mechanism within the TikTec framework. The use of such mechanisms in news classification models the interaction between different parts of the text and different dimensions of the visual representation.

2.3. Modality Dropout and Robustness to Missing Data

The ModDrop technique proposed by Neverova et al. (2016) prevents the network from becoming overly dependent on the dominant modality by setting the probability of a modality to zero during training. Hu et al. (2023) employed similar strategies in image classification frameworks resilient to missing modalities and reported a reduction in performance loss under modality drop conditions during testing.

2.4. The Relationship with Continuous Learning

In the literature on continuous learning, the concept of Catastrophic Forgetting is defined as the model losing information acquired in previous tasks during subsequent tasks (Kirkpatrick et al., 2017). The Modality Switching experiment in this study was conducted on a single task and does not involve sequential tasks, forward/backward transfer measures, or a replay mechanism. Therefore, the findings are presented not as a claim of continuous learning, but as an analysis of resilience against missing modalities.

3. Method

3.1. Dataset

The AG News Topic Classification dataset used in this study was introduced by Zhang et al. (2015) and consists of news headlines and short descriptions across four balanced classes (World, Sports, Business, Sci/Tech). The standard split consists of 120,000 training and 7,600 test examples, with each class containing 30,000 examples in the training set and 1,900 in the test set (Table 1). The dataset is openly available via the Hugging Face datasets library (datasets.load_dataset("ag_news")), Kaggle, and TensorFlow Datasets.

To ensure the reproducibility of the experiments, the random seed (random_state / seed) was set to 42 during data loading and shuffling. For each example, the "Title" and "Description" fields were concatenated using the " - " separator to form a single text string, and class labels were converted to integers starting from zero.

Table 1. Sample distribution of the AG News dataset by class and partition

Class	Training Count	Test Count	Total
World	30.000	1.900	31.900
Sports	30.000	1.900	31.900
Business	30.000	1.900	31.900
Sci/Tech	30.000	1.900	31.900
Total	**120.000**	**7.600**	**127.600**

3.2. Textual Modality

Four different pre-trained Transformer-based language models were used as text encoders: RoBERTa-base (Liu et al., 2019), BERT-base-uncased (Devlin et al., 2019), DistilBERT-base-uncased (Sanh et al., 2019), and ALBERT-base-v2 (Lan et al., 2020). For each model, the model's own tokenizer was used, truncation and padding were applied to a maximum length of 128 tokens, and the [CLS] token ([:,0,:]) of the final layer was selected as the sentence-level representation. This representation is a 768-dimensional vector.

3.3. Visual Modality: Embedding of Synthetic Images Derived from the CLIP Text Encoder

The AG News dataset does not contain actual images for every news article. Therefore, in this study, images were not generated using a text-to-image model or stock image sources. Instead, we leveraged the ability of the CLIP (Radford et al., 2021) model to align text and image spaces within a common contrastive embedding space. The CLIP ViT-B/32 text encoder (openai/clip-vit-base-patch32) was applied to each news text, and the resulting 512-dimensional pooler_output vectors were used as "pseudo-visual features." This representation indicates a position aligned with the visual space due to CLIP being trained on 400 million image-text pairs; its use here serves as a proxy for the feature that a real ViT image encoder would generate when taking an image as input. Therefore, this representation is referred to as "pseudo-visual" throughout the remainder of the paper.

Algorithm 1 summarizes the workflow for extracting pseudo-visual features. The textual input is first truncated to a maximum of 77 tokens using the CLIP tokenizer; it is then passed through a frozen (without gradient calculation) CLIPTextModel to obtain the pooler_output vector. This process is applied to all training and test examples, and the output tensors are stored on disk.

Algorithm 1. Extraction of pseudo-visual features

```
input: text list T = {t_1, ..., t_N}

1. tokenizer ← CLIPTokenizer("openai/clip-vit-base-patch32")

2. model ← CLIPTextModel("openai/clip-vit-base-patch32");
model.eval()

3. F ← [ ]

4. for i in range(0, N, 64):

       batch ← T[i : i+64]

       enc ← tokenizer(batch, padding=True, truncation=True,
max_length=77)
```

```
        with torch.no_grad(): out ← model(**enc)
        F.append(out.pooler_output.cpu())
 5. return torch.cat(F, dim=0)     # şekil: (N, 512)
```

3.4. Fusion Strategies

The 768-dimensional representation obtained from the text encoder and the 512-dimensional pseudo-visual representation obtained from the CLIP text encoder were mapped to a shared 256-dimensional latent space using a separate two-layer MLP (Linear-ReLU-Dropout) for each modality. Three fusion strategies were compared:

• Early Fusion: The 256-dimensional text and pseudo-visual representations were concatenated to obtain a 512-dimensional combined representation, which was then fed into a two-layer classifier.

• Late Fusion: Class logits were generated using a separate linear head for each modality, and a softmax was applied to the sum of these logits.

• Cross-Attention: The text representation was used as the query, while the pseudo-visual representation served as the key and value. A four-head multi-head attention (nn.MultiheadAttention, num_heads = 4) was applied, followed by LayerNorm. The resulting output was then combined with the original text representation and fed into the classifier.

Additionally, text-only and vision-only baselines were trained using the same classifier architecture.

3.5. Modality Release and Resilience to Incomplete Modality

The Modality Dropout technique was implemented during training (only in the model.train() mode) by sampling a random value r from the interval [0,1). If $r < p$, the text representation is set to zero; if $p \leq r < 2p$, the pseudo-visual representation is set to zero. In this study, $p = 0.10$ was selected; that is, the text modality is disabled in approximately 10% of the training examples, and the pseudo-visual modality is disabled in approximately 10%. Modality dropout was disabled during evaluation. Since this design does not include sequential task definition, forward/backward transfer measurement, or experience replay, it is not presented as a continuous learning experiment; rather, it is reported as an analysis of robustness to missing modality.

3.6. Training Protocol, Hardware, and Library Versions

The experiments were conducted in a Google Colab environment using an NVIDIA Tesla T4 GPU. The application was implemented using the PyTorch and Hugging Face transformers libraries. Metrics were calculated using scikit-learn. Optimization was performed using the AdamW algorithm (lr = 1e-4, weight_decay = 1e-4); the learning rate was updated using ReduceLROnPlateau, which multiplies the learning rate by 0.5 when the validation macro F1 score did not improve for 2 epochs. Training was run for a maximum of 15 epochs, and early stopping was applied when the macro F1 score did not improve for 3 consecutive epochs. The gradient norm clipping value was set to 1.0. The random seed is 42. Table 2 summarizes all hyperparameters.

Table 2. Training environment, libraries, and hyperparameters.

Parametre	Değer
Hardware	NVIDIA Tesla T4 GPU (Google Colab, CUDA 12.x)
PyTorch	2.x (torch.nn, torch.utils.data)
Transformers	Hugging Face transformers (AutoModel, AutoTokenizer, CLIPTextModel)
Scikit-learn	1.x (classification_report, confusion_matrix, cohen_kappa_score, matthews_corrcoef)
Optimizer	AdamW (weight_decay = 1e-4)
Learning Rate (lr)	1e-4
Learning Rate Timer	ReduceLROnPlateau (patience=2, factor=0.5)
Loss Function	CrossEntropyLoss
Batch Size	256
Epoch (max)	15
Early Stopping Sabrı (patience)	3 epoch
Gradient Clipping Normu	1.0
Gizli Katman Boyutu (hidden)	256

Parametre	Değer
MLP Dropout	0.3 → 0.2 → 0.3
Cross-Attention Başlık Sayısı	4
Modality Dropout Olasılığı (p)	0.10
Maximum Token Length	128
CLIP Maximum Token Length	77
Random Seed	42

3.7. Evaluation Metrics

Model performance on the test set is reported using Accuracy and Macro F1 score, as well as Precision (macro), Recall (macro), Cohen's Kappa, Matthews Correlation Coefficient (MCC), Balanced Accuracy, Jaccard Index (macro), F0.5, and F2 scores. Since the AG News test set is perfectly balanced across classes (1,900 examples per class), balanced accuracy and overall accuracy yield the same value. This has been reported in all tables for numerical consistency.

4. Result

This section presents the results obtained from four language models and eight experimental configurations (a total of 32 experiments). All values were obtained using a single seed (seed = 42); in the current version of the study, it was not possible to report the mean ± standard deviation across multiple seeds. This limitation is addressed again in the Limitations section.

4.1. Overall Accuracy and Macro F1 Results

Table 3 presents the Precision and Macro F1 scores on the test set for all model and fusion configurations. The highest scores were achieved with ALBERT-base-v2 and the Cross-Attention architecture (Precision 0.9461; Macro F1 0.9460). BERT-base-uncased with Cross-Attention achieved 0.9412; RoBERTa-base Cross-Attention + Dropout achieved 0.9420; and DistilBERT-base-uncased with Late Fusion achieved a score of 0.9408.

Table 3. Accuracy and Macro F1 scores for all fusion and modality dropout configurations across four language models (test set, seed = 42).

Experiment	ALBERT	BERT	DistilBERT	RoBERTa
Baseline Text	0,9078 / 0,9079	0,9167 / 0,9165	0,9175 / 0,9174	0,9204 / 0,9203
Baseline Vision	0,9354 / 0,9353	0,9354 / 0,9353	0,9354 / 0,9353	0,9354 / 0,9353
Early Fusion	0,9451 / 0,9451	0,9407 / 0,9407	0,9388 / 0,9388	0,9400 / 0,9400
Late Fusion	0,9429 / 0,9429	0,9401 / 0,9402	0,9408 / 0,9408	0,9412 / 0,9411
Cross-Attention	0,9461 / 0,9460	0,9412 / 0,9412	0,9371 / 0,9371	0,9413 / 0,9412
Early + Dropout	0,9429 / 0,9429	0,9405 / 0,9405	0,9388 / 0,9388	0,9387 / 0,9386
Late + Dropout	0,9421 / 0,9421	0,9379 / 0,9378	0,9366 / 0,9366	0,9386 / 0,9386
Cross-Attn + Dropout	0,9436 / 0,9436	0,9412 / 0,9412	0,9382 / 0,9382	0,9420 / 0,9419

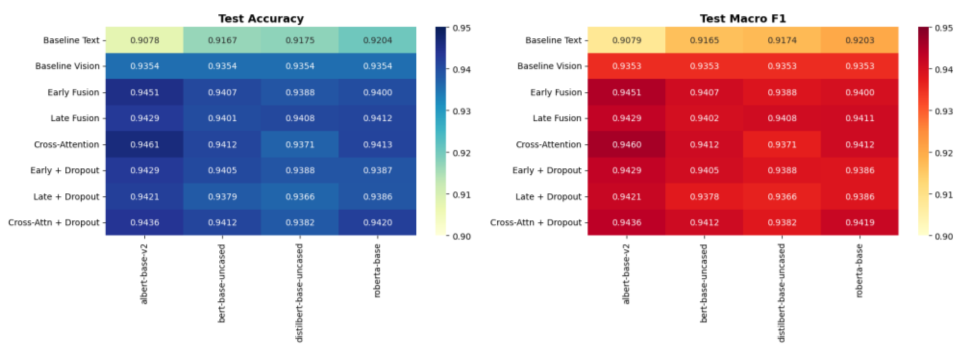

Figure 1. Matrix of Accuracy (heatmap on the left) and Macro F1 (heatmap on the right) scores for four language models and eight experimental configurations. Darker colors indicate higher scores; the color bars are normalized to the range 0.90–0.95.

4.2. The Transition from Plain Text to Multimodal Architecture

The highest Macro F1 score in plain text configurations was achieved with RoBERTa-base (0.9203). In the same model, the application of Cross-Attention increased the Macro F1 to 0.9412 ($\Delta = +2.09$ points). For ALBERT-base-v2, this difference is more pronounced: 0.9079 → 0.9460 ($\Delta = +3.81$ points). For BERT and DistilBERT, the increases were measured as 0.9165 → 0.9412 ($\Delta = +2.47$) and 0.9174 → 0.9371 ($\Delta = +1.97$), respectively. Figure 2 presents a comparison of fusion strategies per model.

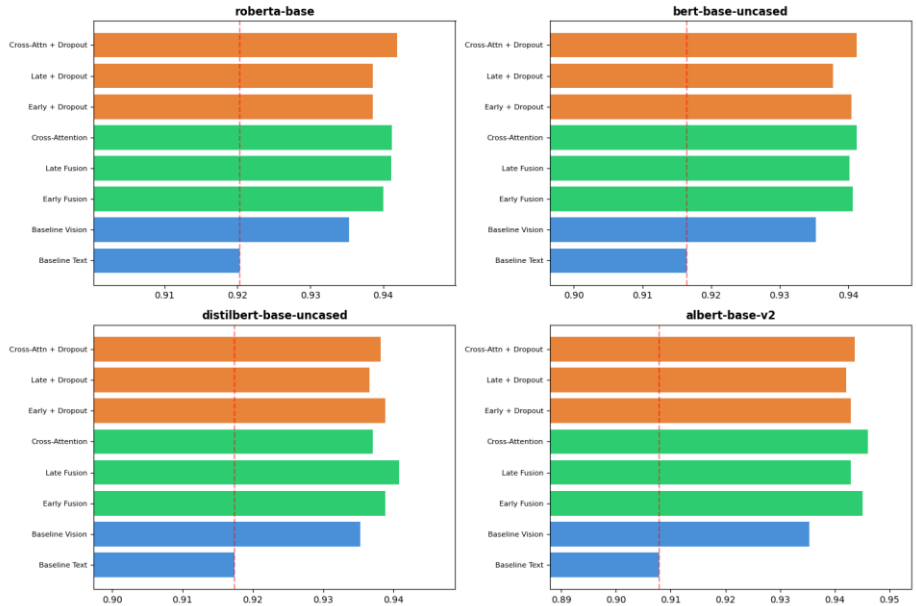

Figure 2. Macro F1 performance for eight experimental configurations for each language model. The red dashed line represents the text-only baseline for the respective model. The blue bars represent the baseline, the green bars represent the fusion configuration, and the orange bars represent the Modality Dropout configuration.

4.3. A Comparative Analysis of Fusion Strategies

Across all four language models, the Cross-Attention architecture yielded an increase in the Macro F1 score compared to the plain text baseline. For ALBERT and BERT, Cross-Attention produced higher scores than Early and Late Fusion; for DistilBERT, Late Fusion remained 0.0037 points higher than this architecture. Figure 3 presents the performance of the Cross-Attention strategy across the four language models, compared to Early/Late Fusion and the plain baseline, using a radar plot.

Figure 3. Academic evaluation plot: (top) radar (spider) plot of Macro F1 scores for four language models across five fusion configurations; (bottom left) absolute gain provided by the Cross-Attention architecture relative to the plain text baseline; (bottom right) class-wise Macro F1 transformation for ALBERT

4.4. Class-Based Analysis

In the class-based analysis, the Sports class achieved the highest F1 score in both text-only and multimodal configurations (e.g., DistilBERT Text-only 0.9774, ALBERT Cross-Attention 0.9869). The lowest F1 score in the text-only setting was observed in the Business class (ALBERT Baseline Text 0.8661). The Cross-Attention application improved the F1 score for the Business class from 0.8661 to 0.9178 (Δ = +5.17 points) and for the Sci/Tech class from 0.8733 to 0.9243 (Δ = +5.10 points) for ALBERT. Since these two classes cover topics that partially overlap in terms of vocabulary, the discriminative contribution of the pseudo-visual modality is more pronounced in these classes (Table 5).

Table 5. Class-based Macro F1 scores (test set, seed = 42)

Model	Yapılandırma	World	Sports	Business	Sci/Tech
ALBERT	Baseline Text	0,9217	0,9704	0,8661	0,8733
ALBERT	Cross-Attention	0,9550	0,9869	0,9178	0,9243
BERT	Baseline Text	0,9237	0,9720	0,8798	0,8905
BERT	Cross-Attention	0,9483	0,9853	0,9125	0,9187
DistilBERT	Baseline Text	0,9248	0,9774	0,8752	0,8921
DistilBERT	Cross-Attention	0,9467	0,9827	0,9064	0,9125
RoBERTa	Baseline Text	0,9248	0,9765	0,8854	0,8945
RoBERTa	Cross-Attention	0,9497	0,9835	0,9112	0,9205

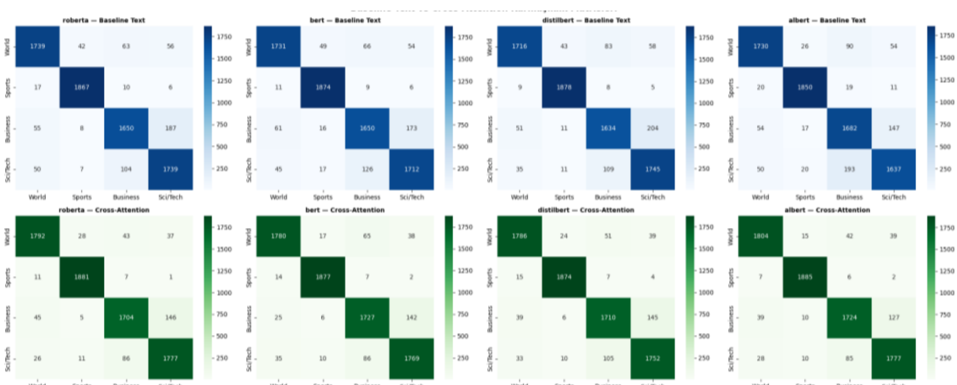

Figure 4. Confusion matrices for four language models under the plain text (top row) and Cross-Attention (bottom row) configurations. Rows represent the true class, and columns represent the predicted class; confusion between the Business and Sci/Tech classes decreases significantly when Cross-Attention is applied.

4.5. Expanded Performance Metrics

Table 4 presents eight extended metrics for the plain text and Cross-Attention configurations. The Cohen's Kappa and MCC values for the ALBERT model have increased from 0.8770–0.8771 to 0.9281. The Jaccard Index shows an increase from 0.8340 to 0.8988. These values indicate that the models' actual predictive power is not due to chance in the test set, where the four classes are evenly distributed.

Table 4. Extended evaluation metrics for plain text and Cross-Attention configurations (test set, seed = 42).

Model	**Experiment**	**Precision**	**Recall**	**Kappa**	**MCC**	**Bal. Acc**	**Jaccard**
ALBERT	Baseline Text	0,9083	0,9078	0,8770	0,8771	0,9078	0,8340
ALBERT	Cross-Attention	0,9461	0,9461	0,9281	0,9281	0,9461	0,8988
BERT	Baseline Text	0,9166	0,9167	0,8889	0,8891	0,9167	0,8479
BERT	Cross-Attention	0,9414	0,9412	0,9216	0,9216	0,9412	0,8904
DistilBERT	Baseline Text	0,9181	0,9175	0,8900	0,8903	0,9175	0,8498
DistilBERT	Cross-Attention	0,9372	0,9371	0,9161	0,9162	0,9371	0,8832
RoBERTa	Baseline Text	0,9207	0,9204	0,8939	0,8940	0,9204	0,8544
RoBERTa	Cross-Attention	0,9414	0,9413	0,9218	0,9218	0,9413	0,8903

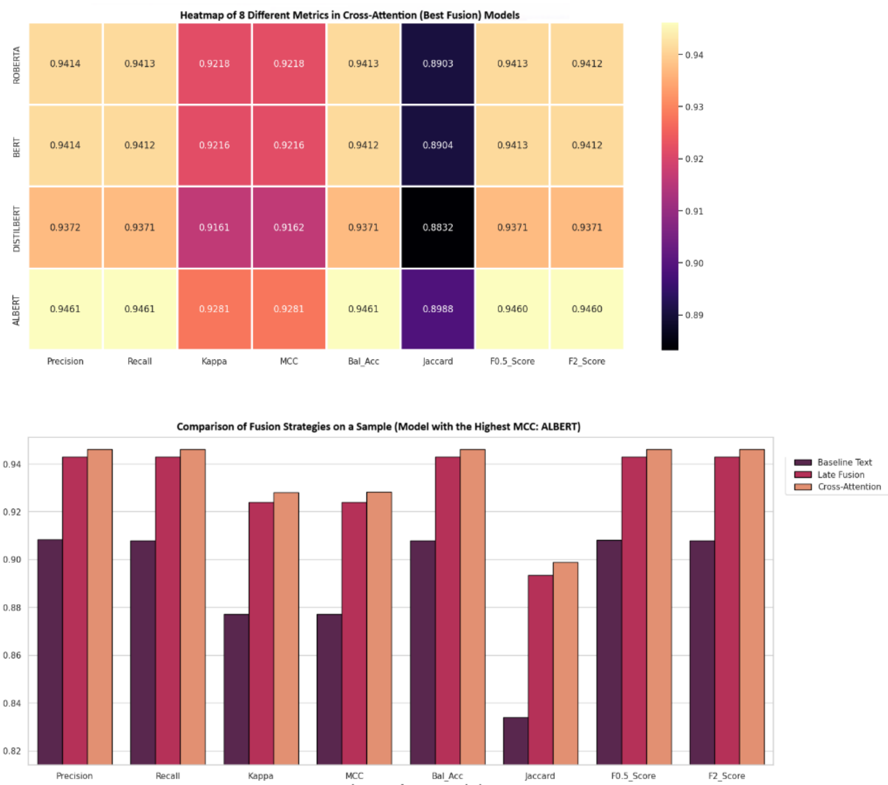

Figure 5. Comparison of four language models across eight different metrics in the Cross-Attention architecture. The values in the heatmap are metric scores normalized to the range 0–1.

4.6. The Effect of Letting Go of Modality

The Modality Dropout (p = 0.10) implementation resulted in a decrease in test accuracy for ALBERT in the cross-attention architecture from 0.9461 to 0.9436, 0.9412 → 0.9412 for BERT, 0.9371 → 0.9382 for DistilBERT, and 0.9413 → 0.9420 for RoBERTa. These small changes indicate that modality dropout does not significantly reduce average performance under clean (full modality) test conditions. The primary benefit of the technique is its robustness against missing modalities; this evaluation requires an additional protocol involving the resetting of a single modality at test time, which falls outside the scope of the results reported in this study and is addressed in the Future Work section.

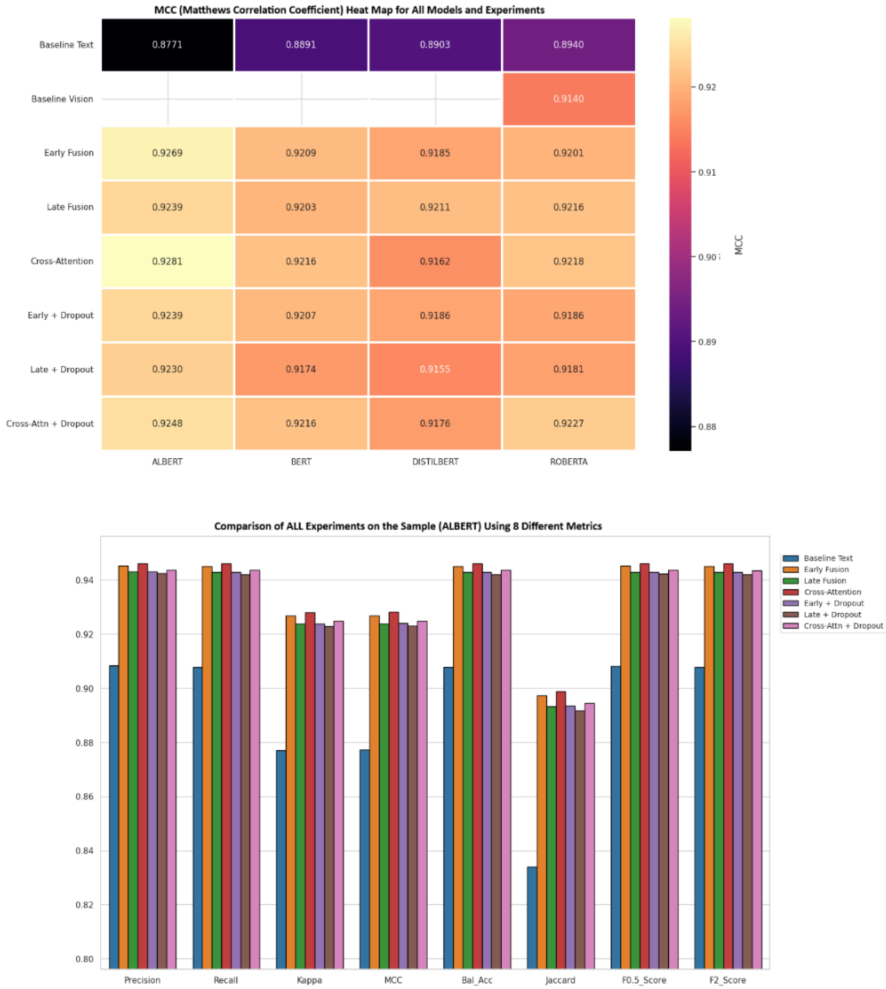

Figure 6. Matthews Correlation Coefficient (MCC) heatmap and additional metric comparisons for all experiments (four models × eight configurations). The MCC provides an assessment of prediction quality that is independent of chance, even on a balanced test set.

4.7. Limitations

- The visual modality was derived not from real images but from synthetic visual embeddings generated by the CLIP text encoder. This makes it difficult to distinguish whether the method's advantage stems from incorporating a text-conditional CLIP representation or from independent visual information.
- Experiments were conducted with a single seed (seed = 42); therefore, variance (mean ± standard deviation) is not reported. Repeating the ablation with at least three different seeds would allow for the calculation of a confidence interval.
- The Modality Dropout effect is reported only under full-modality test conditions; the protocol of removing a single modality during testing (T-only / V-only) was not applied.
- An evaluation of continuous learning requires a separate experimental setup involving sequential tasks, forward/backward transfer metrics, and a replay mechanism; this study does not provide such an evaluation.

5. Discussion and Conclusion

5.1. Discussion

The results show that the pseudo-visual representation derived from the CLIP text encoder achieves a consistent improvement over the text-only baseline in the AG News news classification task. The most significant gains were observed between the Business and Sci/Tech classes; this suggests that the cross-modal attention mechanism plays a distinctive role for classes with partially overlapping vocabularies. The observation of a similar trend across four different language models indicates that the method contributes regardless of the chosen language model. However, a significant portion of the observed improvement may stem from CLIP's text-conditional embedding; therefore, the results should be supported by a comparative study using real image data.

5.2. Conclusion

In this study, pseudo-visual embeddings derived from the CLIP text encoder using four pre-trained language models (RoBERTa, BERT, DistilBERT, ALBERT) were compared on the AG News dataset under three different fusion strategies (Early Fusion, Late Fusion, Cross-Attention) and Modality Dropout configurations. The overall best results were obtained with the Cross-Attention architecture based on ALBERT-base-v2 (Accuracy 0.9461; Macro F1 0.9460; Kappa 0.9281; MCC 0.9281). The Modality Dropout approach was presented as a promising regularizer that maintains performance under clean test conditions while providing robustness against missing modalities.

5.3. Future Work

- Decomposing the effect of text-conditional image embedding by synthetically generating real image data (e.g., Stable Diffusion) or using real stock images related to news stories.
- Comparative robustness evaluation under modality-dropout protocols (T-only and V-only) during testing.
- Reruns with multiple random seeds and mean ± standard deviation tables.
- A comprehensive continuous learning evaluation incorporating sequential task definitions, forward/backward transfer metrics, and experience replay mechanisms.

6. References

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019, 4171–4186. https://doi.org/10.18653/v1/N19-1423

Kiela, D., Bhooshan, S., Firooz, H., & Testuggine, D. (2019). Supervised multimodal bitransformers for classifying images and text. arXiv preprint arXiv:1909.02950. https://doi.org/10.48550/arXiv.1909.02950

Eunjee Choi, Junhyun Ahn, XinYu Piao, Jong-Kook Kim. Atıf "Choi, E., Ahn, J., Piao, X., & Kim, J.-K. (2025). CroMe: Multimodal fake news detection using cross-modal tri-transformer and metric learning. arXiv preprint arXiv:2501.12422. https://doi.org/10.48550/arXiv.2501.12422

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences, 114(13), 3521–3526. https://doi.org/10.1073/pnas.1611835114

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A lite BERT for self-supervised learning of language representations. Proceedings of the International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1909.11942

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. https://doi.org/10.48550/arXiv.1907.11692

Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. Advances in Neural Information Processing Systems, 32, 13–23. https://doi.org/10.48550/arXiv.1908.02265

Neverova, N., Wolf, C., Taylor, G., & Nebout, F. (2016). ModDrop: Adaptive multi-modal gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(8), 1692–1706. https://doi.org/10.1109/TPAMI.2015.2461544

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision.

Proceedings of the 38th International Conference on Machine Learning, 139, 8748–8763. https://doi.org/10.48550/arXiv.2103.00020

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108. https://doi.org/10.48550/arXiv.1910.01108

Shang, L., Kou, Z., Zhang, Y., & Wang, D. (2021). A multimodal misinformation detector for COVID-19 short videos on TikTok. Proceedings of the IEEE International Conference on Big Data, 899–908. https://doi.org/10.1109/BigData52589.2021.9671928

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. Advances in Neural Information Processing Systems, 28, 649–657. https://doi.org/10.48550/arXiv.1509.01626

CHAPTER 16

Chemical Composition, Chemical Elements and Vitamins in Milk

***Julijana Tomovska*[1] *& Ilmije Vllasaku*[2] *& Kristina Tomska*[3]**

1. Introduction

Milk is one of the oldest products used by the population in his diet. Milk is a rich source of all necessary nutrients that are of particular importance for the human body. Milk is a colloidal solution of proteins, milk fat, milk sugar and mineral substances. This type of milk is a complete food because it contains all the necessary nutrients for the normal development not only of the baby, but also for humans, and that is why it is the most valuable food product. Today, there are a large number of breeds for milk production, but the most widespread breed is the Holstein-Friesian breed. In our country, the most cultivated breed is the black and white Holstein-Friesian breed, which is represented in almost all small and large farms and is grown in all regions of our country, characterized by good milk yield, but in recent times, to improve milk yield, crossbreeding with Holstein - the Friesian breed has been carried out. The quality of milk from the black and white breed is standard. The quality of the milk is greatly influenced by the diet of the cows and the use of fodder crops. The milk quality also depends on the period of lactation in which the cows are. Cows that are at the beginning of lactation have milk with a lower fat content, and at the end of lactation it increases. The use of coarse fodder for feeding cows increases fat content, while concentrated feed mixes reduce it. The type of forage has little effect on protein content and no effect on lactose content [1].

Definition and general properties of milk

Milk as a product of the mammary gland is characterized by a certain nutritional value, but it also contains a large amount of other bio stimulating substances that make it an indispensable food for all age categories of people. Humans mostly consume cow's milk in their daily diet [1], [2], [3].

Milk has long been known as "medicine and food for long life". People in developed countries consume a large amount of milk and meat. For a normal diet, a person with a body mass of 75 kg should consume 0.9 g of protein/kg of body mass daily, or expressed in absolute values, it amounts to 67 g, of which 1/3 should be of animal origin [4].

[1] University „St. Kliment Ohridski" - Bitola, Faculty of Biotechnical Sciences- Bitola, Partizanska b.b, (Kompleks Kassarna) 7000 Bitola, North Macedonia
[2] UBT - Higher Education Institution, Kalabrija nn, 10000 Prishtina, Kosovo
[3] University „St. Kliment Ohridski" - Bitola, Faculty of Biotechnical Sciences- Bitola, Partizanska b.b, (Kompleks Kassarna) 7000 Bitola, North Macedonia

Milk as a food product is characterized by easy digestibility and utilization of all nutrients, including mineral substances. Milk is a rich source of all necessary amino acids, fats, some of the necessary vitamins and a very rich source of calcium characterized by easy digestibility and utilization [5].

Milk production in R. Macedonia is decreasing year by year as a consequence of the decrease in the number of cows. According to statistical data, in R. Macedonia has a total of 212,000 cattle, of which about 90,000 are dairy cows. Of these, the largest percentage belong to the black and white (Holstein or Holstein-Friesian) breed of cattle and its hybrids. About 350,000 tons of milk are produced in our country, of which 90% is cow's milk [6].

The amount of milk per dairy cow according to numerous studies is a breed characteristic that largely depends on the conditions of rearing (accommodation, quality of roughage and quality of fodder for dairy cows). Under normal housing conditions and quality coarse feed, a large production of milk of excellent quality can be expected [7].

According to numerous tests that have been carried out in the country and abroad, the food used to feed dairy cows should be of known quality, i.e. chemical composition so that the daily ration can be composed and thus provide all the necessary nutrients for good health of cows and high production. Mineral substances are divided into two groups, macro and microelements. Calcium is one of the most important elements that milk contains the most and plays a role in the mineralization of bones and teeth, muscle contraction and relaxation, as well as in the functioning of nerves and blood clotting. Magnesium is also present in milk, an important element that participates in bone mineralization, protein building, for normal muscle function, transmission of nerve impulses, proper immune function and maintenance of teeth. Phosphorus as an element is important because it is involved in the mineralization of bones and teeth, it is an important element in the construction of genes that are important for the transmission of genetic traits from parents to offspring. Potassium as a chemical element participates and facilitates reactions, including the creation of proteins, and together with sodium is an important element in maintaining the balance of fluids and electrolytes, supporting cellular integrity, transmitting nerve impulses and muscle contraction, including the heart. Zinc is a trace element that is important for the normal functioning of hormones, it is needed for many enzymes, it is involved in the creation of genetic material and proteins, the activation of immune cells, the transport of vitamin A, etc. [10][15].

Tests of milk ware done from three farms, A, B, C, which were fed with alfalfa hay and fodder mixtures. The latest studies show that the higher milk production is the result of the diet, the large intake of easily digestible nutrients consumed through feed mixtures, but also as a result of the mobilization of energy, i.e. body fat from the body, but not of the increase in partial efficiency in the use of energy during lactation [11].

Nutrition of dairy cows

High milk production requires quality and properly balanced feed for dairy cows. The coarse roughage used to feed the cows should be of excellent quality, providing about 70% of the daily required nutrients, and 30% should be provided through the forage compounds for dairy cows. With proper nutrition, dairy cows should be provided with all the necessary nutrients, both in quantity and quality. To achieve this, the daily ratio of the dairy cows should be matched to the daily milk production.

The quantity and quality of the daily ration of dairy cows varies depending on the stage of lactation. During the first 100 days or the so-called early lactation phase, dairy cows should be given quality coarse and concentrated feed to meet increased needs. Although in the early stage of lactation, quality food rich in easily digestible nutrients is given, still there is a mobilization of body reserves to meet energy needs. In the second phase, when milk production decreases, feeding intensity should not be reduced. During this period, the lost body mass should be replenished and a sufficient amount of nutrients should be provided for milk production and for the fetus (the new calf that is expected) [15].

In the third phase, when the cows are in a dry period, that is, when they are not milked, nutrients should be provided for the development of the fetus - the calf and maintenance of their normal body mass. The amount of milk per dairy cow according to numerous studies is a breed characteristic that largely depends on the conditions of rearing (accommodation, quality of roughage and quality of fodder for dairy cows). Under normal housing conditions and quality coarse feed, a large production of milk of excellent quality can be expected [7].

Composition of milk

Cow's milk as a food product is characterized by a complex chemical composition. The average chemical composition of cow's milk is about 87% water and 13% dry matter. Dry matter in milk consists of proteins, which are represented by about 3.2%, milk fat by about 3.6%, milk sugar (lactose) by about 4.6% and mineral substances by about 1%. Proteins in milk are the most important nutrients by which the quality of milk is evaluated and they are represented by about 3.2%, dominated by casein which is the main protein, lactoalbumins and lactoglobulins. Proteins in milk are quite constant and do not show much variability as is the case with milk fat [1].

Milk fat in milk is the second most important ingredient that represents the main source of energy. Fat-soluble vitamins, especially vitamin A, are also dissolved in it. The content of vitamin A in the milk also depends on the color of the milk fat. The milk fat in milk is quite variable and it is represented in smaller quantities at the beginning of lactation when milk production is the highest, and at the end of lactation it increases and is the most represented. The composition of the daily ration of dairy cows has a great influence on the content of milk fat in milk. If there is roughage with a higher cellulose content in the ration of dairy cows, the milk fat content will be higher. Milk sugar (lactose) in milk is the second energy nutrient which is represented by 4.6%. Lactose, like proteins, shows the least variability. Mineral substances are important both for the quality of milk and for its nutritional value. Mineral substances from milk are easily used by the human body, because they are easily digestible and usable. Calcium (up to 1200 mg/L), phosphorus (up to 950 mg/L), potassium (up to 1200 mg/L) and magnesium (up to 110 mg/L) are the most abundant minerals in milk. Milk as a food product is quite poor in iron, cobalt and other trace elements [8], [9], [15].

According to Douglas (2014), the chemical composition of milk varies greatly among different species of animals. According to numerous studies carried out by Feskanich, et all., in 1997 and Harleys book 2009, cow's milk contains on average 3.4% protein, 3.6% milk fat and 4.6% lactose, 0.7 %, minerals (92) with an energy value of 66 kcal of energy in 100 grams [20], [12], [23].

In general, the protein and milk fat content of milk is different in different breeds. The content of proteins in milk, as well as milk sugar, are quite constant, that is, they show little variability, while milk fat varies widely. Thus, milk fat largely depends on both genetic and paragenetic factors. The breed as genetic factors has its own positive influence on the content of milk fat in the milk and as a result there are breeds of cows that give milk with a higher fat content (Jersey and Holstein). Of the paragenetic factors, the nutrition of dairy cows has a great influence. Cows that are fed with coarse roughage give milk with a higher content of milk fat, compared to cows that are fed with concentrated fodder mixtures. During the summer, cows give milk with a lower content of milk fat as a consequence of the high temperatures, and during the winter with a higher content of milk fat and are a response to dairy cattle [11], [13].

However, different breeds of cows have different chemical compositions of milk and differ in the average values of the components. Black and white cows have the lowest milk fat and protein content, while Jersey cows have the highest milk fat content [14].

Milk quality

According to the tests conducted at Cornell University (2018) in the paper "Composition of milk", raw cow's milk contains: water 87.3% (85.5 – 88.7%); fat 3.9% (2.4-5.5%); proteins 3.25% (2.3 – 4.4%); casein 2.6% (1.7 – 3.5%); serum proteins, small proteins and carbohydrates 4.6% (3.8 – 5.3%); organic acids 0.18 % (0.13 – 0.22 %), such as citric, lactic, formic, acetic, oxalic acid, minerals, enzymes. It also contains vitamins: A, C, D, thiamin, riboflavin; minerals: 0.65% (0.53–0.80%); cations, K, Ca, Mg, K, anions: chlorides, phosphates, citrates, carbonates; enzymes: peroxidase, catalase, phosphatase, lipase; gases: CO_2, N_2, O_2 [4], [15]

To evaluate the quality of the milk, the effects of the exposure of the vitamins and minerals in the milk to heat and light were investigated. The tests of the raw milk with heat treatment showed that the quality is lost. Higher ultra-high temperature (UHT) heat treatment, i.e. pasteurization for extended use combined with increased storage time of these products, which is caused by some water-soluble vitamins, leads to a reduction of vitamins and minerals [13], [18].

Milk vitamins and their properties

Vitamins are generally classified as water-soluble, which easily dissolves in water and are readily excreted from the body, and as fat-soluble, which are absorbed through the intestinal tract with lipids (fats). Vitamins A, D, E, and K are fat-soluble vitamins present in human body, while the vitamin C and the vitamins of the B group are water-soluble. Vitamins C and E function as antioxidants [22]. Milk also contains a sufficient amount of bio-stimulants that are necessary for the proper development of newborn animals. Fat-soluble vitamins A, D, E and K are mainly found in milk fat [23]. Milk has limited amounts of vitamin K, and B vitamins is found in the aqueous phase of the milk. Vitamin A has a role in the growth and development of cells, mucous membranes, skin, bones and the health of teeth, eyes, reproduction and immunity. Vitamin A is a bio-stimulant and nutrient and it is known in the forms of all-*trans*-retinol, all-*trans*-retinyl-esters, as well as all-*trans*-beta-carotene and other provitamin A carotenoids [24], [25], [26], [27], [28].

There are eight groups of related molecules (vitamers) of vitamin E-four tocopherols and four tocotrienols [29]. Vitamin E is an antioxidant that protects the cells from damage and supports the function of the immune system. Tocotrienol and tocopherol protect lipids and have significant antioxidant properties. Vitamin E protects lipids, and it is present in milk at a very low level [23]. The combination of selenium (Se) with vitamins A and E has an important role in normal fetal development during pregnancy. The combination of vitamins A and E is also required for the development of antibodies that are present in large amounts in colostrum [30]. Vitamin E acts as an antioxidant and helps lipid membranes to be intact. Vitamin E reacts with selenium (Se), which is an excellent combination of antioxidants. If one of the two nutrients is deficient, the metabolic activity is impaired and therefore its addition is necessary. Products that have sufficient content of vitamin E and selenium (Se) that characterized by greater sustainability due to the antioxidant activity [31], [26]. Vitamin C (ascorbic acid) contributes to collagen forming, and as antioxidant restores the vitamin E to its active form, helps also hormone synthesis, supports immune function, and aids iron absorption [32].

Research objectives

Analysis of the chemical composition of raw milk in fresh milk: proteins, fats, lactose, dry matter and density of milk and analysis of 21 chemical elements, which 13 elements: aluminum (Al), arsenic (As), boron (B), barium (Ba), calcium (Ca), cadmium (Cd), cobalt (Co), chromium (Cr), copper (Cu), iron (Fe), potassium (K), lithium (Li), magnesium (Mg), manganese (Mn), sodium (Na), nickel (Ni), phosphorus (P), lead (Pb), strontium (Sr), vanadium (V) and zinc (Zn) the value were determined. Also were examined the amount of vitamins A, E and C.

2. Materials and Methods

Materials were samples of raw milk taken from three farms A, B, C from different regions our country. Milk was from the breed Holstein and the total daily meal used by the cattle was different (20kg farm A, 28kg Farm B and 31kg Farm C per day). Milk was analyzed immediately after the milking. The samples of animal feed included: two types of concentrates (fodder mixture for dairy cows with the lowest crude protein concentration of 18%), alfalfa hay and straw. The cow breeds from the three farms were of the Holstein-Friesian breed. The cows from farm A and from farm C has fed with two types of concentrate (K_1 and K_2) in the morning, with alfalfa and straw at lunchtime and with concentrate in the evening. The cows from farm B have been fed with concentrate K_2 and alfalfa. Cow breeds ware milking done by a milking machine.

Chemical composition in raw milk: water, fats, proteins, milk sugar and mineral substances were made using standard and accredited analysis methods, and which are still in use the Law on Standardization [16], [17]. The water in the milk was tested by the standard drying method, the milk fat content by the Gerber method, the proteins by the Kjeldahl method, the milk sugar-lactose content by the iodometric method, the mineral substances by the combustion method and examination of the relative density, using Lactoscan. For the investigations of some chemical components in milk, extracts obtain with extraction which was done with trichloroacetic acid. The examination of the mineral composition and the presence of macro-micro elements were performed use Atomic Emission Spectrometry with Inductively Coupled Plasma (ICP-AES). The content of heavy metals (arsenic, cadmium, cobalt, chromium, copper, iron, manganese, nickel, lead, vanadium and zinc) in milk and useful metals needed in food, such as calcium, magnesium, potassium, sodium, etc., a total of 21 elements in milk. Vitamin A (Retinol) and vitamin E (Tocopherol) were analyzed by HPLC technique, using Perkin Elmer apparatous with pump: series 200LC, autosampler; ISS - 200, detector LC - 135 / LC -235 C DA. Vitamin C (Ascorbic acid) was analyzed spectrophotometrically (Spectroquant Pharo 300 - Merck) at 520 nm.

2.1 Methods for determined the chemical composition of milk

2.1.1 Fat determination

Fat determination in milk is very important for scientific research work and practical purposes. Today, this test is necessary because it is possible to achieve control of the fat in the milk of individual cows or herds, which makes it possible to implement the rules of rationed nutrition and correct selection, especially in the more advanced dairy countries, in which payment for milk has been applied for a long time according to the fat percentage content. That is why it is necessary to control the fat in milk, whether it is from own production or from sale. Control of the production of cheese and butter is impossible without data on the fats from which they are created in the milk. Various technical methods are used to determine the percentage of fat in milk, the Gerber method according many authors as standard method [33].

2.1.2 Gerber's method for determination fat

This method was first proposed by Dr. Gerber in 1892, later (1893 and 1895), it received the name acidobutrymetric method because it uses sulfuric acid and works in butyrometers. The Gerber method is relatively inexpensive and quite accurate, giving results with an accuracy of 0.1% fat. The method is based on the dissolution of proteins in milk under the influence of concentrated H_2SO_4 with a specific gravity of 1.815 – 1.825 at a temperature of 20 ^{0}C, which burns organic substances (burning by water without heating). Dilute H_2SO_4 coagulates casein, at a certain concentration it can dissolve it, and it is not strong enough to cause protein charring. In doing so, the fat droplets will remain suspended in the strongly acidic environment and can be separated under the action of centrifugal force.

Procedure – The sample number is noted on the polished circle from the pear-shaped part of the butyrometer, then the butyrometer is placed in a rack in the same order as the samples are arranged. After adjustment, the neck of the butyrometer is turned down, carefully measuring 10 ml of H_2SO_4 with spec. t. 1.815 – 1.825 and 11 ml of milk sample, well mixed at a temperature of 20 ^{0}C. The tip of the pipette is wiped off the edge of the glass container to remove excess droplets and the milk is strained into the butyrometer. Stirring is done by pouring the milk into the inner side surface of the butyrimeter which we touched the tip of the pipette and gradually releasing the finger from the other end of the pipette. Stirring should be slower at first, and when the sulfuric acid forms a solid layer with the contact surface of the milk, it can flow faster. The touch layer should not be allowed to darken as not only can a wrong result be obtained but also the reading will be difficult. Then 1 ml of $C_5H_{12}OH$ (amyl alcohol) is carefully added to the surface of the milk so as not to scratch the inner surface of the neck of the butyrometer as the stopper may fly out. Then, the butyrometer is closed, stirred and heat is generated from the H_2SO_4 which binds the water in the milk. The butyrometer is left for 5 minutes in a water bath at a temperature of 65^{0}C, centrifuged 5 minutes at a speed of 1000-1200 rpm and then left in a water bath at a temperature of 65^{0}C. While reading, the butyrometer is held and the reading is taken as quickly as possible, accurate to within 0.05 % which corresponds to half of ten divisions of the scale. After the reading, the butyrometer is again placed in a water bath, but this time the stopper should be turned upwards and should be carefully washed.

2.1.3 Kjedahl method for determination of total nitrogen substances – proteins

Procedure – Taken 10 cm^3 of the test milk and it placed in a measuring vessel, weighed on an analytical balance, and then placed in a Kjeldahl flask for combustion. The vessel together with the remaining milk is weighed and the difference is used to determine the amount of milk that has been strained. 20 cm^3 of H_2SO_4 is placed in the flask. a. with spec. t. 1.84, a little more than sulfuric acid can be added, but this should be noted because the amount of milk will depend on it, then 0.5 g of copper sulfate is added as a catalyst consisting of copper sulfate, sodium sulfate and selenium in the ratio 1:1:0.2. When the organic substance burns (oxidizes), the liquid becomes clear and transparent. The yellow color of the squash is a sign that the combustion is not complete. Combustion should continue for 20-30 minutes from the moment the contents of the flask become clear. Then, the contents of the flask are cooled and diluted with 50 cm^3 of distilled water and poured into a distillation flask. Then, 30 cm^3 of n/10 H_2SO_4 (or n/10 HCL) is placed in an Erlenmeyer flask and the flask is positioned so that the top of the condenser tube is in the acid. About 80 cm^3 of NaOH with spec. t 1.42 (38 – 40 %) and the distillation flask is connected to the distillation apparatus.

The distillate lasts 40 – 50 minutes and usually ends after the passage of 2/3 of the content in the distillation flask. After the end of the distillation, the content of unbound sulfuric acid is determined in a receiving flask. The contents of the flask are titrated with n/10 NaOH with methyl red indicator, which changes color to yellow at the end of the titration, and then n/10 H2SO4 remaining after distillation is added. The specific amount titrated with n/10 NaOH is calculated as the number of grams of nitrogen in the examined amount of milk and multiplied by 0.0014. To get the proteins, the nitrogen result must be multiplied by 6.38. In this way, the number of proteins in the examined milk is calculated.

2.1.4 Iodometric method for determination sugar-lactose in milk

The following reagents are required for this method: Fehling's solution A – blue solution of copper (II) sulfate; Fehling's solution B – colorless aqueous solution of potassium sodium tartrate (known as Rochell's salt); strong alkali (usually NaOH); 0.5% starch solution dissolved in water; 2% alcoholic solution of phenolphthalein and n/10 solution of iodine.

Procedure – Taken 10 cm^3 of milk is placed in a measuring flask, which is diluted with distilled water to 1/3 of the volume of the flask, n/10 NaOH and 2 drops of 2% alcoholic solution of phenolphthalein. Fehling's solution A is added to the flask until the violet color changes to light blue, then it is topped up with distilled water up to the mark. The contents of the flask are mixed and the serum is allowed to separate from the coagulated proteins, then filtered through a pleated filter into a dry flask. Taken 10 cm^3 of the filtrate, 10 cm^3 of Fehling's solution B and 20 cm^3 of distilled water are placed in a 300 cm^3 Erlenmeyer flask, then the mixture is heated and allowed to boil for 6 minutes from the moment of boiling. Then the entire contents are rapidly cooled under a tap and 3 – 4 drops of alcoholic phenolphthalein solution are added and titrated with sulfuric acid solution (25 cm^3 H_2SO_4/L) until the pale pink color disappears. To achieve a weakly acidic reaction, add another 3 cm^3 of the same acid, and after adding 20 cm^3 of n/10 iodine solution, stir for 3 to 5 minutes, until the precipitate disappears. The determination of the amount of unbound iodine is carried out by titration with n/10 $Na_2S_2O_3$ in the presence of starch as an indicator. Starch is added before the end of the titration.

2.1.5 Direct method for determination dry matter

The direct method for determining the dry substance is based on drying the milk at a temperature of 105 ^{0}C to a constant weight. The obtained result represents a dry substance and is expressed in percent (%). This is the most precise method used for scientific research.

Procedure – The weighing pan will be removed from the desiccator and 25 g of heated quartz sand and small glass rods will be placed in it and weighed together on an analytical balance. 20 cm^3 of milk is added and weighed again, so that the exact weight of the milk is obtained from the difference of the two measurements. The measurement should be done as quickly as possible so that it does not evaporate and is not affected by light, therefore the contents of the container are first dried over a water bath, then the container is placed in a drying oven at a temperature of 105^{0}C, it stands for 2 hours and finally the glass is cool for 30 minutes. If the difference between the previous and the last measurement is less than 0.001 g, drying is considered complete. If the difference is greater than those values, the drying is repeated as many times every half an hour until the maximum limit of 0.001 g is reached, so the drying is carried out to a constant weight. Dry matter is the difference between the weight of the cup with the dried milk and the weight without milk. The percentage of dry matter is calculated based on the measurement data with a simple rule of thumb. This way of determining the dry matter looks very simple, but it requires a lot of experience, as well as a quick and skillful measurement because the error in the incorrect measurement brings a five times greater error in the result. Due to accelerated drying, and in order to avoid the formation of a greasy crust, it is recommended to add absolute alcohol to the container. During drying, it is necessary to maintain a constant temperature of 105 ^{0}C. This is achieved if a mixture of water and glycerin with a specific gravity of 1.143, whose boiling temperature is 105^{0}C, is poured between the walls of the drying oven, and to avoid the concentration change of the glycerin, the evaporation opening is combined with a return condenser.

2.1.6 Determination of specific gravity- density from Lactoscan

The specific gravity of milk is a number that tells how many times milk at a temperature of 15 ^{0}C is heavier than distilled water of the same volume and temperature. The specific gravity of milk varies and depends on the quantitative ratio of its components, such as mineral substances and lactose. If there is many fat in the milk with the same ratio of other ingredients, it will have a lower specific gravity and vice versa. Depending on the influence of different personal factors changes the specific gravity of milk. The dilution and collection of milk lead to a change in its specific gravity. In the first case, it decreases, in the second it increases. By determining the specific gravity, it can often be seen whether water has been added to the milk or fat has been removed.

2.2 Method for testing chemical elements in animal feed and milk

2.2.1 Sample preparation and digestion

In the laboratory, the samples were cleaned, but the samples were not subjected to any further cleaning and drying to constant weight for 48 hours at room temperature. Samples (0.5000 g) were placed in Teflon crucibles, 7 ml of pure trace HNO_3 (Merck, Germany) and 2 ml of H_2O_2 p were added. a. (Alkaloid, Macedonia), then the vessels are hermetically sealed and placed in the rotor of a microwave oven for mineralization – Marsmicrowave digestion apparatus (CEM, USA). Plant samples are boiled at a temperature of 180 ºC. After cooling, the samples were transferred to 25 ml calibration flasks.

2.2.2 Reagents and standards

Standard solutions (11355-ICP Multi Element Standard IV, Merck) of 20 elements with a concentration of 1000 mg L-1 were prepared for further dilution. All chemical reagents are analytically of high purity: HNO_3, p. a. (Merck, Germany), H_2O_2, r. a. (Merck, Germany). All vessels used were pre-cleaned by leaching 1 part HNO_3 and 3 parts HCl in the appropriate proportions for 24 hours, followed by double rinsing with distilled water.

2.2.3 Instrumentation

All analyzed elements (Al, As, B, Ba, Ca, Cd, Co, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Pb, Sr, V and Zn) were determined using ICP – AES (Varian, 715-ES) and using a CETAC ultrasonic nebulizer (ICP/U-5000AT+) for better sensitivity. The optimal instrumental parameters for these techniques are given by Balabanova et al., (2010) [35].

2.2.4 Quality control

Quality control was ensured using standard reference materials M2 and M3, which were prepared for the European Moth Survey (Mr Harry Harmens, 2010), [34]. There were good matches between the measured concentrations and the recommended values. The standard addition method was also applied, achieving quantitative reverse agreement of results for most elements.

2.2.5 Milk extraction

Samples of raw cow's milk from cattle that were fed with the tested food, from three farms (by random selection) from three different places on the territory of our society, were extracted with 6% trichloroacetic acid (CCl_3COOH 99%, Sigma Aldrich).

Preparation of the filtrate (whey fraction) - Place 15 ml of 6% trichloroacetic acid (TCA) in a 50 ml vial, add 5 ml of milk and stir with a glass rod until a fine suspension is obtained. Let it stand at room temperature for 5 minutes. Then the supernatant is separated by centrifugation at 7550 rpm-1 (RCF=5410g) in a Hettich Universal 320R centrifuge (Andreas Hettich GmbH – Germany) for 10 min. The resulting supernatant was filtered through Whatman No.1 filter paper.

2.2.6 Methods for determination of vitamins

2.2.6.1 Methods for the analysis of vitamin A (Retinol) and vitamin E (Tocopherol) in milk

- Extraction method 1 (Chem Elut)

Weigh 20 g of sample (to the nearest 0.01 g) in a 500 ml Erlenmeyer flask and add 1 g of ascorbic acid, 150 ml of ethanol (95%) and 40 ml of 50% aqueous potassium hydroxide (KOH). The condenser is added to the flask and placed in a water bath (approximately t = 95 ^{0}C). Hydrolysis occurs 30 minutes after starting the reaction. During hydrolysis, the sample is shaken twice. After complete hydrolysis, the sample is cooled to room temperature. Add 50 ml of distilled water. The hydrolyzate is transferred to a 500 ml volumetric flask and diluted to the mark with ethanol (50%). Transfer 10.0 ml to a Chem Elut column (of 20 ml) and wait about 10 min. The sample is eluted with 100 ml of n-hexane. Then, collect the eluate in a 500 ml flask. Evaporation (evaporation to dryness) is done with some BHT granules. The sample is dissolved in n-heptane and transferred to a volumetric flask (5.0 ml). Dilution is made to the mark with n-heptane.

- Extraction method 2 (separatory funnel)

Weigh 20 g of sample (to the nearest 0.01 g) into a 500 ml Erlenmeyer flask and add 1 g of ascorbic acid, 150 ml of ethanol (95%) and 40 ml of a 50 % solution of potassium hydroxide in water. A condenser is added to the flask and placed in a water bath (95^{0}C). Hydrolysis lasts 30 minutes from the start of the reaction. During hydrolysis, the sample flask should be shaken twice. After complete hydrolysis, the sample is cooled to room temperature and 50 ml of distillate is added on water. Then, transfer the hydrolyzate to a 500 ml volumetric flask and dilute to the mark with 50% ethanol. Transfer 20 ml to a separatory funnel and dilute with 100 ml of n-hexane. The funnel is shaken for 1 minute, cleaned, washed (hexane phase, 2 x 50 ml 1 M potassium hydroxide in ethanol (40 %) and (2 x 50 ml) with distilled water). Next, evaporation, evaporation to dryness of the hexane phase with a few granules of BHT and about 8 ml of ethanol (99 %). The sample is dissolved in n-heptane and transferred to a volumetric flask (5.0 ml). Dilution is made to the mark with n-heptane. The analyses are chromatographed on the equipment used – HPLC – Perkin Elmer, pump: series 200LC, autosampler; ISS – 200, detector LC – 135/LC -235 C DA.

2.2.6.2 Method for determination of vitamin C (Ascorbic acid) in milk

In this study, a colorimetric method with 2,4-dinitrophenylhydrazine, which is widely used to determine the level of ascorbic acid in biological fluids, was adapted to estimate the total content of vitamin C in animal feed and milk extracts. The procedure depends on the principle of oxidation of L-ascorbic acid into dehydroascorbic acid and 2,3-diketogulonic acid, followed by a reaction with 2,4-dinitrophenylhydrazine. After treatment with sulfuric acid, a colored product is formed, which absorbs at 520 nm.

Using an ascorbic acid standard at various concentrations, Lambert Beer's law was applied up to 20 mg dL^{-1} concentration range, which resulted in the following linear calibration curve equation: $y = 0.0344x + 0.0519$, $R_2 = 0.9709$. The most critical and decisive step in the accurate reading procedure is the preparation of a final and clear extract of the food and milk. The analysis of the content of vitamin C in the selected food and milk was done by spectrophotometric method. Standard curve of ascorbic acid according Burgoes, (2014).

Preparation of a basic solution of ascorbic acid (1000 μg/ml). For this purpose, 100 mg of ascorbic acid is measured in a beaker and dissolved in 50 ml of extraction solution (methanol). The solution is transferred to a volumetric flask protected from light and made up to 100 ml with extraction solution (methanol).

Prepare standard solutions of ascorbic acid (5, 10, 20, 30, 40 and 50 μg/ml) by taking aliquots of 0.25, 0.5, 1.0, 1.5, 2.0 and 2.5 ml of the basic solution of ascorbic acid and make up to 50 ml with the extraction solution. Mix 1 ml of the extraction solution with 9 ml of dissolved 2,4-dinitrophenylhydrazine (2,4-dinitrophenylhydrazine, DNP, Sigma Aldrich) and read the absorbance (reagent + sample) after 1 minute at 520 nm. Mix 1 ml of each standard solution and 9 ml of the diluted 2,4-dinitrophenylhydrazine solution and read the absorbance (absorbance of the standard solution) at 520 nm (three replicates) after 1 minute. Then, mix 1 ml of the standard solution with 9 ml of distilled water and read the absorbance of the standard at 520 nm after 1 minute. Subtract the blank (water) from the standard absorbance and then subtract the value from the reagent (empty cuvette without sample with only reagent). This is the final value (the actual absorbance) and is used to make a standard curve: the actual absorbance (A) versus concentration (μg/ml).

Procedure - With slight modifications to the conditions and reagents used in the method of Al-Ani, et al., (2007), in this labor the same method was applied [37]. For one test, pipette 1 ml in triplicate of the clear supernatant, 1 ml of water as a blank, and 1 ml of the standard solutions into capped tubes. 1 ml of 6% trichloroacetic acid (TCA) and 0.4 ml of 2,4-dinitrophenylhydrazine (DNP) are added to the same test tubes. The test tubes are closed and incubated in a water bath at 37 0C for 3 hours. The tubes are then cooled in a cold bath for 10 minutes. 2 ml of cold sulfuric acid (12 mmol) is added to the cooled solutions and the tubes are closed and vortexed. The temperature of the mixture should not exceed the ambient temperature for the spectrophotometer reading. Zero the spectrophotometer (Spectroquant Pharo 300 – Merck) at 520 nm and read the absorbance first of the standard solutions to compile a calibration curve, then read the absorbance of the samples and read the concentration values from the calibration curve.

3. Results and Discussion

3.1 Milk extraction

Raw cow's milk samples were extracted with 6% trichloroacetic acid (CCl_3COOH 99 %, Sigma Aldrich). Preparation of the filtrate (whey fraction): Place 15 ml of 6% trichloroacetic acid (TCA) in a 50 ml vial, add 5 ml of milk and stir with a glass rod until a fine suspension is obtained. Let it stand at room temperature for 5 minutes. The supernatant is then separated by centrifugation at 7550 rpm-1 (RCF=5410g) in a Hettich Universal 320R centrifuge (Andreas Hettich GmbH – Germany) for 10 minutes. The resulting supernatant filtered through Whatman No. filter paper. 1.

3.2 Chemical composition of milk

The results of the analysis chemical composition of milk from farms A, B and C: dry matter, fat, protein, milk lactose and milk density is presented in Table 1.

Table 1: The chemical composition of the milk from the three farms

Parameters (w) / (%)	**Farm A**	**Farm B**	**Farm C**
Dry matter	7.98	8.22	8.21
Fat	3.20	3.65	3.32
Proteins	3.05	3.07	3.10
Lactoza	4.38	4.49	4.46
Density	28.05	28.49	28.56
Relative Volume mass / (g/cm^3)	1,002805	1,002849	1,002856

The obtained results for dry matter are lower, so, due to the fact that the tests were done on skimmed milk and the obtained value of dry matter tested on skim milk is added to the value of fat and their sum gives the total dry matter for raw milk.

The results for dry matter are similar in the milk from the three farms from the three areas, namely: in farm B, 8.22% was measured, in farm C, 8.21%, while in farm A, 7.98 %. The concentration of milk fat is similar, in farm B high values were measured (3.65%), while in farm C it was measured (3.3%), and in farm A the values were lower (3.20 %). An approximate value also was measured for proteins in milk from the three regions. The highest value is in the milk from farm C (3.10%), then in farm B (3.07%), and the lowest in farm A (3.05%). The highest values for lactose were measured in farm B (4.49%), followed by milk from farm C (4.46%), and the lowest in milk from farm A (4.38%). For milk density, it is also noted that the highest values were measured in farm C (28.56%), followed by milk from farm B (28.49%), and the lowest in milk from farm A (28.05%). Clear value for the chemical composition of milk, it can be noted that the three parameters: dry matter (8.22 %), fat (3.65%) and lactose (4.49%) dominate in the milk from farm B, while in the milk from farm C the following parameters dominate: protein (3.10 %) and milk density (28.56%) are shown in Fig.1. The lowest values for the chemical composition of milk were measured in milk from farm A.

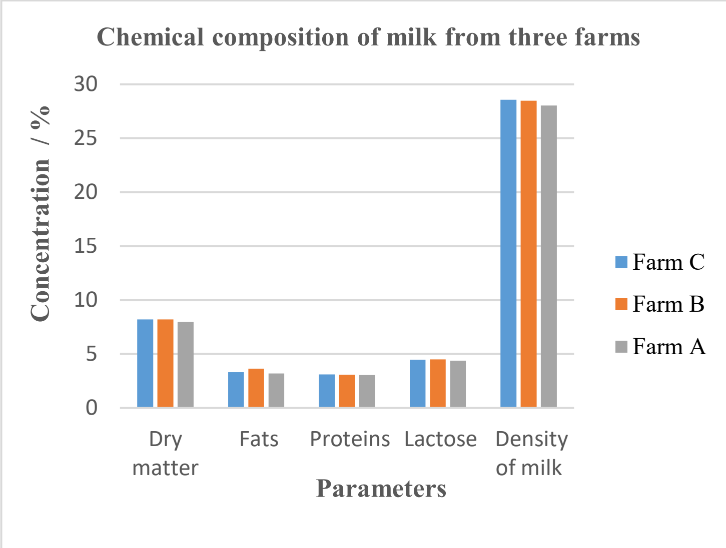

Fig. 1. The chemical composition of the milk from the three farms

3.3 Chemical elements in milk

Table 2 and Figure 2 present the measured values amount of chemical elements wi / (mg/L), present in the milk from the three farms A, B and C. The elements: As, Cd, Co, Cr, Li, Ni, Pb, V are present in less than 1 mg/L, (< 1) in milk from all three farms and were not measured.

Chemical elements w_i / (mg/L) in milk from three farms														
Milk from three Farms	**Al**	**B**	**Ba**	**Ca**	**Cu**	**Fe**	**K**	**Mg**	**Mn**	**Na**	**P**	**Sr**	**Zn**	**Total amounts**
A	**1.00**	**0.56**	**0.26**	**1122**	**2.76**	**3.82**	**1174**	**80.8**	**0.039**	**255**	**786**	**0.95**	**3.17**	**2652.2**
B	**0.72**	**0.35**	**0.10**	**697**	**2.15**	**2.08**	**595**	**74.8**	**0.028**	**489**	**504**	**0.2**	**2.27**	**2367.7**
C	**1.24**	**0.69**	**0.14**	**947**	**2.42**	**1.74**	**724**	**71.0**	**0.025**	**235**	**634**	**0.63**	**3.28**	**2621.2**
Total (A+B+C)	**2.96**	**1.6**	**0.5**	**2766**	**7.33**	**7.64**	**2493**	**226.6**	**0.092**	**979**	**1924**	**1.78**	**8.72**	**7641.1**

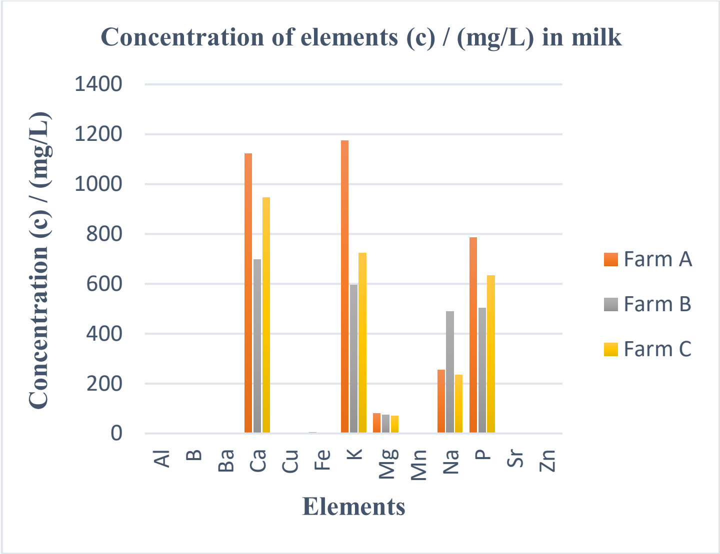

Fig. 2. Chemical elements in milk from the three farms

The percentage representation of each of the elements in 100% milk from all farms can be seen of Figure 3.

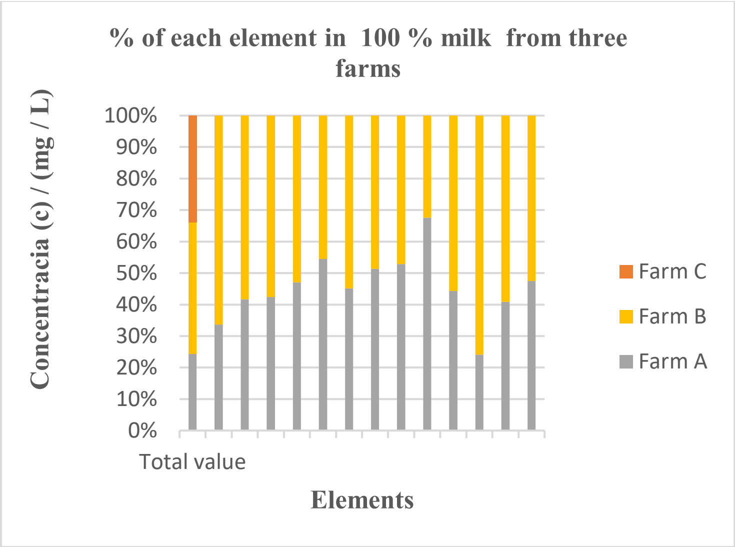

Fig. 3. The representation of chemical elements in percentage (%) present in milk from three farms

3.4 Comparison of total values of the amount of chemical elements in the animal feed and in the milk of the three farms

A comparison was made of the measured amounts of chemical elements in the total animal feed and in the milk from the three farms from the three regions, which were taken as samples for testing. The obtained values is presented in Table 3. The total daily meal of the cattle was mixed alfalfa hay and concentrate in different weights 20kg farm A, 28kg Farm B and 31kg Farm C per day.

Table 3. Amount of chemical elements in animal feed and milk

The amount of chemical elements in animal feed (mg/kg) and milk (mg/L) from the three farms													
Animal feed	**Al**	**B**	**Ba**	**Cu**	**Fe**	**Mn**	**Na**	**Sr**	**Zn**	**Ca**	**K**	**Mg**	**P**
Farm A	89.08	22.11	22.31	11.61	95.33	48.69	64.68	4.76	17.48	1903.2	9672.25	1183.5	2110.0
Farm B	858.3	11.9	20.16	24.47	837.87	106.23	3318.6	14.79	133.9	11372	10273.3	2287.3	7552.67
Farm C	94.95	6.54	13.65	2418	20.3	122	298.5	10.76	37.5	2418	11593.5	2043.0	4266.5
Milk													
Farm A	1	0.56	0.26	2.76	3.82	0.039	255	0.95	3.17	1122	1174	80.8	7.86
Farm B	0.72	0.35	0.1	2.15	2.08	0.028	489	0.2	2.27	697	595	74.8	504
Farm C	1.24	0.69	0.14	2.42	1.74	0.025	235	0.63	3.28	947	724	71	634

From the obtained values of the amount of chemical elements in animal feed and in milk, Table 3, it is observe that the amount of elements in milk is much lower than their presence in animal feed. The values for aluminum (Al), copper (Cu), iron (Fe), sodium (Na), calcium (Ca) and magnesium (Mg) are the highest in the feed of farm B, and for potassium (K) they are the highest in the feed of the farm C. The best and highest values for chemical elements in milk were observed for calcium (Ca), potassium (K) and magnesium (Mg) in milk from farm B, while the same elements have lowest and approximate values in milk from farms A and C.

Examinations of the chemical composition in milk correspond (when compared) with the latest research from Cornell University (2018) in the paper "Composition of milk": raw cow's milk contains: water 87.3% (85.5 – 88, 7 %), fat 3.9 % (2.4 – 5.5 %), protein 3.25 % (2.3 – 4.4 %), casein 2.6 % (1.7 – 3.5), serum proteins, small proteins, carbohydrates 4.6% (3.8 – 5.3%) [19]. Results also are consistent with the results of research by Douglas (2014) [20].

According to the composition, the milk from Farm B is the best, obtained dry matter 8.22%, milk fat 3.65%, proteins 3.07%, lactose 4.49%, density is 28.49%.

Man can consume milk from various mammals, but he mostly consumes cow's milk. That's why the research of nutrients was done in milk from cows, which coincides with the other research of (Penny - Penny, 1995; Samuel - Samuel, 1976) [2], [3]. The following results were obtained for presence of chemical elements:

From Farm A obtained the highest values for potassium 1174 mg/L, than calcium 1122 mg/L, phosphorus 7.86 mg/L and sodium 255 mg/L, but the lowest value for manganese is 0.039 mg/L. In Farm B has the highest values for calcium 697 mg/L, potassium 595 mg/L, phosphorus 504 mg/kg and sodium 489 mg/L, but the lowest value for manganese 0.028 mg/L. Farm C, the highest values have for calcium 947 mg/L, than potassium 724 mg/L, phosphorus 634 mg/L, sodium 235 mg/L, but the lowest value for manganese 0.025 mg/L.

From this research if compare these values with those already published by the research of Feskanich et all. (2011) are similar values [12].

3.5 Determination of vitamins A and E

The determinate concentrations of vitamins A and E in the samples of raw cow milk that were analyzed in this investigation are present in Figure 4 and 5, respectively.

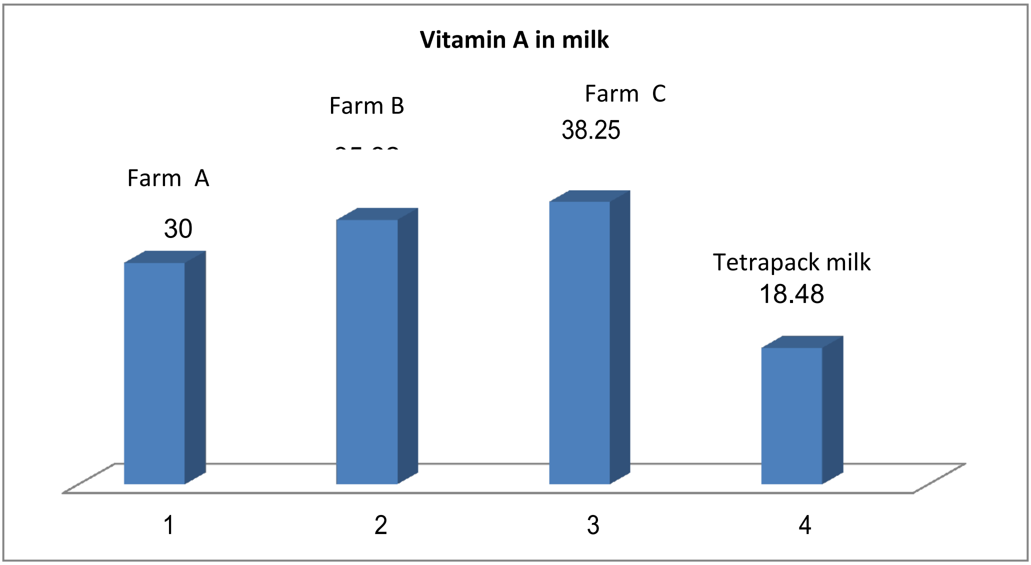

Fig. 4: Concentrations of vitamin A in raw cow milk from three farms and pasteurized milk (tetrapack)

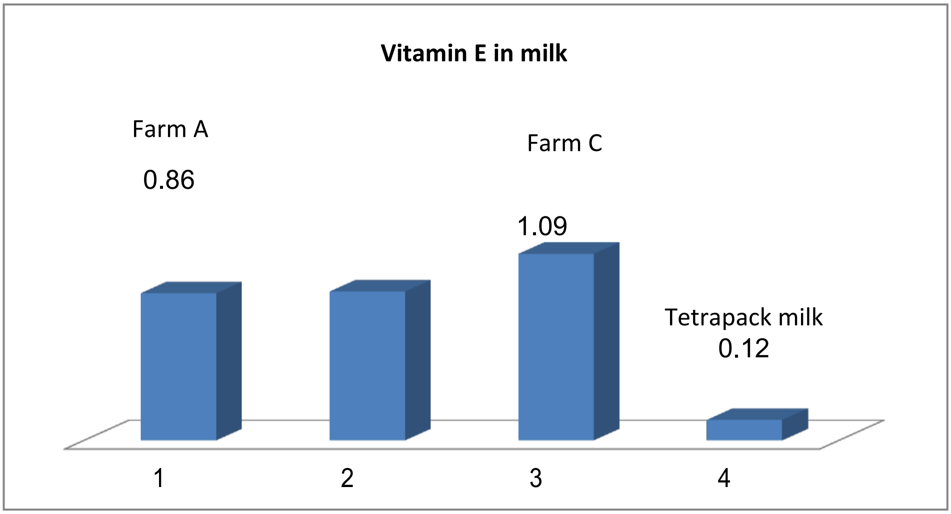

Fig. 5: Concentrations of vitamin E in raw cowmilk from three farms and pasteurized milk (tetrapack)

The milk from farm C showed the highest values of concentration of vitamin A (38,25 µg/100g), the lowest values were found in milk from the farm B (35,8 µg/100g) and the lowest values were from farm A (30 µg/ 100g). Vitamin A has shown higher values than that found in the commercial milk-tetrapack (18,48 µg/100g). Vitamin E had higher values in the milk from the farm C (1,09 µg/100g), followed by the milk from farm B (0,87 µg/ 100g) and farm A (0,86 µg во 100 g). It was also shown that the values of vitamin E were higher than that in the milk of tetrapack (0,12 µg/100g).

A comparison among the concentrations of vitamins A and E determined in the samples of cow milk was made using a statistic ANOVA test. The results are presented in Table 4 where a significant difference was noticed between the value of vitamin A in milk from farm B (35,82 µg/100g) and the value of vitamin E in milk from farm B (0,87 µg/100g). There was also a significant difference between the values of vitamin A in milk from farm C (38,25 µg/100g) and the value of vitamin E from the same farm (1,09 µg/100g).

Table 4. Statistical analysis by ANOVA test of cow milk vitamins A and E

Parameters of comparison (Vitamins A and E) in milk	t = test	Arithmetic mean– $\bar{x}$	Standard Deviation – s	Significance – P
Vit. A Farm A: Vit. A Farm B, 30,0:35,82	t = -1.006	15.000 17.910	2.828 2.956	Ns, P= 0.420
Vit. A Farm A: Vit. A Farm V, 30,0:38,25	t = -1.890	15.000 19.125	2.828 1.237	Ns, P= 0.199
Vit. A Farm B: Vit. A Farm V, 35,82:38,25	t = -0.536	17.910 19.125	2.956 1.237	Ns, P= 0.645
Vit. A Farm A: Vit. A Terapack milk, 30,0 :18,48,	t = -5.760	9.240 15.000	0.000 1.414	S, P= 0.029
Vit. A Farm B: Vit. A Tetrapak milk, 35,82:18,48	t = -96.333	9.240 17.910	0.000 0.127	S, P= < 0.001
Vit. A Farm V: Vit. A Tetrapack milk, 38,25:18,48	t = -79.080	9.240 19.125	0.000 0.177	S, P= < 0.001
Vit. A Farm A: Vit. E Tetrapack milk, 30,00:0,86	t = (+inf)	15.000 0.430	0.000 0.000	S, P= < 0.001
Vit. A Farm B: Vit. E Tetrapack milk, 35,82:0,87	t = (+inf)	17.910 0.435	0.000 0.000	S, P= < 0.001
Vit. A Farm V: Vit. E Tetrapack milk, 38,25:1,09	t = (+inf)	19.125 0.950	0.000 0.000	S, P= < 0.001

Ns – Insignificant value (a negligible value), S – Significant value

3.6 Determination of Vitamin C

Determination of the concentrations of vitamin C in the cow milk samples investigated was performed by using a standard curve of ascorbic acid. The results are present in Figure 6.

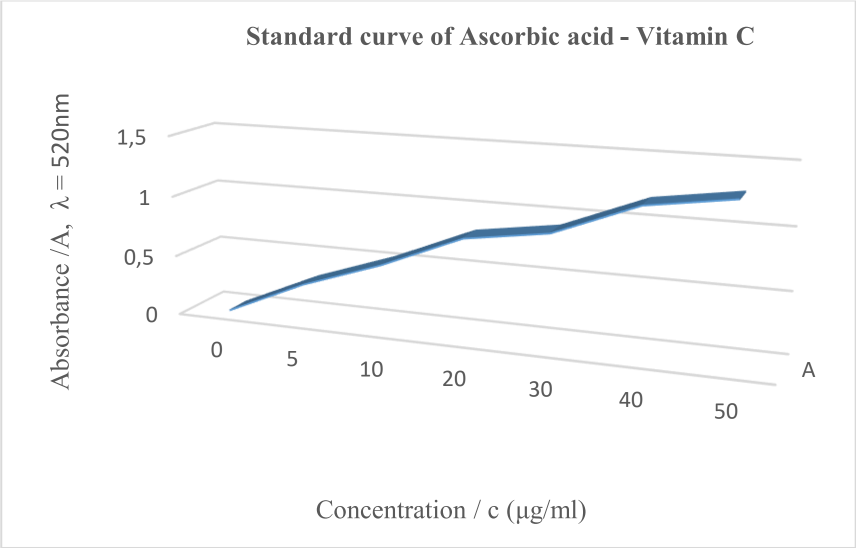

Fig.6. Standard curve of ascorbic acid

The values of the standard curve of ascorbic acid were in the measuring range between 0,00 and 40,00 μg/ml. The measured values of absorbance A for the milk sample analyses are present in Table 5. The absorbance values have transferred to a calibration curve for ascorbic acid from which the values of concentrations (μg/ml) have been read. Figure 7 presents the determination concentrations of vitamin C in milk samples from the three farms.

Table 5. Measured concentrations and absorbance for standard curve of ascorbic acid

Concentration (c) μg/ml	Absorbance (A) λ =520nm
0	0.006
5	0.30
10	0.53
30	0.91
40	1.18
50	1.28

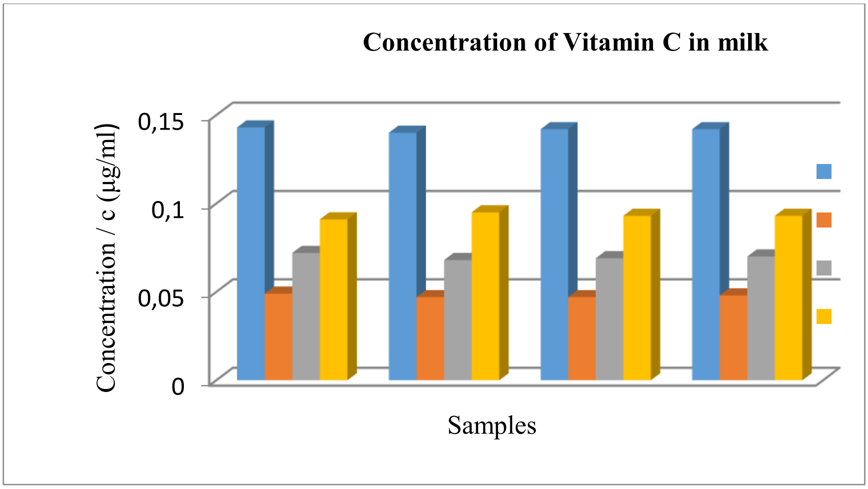

Fig.7: Concentrations of vitamin C in row caw milk from three farms

The highest value for vitamin C was determined in milk from farm A (2,8 µg/ml), while the lowest was in the milk from farm B (1,4 µg/ml). The values of the content of vitamins A and E were higher in raw milk compared to those found in pasteurized milk, while the content of vitamin C was lower. Statistical analysis of absorbance and concentrations of vitamin C in cow milk samples is present in Table 6.

Table 6: Statistical analysis of absorbance and concentrations of vitamin C in milk samples

Number of measurements	**Number of samples Absorbance (A), λ = 520nm**			
	1	2	3	4
1	0.143	0.049	0.072	0.091
2	0.142	0.047	0.068	0.095
3	0.142	0.047	0.069	0.093
n = 3				
$\bar{x}$ =	0.142	0.048	0.070	0.093

s =	0.001	0.001	0.002	0.002
RSD	0.0040	0.0242	0.0298	0.0215
(c)/ (μg/ml)	2.8	1.35	1.4	2.3

Several types of milk were studied for the presence of vitamins such as cow and sheep milk, sterilized whole milk, skim milk, milk with added vitamins and chocolate milk. It has shown that the content of vitamin C in a vitaminized milk was the highest of 1.20 mg/dL. Chocolate milk also contains a significant amount of vitamin C, up to a maximum of 0.22 mg/dL, which is probably due to the cocoa used as a basic supplement in the production of chocolate milk [36]. Skim milk has the lowest content of vitamin C, up to a maximum of 0.25 mg/dL, and a minimum of 0.1 mg/dL from Al-Ani et al., (2007), [37]. The adding of vitamin C in milk also improves milk antioxidant capacity, so the milk with supplements has the status of a functional food, providing a number of health benefits to consumers. The whole milk has a maximum content of vitamin C up to 0.35 mg/dL, which is higher than that in the chocolate milk. Skim milk, which contains 1% milk fat commonly used in the production of chocolate milk, has a small content of vitamin C. This is probably due to the degreasing processes, where along with the fat, the larger amount of vitamin C probably removed. This closely related to vitamin E, which, in turn, is the most concentrated in the membranes of the fat drops.

4. Conclusion

According to the composition of milk from three farms we concluded that Farm B has the best composition obtained dry matter 8.22%, milk fat 3.65%, proteins 3.07%, lactose 4.49%, milk density 28.49%.

About the elements, compared the total values of the content of chemical elements from farm A in each type of food, we noticed that alfalfa has the highest concentration of chemical elements per total food, namely: 200.612mg in alfalfa in 8kg, then in straw 105064.82mg in 7 kg, KMK1 is 48754.256mg in 8 kg, and least in concentrate 1, 47274.48mg in 8 kg. The total value of the content of chemical elements in four types of food used is 401705.556 mg/kg.

When comparing the total values of the content of chemical elements from farm B for each type of food, we notice that in concentrate 1 is the highest concentration of chemical elements per 8kg of total food, namely used: 383567.9mg in 10 kg, in concentrate 2 is 242031.8mg in 10 kg, and the least in alfalfa is 154922.4mg in 8kg. The total value of chemical elements in three types of food used is 780522.14 mg/kg.

When comparing the total values of the content of chemical elements from farm C for each type of food, we notice that in alfalfa there is a higher concentration of chemical elements 9153.95 mg in 10 kg, and in the concentrate

35153.5 mg in 10 kg. The total value of chemical elements in 2 types of food is 44307.45 mg/10 kg.

From the comparison of the values of the content of chemical elements between the different types of food in relation to the used daily meal (20, 28 and 31 kg) from the three farms, it is noted that 20 kg is used daily on farm A, on farm B 28 kg and on farm C 31 kg per day. It is concluded that the richest food with elements is the food from farm B. Regarding the average value of the content of chemical elements in milk, it can be concluded that the highest values were measured in farm A, 263.85 mg/L, and the lowest value was measured in farm B of 183.13 mg/L.

From the analysis of 21 chemical elements in animal feed, it is concluded that the highest value for potassium of 17205 mg/kg was measured in farm B, in alfalfa, and the lowest value is strontium of 0.77 mg/kg in KMK 1 from farm A.

The total percentage representation of the most abundant chemical elements in the total food used on the three farms is: 63% potassium, 14% phosphorus, 13% calcium and 8% magnesium. From the analysis of 21 chemical elements in the milk, it can be seen that the highest value is potassium with 1174 mg/L in the milk from farm A, and the lowest value is barium at 0.10 mg/L.

The investigation on the presence of vitamins A, E and C in cow milk originating from three farms has shown that results was a significant difference in the concentrations of particulate analyses of vitamins in the milk samples analyses. It was also determined that the concentrations of vitamins were higher in the milk samples which were taken immediately after the milking, than in those that were determined in the homogenized and pasteurized commercial milk - tetrapack. It can be also concluded that the composition of the animal feed: concentrates, alfalfa and straw in the feed samples from three different farms, as well as the cow breed, might have an influence on the concentrations of the vitamins in the cow milk investigated, and consequently on the milk antioxidant activity.

The world's milk production aims to produce more milk with better quality, including more essential nutrients having a high antioxidant activity that is important in daily human food and animal feed for the achievement of human and animal wellbeing. The feed intended for dairy cows should have excellent quality and safety, and to contain the necessary all nutrients in quantity and ratio, in order they be easily digested and used by the animal organism.

5. References

1. Nutritive Value of commonly available feeds and fooders in India, Animal Nutrition Group, Nationaly Dairy Development Board Annand - 3888-001, (July 2012). Nutritive value of commonly available feeds and fodders in India. Animal nutrition group (2012). http://www.nddb.org/sites/default/files/pdfs/Animal-Nutrition-booklet.pdf

2. Samuel, Radbill, X. (1976). The Role of Animals in Infant Feeding. In Hand, Wayland D. *American Folk Medicine: A Symposium*. University of California Press. ISBN 9780520040939.

3. Penny, Van E. (1995). The Politics of Breastfeeding". In Stuart-Macadam, Patricia; Dettwyler, Katherine Ann. Breastfeeding: Biocultural Perspectives. Aldine. ISBN 9780202011929

4. Fox, P. F. (1995). Advanced Dairy Chemistry, Vol. 3: Lactose, Water, Salts and Vitamins. 2nd ed. Chapman and Hall: New York, 1995.

5. Sarwar, M., M. A. Khan and Z. Iqbal. (2002b). Feed Resources for Livestock in Pakistan. Status Paper. *Intl. J. Agri. Biol., 4*, pp. 186-192.

6. Trajkovski, T., Bunevski, Gj. (2007). Horse breeding (book), 1-271, Skopje.https://journals.ukim.mk/index.php/jafes/article/view/1008

7. Kellems, O. R. & Church, D.C. (2010). Livestock Feeds and Feeding 6thedition University of Pennsylvania Extension. „ Monitoring Dairy Heifer Growth‖. Holstein Foundation Workbook: World of Dairy Cattle Nutrition.https://www.ag.k-state.edu/ag-experiences/ffa/DairyCattleNutrition.pdf

8. Barba, D., Margarit, G. L., Toma, R., Constantinescu-Groposila, D. (2017) The Repeatability Study of Characters for Development, Reproduction and Milk Production at the Active Romanian Black Spooted Population from Pantelimon and Mogosoaia Farms. Scientific Papers. Series D. *Animal Science, vol. LX*, pp. 11-13.

9. Caisin L, Vasile H., & Vasile V. (2012). Chemical Composition and Nutritional Value of the Fooder Grown in the Conditions of the Republic of Moldava, Scientific papers, series D. Animal Science, vol. LVI.

10. Scientific Papers. Series D. Animal Science, Vol. LX, 2017. https://animalsciencejournal.usamv.ro/pdf/2023/issue_2/vol2023_2.pdf

11. Jensen, Robert G. (1995). Handbook of milk composition. San Diego, CA: Academic Press, Inc. Pages: 54, 55, 82, 83.

12. Feskanich, D., Willett, W. C., Stampfer, M . J., Colditz, G. A. (1997). Milk, dietary calcium, and bone fractures in women: a 12-year prospective study. *Am J Public Health., 87 (6)*, pp. 992-7.

13. Jones, C. M., and A. J. Heinrichs. (2017). Milk components: Understanding milk fat and protein variation in dairy herd, Description of variation found in production of milk components, factors that contribute to this variation, and strategies to improve component production. PennState Extension, 8/17/2018. https://extension.psu.edu/milk-components-understanding-milk-fat-and-protein- variation-in-your-dairy-herd

14. Bailey, K. E., C. M. Jones, and A. J. Heinrichs. (2005). Economic returns to Holstein and Jersey herds under multiple component pricing. *J. Dairy Sci., 88*, pp. 2269-2280.

15. FAO. (2011). Selected indicators of Food and Agriculture Development in the Asia – Pacific Region 2000-2010. Food and Agriculture Organization of the United Nations, Regional Office for Asia and the Pacific, Bangkok.

16. Food and Veterinary Agency, RM. (2016). Labeling of milk and milk products - OPM, Consumers' Organization of Macedonia, OPM. https://fva.gov.mk/en

17. Rulebook on sampling methods and methods for physical, chemical and microbiological analyzes of KMK 2. "Official Gazette of SFRJ", no. 15 p. 423–427, 1987. http://www.slvesnik.com.mk/ http://www.slvesnik.com.mk/Issues/66b0357f6d894050b73b842cce2770bc.pdf

18. Stajic, D. (2012). За обезбедување здраво млеко на пазарот, контролите мора да почнат од сточната храна, *Академик*. http://www.akademik.mk/za-obezbeduvane-zdravo-mleko-na-pazarot-kontrolite-mora-da-pochnat-od-stochnata- hrana

19. Cornell, C., Cornell Universitety. (2018). College of Agriculture and Life Sciences, Dairy Extension, Food Safety Laboratory and milk quality Improvement program, Composition of milk pp.1-5, 2012. https://foodsafety.foodscience.cornell.edu/mqip/ (https://dairyextension.foodscience.cornell.edu/sites/dairyextension.foodscience.cornell.edu/files/shared/Composition%20of%20Milk.pdf).

20. Douglas D., (2014). The basis of Structure in Dairy – Based Foods: Casein Micelles and their Propperties. Book Chapter. Elsivier Inc., (2014), 83-105.

21. Sarwar, M., M. A. Shahzad, M. Nisa, S. A. Bhatti and N. A. Tauqir. (2012). Enhancing Buffalo Productivity Through Usage of Low Quality Feed Stuffs. Institute of Animal Nutrition and Feed Technology, University of Agriculture, Faisalabad, Pakistan. *The Journal of Animal and Plant Sciences, 22 (3 Suppl.)*, pp. 128-132. ISSN: 1018-7081.

22. Bender, D.A. (2003). *Nutritional biochemistry of the vitamins*. Cambridge, U.K. Cambridge University Press UK.

23. Hurley, W., L. (2009). Milk Composition and Characteristics, *Milk composition & Synthesis.* University Illinois, Urbana, Champaign.

24. Palozza, P., Calviello, G., Maggiano, N., Lanza, P., Ranelletti, F. O., & Bartoli, G. M. (2000). Beta-carotene antagonizes the effects of eicosapentaenoic acid on cell growth and lipid peroxidation in Wi Dr adenocarcinoma cells. *Free Radic. Biol. Med., 28*, 228-234.

25. World Health Organization (2005). *Vitamin and mineral requirements in human nutrition.* 2nd ed. World Health Organization. https://apps.who.int/iris/handle/10665/42716.

26. Naziroglu, M., Simsek, M., Simsek, H., Aydilek, N., Özcan, Z., & Atılgan, R. (2004). The effects of hormone replacement therapy combined with vitamins C and E on antioxidants levels and lipid profiles in postmenopausal women with type 2 diabetes. *Clin. Chem. Acta*., *344*, 63-71.

27. Ross, A. (2006a). Vitamin A and Carotenoids. In: Shils, M., Shike, M., Ross, A., Caballero, B., & Cousins R. (eds.), *Modern Nutrition in Health and Disease* (pp. 351-75). 10th ed. Baltimore, MD: Lippincott Williams & Wilkins.

28. Ross, C. A. (2010b). Vitamin A. In Coates P., M, Betz, J., M, Blackman, M., R. et al., (Eds.) *Encyclopedia of Dietary Supplements* (pp. 778-91). 2nd ed. London and New York: Informa Healthcare.

29. Harvard Health (2009). *Listing of vitamins.* Harvard Health Publishing.

30. Alberta Agriculture & Rural Development (ARD). (2012). *Nutrition and management types and sources of protein*. http://www1.agric.gov.ab.ca/$department/deptdocs.nsf/all/beef11678.

31. Surai, P., F. (2003, January). Selenium-vitamin E interactions: Does 1 + 1 equal more than 2? In *Conference paper,* Antioxidant and Lipid Research Group, Avian Science Research Centre, SAC, Scotland, UK. https://www.researchgate.net/publication/236658877_Selenium-vitamin_E_interactions_does_11_equal_more_than_2

32. Vaughan, J. G. Geissler, C. A., 2009. The new Oxford Book of food plants. Oxford University Press, UK. Animal feed resources information system.

33. EAC 2006 Second Edition 2006 East African Standard Milk. Determination of fat content (Routine method) EAS 164:2006 ICS 67.100.10. https://law.resource.org/pub/eac/ibr/eas.164.2006.pdf

34. Mr Harry Harmens (2010). HEAVY METALS IN EUROPEAN. Centre for Ecology and Hydrology Environment Centre Wales Deiniol Road Bangor Gwynedd LL57 2UW United Kingdom MOSSES: 2010 SURVEY. https://core.ac.uk/download/pdf/57335.pdf

35. Nikolovska, M., Arsovska, F. (2014). Assessment of persistence of lactation in dairy cows. *Journal of Agricultural, Food and Environmental Sciences, vol 64*, pp. 121-125. (UDC: 636.234.034(497.7)). https://journals.ukim.mk/index.php/jafes/index

36. Tomovska, J., Menkovska M., Ahmad M. A. (2018). Determination of vitamin C in different types of milk, *The International Journal of Engineering and Science (IJES*), vol. 7, issue 5, ver. 2, pp. 77-82.

37. Al-Ani, M., Linus U. Opara, L.U., Al-Bahri D., Al-Rahbi N. (2007). Spectrophotometric quantification of ascorbic acid contents of fruit and vegetables using the 2,4-dinitrophenylhydrazine method. *Journal of Food, Agriculture & Environment, vol. 5 (3&4)*, pp. 165-168.

CHAPTER 17

GLONASS Signal Availability Analysis in Global GNSS Networks

Gurkan Oztan[1] *& Sermet Ogutcu*[2] *& Salih Alcay*[3]

INTRODUCTION

The Global Navigation Satellite System (GNSS) includes multiple satellite constellations, with the Global Positioning System (GPS) serving as the foundational architecture. The GLONASS (Globalnaya Navigatsionnaya Sputnikovaya Sistema) constellation, constituting the second system to achieve global coverage, comprises 23 satellites with an orbital period of 11 hours, 15 minutes, 44 seconds (GLONASS constellation status, 2026). Contemporary GNSS applications necessitate high-precision positioning accuracy across diverse domains. The operational reliability and accuracy of positioning services are substantially enhanced through the integrated utilization of multiple satellite constellations (Teunissen, 2019; Zaminpardaz et al., 2021). The signal performance characteristics transmitted by GLONASS exhibit considerable heterogeneity attributable to several fundamental factors. The GLONASS constellation comprises multiple satellite generations (GLONASS-M, K1, K1+, K2), each possessing distinct technical capabilities (Gurtner & Estey, 2013), variable receiver hardware and firmware support across geographic networks, and incomplete constellation-wide deployment of modernized signal types (L3OC, L1OC, L2OC)(Montenbruck, 2024). The GLONASS modernization initiative represents a significant transition from legacy frequency-division architectures. Historically, the system has employed Frequency Division Multiple Access (FDMA) modulation schemes; contemporary developments have introduced Code Division Multiple Access (CDMA) modulation technologies, thereby establishing a dual-access methodology within the satellite constellation (Teunissen, 2019; Zaminpardaz et al., 2021).

The Receiver Independent Exchange Format (RINEX) standard, as the primary mechanism for archiving and exchanging GNSS observational data, constitutes an essential component of contemporary positioning and navigation infrastructures. The RINEX 3.xx format specification supports multi-frequency

[1] Dr., Nigde Omer Halisdemir University, Bor Vocational School, Department of Land Registry and Cadastre, Bor, Nigde, ORCID: 0000-0002-7629-4629
[2] Prof. Dr., Necmettin Erbakan University, Geomatics Engineering Department, Konya, Turkey, ORCID: 0000-0001-5669-7247
[3] Prof. Dr., Necmettin Erbakan University, Geomatics Engineering Department, Konya, Turkey, ORCID: 0000-0002-2680-1856

and multi-constellation observations, thereby enabling comprehensive representation of modernized GNSS signal architectures (Gurtner & Estey, 2013). Recent standardization efforts, specifically RINEX versions 4.00 and 4.01, have extended systematic support across all satellite constellations and signal type categories (Gini, 2023). Signal availability analysis refers to the assessment of the proportion of time during which a specified signal type remains successfully tracked at a receiver station throughout a defined observation interval. This metric is useful in assessing receiver hardware signal-tracking capacity and satellite constellation technical capability. The diversity of signal transmission characteristics across GNSS constellations necessitates receiver hardware capable of supporting a wide range of signal types. Such capability is particularly important in the context of the International GNSS Service Multi-GNSS Experiment (IGS-MGEX) processing initiatives (Ögütcü et al., 2023).

DATA AND METHODS

Within the RINEX specification framework, different signal components within identical frequency bands present distinct characteristics. The most significant signal components for GLONASS include C (coarse/acquisition code, FDMA modulation), P (precision code, FDMA modulation), and Q and X designations (CDMA components)(Gurtner & Estey, 2013).

RINEX Version 3.XX observational records for selected permanent GNSS stations were systematically retrieved from multiple global archive repositories representing four geographically distributed network infrastructures—AUSTRALIA (Geoscience Australia), CDDIS (NASA GSFC), EPN (EUREF Permanent Network), and TUSAGA-Aktif (Turkish National Network) (Table 1)—for the extended observation interval spanning July 19, 2025 through August 18, 2025. All retrieved RINEX data records conform to standard specification requirements: 24-hour observation duration per station-day with 30-second sampling intervals. Computational processing of RINEX observational records via custom-developed software applications enabled systematic extraction and quantification of frequency-specific signal availability metrics for participating GNSS constellation systems.

Table 1 GNSS Network

GNSS Network
AUSTRALIA (Geoscience Australia)
CDDIS (NASA GSFC)
EPN (EUREF Permanent Network)
TUSAGA-Aktif (Turkish National Network)

This investigation conducted a comprehensive signal availability analysis of GLONASS observational records obtained from four network infrastructures:

CDDIS (Crustal Dynamics Data Information System, NASA GSFC), EPN (EUREF Permanent Network), Geoscience Australia, and TUSAGA-Aktif (Turkish National Permanent GNSS Stations Network). RINEX observational data from selected IGS-MGEX station networks were processed using software developed in the C# programming framework (Ozdemir, 2025). Signal availability quantification for GLONASS satellites was performed by estimating the percentage availability of pseudorange (code) and carrier-phase observations across the signal frequencies.

RESULTS AND DISCUSSION

A comprehensive analysis of GLONASS signal availability across six distinct frequency categories (C1C, L1C, C2C, L2C, C2P, L2P) obtained from four global GNSS station networks reveals pronounced heterogeneity in signal reception and tracking characteristics (Figure 1). Primary frequency code and phase measurements (C1C and L1C) demonstrate consistently elevated availability percentages exceeding 98% across all analyzed networks, substantiating the operational maturity and reliability of fundamental GLONASS open-service signal architecture. Specifically, C1C pseudorange measurements exhibit availability ranging from 99.47% (CDDIS) to 99.99% (TUSAGA-Aktif), whereas L1C carrier phase observations demonstrate comparable performance within the 98.35%–99.99% range, confirming robust receiver signal-tracking capability across heterogeneous geographic and institutional network contexts.

Secondary frequency FDMA observations (C2C and L2C) manifest substantially attenuated availability characteristics relative to primary frequency measurements, constrained within the 73.00%–79.01% range across all analyzed networks. TUSAGA-Aktif network exhibits notably reduced secondary frequency availability (73.00% for both C2C and L2C), whereas Geoscience Australia and EPN networks demonstrate secondary frequency availability of approximately 78% across both code and phase measurements.

Precision-code secondary frequency measurements (C2P and L2P) demonstrate availability percentages within the 77.18%–78.54% range across network infrastructures. EPN network exhibits marginally elevated C2P and L2P availability (78.54% and 78.18% respectively), whereas TUSAGA-Aktif demonstrates reduced secondary frequency precision-code availability (77.18%).

The multi-frequency signal availability analysis demonstrates the critical importance of comprehensive frequency-specific performance quantification in contemporary multi-constellation GNSS applications. The availability differential between primary frequencies (>99%) and secondary frequencies (~77%) substantiates the necessity for dual-frequency capable receiver systems and appropriate processing algorithm implementations to maximize ionospheric delay mitigation and positioning solution robustness. Network operators and receiver manufacturers should prioritize comprehensive frequency-specific monitoring protocols to identify constellation-level architectural constraints and facilitate informed infrastructure modernization and receiver upgrade scheduling decisions within evolving multi-GNSS operational environments.

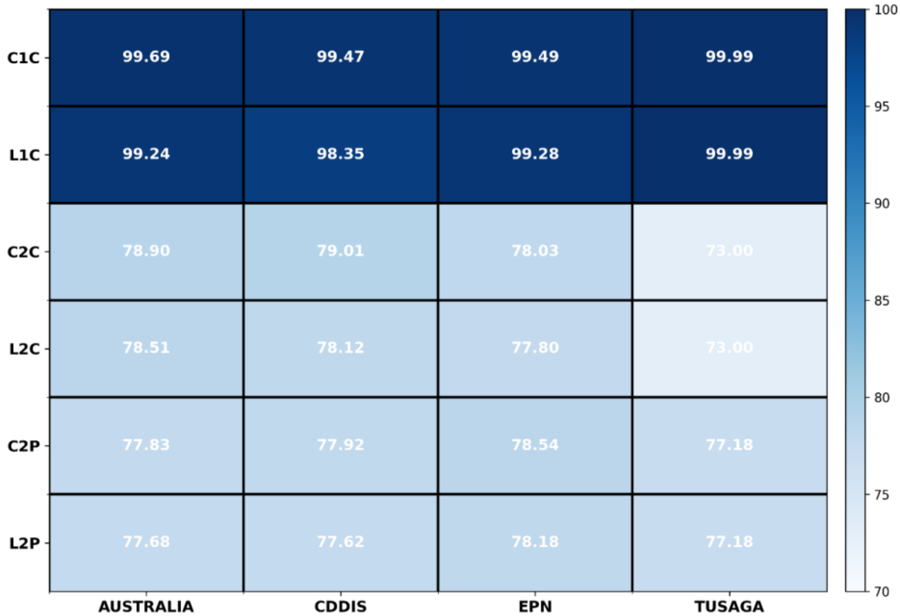

Figure 1 GLONASS Signal Availability Percentages (C1C, L1C, C2C, L2C, C2P, L2P) Across Global GNSS Networks

The comprehensive analysis across six distinct GLONASS signal frequencies (C1C, L1C, C2C, L2C, C2P, L2P) exhibited that primary frequency FDMA open-service signals (C1C and L1C) consistently exhibit availability exceeding 98% across all analyzed network infrastructures (C1C: 99.47%–99.99%, L1C: 98.35%–99.99%), substantiating the operational maturity and reliability of fundamental GLONASS constellation architecture. Conversely, secondary frequency FDMA observations (C2C/L2C) manifest substantially depressed availability constrained within the 73.00%–79.01% range across global GNSS networks, a mechanically-imposed constellation architecture limitation attributable to incomplete L2 signal transmission by first-generation GLONASS satellite architecture (satellite slots R01, R06, R10, R13, and R23). Systematic network-specific analysis of L2 signal absence frequencies reveals pronounced inter-network heterogeneity, as documented in Table 2: AUSTRALIA network exhibits the highest absence rates for R06, followed by R10, R13, R23, and R01; CDDIS and EPN demonstrate comparable degradation patterns with R06 consistently registering the highest absence frequency across both networks; whereas TUSAGA-Aktif network exhibits markedly elevated and persistent L2 signal constraints with near-universal absence across all five satellite slots throughout the observational dataset. Precision-code secondary frequency measurements (C2P and L2P) demonstrate consistent availability within the 77.18%–78.54% range across networks.

Table 2 L2 Signal Absence Analysis by Network and Satellite Slot

Network	Satellite	Zero-Value Rate (%)
AUSTRALIA	R01	89.0
	R06	98.2
	R10	95.4
	R13	91.8
	R23	91.8
CDDIS	R01	85.5
	R06	92.4
	R10	90.5
	R13	88.4
	R23	87.4
EPN	R01	87.3
	R06	93.1
	R10	91.0
	R13	90.0
	R23	88.9
TUSAGA-Aktif	R01	98.4
	R06	98.4
	R10	98.4
	R13	98.4
	R23	98.4

CONCLUSIONS

The pronounced availability differential between primary (>99%) and secondary (~77–78%) GLONASS frequency bands carries significant implications for high-precision geodetic applications, particularly Precise Point Positioning (PPP) techniques that fundamentally rely on dual-frequency observations to form ionosphere-free linear combinations for tropospheric and ionospheric delay elimination. The systematic L2 signal absence documented across satellite slots R01, R06, R10, R13, and R23—reaching absence rates of 98.4% in TUSAGA-Aktif and exceeding 89% across AUSTRALIA, CDDIS, and EPN infrastructures—directly degrades the geometric strength and convergence performance of dual-frequency PPP solutions, potentially extending convergence times and compromising centimeter-level positioning accuracy in multi-constellation operational environments. Furthermore, GLONASS FDMA inter-frequency code and phase biases inherent to the L1/L2 signal architecture introduce additional systematic error sources that must be rigorously calibrated within PPP processing algorithms, underscoring the critical necessity for receiver hardware modernization, firmware upgrade protocols prioritizing secondary frequency reception capability, and constellation-wide architectural advancement toward CDMA-modulated signal transmission to sustain the integrity of contemporary geodetic positioning infrastructures.

REFERENCES

Gini, F., (2023). RINEX, The Receiver Independent Exchange Format Version 4.01., July.

GLONASS constellation status. (2026, May 21). Retrieved from https://glonass-iac.ru/en/sostavOG/

Gurtner, W., & Estey, L., (2013). RINEX The Receiver Independent Exchange Format Version 3.02., 104.

Montenbruck, P. S. O., (2024). Characterization of the GLONASS - K1 + atomic frequency standard. *GPS Solutions*, 28(4), 1–8. https://doi.org/10.1007/s10291-024-01736-1

Öğütcü, S., Özdemir, B. N., S., A., & Buğdaycı, İ., (2023). GPS, GLONASS, Galileo ve BeiDou GNSS Sistemlerinin 1. ve 2. Temel Frekanslarının Doluluk Analizi. *Necmettin Erbakan Universitesi Fen ve Mühendislik Bilimleri Dergisi*, 5(1), 14–22. https://doi.org/10.47112/neufmbd.2023.6

Ozdemir, B. N., (2025). SOFT WARE RINEX SCAN : open-source RINEX epoch and frequency scanning software. *Earth Science Informatics*, 18(3), 290. https://doi.org/10.1007/s12145-025-01810-3

Teunissen, P. J. G., (2019). GNSS IN PROGRESS A new GLONASS FDMA model. *GPS Solutions*, 23(4), 1–19. https://doi.org/10.1007/s10291-019-0889-0

Zaminpardaz, S., Teunissen, P. J. G., & Khodabandeh, A., (2021). GLONASS – only FDMA + CDMA RTK : Performance and outlook. *GPS Solutions*, 25(3), 1–12. https://doi.org/10.1007/s10291-021-01132-z