TÜRKİYE VE DÜNYADA BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ

Editör: Doç.Dr. Bekir PARLAK

yaz
yayınları

Türkiye ve Dünyada
Bilgisayar Bilimleri ve Mühendisliği

Editör

Doç.Dr. Bekir PARLAK

yaz yayınları

2025

yaz yayınları

Türkiye ve Dünyada Bilgisayar Bilimleri ve Mühendisliği

Editör: Doç.Dr. Bekir PARLAK

E_ISBN 978-625-8678-38-3

Aralık 2025 – Afyonkarahisar

Dizgi/Mizanpaj: YAZ Yayınları

Kapak Tasarım: YAZ Yayınları

YAZ Yayınları. Yayıncı Sertifika No: 73086

M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar/AFYONKARAHİSAR

www.yazyayinlari.com

yazyayinlari@gmail.com

İÇİNDEKİLER

THE ART OF EFFICIENCY IN LARGE LANGUAGE MODELS: DOMAIN-SPECIFIC SPECIALIZATION WITH LORA, QLORA, AND BEYOND

Seda BAYAT TOKSOZ[1]

Gultekin ISIK[2]

1. INTRODUCTION

In recent years, language models have grown rapidly in size, while the computational resources required to adapt these models to specific tasks have been increasing exponentially. Traditional fine-tuning methods, which require updating all parameters of models with billions of parameters, have become unattainable for most research groups and companies. This problem has led researchers to develop more efficient methods. Parameter-efficient fine-tuning methods were born out of this need and are widely used today.

In this section, we will explore in detail the LoRA (Low-Rank Adaptation) and QLoRA (Quantized LoRA) methods, which have been developed to adapt large language models efficiently. LoRA was introduced by Hu et al. in 2021 and is able to achieve near-full fine-tuning performance by updating only a small portion of the model. QLoRA, on the other hand, is a method developed by Dettmers, Pagnoni, Holtzman, &

[1] Research Assistant, Iğdır University, Faculty of Engineering, Computer Engineering, ORCID: 0000-0002-8427-9971.

[2] Associate Professor Doctor, Iğdır University, Faculty of Engineering, Computer Engineering, ORCID: 0000-0003-3037-5586.

Zettlemoyer (2023) that adds quantization techniques to LoRA, further reducing memory usage.

In traditional fine-tuning methods, all the parameters of a pre-trained model are updated. For example, for a model with 7 billion parameters, this means that each of the 7 billion weights is adjusted for the new task. This process requires both large amounts of GPU memory and long training times. Additionally, we have to keep an exact copy of the model for each new task, which increases storage costs.

The basic idea of LoRA is that the weight updates made during fine-tuning actually take place in a low-dimensional space. In other words, while billions of parameters appear to change, these changes are actually happening in a much smaller number of independent directions. From this observation, LoRA expresses the weight update matrix as the product of two smaller matrices. This greatly reduces the number of parameters that need to be updated.

The purpose of this book chapter is to comprehensively cover the LoRA and QLoRA methods from both theoretical and practical perspectives. First, we will explain the mathematical foundations of these methods. Then, we will discuss the challenges faced in practical applications and how they can be overcome. Finally, we will consider future research directions and open problems in this field.

2. MATHEMATICAL FOUNDATIONS OF LORA

To understand the basic mathematical idea behind LoRA, let's first review the traditional fine-tuning process. In a pre-trained neural network, each layer has weight matrices. Let us denote these matrices with W_0. During fine-tuning, these matrices are updated, and the new weights are expressed as $W = W_0 +$

ΔW. Here, ΔW represents the changes learned during fine-tuning.

The basic assumption of LoRA is that the ΔW matrix has a low rank. Mathematically, this means that a matrix of size ×k ΔW can be written as the product of two matrices of much smaller size. Specifically, LoRA expresses this matrix as $\Delta W = BA$, where matrix B is of dimension d×r and matrix A is of dimension r×k, and the condition of $r << min(d, k)$ is satisfied.

The practical meaning of this decomposition is that if the original weight matrix has $d \times k = 10{,}000 \times 10{,}000 = 100$ million parameters, and we choose r=16, then the total number of parameters of matrices B and A is only $(10{,}000 \times 16) + (16 \times 10{,}000) = 320{,}000$. In other words, we will reduce the number of parameters from 100 million to 320 thousand, approximately 300 times.

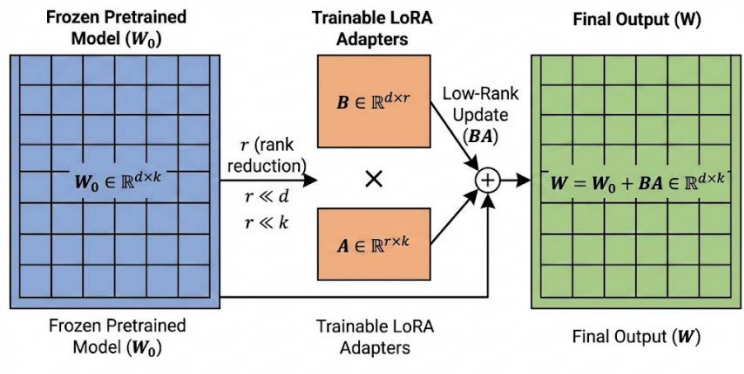


Figure 1. LoRA Architecture and Weight Update Mechanism

2.1. Why Does the Low Rank Assumption Apply?

There are several theoretical and experimental reasons why the low-order assumption is valid. First, pre-trained models have already learned strong feature representations. Changes made during fine-tuning only adjust these representations in certain directions, rather than completely reshaping them. In their

study, Li, Farkhoor, Liu, and Yosinski (2018) showed that optimization actually takes place in a very low-dimensional subspace during the training of neural networks.

Second, from a Singular Value Decomposition (SVD) perspective, singular values diminish rapidly in most real-world matrices. This means that the effective rank of the matrix is much lower than its apparent size. Experimental studies, when examining the SVD of weight updates during fine-tuning, have shown that the top 10-20 singular values explain more than 90% of the total variance.

Table 1. LoRA Parameters and Their Meanings

Parameter	Explanation	Typical Values	Memory Effect
W_0	Pre-trained weight matrix	$d \times k$	Frozen
B	Output projection matrix	$d \times r$	Trainable
A	Input projection matrix	$r \times k$	Trainable
r	Rank parameter	4, 8, 16, 32	$r\uparrow$ = Memory$\uparrow$
α	Scaling factor	16, 32	Not changing

As seen in Table 1, the most important hyperparameter to be tuned in LoRA is the r value. The larger this value, the greater the expressiveness of the model, but also the memory usage and computational cost increase. In practice, even small values such as r=4 or r=8 provide sufficient performance for many tasks.

3. QLORA: QUANTIZATION-ENHANCED LORA

QLoRA is a method developed to further improve the memory efficiency of LoRA. The basic idea is to store the pre-trained model quantized to 4-bit precision, but keep the LoRA adapters and gradients at high precision (16-bit or 32-bit). This approach makes it possible to fine-tune a model with 65 billion parameters on a single GPU with 48GB of memory.

QLoRA has three key innovations. The first is a new quantization data type called NormalFloat 4-bit (NF4). This data type is optimized for normally distributed weights and minimizes information loss. The second is the double quantization technique. In this technique, quantization constants are also quantized, thus saving additional memory. The third is paged optimizers.

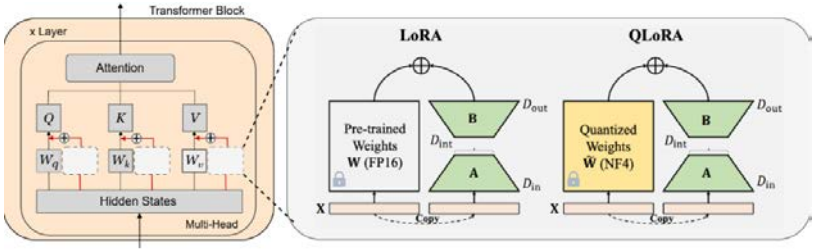


Figure 2. LoRa- QLoRa Architecture

3.1. NormalFloat 4-bit Quantization

NF4 quantization is specifically designed for data with a normal distribution. It starts from the observation that neural network weights are generally close to the zero-centered normal distribution. Instead of traditional uniform quantization, NF4 determines quantization levels using quantiles of the normal distribution. This results in higher resolution in areas where weights are concentrated.

Mathematically, NF4 quantization works like this: First, the weights are normalized, then scaled to the range $[-1, 1]$. Then, this interval is divided into 16 equally probable regions ($2^4 = 16$ for 4-bits). The center of each region becomes the quantization value that represents that region. During recycling, each quantization value is converted to the center value of its region.

3.2. Memory Saving Calculation

Let's explain the memory savings provided by QLoRA with a concrete example. Let's consider a model with 7 billion parameters. In traditional 16-bit semi-precision, this model uses $7B \times 2\,bytes = 14\,GB$ of memory. When using 4-bit quantization with QLoRA, the base model uses only $7B \times 0.5\,bytes = 3.5\,GB$ of memory. For LoRA adapters, an additional approx. 200-500 MB is required. The total memory usage is around 4 GB, which means that more than 70% is saved.

4. ADVANCED LORA VARIANTS

The success of LoRA has led researchers to further refine this method. In this section, we will explore the key variants of LoRA and the innovations each brings.

4.1. AdaLoRA: Adaptive Low-Rank Adaptation

AdaLoRA (Adaptive Budget Allocation for Parameter-Efficient Fine-Tuning) was developed by Zhang et al. (2023). In standard LoRA, all layers are assigned the same rank (r) value, while AdaLoRA assigns different rank values based on the importance of each layer. More critical layers are given high rank, and less important layers are given low rank. This allocation is dynamically updated during training.

AdaLoRA uses the magnitude of singular values as a measure of importance. During training, the singular values of the LoRA matrices in each layer are calculated. Components with small singular values are pruned, so that the parameter budget is used more efficiently. The experimental results revealed that AdaLoRA outperforms standard LoRA with the same parameter budget.

4.2. DoRA: Adaptation with Weight Separation

DoRA (Weight-Decomposed Low-Rank Adaptation) was introduced by Liu et al. (2024). The main innovation of DoRA is that it breaks down weight matrices into magnitude and direction components. LoRA is applied only to the direction component, while the magnitude component is learned separately. This decoupling improves the model's learning dynamics and ensures more stable training.

Mathematically, DoRA defines the weight matrix as $W = m \times (W_0 + BA) / ||W_0 + BA||$ expresses as follows. where m is the scalar magnitude parameter, and $W_0 + BA$ is the normalized direction vector. This formulation has given better results than standard LoRA, especially in image models and multimodal (multimodal) systems.

4.3. LoRA+: Improved Learning Rate Strategy

LoRA+ (Hayou, Ghosh, & Yu, 2024) is a method that proposes applying different learning rates to matrices A and B. In standard LoRA, both matrices are updated with the same learning rate, but LoRA+ applies higher learning rate to the B matrix than to the A matrix. This asymmetrical approach improves optimization dynamics.

Theoretical analysis shows that matrix B operates in output space and matrix A operates in input space, and these spaces have distinct learning dynamics. LoRA+ typically uses 10-100 times higher learning rate for matrix B than matrix A. This simple change significantly increases the speed of convergence, especially in large models.

5. PRACTICAL APPLICATIONS AND PERFORMANCE ANALYSIS

To understand the performance of LoRA and its variants in real-world applications, we will explore comparative studies conducted across various tasks. In this section, we will cover the implications in natural language processing, image generation, and multimodal tasks.

5.1. Natural Language Processing Tasks

In the realm of natural language processing, LoRA is most commonly used in tailoring large language models to specific tasks. For example, a general-purpose language model can be adapted to specialized domains such as medical text analysis, legal document processing, or customer service chatbot using LoRA. The study by Hu et al. (2021) demonstrated that in the GPT-3 175B model, LoRA achieved similar performance by updating only 0.01% parameters compared to full fine-tuning.

Tests on the GLUE benchmark revealed that with r=8, LoRA performed 97-99% of full fine-tuning. Especially in classification tasks (sentiment analysis, true/false detection), LoRA experiences almost no performance loss. In productive tasks (summarizing, translation), the performance loss is slightly higher, but it is usually below 5%.

Table 2. LoRA Performance in Different Tasks

Task	Full Fine-Tuning	LoRA (r=4)	LoRA (r=8)	QLoRA	Parameter %
SST-2 (Sentiment)	94.8	93.9	94.5	94.2	0.06%
MNLI (NLI)	87.6	86.1	87.2	86.9	0.12%
SQuAD (QA)	91.2	89.3	90.8	90.4	0.12%
XSum (Summarizing)	45.1	42.8	44.5	44.1	0.12%

5.2.Image Generation and Editing

LoRA is also widely used in image generation models such as Stable Diffusion. It is particularly effective in customizing the model for specific art styles, characters, or objects. A typical Stable Diffusion LoRA model is only 50-200 MB in size, while the full model takes up 4-7 GB. This compact size allows users to easily switch and combine different styles.

One advantage of LoRA in image models is that multiple LoRAs can be used simultaneously. For example, complex images can be created by combining a character LoRA, a style LoRA, and a background LoRA. By giving different weights to each LoRA, one can control which feature will be dominant and how much in the final image.

6. DOMAIN-BASED CASE STUDIES

To understand the success of LoRA and QLoRA in real-world applications, it is crucial to examine case studies in specific areas. In this section, we will discuss in detail the pioneering studies conducted in the fields of medicine and law, the data sets used, the challenges encountered and the results obtained. Both fields are characterized by their unique terminology, precision requirements, and ethical considerations, making the application of parameter-efficient fine-tuning techniques in these fields particularly interesting.

6.1. Applications of LoRA in the Medical Field

In the medical field, natural language processing is used for critical tasks such as analyzing clinical notes, diagnosing diseases, detecting drug interactions, and summarizing medical literature. One of the pioneering studies in this field is the BioBERT model developed by Lee et al. (2020). BioBERT is pre-trained on an 18 billion word corpus of PubMed and PMC

articles, resulting in significant performance improvements compared to BERT in biomedical text mining tasks.

In a recent study, the Clinical LLaMA-LoRA model was introduced (2024). This model has been fine-tuned on the MIMIC-III and MIMIC-IV datasets by adding LoRA adapters to the LLaMA base model. By updating only 0.12% of the model, the researchers achieved near-full fine-tuning performance on clinical note classification tasks. The model achieved an AUROC score of 76.07% on the mortality prediction task, a 3.37% improvement from the best clinical language model available.

The BioInstruct study by Tran, Yang, Yao, & Yu (2024) highlights the importance of instruction tuning in adapting large language models to biomedical tasks. The researchers created a dataset containing 25,005 instructions using GPT-4 and fine-tuned the LLaMA models (7B and 13B versions) using LoRA. BioInstruct has demonstrated significant success in tasks such as medical question answering, clinical trial eligibility assessment, and differential diagnosis determination.

The study by Liu et al. (2024), published in the journal JMIR Medical Informatics, demonstrates the effectiveness of LoRA in the Medical Entity Recognition task. In the study, transformer-based models such as BioBERT, RoBERTa, and DeBERTa were fine-tuned on BC5CDR and Revised JNLPBA datasets using LoRA. Notably, the Gemma LLM model achieved the highest accuracy in the BC5CDR dataset when fine-tuned with the LoRA technique. This study showed that macro factors (entity sentence length, entity word count) significantly affect model performance, and LoRA increases the model's resilience to these factors.

The Clinical ModernBERT study (Warner et al., 2024) presents a novel approach to processing medical texts with support for a long context window. The model is trained on a

wide-ranging dataset, including PubMed abstracts, MIMIC-IV clinical data, and medical codes and their textual descriptions. Built on the ModernBERT architecture, it supports a context length of 8,192 tokens and employs modern techniques such as rotary positional coding (RoPE), GeGLU activation functions, and Flash Attention. Clinical ModernBERT outperformed BioClinicalBERT in both standard biomedical NLP benchmarks and clinical tasks requiring long context.

6.2. Applications of LoRA in the Legal Field

In the legal field, the use of natural language processing technologies is increasing in tasks such as contract analysis, legal document classification, case outcome prediction, and legal information extraction. Developed by Chalkidis et al. (2020), LegalBERT is one of the first large-scale language models dedicated to the legal field. The model was trained on a 12GB corpus of diverse legal texts, significantly outperforming the general-purpose BERT model in legal text classification tasks.

The InLegalBERT study presented by Paul et al. (2023) at ICAIL 2023 developed a customized language model for the Indian legal system. This study emphasizes the importance of field-specific pre-training and the specific terminology and structures of different legal systems. InLegalBERT is built on LegalBERT and fine-tuned using LoRA for 300,000 steps on Indian legal texts. The model has been tested on both Indian and non-Indian datasets in the tasks of statutory judgment identification, semantic segmentation of court judgment documents, and appellate judgment prediction.

Bernsohn et al.'s (2024) study in LegalLens leverages large language models for the detection of legal violations in unstructured text. In this study, various LLMs (BERT, RoBERTa, DeBERTa) were adapted for legal text classification and entity recognition tasks using the LoRA technique. By combining

techniques such as instruction adjustment, reinforcement learning from human feedback, and in-context learning, researchers have developed models that specialize in understanding legal contexts. The LegalLens shared task was a significant event, with 87 participants organized into 38 teams, demonstrating community interest in the problem of legal violation detection.

The LegalPro-BERT study (2024) is a system built on the BERT-large model for the classification of statutory provisions into predefined taxonomic categories. In this study, only certain layers of the model were fine-tuned using LoRA, thereby reducing the overall training time and improving the prediction performance. In tests conducted on the LexGLUE LEDGAR benchmark, LegalPro-BERT has shown significant performance improvements compared to standard BERT. The researchers have examined the effects of data preprocessing techniques on model accuracy and highlighted the importance of preprocessing strategies specific to legal texts.

In the "Generalization to Specialization" report published by Infosys in 2024, the practical applications of using LoRA in the legal field are detailed. The report addresses the challenges faced by general-purpose LLMs in the legal domain: archaic language usage, Latin terms, and context-dependent interpretations. The study evaluated the performance of models like LLaMA, Falcon, and Flan-T5 in legal tasks, demonstrating that domain adaptation using LoRA yielded significant improvements in contract analysis, legal research, and compliance check tasks. In particular, it has been reported that with the r=16 rank value and selective updating of the target modules (query, key, value, output projection layers), memory savings of up to 70% are achieved and performance loss is minimal.

6.3. Domain-Based Comparative Analysis

The comparative analysis of LoRA applications in the medical and legal fields reveals the unique characteristics and shared challenges of both fields. Both fields require high precision, involve critical decisions with low fault tolerance, and use domain-specific terminology. However, there are significant differences in terms of data accessibility, privacy requirements, and model evaluation criteria.

In the medical field, regulations such as HIPAA (Health Insurance Portability and Accountability Act) and GDPR (General Data Protection Regulation) strictly control the use of patient data. Therefore, training medical LoRA models is often done on anonymized datasets. While publicly available sources like MIMIC datasets enable researchers to develop models, deployment in real clinical settings requires additional security and privacy measures. In the legal field, court decisions and legal documents are generally public, but client-lawyer confidentiality and information about sensitive cases require special protection.

In terms of performance evaluation, precision and recall metrics are critical in the medical field, as false-positive or false-negative results can directly impact patient health. In the study by Liu et al. (2024), fine-tuning BioBERT with LoRA in the medical entity recognition task demonstrated up to 92% performance on the F1 score. In the field of law, accuracy and consistency are important. LegalBERT's success in the legal document classification task was measured with an accuracy rate of 89%.

In terms of memory and computational efficiency, the advantages of LoRA in both areas are critical. In the medical field, models to be integrated into hospital information systems often have to operate with limited computational resources. The fact that Clinical LLaMA-LoRA achieves high performance with only 0.12% parameter updates is an important step in overcoming

these constraints. Similarly, systems used in law firms must be efficient when handling large collections of documents. Fine-tuning InLegalBERT with LoRA has produced a model that is light enough to run on standard hardware.

6.4. Future Perspectives for Domain-Based Applications

The future of domain-specific LoRA applications will be shaped by the development of multimodal systems, the adoption of continuous learning approaches, and the integration of ethical AI principles. In the medical field, the integration of medical images (X-rays, MRI, CTs) and genomic data, alongside text data, will enable the development of systems that can provide more comprehensive diagnostic and treatment recommendations.

In the field of law, one of the primary goals is to develop multilingual and multi-jurisdictional models that can process legal texts in different jurisdictional systems and languages. Additionally, the explainability and transparency of legal reasoning are critical for the right to a fair trial. The parameter efficiency provided by LoRA holds the potential to enhance the interpretability of these models.

A common area of development for both fields is the combination of federated learning and LoRA. This approach allows for model updates to be made locally, without sending sensitive data to centralized servers. Inter-hospital collaboration or consortia of law firms can carry out joint model development projects while protecting data privacy. Furthermore, dynamic LoRA rank allocation and adaptive learning rate strategies will contribute to the development of systems that can better adapt to the variable nature of domain-specific tasks.

7. CHALLENGES IN REAL-WORLD APPLICATIONS

While the theoretical advantages of LoRA and its variants are clear, several challenges arise in real-world applications. In this section, we will discuss the main problems encountered in practical applications and suggestions for solutions.

7.1. Hyperparameter Selection

In LoRA, the most critical hyperparameter is the rank (r) value. Too low an r value leads to insufficient expressiveness, and too high an r value leads to unnecessary computational cost. In practice, the following approach is recommended for determining the r-value: Start with a small r-value (e.g., 4) and gradually increase it until performance reaches a plateau. For most tasks, $r = 8$ or $r = 16$ is sufficient.

Which layers to apply LoRA to is also an important decision. Applying LoRA to all layers doesn't always work best. As a general rule, applying LoRA to attention layers is more effective than to feed-forward layers. Some studies have shown that it is sufficient to apply LoRA only to query and value projections.

7.2. Performance Optimization

The impact of LoRA on inference speed is generally negligible, but optimization becomes essential in large-scale deployments. Fusion of matrix products is an effective way to speed up LoRA calculations. Modern deep learning frameworks offer dedicated kernels for LoRA.

For memory optimization, LoRA weights can be stored at 8-bit or 4-bit precision. During inference, these weights are converted to high precision. This approach significantly reduces memory usage, especially in systems with a large number of LoRA adapters.

8. CONCLUSION AND FUTURE PERSPECTIVES

In this section, we have extensively explored LoRA and its variants, which have been developed for the efficient adaptation of large language models. QLoRA's further enhancement of this efficiency through quantization techniques has made previously unattainable models available on standard hardware.

Domain-based case studies, particularly applications in the fields of medicine and law, clearly demonstrate the practical value of LoRA. In the medical field, the success of models like Clinical LLaMA-LoRA and BioBERT holds immense potential for improving patient care and supporting clinical decision-making. In the legal field, models like LegalBERT and InLegalBERT are enhancing the efficiency of the justice system by accelerating legal research and document analysis.

Several important directions have been identified for future research. Automated sequence determination methods, multimodal model adaptation, integration of federated learning and LoRA, and the incorporation of explainable AI principles are among the areas of focus in the coming period. Furthermore, managing LoRA adapters in continuous learning scenarios and maintaining performance on older tasks while adding new ones are important problems to solve.

In conclusion, LoRA and related technologies play a central role in democratizing and making large language models accessible. Parameter-efficient fine-tuning methods not only reduce computational costs but also contribute to the development of more sustainable and environmentally friendly AI systems.

REFERENCES

Aghajanyan, A., Zettlemoyer, L., & Gupta, S. (2021). Intrinsic dimensionality explains the effectiveness of language model fine-tuning. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, 7319-7328. https://doi.org/10.18653/v1/2021.acl-long.568

Alsentzer, E., Murphy, J., Boag, W., Weng, W., Jindi, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 72-78. https://doi.org/10.18653/v1/W19-1909

Bernsohn, D., Semo, G., Vazana, Y., Hayat, G., Hagag, B., Niklaus, J., Saha, R., & Truskovskyi, K. (2024). LegalLens: Leveraging LLMs for legal violation identification in unstructured text. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics*, 2129-2145.

Chalkidis, I., Fergadiotis, M., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). LEGAL-BERT: The muppets straight out of law school. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2898-2904. https://doi.org/10.18653/v1/2020.findings-emnlp.261

Clinical LLaMA-LoRA Team. (2024). Parameter-efficient fine-tuning of LLaMA for the clinical domain. *Proceedings of the Clinical NLP Workshop*. Association for Computational Linguistics.

Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2023). QLoRA: Efficient finetuning of quantized LLMs. *Advances in Neural Information Processing Systems, 36*, 10088-10115. https://doi.org/10.48550/arXiv.2305.14314

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C. M., Chen, W., Yi, J., Zhao, W., Wang, X., Liu, Z., Zheng, H. T., Chen, J., Liu, Y., Tang, J., Li, J., & Sun, M. (2023). Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence, 5*(3), 220-235. https://doi.org/10.1038/s42256-023-00626-4

Hayou, S., Ghosh, N., & Yu, B. (2024). LoRA+: Efficient low rank adaptation of large models. *arXiv preprint arXiv:2402.12354*. https://doi.org/10.48550/arXiv.2402.12354

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., & Gelly, S. (2019). Parameter-efficient transfer learning for NLP. *Proceedings of the 36th International Conference on Machine Learning*, 2790-2799. PMLR.

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). LoRA: Low-rank adaptation of large language models. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2106.09685

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics, 36*(4), 1234-1240. https://doi.org/10.1093/bioinformatics/btz682

Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3045-3059. https://doi.org/10.18653/v1/2021.emnlp-main.243

Li, C., Farkhoor, H., Liu, R., & Yosinski, J. (2018). Measuring the intrinsic dimension of objective landscapes. *International Conference on Learning Representations*.

Liu, S., Wang, A., Xiu, X., Zhong, M., & Wu, S. (2024). Evaluating medical entity recognition in health care: Entity model quantitative study. *JMIR Medical Informatics, 12,* e59782. https://doi.org/10.2196/59782

Liu, S., Zheng, C., Demeo, G., Li, C., Wang, L., & Yang, Y. (2024). DoRA: Weight-decomposed low-rank adaptation. *International Conference on Machine Learning*. https://doi.org/10.48550/arXiv.2402.09353

Mangrulkar, S., Gugger, S., Debut, L., Belkada, Y., Paul, S., & Bossan, B. (2022). *PEFT: State-of-the-art parameter-efficient fine-tuning methods* [GitHub repository]. https://github.com/huggingface/peft

Paul, S., Mandal, A., Goyal, P., & Ghosh, S. (2023). Pre-trained language models for the legal domain: A case study on Indian law. *Proceedings of the 19th International Conference on Artificial Intelligence and Law (ICAIL 2023)*. https://doi.org/10.1145/3594536.3595165

Tran, H., Yang, Z., Yao, Z., & Yu, H. (2024). BioInstruct: Instruction tuning of large language models for biomedical natural language processing. *Journal of the American Medical Informatics Association, 31*(9), 1821-1832. https://doi.org/10.1093/jamia/ocae122

Valipour, M., Rezagholizadeh, M., Kobyzev, I., & Ghodsi, A. (2023). DyLoRA: Parameter efficient tuning of pre-trained models using dynamic search-free low-rank adaptation. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational*

Linguistics, 3274-3287. https://doi.org/10.18653/v1/2023.eacl-main.239

Warner, J., et al. (2024). Clinical ModernBERT: An efficient and long context encoder for biomedical text. *arXiv preprint arXiv:2504.03964*. https://doi.org/10.48550/arXiv.2504.03964

Xu, Y., Xie, L., Gu, X., Chen, X., Chang, H., Zhang, H., Chen, Z., Zhang, X., & Tian, Q. (2024). QA-LoRA: Quantization-aware low-rank adaptation of large language models. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2309.14717

Zhang, Q., Chen, M., Liu, Z., Lin, F., & Zhang, S. (2023). AdaLoRA: Adaptive budget allocation for parameter-efficient fine-tuning. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2303.10512

Zheng, L., Chiang, W., Sheng, Y., Li, D., Zhuang, S., Wu, Z., Shen, Y., Lin, Q., Li, Y., Xing, E., Zhang, H., Gonzalez, J., & Stoica, I. (2021). CaseLaw-BERT: Pre-training language models with domain-specific vocabulary for legal language understanding. *arXiv preprint arXiv:2104.08671*. https://doi.org/10.48550/arXiv.2104.08671

ÇEVRİM İÇİ İZLEME SİSTEMLERİ İLE ÜRETİM SÜREÇLERİNİN DİJİTAL TAKİBİ

Emin ÖZTÜRK[1]

Aziz Kubilay OVACIKLI[2]

1. GİRİŞ

Gelişen teknoloji ile birlikte fabrikalar sadece makinelerin çalıştığı alanlar olmaktan çıkarak sürekli veri üreten ortamlar haline gelmiştir. Makinelerin çalışması sırasında ürettikleri veriler üretim süreçleri hakkında kritik bilgiler sağlamaktadır. Makinelere yerleştirilen sensörler aracılığıyla toplanan sıcaklık, titreşim, basınç, ses, devir gibi verilerin doğru şekilde bir araya getirilmesi ile üretim süreçlerinde oluşan ve erken aşamada fark edilmeyen sorunlar tespit edilebilmektedir (Soori ve ark., 2023).

Endüstri 4.0 ile birlikte gözleme dayalı üretim süreçlerinden veri temelli analiz ve bakımların yapıldığı üretim süreçlerine geçilmiştir. Son yıllarda nesnelerin interneti (IoT) alanındaki gelişmeler fabrikaların dijital bir dönüşüm süreci içerisine girmesine olanak sağlamıştır (Javaid ve ark., 2021; Witczak ve Szymoniak, 2024). Dönüşüm ile birlikte artan rekabet ortamında makinelerin anlık olarak izlenmesini sağlayan çevrim içi izleme (online monitoring) sistemleri endüstrinin en önemli unsurlarından biri haline gelmiştir.

Çevrim içi izleme, makinelerden toplanan verilerin web tabanlı arayüzlerde görselleştirilmesini sağlayan sistemlerdir

[1] Bilgisayar Mühendisi, Environics Uygulamalı Bilimler A.Ş, Namık Kemal Üniversitesi Teknopark, ORCID: 0009-0008-6107-5993.

[2] Dr. Öğr. Üyesi, İstanbul Arel Üniversitesi, Mühendislik Mimarlık Fakültesi, Yazılım Mühendisliği Bölümü, ORCID: 0000-0001-6687-7794.

(Kumar ve ark., 2018). Bu sistemler sayesinde teknisyenler, mühendisler, bakım ekipleri ve yöneticiler makineleri sürekli olarak izleyebilir, oluşan sorunları öngörebilir ve gerektiğinde hızlı bir şekilde müdahale edebilirler. Böylece makinelerin daha performanslı çalışması sağlanırken ürün kalitesinin de gerçek zamanlı takibi sağlanır.

Sistemi manuel olarak sürekli takip etmek mümkün olmadığından modern izleme platformları belirlenen eşik değerin aşılması durumunda otomatik e-posta göndererek kullanıcıları bilgilendirmektedir. Makineler için eşik değerler kullanıcılar tarafından makineye özgü tamamlanabileceği gibi geçmiş veriler analiz edilerek sistem tarafından otomatik olarak da tanımlanabilir. Çevrim içi izleme sistemlerinde önemli bir yeri olan "auto threshold" (otomatik eşik değer belirleme), geçmiş verilerin analiz edilmesi ile eşik değerlerin otomatik olarak belirlenmesini sağlamaktadır (Burr ve ark., 2013).

Bu bölümde, endüstriyel sensörlerden veri toplama aşamaları, çevrim içi izleme sistemleri ve alarm yönetimi konuları ele alınacaktır. Çalışmanın amacı, çevrim içi izleme sistemlerinin işletmelere sağladığı avantajları açıklamaktır.

2. ÜRETİM SÜREÇLERİNİN DİJİTAL TAKİBİ

İşletmelerde ve üretim süreçlerinde yaşanan dijital dönüşüm ile birlikte endüstriyel veri önemli bir unsur haline gelmiştir. Öyle ki makinelerden toplanan verinin işlenmesi işletmede bulunan makine sayısı ve kapasitesinden daha önemli bir hale gelmiştir. Makinelerden toplanan sıcaklık, titreşim, ses, basınç gibi değerlerin analizi ve düzenli takibi sayesinde üretim süreçleri çok daha etkin bir şekilde yönetilmektedir.

Üretim süreçlerinin dijitalleşmesi yalnızca makinelerden veri toplamayı değil (Sağbaş ve Gülseren, 2019), aynı zamanda

bu verilerin doğru yöntemler ile bir araya getirilmesini, görselleştirilmesini ve analiz edilmesini de kapsamaktadır. Toplanmak istenilen veriye göre makinelerin üzerine farklı türde sensörler yerleştirilir (Çelebi ve Koda, 2021). Her bir sensör bulunduğu konumundan sorumlu olup kendi ölçüm noktasına ilişkin veriler toplar. Bu nedenle farklı konumlardan toplanan verilerin birlikte değerlendirilmesi gerekmektedir. (Çengiz ve Daş, 2022). Bu değerlendirmeyi yapmanın en iyi yolu çevrim içi izleme sistemleridir. Sistem sayesinde veriler bir araya getirilerek makinenin bir bütün olarak değerlendirilmesi sağlanır. Böylece verinin zamana göre değişimi, arızaların tespitleri ve belirlenen sınırların aşılıp aşılmadığı net bir şekilde takip edilebilir. Çevrim içi izleme sistemleri, kullanıcıların sahada bulunmadıkları durumlarda dahi üretim süreçlerini anlık olarak kontrol edebilmelerine olanak sağlamaktadır.

Çevrim içi izleme sistemleri, hem bir gözlem aracı hem de karar destek mekanizmasıdır. Toplanan verilerin geçmişe dönük olarak analiz edilmesi sayesine tekrar eden sorunlar ve olası kök nedenleri tespit edilebilir. Bu sayede işletmeler sorunlar oluşmadan önce önlem alabilirler. Gerekli önlemlerin alınması sonucunda makine ömrü uzatılabilir, enerji tüketimi azaltılabilir, üretim ve üretim kalitesinde önemli iyileştirmeler sağlanabilir.

3. VERİ TOPLAMA

Çevrim içi izleme sistemlerinde en önemli unsur veridir. Sensörler aracılığıyla sürekli olarak veri toplanarak sistemin mevcut durumu takip edilir. Toplanacak verinin türüne göre sıcaklık, titreşim ve akustik sensörler gibi farklı türde sensörler kullanılabilir. Elde edilen veriler kablolu veya kablosuz bağlantılar aracılığı ile sensörlerden alınarak SCADA sistemine gönderilir. SCADA, (Supervisory Control and Data Acquisition) verilerin toplandığı ve depolandığı kontrol ve izleme sistemidir

(İbrahim, 2022). Burada sensörlerden gelen veriler ayrı ayrı görüntülenebilir ancak verilerin bir bütün olarak değerlendirilebilmesi için çevrim içi izleme sistemlerine aktarılması gerekmektedir. Geliştirilen ara yazılım sayesine toplanan veriler düzenli aralıklarla çevrim içi izleme sistemlerine aktarılır. Şekil 1'de sensörlerden toplanan verilerin çevrim içi izleme sistemlerine gönderilme süreci gösterilmektedir.

Şekil 1. Veri toplama ve gönderme aşamaları

4. VERİ GÖRSELLEŞTİRME VE ÇEVRİM İÇİ İZLEME

Makinelerden toplanan sıcaklık, titreşim, basınç, devir gibi değerler tek başlarına incelendiğinde karmaşık ve anlamsızdır. Toplanan verilerin anlamlı hale gelmesi, bu verilerin görselleştirilmesi ile mümkün olmaktadır. Görselleştirme işlemi verilerin grafikler ve tablolar gibi çeşitli yöntemlerle ifade edilmesini sağlar. Bu sayede kullanıcılar üretim süreçlerini etkin bir şekilde yorumlayabilir.

Veri görselleştirme işleminde ilk adım verinin işlenmesidir. Sıcaklık, devir gibi değerler sayısal olduğu için herhangi bir ön işlem yapılmasına gerek yoktur. Toplandıkları hali ile anlaşılır verilerdir. Fakat titreşim verisi gibi değerler çok sayıda sayısal değerden oluşan dizilerdir. Toplandıkları hali ile gürültülü ve dolaysıyla karmaşık bir yapıya sahiptirler ve yorumlanmaları oldukça zordur. Bu yüzden görselleştirme işleminden önce elde edilmek istenilen sonuçlara göre farklı

teknikler ile analiz edilir (Öztürk ve Ovacıklı, 2025). Görselleştirmeye hazır hale getirilen veriler çevrim içi izleme sistemleri aracılığıyla kullanıcılara sunulur. Görselleştirme sayesinde makinelerin durumu, performansı, oluşan sorunlar kolayca takip edilebilir.

Çevrim içi izleme, kullanıcılara üretim süreçlerinin dijital bir kopyasını gösterir. Kullanıcılar sahaya inmeden bilgisayar veya mobil cihazlarından üretim süreçlerinin tamamını takip edebilirler. Bu sayede oluşan sorunlara hızlı bir şekilde müdahale edebilir, üretim duruşlarını ve bakım maliyetlerini en aza indirebilirler.

Verilerin görselleştirilmesi sadece durum takibi yapmaya değil aynı zamanda geçmiş verilerin analiz edilmesine de yardımcı olmaktadır. Yapılan analiz ile tekrar eden arızalar tespit edilebilir, kök nedenler araştırılabilir, bakım planlamalarının daha etkili bir hale getirilmesi sağlanabilir.

Çevrim içi izleme sistemlerinde eşik değer tanımlama kritik bir öneme sahiptir. ISO standartları tarafından belirlenmiş veya kullanıcı tarafından makineye özgü olarak belirlenen bu değerler aşıldığında sistem bir alarm üretir ve e-posta yoluyla kullanıcıyı bilgilendirir. Bu sayede kullanıcı sahada veya sistemin başında olmasa bile oluşan sorunlar hemen fark edilebilir. İşletmelerde üretimin sabit olmaması üretim koşullarının da değişken olmasına neden olmaktadır. Değişken üretim koşulları için sabit eşik değer belirlemek gereksiz alarm üretilmesine neden olmaktadır. Bu durumu önlemek için çevrim içi izleme sistemlerinde otomatik eşik belirleme (auto threshold) özelliği bulunmaktadır.

5. OTOMATİK EŞİK BELİRLEME

Üretim süreçlerinde makinelerin çalışma koşulları ortam sıcaklığı, kullanılan ham madde değişikliği, üretim hızı gibi pek çok etkene bağlı olarak değişmektedir. Bu durum sensörler aracılığıyla toplanan verilere de yansımaktadır. Sistemde kullanılan eşik değer çalışma koşullarında oluşan değişikliğe kendini uyarlayamadığı zaman gereksiz ve yanlış alarm üretilmesine neden olacaktır. Gereksiz alarmları önlemek ve kullanıcılara daha doğru bilgi vermek için otomatik eşik belirleme yöntemi kullanılmaktadır.

Otomatik eşik belirlemede amaç geçmiş verilerin analiz edilmesi ile eşik değerlerin üretim koşullarına uygun olarak değişmesini sağlamaktır. Bu işlem için en sık kullanılan yöntemlerden biri ortalamaya ve standart sapmaya dayalı eşik değer belirleme yöntemidir. Hesaplama sayesinde verinin genel davranışı belirlenerek normalin dışındaki durumlar tespit edilir.

İlk olarak (1) numaralı denklem kullanılarak verinin ortalaması ($\underline{x}$) hesaplanır.

$$\underline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad (1)$$

Burada n toplam veri sayısını ifade etmektedir. Ardından (2) numaralı denklem ile verinin standart sapması (σ) hesaplanır.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \underline{x})^2}{n-1}} \quad (2)$$

Elde edilen sonuçlar kullanılarak (3) numaralı denklem ile eşik değer hesaplaması yapılır. Formülde yer alan k sistem tarafından belirlenen katsayısı ifade etmektedir. Bu katsayı eşik değerin, ortalamanın ne kadar üzerinde olacağını belirlemek için kullanılır.

$$Eşik\ değer = \underline{x} + k * \sigma \quad (3)$$

Bu sayede sistem değişen koşullara uygun eşik değere sahip olur. Örneğin bir makinenin çalışma hızı düzenli olarak değişiyorsa sistem bu değişimleri zaman içerisinde öğrenir ve yalnızca anormal durumlarda oluşan artışlar için alarm üretir.

Otomatik eşik belirleme, özellikle yüksek veri üretim kapasitesine sahip işletmelerde önemli faydalar sağlar. Eşik değerin çalışanlar tarafından belirlenmesi hem zaman alıcı hem de hataya açık bir süreçtir. Otomatik eşik belirleme insana olan bağlılığı azaltarak veriye dayalı kararlar verebilen ve öğrenen bir yapı sağlamaktadır.

6. SONUÇ

Endüstri 4.0 ile birlikte işletmelerin dijital dönüşümü de hız kazanmıştır. Dijitalleşme ile birlikte makinelerden toplanan verilerin analizi ve görselleştirilmesi işletmeler için temel bir ihtiyaç haline gelmiştir. Çevrim içi izleme sistemleri oluşan ihtiyaçları karşılayarak üretim süreçlerinin etkin bir şekilde yönetilmesini mümkün hale getirmektedir.

Çevrim içi izleme sistemleri web tabanlı arayüzü sayesinde makinelerin anlık durumunu kullanıcılara göstererek kontrollü bir üretim ortamı sağlar. Üretim süreçlerinde yaşanan sorunlar ve performans düşüşleri tespit edilerek hızlı bir şekilde müdahale edilmesi sağlanır. Bu sayede üretim verimliliği artarken plansız duruşlar azalır. Üretim süreçlerinin sürekli olarak izlenmesi işletmelerin kalite standartlarını korumasına ve müşteri memnuniyetinin artmasına katkı sağlar.

Sistem, içerisinde yer alan otomatik eşik değer belirleme ve alarm üretme özellikleri sayesinde durum izleme aracı olmaktan fazlasını sağlar. Otomatik eşik belirleme geçmiş verileri analiz ederek çalışma koşullarına uygun tehlike seviyelerinin

belirlenmesini sağlarken, alarm sistemi kullanıcıların sistem hakkında otomatik olarak haber almasını sağlamaktadır.

Bir bütün olarak değerlendirildiğinde çevrim içi izleme sistemleri işletmelerde veriye dayalı karar alma kültürünün oluşmasında büyük öneme sahiptir. Bu sistemler yalnızca mevcut durumun izlenmesi sağlamakla kalmaz aynı zamanda geçmiş verileri inceleyerek gelecekte oluşabilecek arızalar hakkında bilgi edinme imkanı da sağlamaktadır.

KAYNAKÇA

Burr, T., Hamada, M.S., Howell, J., Skurikhin, M., Ticknor, L, Weaver, B. (2013). Estimating Alarm Thresholds for Process Monitoring Data under Different Assumptions about the Data Generating Mechanism. *Science and Technology of Nuclear Installations,* 2013(1), 1-18.

Çelebi, A., Koda, D.Y. (2021). Endüstri 4.0 Çerçevesinde Katmanlı İmalatta Sensör Uygulamaları. *Int. J. of 3D Printing Tech. Dig. Ind.,* 5(1), 85-97.

Çengiz, B., Daş, R. (2022). Veri Füzyonu: Veri Kaynakları, Mimariler, Zorluklar ve Çözüm Yaklaşımları. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi,* 34(2), 899-922.

İbrahim, M.H. (2022). SCADA Sistemi: Şehir İçi ve Şehirlerarası Yolların Aydınlatma Sisteminin Kontrolü ve Otomasyonu. *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi,* 37(3), 653-661.

Javaid, M., Haleem, A., Singh, R.P., Rab, S., Suman, R. (2021). Significance of sensors for industry 4.0: Roles, capabilities, and applications. *Sensors International,* 2(1), 1-12.

Kumar, M., Vaishya, R., Parag. (2018). Real-Time Monitoring System to Lean Manufacturing. *2nd International Conference on Materials Manufacturing and Design Engineering,* 135-140.

Öztürk, E., Ovacıklı, A.K., (2025). Endüstride Bulut Bilişim Tabanlı Kestirimci Bakım ve Denetimsiz Titreşim Analizi Uygulaması. *Int. J. of 3D Printing Tech. Dig. Ind.,* 9(3), 448-461.

Sağbaş, A., Gülseren, A. (2019). Endüstri 4.0 Perspektifinde Sanayide Dijital Dönüşüm Ve Dijital Olgunluk

Seviyesinin Değerlendirilmesi. *European Journal of Engineering and Applied Sciences,* 2(2), 1-5.

Soori, M., Arezoo, B., Dastres, R. (2023). Internet of things for smart factories in industry 4.0, a review. *Internet of Things and Cyber-Physical Systems,* 3(1), 192-204.

Witczak, D., Szymoniak, S. (2024). Review of Monitoring and Control Systems Based on Internet of Things. *MDPI,* 14(19), 1-27.

ARCHITECTURES DERIVED FROM THE TRANSFORMER MODEL

Hilal ÇELİK[1]

Ramazan KATIRCI[2]

1. INTRODUCTION

The Transformer architecture (Vaswani et al. 2017) has revolutionized artificial intelligence by replacing traditional recurrent and convolutional models with a fully attention-based approach. Unlike RNN and CNN models, which suffer from sequential processing limitations and vanishing gradient issues in long sequences (Wu et al. 2016**)**, Transformers enable full parallel computation and capture long-range dependencies more efficiently through self-attention.

Since their introduction in 2017, Transformers have not only transformed Neural Machine Translation (NMT) but have also become the foundational architecture for large-scale language models, evolving into distinct paradigms such as encoder-only models like BERT (Devlin et al. 2019) for understanding tasks and decoder-only models like GPT (Radford et al. 2018) for generation, as well as vision models like Vision Transformer (Dosovitskiy et al. 2021).

[1] Research Assistant, Sivas University of Science and Technology, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, 58000, Sivas, Turkey. ORCID: 0000-0001-5428-3411.

[2] Professor, Sivas University of Science and Technology, Faculty of Engineering and Natural Sciences, Department of Computer Engineering, 58000, Sivas, Turkey. ORCID: 0000-0003-2448-011X.

This chapter provides a structured overview of the evolution of Transformer-based architectures—including encoder-only, decoder-only and encoder–decoder models—highlighting their key innovations and domain-specific applications.

2. TRANSFORMER-BASED ARCHITECTURES

Transformer-based architectures represent the foundational design paradigm of modern deep learning models in natural language processing. Over time, it has evolved into several architectural variants—encoder-only, decoder-only and encoder–decoder frameworks—each optimized for distinct language understanding and generation tasks. As illustrated in Figure 1, these architectures form the structural basis for most contemporary models in Artificial Intelligence (AI) systems.

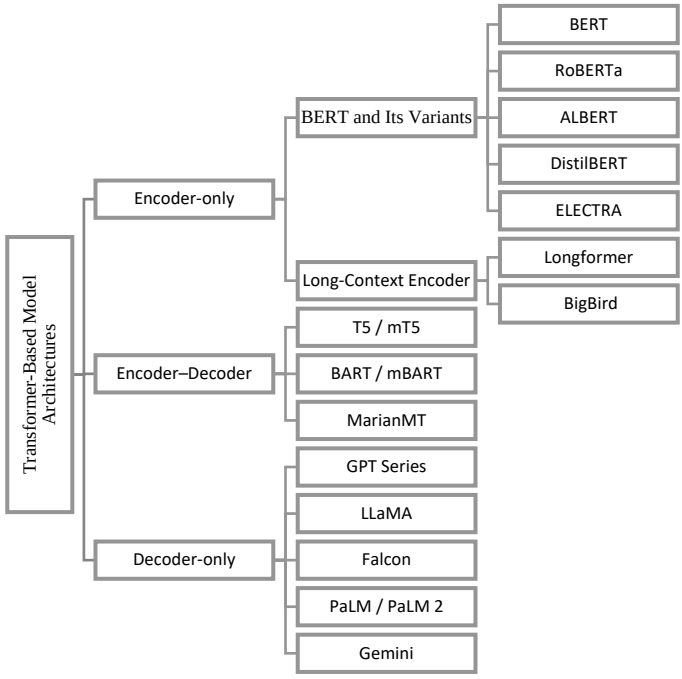


Figure 1. Transformer-Based Model Architectures

2.1. Encoder and Decoder Structure

The encoder–decoder architecture is a neural network framework designed to transform an input sequence into an output sequence by encoding the source information into a fixed or variable-length representation and then decoding it to generate the target output (Cho et al. 2014; Sutskever, Vinyals, and Le 2014; Katırcı and Çelik 2024). In this structure, the encoder processes the input data and compresses it into a contextual representation, while the decoder takes this representation and produces the corresponding output step-by-step. This architecture is particularly well suited to sequence-to-sequence learning problems—such as machine translation (Katırcı and Çelik 2025b, 2025a) and question answering (Çelik, Katırcı, and İşlek 2024; Katırcı and Çelik 2025a)—where the length and structural properties of the input and output sequences may differ.

2.2. Encoder-Based Architectures

Encoder-based architectures focus exclusively on extracting deep contextual representations from input data and are not designed for sequence generation. These models employ bidirectional attention mechanisms that enable the simultaneous processing of information from both preceding and succeeding tokens, allowing them to capture rich semantic dependencies across the entire input (Devlin et al. 2019; Liu et al. 2019; Bucci et al. 2025; Omar et al. 2025). By generating context-aware representations, encoder-based models effectively learn the underlying structure and meaning of language, making them highly suitable for tasks such as text classification, sentiment analysis, named entity recognition and question answering (Sun et al. 2020; Shon et al. 2021). Their ability to process input sequences in parallel further enhances computational efficiency, establishing encoder-only architectures—such as BERT—as the

foundation for modern language understanding systems (Lan et al. 2020; Manning 2020).

2.2.1. BERT and Its Variants

BERT and its subsequent variants represent a family of encoder-based Transformer architectures designed to enhance contextual understanding in natural language processing. Each variant introduces architectural or training optimizations—such as improved efficiency, scalability, or learning objectives—that collectively advance the performance and adaptability of pre-trained language models across diverse NLP tasks (Devlin et al. 2019; İşlek, Katırcı, and Celik 2024).

Bidirectional Encoder Representations from Transformers (BERT)

BERT introduced a bidirectional attention mechanism that considers both left and right contexts simultaneously, marking a significant advancement in language understanding (Clark et al. 2019; Devlin et al. 2019). The model is trained using Masked Language Modeling (MLM), where random tokens are masked and predicted based on surrounding words and Next Sentence Prediction (NSP) to capture sentence-level relationships (Devlin et al. 2019; Liu et al. 2019; Lan et al. 2020). BERT's deep contextualized embeddings significantly outperform previous models in downstream NLP tasks by generating representations that reflect nuanced linguistic meaning. Its encoder-only design enables efficient fine-tuning for various applications using minimal task-specific modifications (Shon et al. 2021; İşlek et al. 2024).

Robustly Optimized BERT Pretraining Approach (RoBERTa)

RoBERTa is an enhanced variant of BERT that focuses on optimizing the pretraining process to achieve superior

performance. It removes the *Next Sentence Prediction (NSP)* objective and employs *dynamic masking*, larger training corpora (such as CommonCrawl and BooksCorpus) and extended training duration. These modifications enable RoBERTa to learn more robust contextual representations and deliver consistent improvements across various natural language understanding benchmarks (Liu et al. 2019; Jurkiewicz et al. 2020).

A Lite BERT (ALBERT)

ALBERT was introduced to improve the parameter efficiency of BERT while maintaining comparable performance levels. It achieves this through *parameter sharing* across layers and *factorized embedding parameterization*, which significantly reduces model size. As a result, ALBERT offers a lightweight yet effective alternative, particularly suitable for environments with limited computational or memory resources (Lan et al. 2020).

DistilBERT

DistilBERT is a compact and faster version of BERT obtained through knowledge distillation, a teacher–student training paradigm. The student model learns to replicate the behavior of the larger BERT model while using fewer parameters. Despite being approximately 40% smaller and 60% faster, DistilBERT retains about 97% of BERT's performance, making it ideal for real-time or resource-constrained applications (Sanh et al. 2020; Fang et al. 2023).

ELECTRA

Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA) introduces a more efficient pretraining objective known as *Replaced Token Detection (RTD).* Instead of predicting masked tokens, the model learns to distinguish between real and synthetically replaced tokens in the input text. This discriminative training paradigm

allows the model to learn from every token rather than only masked positions, resulting in substantially improved training efficiency and competitive performance compared to BERT (Manning 2020; Ni et al. 2024).

2.2.2. Long-Context Encoder Models (Longformer, BigBird)

Long-context encoder models extend the Transformer architecture to efficiently process lengthy input sequences by modifying the self-attention mechanism. Approaches such as Longformer and BigBird employ sparse or attention patterns, significantly reducing computational complexity while preserving contextual understanding, making them highly effective for document-level and long-text natural language tasks.

Longformer

Longformer is a transformer-based architecture specifically designed to efficiently handle long text sequences. It introduces a *sliding-window local attention* mechanism combined with *global attention tokens*, enabling the model to capture both local and global dependencies without the prohibitive computational cost of traditional self-attention (Beltagy, Peters, and Cohan 2020; Hwang et al. 2024). This sparse attention pattern significantly reduces complexity from quadratic to linear with respect to sequence length, allowing the model to scale effectively to documents containing thousands of tokens. As a result, Longformer demonstrates strong performance in long-form document classification, summarization and analytical tasks involving scientific or legal texts. In contrast, BERT relies on a fully dense self-attention mechanism with quadratic complexity, making it computationally expensive and memory-intensive for long sequences (Devlin et al. 2019).

BigBird

BigBird is a transformer-based architecture that extends the Longformer approach by introducing a hybrid sparse attention mechanism combining sliding-window local attention, global attention and random attention patterns. This design enables the model to efficiently capture both nearby and distant contextual dependencies while significantly reducing computational complexity (Zaheer et al. 2020; Kramp, Cassani, and Emmery 2024).

2.3. Decoder-Based Architectures

Decoder-based architectures focus on autoregressive language modeling**,** where the next token is predicted based on previously generated tokens. Unlike encoder-based models that process the entire input bidirectionally, decoder-only models operate in a unidirectional manner, enabling the generation of coherent and contextually relevant text. These architectures form the foundation of modern large language models and are designed primarily for tasks involving text completion, dialogue systems, story generation and code synthesis (Radford et al. 2018). They are widely used in conversational AI agents, code assistants, content generation tools and multimodal systems (GPT-4 Technical Report 2023; Al-Amin et al. 2024). In addition, decoder-based models are central to autonomous AI agents **and** RAG (Retrieval-Augmented Generation) frameworks, where they integrate with external tools and knowledge sources to perform complex reasoning and decision-making tasks (Lewis et al. 2021).

GPT

The Generative Pre-trained Transformer (GPT) series (Radford et al. 2018), introduced by OpenAI, demonstrated that large-scale unsupervised pre-training followed by minimal fine-tuning can achieve state-of-the-art performance across a range of

language tasks. GPT models rely exclusively on stacked decoder blocks using masked self-attention to ensure autoregressive flow. This structure allows the model to generate text token-by-token, maintaining logical consistency and contextual alignment.

LLaMA

Large Language Model Meta AI (LLaMA) is a family of open foundational language models developed by Meta AI, designed to provide efficient and high-performance alternatives to proprietary large models. Introduced in 2023, LLaMA models focus on training efficiency and accessibility, achieving competitive results across benchmarks while using significantly fewer parameters than models like GPT-3(Touvron et al. 2022).

Falcon

Falcon is a high-performance, open-weight autoregressive language model developed by the Technology Innovation Institute (TII), Abu Dhabi**.** Released in 2023, Falcon emphasizes training efficiency, data quality and scalability. It was trained on the *RefinedWeb* dataset (over one trillion tokens), resulting in strong performance across common LLM benchmarks while maintaining open accessibility for research and commercial use (Dhabi 2023).

PaLM

Pathways Language Model (PaLM) is a large-scale decoder-only Transformer developed by Google Research within the Pathways project. Following an autoregressive language modeling objective, it predicts each token from preceding context and is trained on extensive multilingual and multitask data using the Pathways system. With its stacked decoder blocks and masked self-attention, PaLM exemplifies the scalability of modern decoder-based models and achieves strong reasoning,

few-shot learning, and code-generation performance (Chowdhery et al. 2023).

Gemini

Gemini is a multimodal large language model developed by Google DeepMind, built upon the decoder-only Transformer architecture of PaLM 2 (Google 2023). It extends the PaLM family by integrating text, vision and audio processing within a unified framework, enabling advanced multimodal reasoning and tool-use capabilities. Gemini incorporates cross-modal attention mechanisms inspired by models such as Flamingo, allowing it to understand and generate information across multiple modalities. This architecture supports reasoning, planning and interaction with external environments, positioning Gemini as a step toward general-purpose, agentic AI systems (Team and Google 2024, 2025).

2.4. Encoder–Decoder Architectures

Encoder–decoder architectures integrate the strengths of bidirectional contextual understanding from the encoder and autoregressive sequence generation from the decoder, resulting in highly versatile models for a wide range of natural language processing tasks (Jurafsky and Martin 2021; Raffel et al. 2023; Çelik et al. 2024). This dual structure allows the model to first encode the input into a rich contextual representation and then generate an output sequence conditioned on that representation. Such architectures are particularly effective for sequence-to-sequence transformations, including machine translation and question answering (Nielsen, Enevoldsen, and Schneider-kamp 2025)**.**

T5

The Text-to-Text Transfer Transformer (T5) architecture introduced a unified framework in which *every* language task is

reframed as a text-to-text problem(Raffel et al. 2023). This model is pre-trained using a *span corruption objective*, where random spans of text are masked and the decoder is trained to generate the missing tokens. This approach allows T5 to learn rich, task-agnostic representations (Senadeera and Ive 2022; Zha et al. 2023).

BART

Bidirectional and Auto-Regressive Transformer(BART) is a denoising autoencoder designed for pretraining sequence-to-sequence Transformer models. It combines the bidirectional encoding capability of BERT with the left-to-right generative decoding of GPT, effectively bridging language understanding and generation within a unified framework. The model is pretrained by first corrupting input text using various noising strategies—such as sentence shuffling or span infilling—and then learning a model to reconstruct the original text. This objective allows BART to learn robust contextual and generative representations (Lewis et al. 2019)**.**

MarianNMT

MarianNMT is a family of Transformer-based encoder–decoder neural machine translation (NMT) models developed by Microsoft Translator. It was designed to provide efficient multilingual translation across language pairs without relying on external resources or large pipelines. The models are trained on large-scale parallel corpora using the standard Transformer architecture and optimized for both speed and translation quality (Roman et al. 2018; Junczys-dowmunt 2019).

mBART

Multilingual Bidirectional and Auto-Regressive Transformer (mBART) is the multilingual extension of the BART model, designed to handle sequence-to-sequence generation

across multiple languages within a single Transformer architecture. It follows the denoising autoencoding pretraining objective, where text is corrupted by masking or sentence permutation and the model is trained to reconstruct the original text (Liu et al. 2020). Unlike BART, which is trained on English-only data, mBART is pretrained on large-scale monolingual corpora covering dozens of languages, allowing it to capture cross-lingual representations(Liu et al. 2020).

mT5

Multilingual Text-to-Text Transfer Transformer (mT5) is a multilingual extension of the Text-to-Text Transfer Transformer (T5) model that generalizes the text-to-text paradigm beyond English. It is pretrained on a large-scale Common Crawl–based corpus (mC4) covering 101 languages, enabling it to perform a wide range of multilingual NLP tasks within a unified framework. The model maintains the same architecture and span corruption objective as T5 but incorporates modified training strategies to enhance cross-lingual alignment and prevent accidental translation—a phenomenon in zero-shot generation where the model unintentionally switches to another language (Kale 2020).

3. FUTURE DIRECTIONS

Future Transformer developments are expected to evolve toward autonomous, reasoning-driven systems. Approaches like Retrieval-Augmented Generation (RAG) improve factual accuracy and reduce hallucinations by providing dynamic access to external knowledge (Lewis et al. 2021; Katırcı, Çelik, and Oğuz 2024; Oche et al. 2025). Meanwhile, the integration of Transformer models into multi-agent frameworks—such as LangChain and CrewAI—enables tool use, planning, and self-reflection (Derouiche 2025; Singh et al. 2025). Overall, the future

of Transformers points to hybrid architectures that blend generation, retrieval, and autonomous reasoning, bringing language models closer to intelligent agents.

4. CONCLUSION

Transformer-based architectures have redefined modern artificial intelligence by introducing a scalable, attention-driven framework capable of understanding and generating complex information across language, vision and multimodal domains. Encoder models such as BERT revolutionized language comprehension, decoder models like GPT transformed generative AI and encoder–decoder models such as T5 enabled a unified approach to diverse NLP tasks. These developments show a clear evolution from task-specific models to general-purpose AI systems with growing reasoning capabilities. As Transformers continue to advance through efficiency improvements, multimodal integration and tool-augmented intelligence, they are expected to drive the next generation of autonomous and human-aligned AI systems.

REFERENCES

Al-Amin, Md., Mohammad Shazed Ali, Abdus Salam, Arif Khan, Ashraf Ali, Ahsan Ullah, Md Nur Alam, and Shamsul Kabir Chowdhury. 2024. "History of Generative Artificial Intelligence (AI) Chatbots: Past, Present, and Future Development." http://arxiv.org/abs/2402.05122.

Beltagy, Iz, Matthew E. Peters, and Arman Cohan. 2020. "Longformer: The Long-Document Transformer."

Bucci, Giordana De Farias F. B., Juliana Felix, Rogerio Salvini, Hugo Nascimento, and Fabrizzio Soares. 2025. "Encoder-Only Transformer for Detecting Multiple Neurodegenerative Diseases from Gait Analysis." *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)* 907–12. doi:10.1109/COMPSAC65507.2025.00119.

Çelik, Hilal, Ramazan Katırcı, and Berrin İşlek. 2024. "Effect Of Parameters On Performance In Question-Answer Model With Simple Rnn Deep Learning Method." (May). https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7b0QCpsAAAAJ&citation_for_view=7b0QCpsAAAAJ:WF5omc3nYNoC.

Cho, Kyunghyun, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation." *Journal of Clinical Microbiology* 28(4):828–29. doi:10.1128/jcm.28.4.828-829.1990.

Chowdhery, Aakanksha, Emily Reif, Daphne Ippolito, David Luan, Hyeontaek Lim, and Barret Zoph. 2023. "PaLM : Scaling Language Modeling with Pathways." 24:1–113.

Clark, Kevin, Urvashi Khandelwal, Omer Levy, and Christopher D. Manning. 2019. “What Does BERT Look at? An Analysis of BERT’s Attention.”

Derouiche, Hana. 2025. “Agentic AI Frameworks : Architectures , Protocols , and Design Challenges.”

Devlin, Jacob, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. “BERT: Pre-Training of Deep Didirectional Transformers for Language Understanding.” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference* 1(Mlm):4171–86.

Dhabi, Abu. 2023. “The Falcon Series of Open Language Models.” 2(2020):1–57.

Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. “An Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale.” *ICLR 2021 - 9th International Conference on Learning Representations*.

Fang, Li, Qingyu Chen, Chih-Hsuan Wei, Zhiyong Lu, and Kai Wang. 2023. “Bioformer: An Efficient Transformer Language Model for Biomedical Text Mining.” *ArXiv*. http://www.ncbi.nlm.nih.gov/pubmed/36945685%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC10029052.

Google. 2023. “PaLM 2 Technical Report.” (May).

GPT-4 Technical Report. 2023. 4:1–100.

Hwang, Dongseong, Weiran Wang, Zhuoyuan Huo, Khe Chai Sim, and Pedro Moreno Mengibar. 2024. "TransformerFAM: Feedback Attention Is Working Memory." http://arxiv.org/abs/2404.09173.

İşlek, Berrin, Ramazan Katırcı, and Hilal Celik. 2024. "Enhancing Question Answering Systems Through Optimal Hyperparameter Tuning in GRU." *8th International Artificial Intelligence and Data Processing Symposium, IDAP 2024* (September). doi:10.1109/IDAP64064.2024.10710732.

Junczys-dowmunt, Marcin. 2019. "Microsoft Translator at WMT 2019 : Towards Large-Scale Document-Level Neural Machine Translation." 2(Day 1):225–33.

Jurafsky, Daniel, and James Martin. 2021. "Machine Translation and Encoder-Decoder Models."

Jurkiewicz, Dawid, Łukasz Borchmann, Izabela Kosmala, and Filip Graliński. 2020. "ApplicaAI at SemEval-2020 Task 11: On RoBERTa-CRF, Span CLS and Whether Self-Training Helps Them Dawid." *14th International Workshops on Semantic Evaluation, SemEval 2020 - Co-Located 28th International Conference on Computational Linguistics, COLING 2020, Proceedings* 1415–24. doi:10.18653/v1/2020.semeval-1.187.

Kale, Mihir. 2020. "MT5: A Massively Multilingual Pre-Trained Text-to-Text Transformer."

Katırcı, Ramazan, and Hilal Çelik. 2024. "Transformer Mimarisi." Pp. 0–3 in.

Katırcı, Ramazan, and Hilal Çelik. 2025a. "Analyzing The Impact Of Attention Head Count On Transformer-Based Machine Translation." (June). https://scholar.google.com/citations?view_op=view_citat

ion&hl=en&user=7b0QCpsAAAAJ&citation_for_view=7b0QCpsAAAAJ:UebtZRa9Y70C.

Katırcı, Ramazan, and Hilal Çelik. 2025b. "Learning Rate Sensitivity In Transformer Models: A Case Study In." 4(May). doi:https://doi.org/10.5281/zenodo.15769066.

Katırcı, Ramazan, Hilal Çelik, and Taha Oğuz. 2024. "Conversational Ai and Question Answering Systems: Architectures, Training Challenges and Solutions." 47–68. doi:10.5281/zenodo.17449635.

Kramp, Sergey, Giovanni Cassani, and Chris Emmery. 2024. "BigNLI : Native Language Identification with Big Bird Embeddings." 2375–82.

Lan, Zhenzhong, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2020. "Albert: A Lite Bert for Self-Supervised Learning of Language Representations." *8th International Conference on Learning Representations, ICLR 2020* 1–17.

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, Luke Zettlemoyer, and Pre-training Bart. 2019. "BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension."

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen Tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Advances in Neural Information Processing Systems* 2020-Decem.

Liu, Yinhan, Jiatao Gu, Naman Goyal, Xian Li, and Sergey Edunov. 2020. “Multilingual Denoising Pre-Training for Neural Machine Translation.” 8:726–42.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. “RoBERTa: A Robustly Optimized BERT Pretraining Approach.” (1).

Manning, Christopher D. 2020. “ELECTRA: Pre-Training Text Encoders as Discriminators Rather Than Generators.” 1–18.

Ni, Shiwen, Min Yang, Ruifeng Xu, Chengming Li, and Xiping Hu. 2024. “Layer-Wise Regularized Dropout for Neural Language Models.” *2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, LREC-COLING 2024 - Main Conference Proceedings* 10208–18.

Nielsen, Dan Saattrup, Kenneth Enevoldsen, and Peter Schneider-kamp. 2025. “Encoder vs Decoder : Comparative Analysis of Encoder and Decoder Language Models on Multilingual NLU Tasks.” 3:561–72.

Oche, Agada Joseph, Ademola Glory Folashade, Tirthankar Ghosal, and Arpan Biswas. 2025. “A Systematic Review of Key Retrieval-Augmented Generation (RAG) Systems : Progress , Gaps , and Future Directions.” (Llm).

Omar, M. A. S., Fitri Yakub, Senior Member, Andika A. J. I. Wijaya, Inge Dhamanti, and S. H. I. Zhongchao. 2025. “Evaluation of Encoder-Only Transformer for Multi-Step Traffic Flow Prediction.” *IEEE Access* 13(May):106349–68. doi:10.1109/ACCESS.2025.3580500.

Radford, Alec, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. *Improving Language Understanding by*

Generative Pre-Training. https://gluebenchmark.com/leaderboard.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2023. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." *Journal of Machine Learning Research* 21:1–67.

Roman, Marcin Junczys-dowmunt, Grundkiewicz Tomasz, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, and Alham Fikri. 2018. "Marian : Fast Neural Machine Translation in C ++." 116–21.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. "DistilBERT , a Distilled Version of BERT : Smaller , Faster , Cheaper and Lighter." 2–6.

Senadeera, Damith Chamalke, and Julia Ive. 2022. "Controlled Text Generation Using T5 Based Encoder-Decoder Soft Prompt Tuning and Analysis of the Utility of Generated Text in AI." 1–12. http://arxiv.org/abs/2212.02924.

Shon, SuwonLeveraging Pre-trained Language Model for Speech Sentiment Analysis, Pablo Brusco, Jing Pan, Kyu J. Han, and Shinji Watanabe. 2021. "Leveraging Pre-Trained Language Model for Speech Sentiment Analysis." 2–6.

Singh, Aditi, Abul Ehtesham, Saket Kumar, and Tala Talaei Khoei. 2025. "AGENTIC RETRIEVAL-AUGMENTED GENERATION: A SURVEY ON AGENTIC RAG." 1–39. doi:https://doi.org/10.48550/arXiv.2501.09136.

Sun, Chi, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2020. "How to Fine-Tune BERT for Text Classification?" (2).

Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. 2014. "Sequence to Sequence Learning with Neural Networks." *Advances*

in Neural Information Processing Systems 4(January):3104–12.

Team, Gemini, and Google. 2024. "Gemini 1 . 5 : Unlocking Multimodal Understanding across Millions of Tokens of Context." 1–154.

Team, Gemini, and Google. 2025. "Gemini : A Family of Highly Capable Multimodal Models." 1–90.

Touvron, Hugo, Thibaut Lavril, Xavier Martinet, Marie-anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, and Guillaume Lample. 2022. "LLaMA : Open and Efficient Foundation Language Models."

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." *Advances in Neural Information Processing Systems* 2017-Decem(Nips):5999–6009.

Wu, Yonghui, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." 1–23. http://arxiv.org/abs/1609.08144.

Zaheer, Manzil, Avinava Dubey, Joshua Ainslie, Chris Alberti, Santiago Ontanon, Philip Pham, Anirudh Ravula, Qifan

Wang, Li Yang, and Amr Ahmed. 2020. “Big Bird : Transformers for Longer Sequences.” (NeurIPS).

Zha, Yuheng, Yichi Yang, Ruichen Li, and Zhiting Hu. 2023. “Text Alignment Is An Efficient Unified Model for Massive NLP Tasks.” (NeurIPS).

SOSYAL MEDYA VE OYUN VERİLERİ İÇİN HİBRİT VE ÇOK DİLLİ BİR METİN NORMALİZASYON MİMARİSİ

Arzum KARATAŞ[1]

1. GİRİŞ

Sosyal medya ve çevrimiçi oyun platformları, kullanıcıların yoğun etkileşimde bulunduğu ve dilsel çeşitliliğin en belirgin şekilde gözlemlendiği ortamlardır. Bu ortamlardan elde edilen veriler, doğal dil işleme (NLP) araştırmaları için zengin bir kaynak sunarken, aynı zamanda "gürültülü metin (noisy text)" olarak adlandırılan biçimsel bozukluklar, argo ifadeler, emoji kullanımı, harf-rakam karışımı yazım biçimleri (leetspeak) ve diller arası geçiş (code-switching) gibi zorlukları da beraberinde getirir. Baldwin ve arkadaşları (2015) sosyal medya metinlerinin "standart dilin sınırlarını zorlayan, biçimsel olarak bozuk ve kaynaklar arasında büyük farklılıklar gösteren" bir yapıya sahip olduğunu vurgulamaktadır. Eisenstein (2013) ise sosyal medya dilinin standart yazım kurallarını sistematik biçimde ihlal ettiğini ve bu durumun NLP için özel çözümler gerektirdiğini belirtmektedir.

Standart NLP kütüphaneleri (örn. NLTK, SpaCy), ağırlıklı olarak kurallı ve tek dilli metinler üzerinde optimize edildikleri için, sosyal medya ve oyun platformlarında karşılaşılan gayri resmi (informal) dil yapıları karşısında genellikle yetersiz kalmaktadır. Literatürdeki çalışmalar da bu kısıtlılığı destekler niteliktedir; nitekim Solorio ve arkadaşları

[1] Dr. Öğr. Üyesi, Bandırma Onyedi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği, ORCID: 0000-0001-6433-3355.

(2014), “I am coming, bekle beni” örneğinde olduğu gibi dillerin cümle veya öbek düzeyinde ardışık değiştiği kod değiştirme (code-switching) durumlarında, tek dilli modellerin ciddi performans kayıpları yaşadığını raporlamıştır. Benzer bir düzlemde Pratapa, Choudhury ve Bali (2018) ise, “Server-lar down oldu” örneğindeki gibi dillerin sözcük veya ek düzeyinde iç içe geçtiği kod karıştırma (code-mixing) yapılarında, standart yaklaşımların ötesine geçilerek bu yapıya özgü özel stratejilerin geliştirilmesi gerektiğini ortaya koymuştur.

Son yıllardaki literatür, çok dilli ve kod karıştırma (code-mixing) içeren metinlerde veri normalizasyonunun kritik bir ön adım olduğunu göstermektedir. Bu bağlamda Singh, Choudhary ve Shrivastava (2023), sosyal medya metinlerindeki kelime varyasyonlarını otomatik normalize etmenin, duygu ve konu analizi performansını belirgin şekilde artırdığını kanıtlamıştır. Benzer şekilde Sharma ve arkadaşları (2025), yanlış bilgilendirme tespitinde güvenilirliği artırmak adına, çok dilli gönderilerde iddia (claim) normalizasyonunun önemine dikkat çekmektedir.

Normalizasyonun yanı sıra, modelleme mimarilerindeki gelişmeler de göze çarpmaktadır; Winata, Lin ve Fung (2021), kod değiştirme tespitinde transformer tabanlı yaklaşımların klasik yöntemlere kıyasla üstünlüğünü raporlarken, Sterner (2024) çok dilli İngilizce veri setlerinde sözcük düzeyindeki modellerin %4’e varan performans artışı sağladığını ortaya koymuştur.

Sosyal medya metinlerinin işlenmesinde, metin dışı bir unsur olan emojilerin kullanımı hem bir fırsat hem de bir zorluk olarak öne çıkmaktadır. Literatürde bu konuda iki temel görüş hakimdir: Novak ve arkadaşları (2015), emojileri dil sınırlarını aşan evrensel duygu göstergeleri olarak tanımlarken; Zhou ve arkadaşları (2024), bu yaklaşıma önemli bir nüans ekleyerek, kültürler arası kullanım farklılıklarının analizde mutlaka

gözetilmesi gerektiğini savunmaktadır. Tüm bu dinamiklerin ışığında Jahan ve arkadaşları (2024), emojilerin özellikle çapraz-dilli (cross-lingual) çalışmalarda kritik bir köprü görevi gördüğünü ortaya koymuştur. Örneğin, İngilizce veri setleri üzerinde 'kalp (❤️)' emojisinin pozitif duygu yükünü öğrenen bir modelin, bu evrensel ipucunu kullanarak Türkçe bir metni — kelime anlamlarını tam çözemese dahi— doğru duygu sınıfına atayabilmesi, bu aktarıma tipik bir örnektir.

Sosyal medya dilinin kendine has bir diğer zorlu karakteristiği de harflerin görsel benzerlik taşıyan rakam veya sembollerle ikame edildiği 'Leetspeak' (örn. 'nefes' yerine 'n3f3s' veya 'love' yerine 'l0v3') kullanımı ve standart dışı yazım biçimleridir. Bu tür kullanımlar, basit birer yazım hatası olmaktan öte, Crystal (2001)'ın da vurguladığı gibi dijital topluluklarda bir kimlik inşası ve gruba aidiyet göstergesi olarak sosyolojik bir işlev görür. Ancak insanlar tarafından kolayca deşifre edilebilen bu yaratıcı çarpıtmalar, doğal dil işleme araçları için ciddi bir gürültü kaynağıdır. Nitekim Sampath ve arkadaşları (2022), standart sözlük tabanlı yaklaşımların bu noktada yetersiz kaldığını ve 'leetspeak' dönüşümlerini çözmek için sisteme özel kurallar entegre edilmesi gerektiğini göstermiştir. Bu sorunun çözümüne odaklanan Khan ve Lee (2021) ise, kelimenin sadece biçimine değil cümledeki rolüne de odaklanan bağlam duyarlı (context-aware) normalizasyon yöntemlerinin, duygu analizi doğruluğunu anlamlı ölçüde yükselttiğini rapor etmiştir.

Türkçe özelinde, Çöltekin (2020) sosyal medya verilerinde saldırgan dilin normalizasyonu üzerine yaptığı çalışmada, Türkçe'nin morfolojik zenginliği nedeniyle yüzeysel düzeltmelerin yetersiz kaldığını, bağlam duyarlı yöntemlerin gerekliliğini vurgulamaktadır. Sak, Güngör ve Saraçlar (2011) tarafından geliştirilen Türkçe morfolojik ayrıştırıcı, eklemeli yapının NLP için özel çözümler gerektirdiğini göstermektedir.

Tablo 1. Mevcut Araçlar ve Önerilen Mimarinin Karşılaştırması

Özellik	**Standart NLP Araçları (SpaCy, NLTK)**	**Türkçe NLP Araçları (VNLP, Zemberek)**	**Önerilen Mimari (GamerNLP)**
Kod Değiştirme (Code-Switching)	Genellikle Cümle Bazlı	Desteklemez veya Sınırlı	Token Bazlı (Kelime Kelime)
Oyun Jargonu	Gürültü (Noise) olarak siler	Yazım hatası sayar	Anlamlı veriye dönüştürür
Leetspeak	İşleyemez	İşleyemez	Düzenli ifadeler (Regex) ile Normalize Eder
Emoji	Genellikle siler	Metin sonuna atar	Bağlam içinde dönüştürür

Türkçe özelinde yapılan çalışmalara bakıldığında, Gökırmak (2023) sinirsel kodlayıcı-kod çözücü (neural encoder-decoder) tabanlı mimarilerin, sosyal medyada sıkça karşılaşılan standart dışı metinlerin normalizasyonunda umut verici sonuçlar ürettiğini raporlamıştır. Bu algoritmik gelişmelerin yanı sıra, VNLP Araştırma Ekibi (2024) tarafından sunulan kapsamlı araç seti de alandaki önemli bir boşluğu doldurmaktadır. Söz konusu çalışma; normalizasyondan morfolojik çözümlemeye ve duygu analizine kadar uzanan geniş bir yelpazede, araştırmacılara hazır ve açık kaynaklı bir altyapı desteği sağlamaktadır.

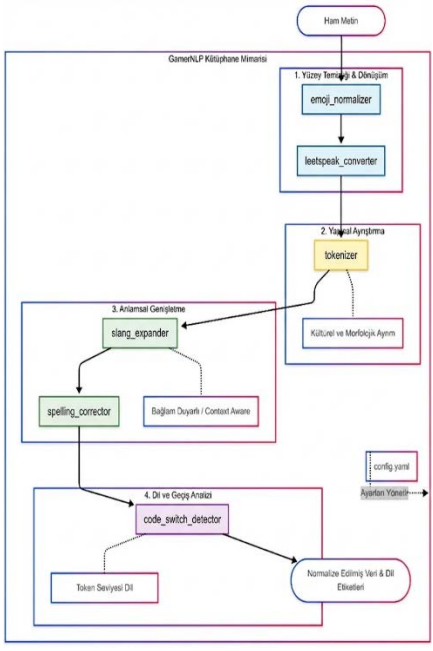
Standart NLP kütüphaneleri (NLTK, SpaCy) ve Türkçe odaklı araçlar (VNLP, Zemberek), resmi metinlerde yüksek başarı gösterse de, oyun verisinin *anlık* ve *gürültülü* doğasında yetersiz kalabilmektedir. Tablo 1, mevcut yaklaşımlar ile önerilen mimarinin karşılaştırmasını sunmaktadır.

Bu çalışmanın temel amacı, mevcut literatürdeki boşluğu doldurarak; Türkçe-İngilizce karma metinler ve oyun jargonu için özelleştirilmiş, hibrit ve modüler bir normalizasyon mimarisi sunmaktır. Önerilen sistem; emoji normalizasyonu, leetspeak dönüşümü, bağlam duyarlı argo genişletme, kod değiştirme (code-switching) tespiti ve kültürel/morfolojik tokenizasyon gibi kritik bileşenleri tek bir çatı altında entegre etmektedir.

Araştırmacılara esnek, özelleştirilebilir ve açık kaynaklı bir ön işleme aracı olarak sunulan bu mimarinin hem akademik çalışmalarda hem de endüstriyel uygulamalarda sosyal medya ve oyun verilerinin analiz kalitesini önemli ölçüde artırması hedeflenmektedir.

2. SİSTEM MİMARİSİ VE İŞLEM AKIŞI

Bu çalışmada geliştirilen ve GamerNLP olarak adlandırılan kütüphane mimarisi, sosyal medya ve çevrimiçi oyun platformlarından elde edilen, yüksek gürültü içeren ve çok dilli (multi-lingual) karakteristik gösteren metin verilerini, makine öğrenmesi algoritmaları için işlenebilir hale getirmek üzere tasarlanmıştır. Sistemin temel çalışma prensibi ve veri işleme boru hattı (pipeline) Şekil 1'de görselleştirilmiştir.



Şekil 1. GamerNLP Kütüphane Mimarisi ve Veri İşleme Akış Şeması

Şekil 1 incelendiğinde görüleceği üzere, sistem doğrusal bir veri temizleme yaklaşımından ziyade, birbirini hiyerarşik olarak besleyen dört temel mantıksal katman üzerine kurgulanmıştır. Literatürdeki "gürültülü metin normalizasyonu (noisy text normalization)" standartlarına (Baldwin vd., 2015; Singh vd., 2023) dayanan bu yapı, özellikle Türkçe-İngilizce kod değiştirme (code-switching) ve oyun jargonu gibi kültürel dinamikleri yakalayacak şekilde özelleştirilmiştir. Katmanların işlevsel detayları aşağıda sunulmuştur:

2.1. Yüzey Temizliği ve Dönüşüm (Surface Cleaning & Conversion)

Veri işleme hattının ilk durağı olan bu katman, metnin yapısal bütünlüğüne dokunmadan, karakter bazlı temsili düzeltmeyi hedefler. Ham metin sisteme girdiğinde sırasıyla şu işlemlerden geçer:

Emoji Normalizasyonu: Sosyal medya metinlerinde duygu yoğunluğunu taşıyan en önemli unsurlardan biri emojilerdir. Ancak standart modeller bu sembolleri genellikle yok sayar. *emoji_normalizer* modülü, emojileri silmek yerine onları anlamsal karşılıklarına (örn. 🥲 → "mutlu", 🔥 → "harika/ateş") dönüştürerek metne entegre eder (Zampieri vd., 2020).

Leetspeak Dönüşümü: Özellikle oyun topluluklarında, bir kimlik göstergesi olarak harflerin rakam veya sembollerle ikame edildiği (örn. Hacker yerine "h4x0r", noob yerine "n00b", veya selam yerine "s3lam") "Leetspeak" alfabesi yaygındır. *leetspeak_converter* modülü, tanımlı düzenli ifade (regex) örüntülerini kullanarak bu bozuk yüzey formlarını standart Latin alfabesine geri döndürür. Bu adım, sonraki aşama olan tokenizasyonun başarısını doğrudan artırır.

2.2. Yapısal Ayrıştırma (Structural Parsing)

Yüzeyi temizlenen metin, dilbilgisel birimlerine ayrılmak üzere bu katmana iletilir.

Kültürel ve Morfolojik Tokenizasyon: Standart boşluk tabanlı ayırıcıların (white-space tokenizers) aksine, bu modül hibrit bir ayrıştırma stratejisi izler. Sistem bir yandan Türkçe gibi sondan eklemeli dillerin yapısına uygun olarak morfolojik kök-ek ayrımı yaparken (örn. "geliyorum" → gel-iyor-um); diğer yandan oyun kültürüne özgü çok kelimeli kalıpları (multi-word expressions) parçalamadan bir bütün halinde işler (örn. "gg wp" yani iyi oyun, "all pick" yani serbest seçim). Bu seçici yaklaşım, dilbilgisel yapıyı analiz ederken kültürel bağlamın parçalanarak kaybolmasını engeller.

2.3. Anlamsal Genişletme (Semantic Expansion)

Sistemin en kritik zekâ katmanı burasıdır. Kelimelerine (tokenlarına) ayrılmış verideki belirsizlikler, bağlam (context) bilgisi kullanılarak çözümlenir.

Bağlam Duyarlı Argo Analizi: Bir kelimenin anlamı, kullanıldığı yere göre değişebilir. *slang_expander* modülü, "Bağlam Duyarlı (Context Aware)" algoritmalar kullanarak çok anlamlı kelimeleri analiz eder. Örneğin "kollamak" kelimesi genel Türkçede "korumak" iken, oyun jargonunda "pusu kurmak/beklemek" anlamına gelebilir. Modül, cümlenin gelişine göre doğru anlamı seçerek genişletme yapar (Solorio vd., 2014).

Yazım Denetimi: Anlamsal genişletmenin hemen ardından *spelling_corrector* devreye girerek, klavye hatalarından kaynaklanan yazım yanlışlarını (typo) düzeltir ve veriyi son analize hazırlar.

2.4. Dil ve Geçiş Analizi (Language Analysis & Code-Switching)

Normalize edilmiş ve zenginleştirilmiş veri, son aşamada dilsel kimliklendirme sürecine tabi tutulur.

Kelime (Token) Seviyesi Dil Tespiti: Çok dilli metinlerde cümlenin tamamına tek bir dil etiketi atamak yanıltıcıdır. *code_switch_detector* modülü, her bir kelimeyi (tokenı) ayrı ayrı analiz ederek Türkçe (TR) veya İngilizce (EN) olarak etiketler. Bu işlem, diller arası geçiş noktalarının (switch-points) tespit edilmesini sağlar ve çapraz dilli (cross-lingual) duygu analizi modelleri için hayati bir girdi oluşturur (Çöltekin, 2020).

2.5. Merkezi Yönetim ve Esneklik

Mimarinin sağ alt köşesinde belirtilen *config.yaml* dosyası, sistemin beyni niteliğindedir. Araştırmacılar; argo genişletme modülünün agresiflik seviyesini, hangi dillerin dikkate alınacağını veya leetspeak dönüşümünün aktif olup olmayacağını bu dosya üzerinden parametrik olarak yönetebilir. Bu esnek yapı, sistemin FPS oyunlarından MMORPG oyunlarına veya farklı sosyal medya platformlarına (Twitter, Discord vb.) kolayca adapte edilmesini sağlar.

3. VAKA TAKDİMİ

Önerilen mimarinin işlevselliğini göstermek amacıyla, Türkçe-İngilizce karışık bir oyun yorumu üzerinde uygulama yapılmıştır. Ham metin olarak seçilen örnek ifade “Bu oyun 1nc1 derecede muth1şt1 👑 GG broooo!” biçimindedir. Bu yorum, sosyal medya ve oyun ortamlarında sıkça karşılaşılan leetspeak yazımları, argo kısaltmalar, emoji kullanımı ve dil geçişlerini bir arada barındırmaktadır. Tablo 2, ham verinin sistemden geçerken uğradığı dönüşümleri adım adım göstermektedir. Bu işlem

sonucunda elde edilen çıktı, duygu analizi modelleri tarafından "Pozitif (olumlu)" olarak sınıflandırılmaya çok daha uygun hale gelmiştir

İlk aşamada, leetspeak dönüştürücü modül rakam-harf karışımlarını standart biçime çevirmiş ve "1nc1" ifadesi "birinci", "muth1şt1" ifadesi ise "muhteşem" olarak normalize edilmiştir. Ardından, emoji normalizasyonu devreye girerek 👑 sembolü bağlama uygun şekilde "taç" veya "kral" olarak yorumlanmıştır. Argo genişletme modülü ise "GG" ifadesini "good game" ve "broooo" ifadesini "brother" biçiminde genişletmiştir.

Sonraki aşamada, code-switching tespiti yapılmış ve metindeki Türkçe-İngilizce geçiş noktaları işaretlenmiştir. Bu sayede "GG" ve "broooo" İngilizce olarak etiketlenirken, diğer kelimeler Türkçe olarak sınıflandırılmıştır. Son olarak, kültürel ve morfolojik tokenizasyon uygulanmış; Türkçe kelimeler kök ve eklerine ayrılmış, İngilizce kelimeler ise lemmatization sürecinden geçirilmiştir.

Bu işlemler sonucunda elde edilen normalize edilmiş çıktı "Bu oyun birinci derecede muhteşem taç good game brother" biçimindedir. Çıktı üzerinde yapılan analiz, yorumun genel olarak pozitif duygu taşıdığını, oyuncunun performansına yönelik övgü ve takdir niyeti içerdiğini ve dil geçişinin oyun bağlamında doğal bir iletişim biçimi olduğunu göstermektedir.

Bu senaryo, önerilen mimarinin çokdilli, informal ve kültürel açıdan zengin metinlerde nasıl çalıştığını somut biçimde ortaya koymaktadır. İlerleyen çalışmalarda, farklı platformlardan (örneğin Twitch, Discord, Reddit) alınan yorumlarla geniş ölçekli testler gerçekleştirilerek mimarinin genellenebilirliği değerlendirilecektir.

Tablo 2. Örnek Metin Üzerinde Normalizasyon Adımları

Adım	Modül	Girdi Verisi	İşlem / Dönüşüm	Çıktı Verisi
1	Leetspeak Converter	1nc1, muth1şt1	Rakam-Harf Değişimi	birinci, muhteşem
2	Emoji Normalizer	👑	Anlamsal Etiketleme	taç (veya kral)
3	Slang Expander	GG, broooo	Kısaltma ve Tekrar Düzeltme	good game, brother
4	Code-Switch Detector	*Tüm Cümle*	Dil Etiketleme (Token Bazlı)	(good game, EN), (birinci, TR)...
5	Tokenizer	muhteşem, brother	Morfolojik Ayrıştırma	muhteşem, brother (Lemma)
SONUÇ	Normalize Çıktı	-	-	"Bu oyun birinci derecede muhteşem taç good game brother"

4. TARTIŞMA VE SONUÇ

Bu çalışmada sunulan **GamerNLP** mimarisi, standart NLP süreçlerinin göz ardı ettiği gürültülü metin problemlerine, alan bilgisine dayalı (domain-specific) ve hibrit bir çözüm getirmektedir.

4.1. Katkılar ve Avantajlar

Çalışmanın en özgün yanı, Türkçe ve İngilizce'nin iç içe geçtiği (code-switching) ve oyun jargonunun yoğun olduğu metinleri, bilgi kaybı olmadan işleyebilmesidir. config.yaml üzerinden sağlanan esneklik, bu mimarinin sadece oyun verilerinde değil, Twitch sohbet günlükleri, Discord sunucuları veya YouTube canlı yayın yorumları gibi benzer dinamiklere sahip platformlarda da kullanılabilmesine olanak tanır.

4.2. Sınırlılıklar ve Gelecek Çalışmalar

Mevcut tasarımın temel kısıtı, argo ve leetspeak modüllerinin statik sözlük yapısına dayanması nedeniyle, dilin dinamik doğasına anlık uyum sağlayamamasıdır. Özellikle oyun dünyasında jargonun baş döndürücü bir hızla değiştiği düşünüldüğünde; bugün literatürde olmayan ancak yarın viral hale gelebilecek "*rizz* (etkileme becerisi)" veya "*diff* (yetenek farkı)" gibi yeni türetilen kelimeler (neologisms), sistem tarafından ilk etapta tanınamayabilir. Bu sınırlılığı aşmak adına, gelecek çalışmalarda sisteme 'kendi kendine öğrenen' (self-learning) veya bağlamdan yola çıkarak 'yeni terim keşfi' (new term discovery) yapabilen makine öğrenmesi modüllerinin entegre edilmesi planlanmaktadır. Ayrıca, geliştirilen mimarinin Python tabanlı açık kaynak kütüphanesi olarak yayınlanması ve böylece veri tabanının topluluk katkısıyla (crowdsourcing) güncel tutulması hedeflenmektedir.

KAYNAKÇA

Baldwin, T., Cook, P., Lui, M., MacKinlay, A., & Wang, L. (2015). Sosyal medya metinlerindeki gürültü ve kaynak farklılıkları. Uluslararası Sosyal Medya ve Web Konferansı Bildirileri (ss. 18–27). AAAI Yayınları.

Crystal, D. (2001). İnternet ve Dil. Cambridge University Press.

Çöltekin, Ç. (2020). Sosyal medyada Türkçe saldırgan dil derlemi. 12. Dil Kaynakları ve Değerlendirme Konferansı Bildirileri (ss. 6174–6184). Avrupa Dil Kaynakları Derneği.

Eisenstein, J. (2013). Sosyal medya metinleri üzerine yapılacaklar. Language and Linguistics Compass, 7(11), 466–477.

Gökırmak, S. (2023). Türkçe sosyal medya için neural metin normalizasyonu. Hacettepe Üniversitesi Teknik Raporu.

Jahan, R. I., Fan, H., Chen, H., & Feng, Y. (2024). Emojiler aracılığıyla çapraz-dilli duygu analizi. arXiv ön baskı arXiv:2412.17255.

Khan, J., & Lee, S. (2021). Sosyal medya informal metinlerinde bağlam duyarlı normalizasyon. Applied Sciences, 11(17), 8172.

Khan, J., Ahmad, K., Jagatheesaperumal, S. K., & Sohn, K. A. (2025). Sosyal medya metinlerindeki biçimsel varyasyonlar: zorluklar ve çözümler. Artificial Intelligence Review, 58(89).

Liu, F., Weng, F., & Jiang, H. (2019). Gürültülü metin normalizasyonu. ACM Transactions on Asian and Low-Resource Language Information Processing, 18(3).

Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Emojilerin duygu analizi üzerindeki etkisi. PLOS ONE, 10(12), e0144296.

Pratapa, A., Choudhury, M., & Bali, K. (2018). Code-mixed metinlerde kelime gömme yöntemleri. 56. Yıllık ACL Konferansı Bildirileri (ss. 602–612). Association for Computational Linguistics.

Sak, H., Güngör, T., & Saraçlar, M. (2011). Türkçe metinlerin morfolojik ayrıştırılması. Computer Speech & Language, 25(1), 45–63.

Sampath, S., ve diğerleri. (2022). Sosyal medya metinlerinde leetspeak normalizasyonu. 29. Uluslararası Hesaplamalı Dilbilim Konferansı Bildirileri (ss. 1234–1245). Uluslararası Hesaplamalı Dilbilim Komitesi.

Sharma, M., Suneesh, A., Jain, M., Rajpoot, P. K., Devadiga, P., Hazarika, B., Shrivastava, A., Gurumurthy, K., Suresh, A. B., & Baliga, U. A. (2025). Gürültülü çokdilli sosyal medya gönderilerinde akıl yürütme ile iddia normalizasyonu. CLEF 2025 Çalışma Notları. CEUR Workshop Proceedings.

Singh, R., Choudhary, N., & Shrivastava, M. (2023). Code-mixed sosyal medya metinlerinde kelime varyasyonlarının otomatik normalizasyonu. Lecture Notes in Computer Science, 13396 (ss. 371–381). Springer.

Solorio, T., Blair, E., Maharjan, S., ve diğerleri. (2014). Code-switched verilerde dil tanımlama üzerine ortak görev. Birinci Code Switching Çalıştayı Bildirileri (ss. 58–72). Association for Computational Linguistics.

Sterner, I. (2024). Çokdilli İngilizce code-switching tespiti. 11. Benzer Diller, Çeşitler ve Lehçeler için NLP Çalıştayı

(VarDial 2024) Bildirileri (ss. 163–173). Association for Computational Linguistics.

AI-DRIVEN PHISHING DEFENSE: PREDICTION, DETECTION, PREVENTION, AND ETHICAL CHALLENGES

Sara Naghib ZADEH[1]

Ebrar KAVALCI[2]

Zeynep Rana KAYIKCI [3]

1. INTRODUCTION

In today's digital environment, data security has become a critical concern as the internet enables both broad opportunities and an expanding landscape of cyber threats. Among these, phishing remains one of the most effective and costly forms of attack, exploiting human psychological weaknesses rather than technical flaws—making traditional signature-based and rule-driven defenses increasingly inadequate (Alkhalil, Hewage, Nawaf, & Khan, 2021).

The rapid evolution of phishing methods—from fake login pages and deceptive SMS messages to AI-generated audio and video deepfakes—allows attackers to outpace classical security tools (Mathur & Singh, 2025). This highlights the limitations of conventional defenses against highly adaptive, personalized phishing campaigns. In response, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as

[1] Dr. Lecture, Halic University, Vocational School, Department of Computer Programming, ORCID: 0009-0005-6959-1165.

[2] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0007-1246-3853.

[3] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0008-4018-0452.

key solutions, employing data-driven analytics, NLP, behavioral modeling, Computer Vision (CV), and SOAR automation to detect threat patterns proactively and identify anomalies in real time, functioning as a “digital immune system” (George, George, & Baskar, 2023).

While much of the existing literature focuses on isolated detection methods such as URL or text analysis, emerging threats,including adversarial AI, LLM-generated phishing content, and deepfake impersonation, require a more holistic and multi-layered perspective. This study addresses that gap by examining AI’s role across the entire phishing defense lifecycle: prediction, detection, prevention, and automated response.

The objective is to analyze how AI supports the identification and mitigation of phishing attacks, clarify underlying mechanisms, review real-world applications, and evaluate associated ethical and technical challenges. The study covers the evolution of phishing, AI-based detection using NLP, ML, and CV, and automated response through SOAR systems, relying on conceptual analysis and threat intelligence rather than detailed mathematical modeling.

The article is structured across four sections: attack vectors and threat modeling, AI-driven detection mechanisms, proactive prevention and automated response, and ethical considerations with future directions. Overall, it argues that the future of cybersecurity depends on explainable AI (XAI), continuous learning, and strong human–machine collaboration to significantly reduce response times and enable next-generation intelligent defense systems.

2. THREAT MODELING AND THE ATTACK PHASE

This section analyzes how phishing has evolved from simple, broad deception attempts into today's sophisticated, multi-layered, and highly targeted campaigns. It also highlights how artificial intelligence supports the prediction of attack vectors and reinforcement of cybersecurity defenses before an incident occurs. The aim is to provide a structured view of phishing architectures, attacker behavior, and the data-driven techniques capable of disrupting the attack cycle.

2.1. Anatomy and Evolution of Phishing Threats (Attack Vectors)

As illustrated in Figure 1 (Evolution and types of phishing attack vectors), phishing began as basic fraudulent messaging but has transformed into a complex ecosystem of multi-stage and highly personalized attacks, driven by advanced tools, expanded digital footprints, and AI-enhanced profiling (Kavya & Sumathi, 2024). Major attack vectors include:

- **Spear Phishing**: Highly targeted messages crafted using victims' behavioral and organizational information, making infiltration into corporate environments significantly more effective (Birthriya, Ahlawat, & Jain, 2025).
- **Whaling**: Targeting executives and decision-makers with formal, high-stakes requests such as urgent payments or confidential document transfers, often resulting in major financial or operational loss (Pienta, Thatcher, & Johnston, 2020).
- **Smishing and Vishing**: SMS-based phishing and voice-based deception, increasingly employing AI-generated voice clones and deepfake audio, fueled by the

widespread use of mobile devices (Debnath, Kar, Biswas, Das, & Das, 2025).

- **Clone Phishing**: Replicating legitimate emails but replacing links or attachments with malicious versions, making detection difficult due to the credibility of the original message (Chaudhuri, 2023).

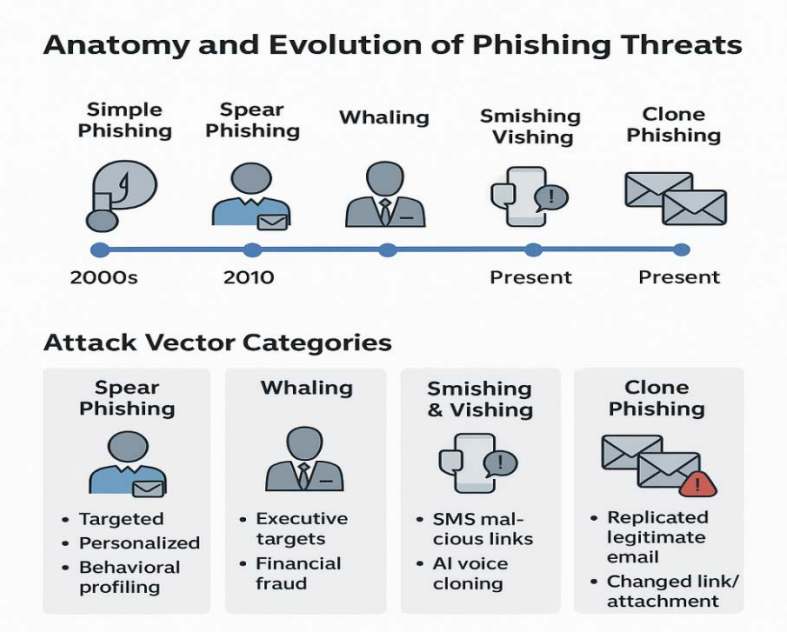


Figure 1. The Evolution and Types of Phishing Attack Vectors

2.2. The Human Factor: The Weakest Link in the Security Chain

A defining characteristic of phishing attacks is that they exploit cognitive and psychological vulnerabilities rather than technical flaws. Attackers manipulate emotions such as fear, urgency, curiosity, trust, or the natural inclination to respond quickly. These human tendencies serve as powerful tools for deception (Hughes-Lartey, Li, Botchey, & Qin, 2021).

Since no technical system can eliminate human error, the human element remains the most vulnerable component of any security architecture. As a result, modern defense strategies increasingly focus on identifying abnormal behavioral patterns and predicting the likelihood of user susceptibility, even before the individual interacts with the malicious message. This shift reflects the growing recognition that human-centric vulnerabilities require data-driven, behavior-aware detection mechanisms (Ghafir et al., 2018).

2.3. The Role of Artificial Intelligence in Predicting Attacks (Predictive Threat Intelligence)

Given the rising complexity, diversity, and velocity of phishing campaigns, organizations require tools capable of analyzing attacker behavior before an attack materializes. Artificial intelligence has emerged as one of the most advanced defensive technologies in this regard.

AI systems process large-scale threat intelligence data drawn from global networks, analyze millions of messages, examine newly registered domains, and compare behavioral patterns across vast datasets. Through this analysis, AI can identify subtle indicators that have not yet escalated into full-scale attacks and issue early warnings accordingly (Sree, Koganti, Kalyana, & Anudeep, 2021).

This capability, known as Predictive Threat Intelligence, enables organizations to:

- Strengthen potential vulnerabilities before attackers exploit them,
- Update behavior-based security policies in real time,
- Transition from reactive cybersecurity to a fully proactive and preventive defense model.

By detecting anomalies and suspicious behavioral signatures before an attack occurs, AI effectively disrupts the attacker's decision-making cycle and significantly reduces the probability of a successful compromise (Hasan et al., 2025).

3. DETECTION PHASE AND AI-DRIVEN MECHANISMS

The detection phase is a critical element of phishing defense, and as illustrated in Figure 2 (AI-Driven Detection Mechanisms in Phishing Defense), weaknesses at this stage can lead to severe operational and financial consequences. With attackers increasingly using advanced deception, identity manipulation, and information tampering, organizations require intelligent, adaptive, and context-aware detection systems (Simone Raponi & Di Pietro, 2021). This section outlines the limitations of traditional detection methods and explains how AI enhances threat identification, behavioral analysis, and anomaly detection.

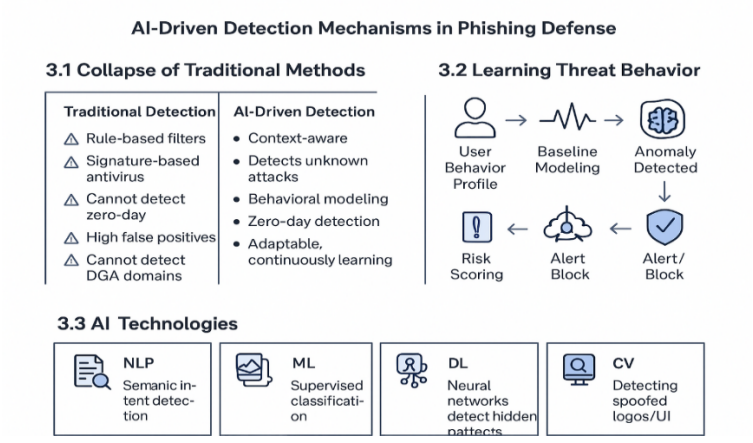


Figure 2. AI-Driven Detection Mechanisms in Phishing Defense

3.1. The Collapse of Traditional Methods

Traditional rule-based and signature-based detection systems operate effectively only when an attack is already known, and its signature is available in a database (Soe, Feng, Santosa, Hartanto, & Sakurai, 2019) . As modern phishing evolves rapidly, these systems cannot detect zero-day threats or creative, emerging techniques.

Domain Generation Algorithms (DGA) pose a major challenge by generating thousands of dynamic domains, making signature-based blocking nearly impossible (Anderson, Woodbridge, & Filar, 2016). Additionally, traditional systems lack contextual understanding—causing high false positives when benign messages contain suspicious keywords and false negatives when malicious messages evade predefined rules. Given the speed and complexity of modern phishing, static methods are insufficient (Brandao, 2025) .

3.2. The Rise of Artificial Intelligence: Learning Threat Behavior

AI introduces a behavior-driven approach, identifying suspicious activity rather than relying on stored signatures, enabling detection of unknown and zero-day attacks (Mohamed, 2025) .

Behavioral Analysis

AI models learn normal user and system behavior by analyzing historical activity, geographic patterns, session data, and interaction habits (Subrahmanyam, 2025) . Deviations—such as unusual login locations, abnormal data transfers, or clicks on unfamiliar links—are flagged as anomalies, enabling more context-aware phishing detection (Hussain, 2024) .

Continuous Learning

AI systems continuously update by analyzing new attacks in real time. Unlike traditional tools requiring manual updates, AI autonomously adapts to new attacker tactics and emerging patterns, keeping pace with evolving threats (Sarker, 2022).

3.3. AI Technologies Used in Phishing Detection

Artificial intelligence integrates multiple advanced technological components to analyze threats from linguistic, behavioral, and visual perspectives. The most prominent technologies used in phishing detection include Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL), and Computer Vision (CV) (Basit et al., 2020).

A) Natural Language Processing (NLP)

Natural Language Processing (NLP) enables security systems to interpret human language and identify suspicious cues within emails, SMS, and chat messages. Through semantic analysis, NLP detects manipulative intents such as urgency or emotional pressure, while stylistic analysis reveals inconsistencies between the message and the legitimate sender's typical writing style. Additionally, grammatical and structural anomaly detection uncovers subtle linguistic irregularities that may not be noticed by human users. Together, these capabilities allow NLP-based models to detect phishing even when traditional keywords are absent (Salloum, Gaber, Vadera, & Shaalan, 2022).

B) Machine Learning and Deep Learning

Machine Learning (ML) and Deep Learning (DL) further enhance phishing detection by learning patterns of malicious communication. Supervised learning relies on labeled datasets to differentiate legitimate from malicious

messages, whereas unsupervised learning identifies anomalies and previously unknown threats without the need for labeled data. Deep learning models, with their ability to analyze complex, multi-dimensional features, capture hidden manipulation patterns that conventional methods might miss. These techniques significantly improve detection accuracy compared to traditional rule-based approaches (Gandotra & Gupta, 2021).

C) Computer Vision (CV)

Computer Vision (CV) also plays a crucial role in identifying phishing attacks that rely on visual imitation of legitimate websites. CV algorithms analyze graphical elements to detect spoofed or modified logos, inconsistencies in layout, color schemes, and design structures, and they can automatically compare webpage screenshots with official templates. Moreover, CV systems are capable of identifying micro-level visual artifacts that indicate manipulation. This makes CV particularly effective in detecting image-based phishing threats (Zhao, Masood, & Seneviratne, 2021; Kultan et al., 2025).

4. RESPONSE AND PREVENTION MECHANISMS

The response and prevention phase is a crucial part of the cybersecurity defense lifecycle. As illustrated in Figure 3 (AI-Driven Response and Prevention Mechanisms in Cybersecurity), it plays a central role in stopping the spread of attacks, reducing potential damage, and restoring systems to a secure state (Kim et al., 2023). With the advancement of artificial intelligence, this phase has transformed from manual, slow, and reactive approaches into automated, rapid, and proactively adaptive processes. This section explores AI-driven mechanisms that

prevent phishing attacks before they reach users and highlights the role of AI in autonomously managing the impact of ongoing or successful intrusions.

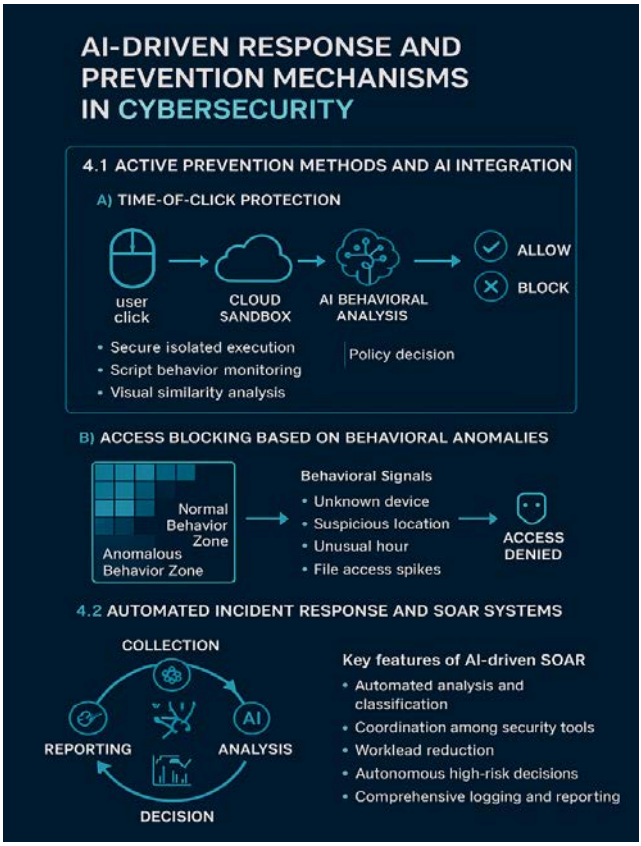


Figure 3. AI-Driven Response and Prevention Mechanisms in Cybersecurity

4.1. Active Prevention Methods and the Role of Artificial Intelligence

Active prevention includes measures designed to detect and stop threats before a system or user is compromised (Doherty, 2016). Integrating artificial intelligence into this phase improves detection accuracy, reduces human error, and enables rapid, automated responses.

One key AI-driven method is time-of-click protection, which evaluates URLs at the exact moment a user selects them (Thomas et al., 2011). The link is first opened in a secure cloud sandbox, then assessed through AI-based behavioral analysis that monitors suspicious scripts, cookie-stealing attempts, unusual interaction requirements, or visual similarity to known phishing sites. If malicious activity is detected, the system blocks the connection; otherwise, the user is safely redirected. This technique is particularly effective against delayed-activation phishing attacks, where links turn harmful after delivery (Afzal et al., 2021).

Another preventive approach is access blocking based on behavioral anomalies, in which AI analyzes contextual login behavior instead of relying solely on passwords (Seo, Baek, & Ko, 2024). Access may be denied when unusual conditions are detected, such as unexpected login locations, unfamiliar devices, repeated attempts to access sensitive files, or activity during atypical hours (Tiwari, 2021). This method protects not only against credential-theft phishing attacks but also against insider threats and unauthorized movement within networks.

4.2. Automated Incident Response and SOAR Systems

SOAR platforms (Security Orchestration, Automation, and Response) represent a major advancement in cybersecurity by integrating artificial intelligence with automated workflows and incident management. These systems can fully automate the response to security events, reducing incident resolution time from hours in manual processes to just seconds (Rathore & Kolhe, 2026).

AI-driven SOAR solutions perform automated analysis and classification of incidents based on risk levels, coordinate actions across various security tools such as SIEM, firewalls, anti-malware, and identity management systems, and significantly

reduce the workload of security teams by handling repetitive tasks. They can also make autonomous decisions during high-risk situations and generate detailed logs and reports for auditing and post-incident analysis (Mir & Ramachandran, 2021).

5. ETHICAL CONSIDERATIONS, CHALLENGES, AND THE FUTURE OF AI SECURITY

Despite the significant benefits of artificial intelligence in strengthening cybersecurity, its use introduces critical challenges related to ethics, privacy, transparency, and misuse. At the same time, attackers increasingly exploit advanced AI technologies to enhance the sophistication of their attacks. This section outlines the main ethical and operational concerns as well as the emerging AI-driven threat landscape.

5.1. Ethical and Operational Challenges in AI-Driven Security

As illustrated in Figure 4 (Ethical and Operational Challenges in AI-Driven Security), one major issue is balancing privacy and surveillance, as AI-based monitoring requires continuous analysis of network traffic, user behavior, and internal data, which may conflict with regulations such as GDPR and KVKK. To maintain both privacy and security, organizations must employ strategies like anonymization, access control, encryption, and data minimization (Mosa et al., 2024).

AI systems also face false positives and false negatives, where benign activities may be flagged as malicious or genuine threats may be misclassified as safe. Such errors can disrupt operations or allow attackers to infiltrate systems. Therefore, many organizations adopt hybrid models combining rule-based and AI-driven detection to improve accuracy (Badoni et al., 2024).

Another major challenge is the black box nature of deep learning models, which makes their decision-making difficult to interpret (Buhrmester, Münch, & Arens, 2021). Limited transparency reduces trust, complicates auditing, and introduces risks in critical sectors. As a result, Explainable AI (XAI) has become essential for ensuring clarity, accountability, and reliable threat assessment (Vainio-Pekka et al., 2023).

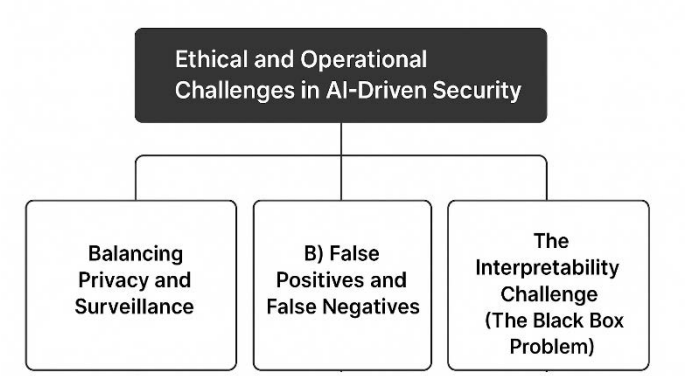


Figure 4. Ethical and Operational Challenges in AI-Driven Security

5.2. The Future Battlefield: Adversarial AI and Deepfake-Driven Threats

AI is a double-edged sword: while enhancing defense, it also empowers attackers to develop more advanced, stealthy techniques (Tounsi & Rais, 2018). As illustrated in Figure 5 (The Future Battlefield: Adversarial AI and Deepfake-Driven Threats), these emerging threats increasingly leverage adversarial manipulation, generative models, deepfake technologies, and data poisoning to bypass modern security systems.

Adversarial AI allows attackers to manipulate detection models by injecting malicious training data or subtly modifying inputs, causing systems to misclassify phishing pages, emails, or malware as safe (Lopes Antunes & Llopis Sanchez, 2023).

Generative AI tools—including LLMs such as ChatGPT, LLaMA, FraudGPT, and WormGPT—enable attackers to craft highly personalized, grammatically flawless phishing messages and automated campaigns that closely mimic real communication patterns, making them difficult even for experts to detect (Falade, 2023; Jabir, Le, & Nguyen, 2025).

Deepfake audio and video technologies allow cybercriminals to impersonate executives, create fake video calls, or spoof identities with high realism, making detection extremely challenging without specialized tools (Hashmi et al., 2024; Nailwal, Singh, Raza, & Singhal, 2023).

Finally, data poisoning attacks pose a severe risk by inserting corrupted data into AI training sets, degrading model accuracy and causing dangerous misclassifications. This threat is particularly serious in autonomous AI-driven security environments (Kure et al., 2025; Wang et al., 2023).

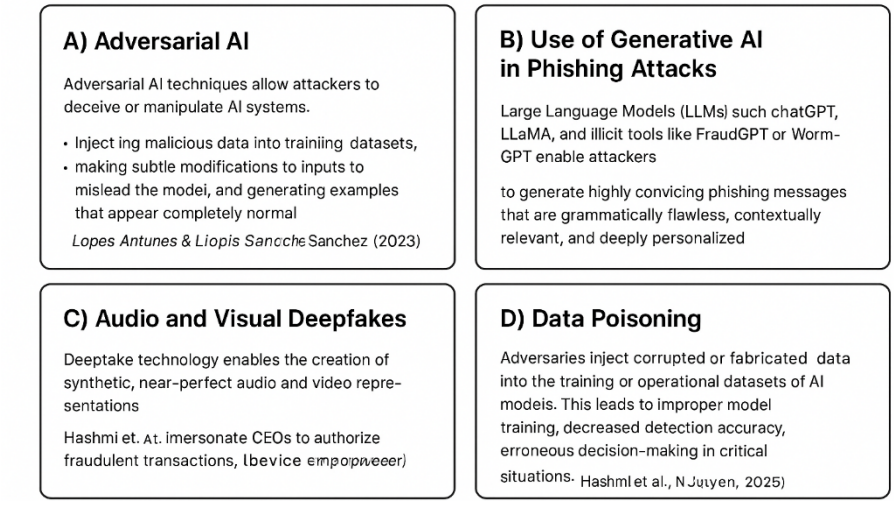


Figure 5. The Future Battlefield: Adversarial AI and Deepfake-Driven Threats

6. CONCLUSION

Phishing has become a highly adaptive and sophisticated threat that exploits both technical vulnerabilities and human behavior. This study shows that traditional, signature-based defenses are no longer adequate against modern, AI-enabled phishing techniques. In contrast, artificial intelligence now plays a central role in next-generation cybersecurity by enabling proactive threat prediction, behavioral monitoring, context-aware detection, and rapid automated response.

The analysis demonstrates that AI's power lies in learning from large and diverse datasets, recognizing linguistic and visual manipulation, and detecting subtle anomalies that humans may miss. Technologies such as NLP, Machine Learning, Deep Learning, Computer Vision, and SOAR-based automated response together form a multi-layered defense model that reduces errors and accelerates incident handling.

At the same time, integrating AI into security systems introduces new risks, including privacy concerns, lack of model transparency, algorithmic bias, and vulnerabilities like adversarial attacks and data poisoning. The emergence of deepfakes and generative AI–driven phishing further shows that attackers are also using advanced AI to enhance the realism and scale of their operations.

Overall, the findings indicate that the future of phishing defense requires continuous learning, explainable and trustworthy AI, strong human–machine collaboration, and adaptive, risk-aware policies. Organizations must combine advanced technical solutions with user training, ethical governance, and resilient data infrastructures. By doing so, AI can transform cybersecurity into a predictive, autonomous, and robust ecosystem capable of confronting increasingly intelligent digital threats

REFERENCES

Afzal, S., Asim, M., Javed, A. R., Beg, M. O., & Baker, T. (2021). URLdeepDetect: A Deep Learning Approach for Detecting Malicious URLs Using Semantic Vector Models. *Journal of Network and Systems Management 2021 29:3*, *29*(3), 21-. doi:10.1007/S10922-021-09587-8

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, *3*, 563060. doi:10.3389/FCOMP.2021.563060/BIBTEX

Anderson, H. S., Woodbridge, J., & Filar, B. (2016). DeepDGA: Adversarially-tuned domain generation and detection. *AISec 2016 - Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security, Co-Located with CCS 2016*, 13–21. doi:10.1145/2996758.2996767;GROUPTOPIC:TOPIC:ACM-PUBTYPE

Badoni, P., Wadhwa, M., Shrimal, V. M., & Dutta, N. (2024). Transformative Potential and Ethical Challenges: AI Driven Innovations in Cyber Security. *Proceedings - 2024 2nd International Conference on Advanced Computing and Communication Technologies, ICACCTech 2024*, 155–160. doi:10.1109/ICACCTECH65084.2024.00035

Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2020). A comprehensive survey of AI-enabled phishing attacks detection techniques. *Telecommunication Systems 2020 76:1*, *76*(1), 139–154. doi:10.1007/S11235-020-00733-2

Birthriya, S. K., Ahlawat, P., & Jain, A. K. (2025). Detection and prevention of spear phishing attacks: A comprehensive

survey. *Computers & Security*, *151*, 104317. doi:10.1016/J.COSE.2025.104317

Brandao, P. R. (2025). The Impact of Artificial Intelligence on Modern Society. *AI 2025, Vol. 6, Page 190*, *6*(8), 190. doi:10.3390/AI6080190

Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *Machine Learning and Knowledge Extraction 2021, Vol. 3, Pages 966-989*, *3*(4), 966–989. doi:10.3390/MAKE3040048

Chaudhuri, A. (2023). Clone Phishing: Attacks and Defenses. *International Journal of Scientific and Research Publications*, *13*(4), 180. doi:10.29322/IJSRP.13.04.2023.p13626

Debnath, B., Kar, N., Biswas, P., Das, N., & Das, A. (2025). A comprehensive assessment on phishing, smishing and vishing. *Data Science & Exploration in Artificial Intelligence*, 274–282. doi:10.1201/9781003589273-41

Doherty, M. (2016). From protective intelligence to threat assessment: Strategies critical to preventing targeted violence and the active shooter. *Journal of Business Continuity & Emergency Planning*, *10*(1), 9. doi:10.69554/OTRA9033

Falade, P. V. (2023). Decoding the Threat Landscape : ChatGPT, FraudGPT, and WormGPT in Social Engineering Attacks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *9*(5), 185–198. doi:10.32628/CSEIT2390533

Gandotra, E., & Gupta, D. (2021). An Efficient Approach for Phishing Detection using Machine Learning, 239–253. doi:10.1007/978-981-15-8711-5_12

George, Dr. A. S., George, A. S. H., & Baskar, Dr. T. (2023). Digitally Immune Systems: Building Robust Defences in the Age of Cyber Threats. *Partners Universal International Innovation Journal*, *1*(4), 155–172. doi:10.5281/ZENODO.8274514

Ghafir, I., Saleem, J., Hammoudeh, M., Faour, H., Prenosil, V., Jaf, S., … Baker, T. (2018). Security threats to critical infrastructure: the human factor. *The Journal of Supercomputing 2018 74:10*, *74*(10), 4986–5002. doi:10.1007/S11227-018-2337-2

Hasan, K., Hossain, F., Amin, A., Sutradhar, Y., Jeny, I. J., & Mahmud, S. (2025). Enhancing Proactive Cyber Defense: A Theoretical Framework for AI-Driven Predictive Cyber Threat Intelligence. *Journal of Technologies Information and Communication*, *5*(1), 33122. doi:10.55267/RTIC/16176

Hashmi, A., Adil Shahzad, S., Lin, C.-W., Tsao, Y., Member, S., & Wang, H.-M. (2024). Understanding Audiovisual Deepfake Detection: Techniques, Challenges, Human Factors and Perceptual Insights. Retrieved from https://arxiv.org/pdf/2411.07650

Hughes-Lartey, K., Li, M., Botchey, F. E., & Qin, Z. (2021). Human factor, a critical weak point in the information security of an organization's Internet of things. *Heliyon*, *7*(3), e06522. doi:10.1016/j.heliyon.2021.e06522

Hussain, M. J. (2024). A Survey Based on Behavior Analysis of Artificial Intelligence Using Machine Learning Process. *4th International Conference on Sustainable Expert Systems, ICSES 2024 - Proceedings*, 1694–1701. doi:10.1109/ICSES63445.2024.10763264

Jabir, R., Le, J., & Nguyen, C. (2025). Phishing Attacks in the Age of Generative Artificial Intelligence: A Systematic Review of Human Factors. *AI 2025, Vol. 6, Page 174*, *6*(8), 174. doi:10.3390/AI6080174

Kavya, S., & Sumathi, D. (2024). Staying ahead of phishers: a review of recent advances and emerging methodologies in phishing detection. *Artificial Intelligence Review 2024 58:2*, *58*(2), 50-. doi:10.1007/S10462-024-11055-Z

Kim, D., Ahn, M. K., Lee, S., Lee, D., Park, M., & Shin, D. (2023). Improved Cyber Defense Modeling Framework for Modeling and Simulating the Lifecycle of Cyber Defense Activities. *IEEE Access*, *11*, 114187–114200. doi:10.1109/ACCESS.2023.3324901

Kultan, J., Meruyert, S., Danara, T., Nassipzhan, D., & Gumilyov, E. N. (2025). Application of Computer Vision Methods for Information Security. *R&E-SOURCE*, *12*(1), 161–177. doi:10.53349/RE-SOURCE.2025.IS1.A1389

Kure, H. I., Sarkar, P., Ndanusa, A. B., & Nwajana, A. O. (2025). Detecting and Preventing Data Poisoning Attacks on AI Models, 4–8. Retrieved from https://arxiv.org/pdf/2503.09302

Lopes Antunes, D., & Llopis Sanchez, S. (2023). The Age of fighting machines: The use of cyber deception for Adversarial Artificial Intelligence in Cyber Defence. *ACM International Conference Proceeding Series*. doi:10.1145/3600160.3605077;ISSUE:ISSUE:DOI

Mathur, S., & Singh, A. (2025). Cybersecurity and Forensics for Multimedia Investigation. *Cyber Security, Forensics and National Security*, 244–281. doi:10.1201/9781003497868-13/CYBERSECURITY-

FORENSICS-MULTIMEDIA-INVESTIGATION-SURBHI-MATHUR-ANUBHAV-SINGH

Mir, A. W., & Ramachandran, R. K. (2021). Implementation of Security Orchestration, Automation and Response (SOAR) in Smart Grid-Based SCADA Systems, 157–169. doi:10.1007/978-981-16-1335-7_14

Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity: a deep dive into state-of-the-art techniques and future paradigms. *Knowledge and Information Systems 2025 67:8*, *67*(8), 6969–7055. doi:10.1007/S10115-025-02429-Y

Mosa, M. J., Barhoom, A. M., Alhabbash, M. I., Harara, F. E. S., Abu-Nasser, B. S., & Abu-Naser, S. S. (2024). AI and Ethics in Surveillance: Balancing Security and Privacy in a Digital World. *International Journal of Academic Engineering Research*. Retrieved from https://philpapers.org/rec/MOSAAE-2

Nailwal, S., Singh, N. T., Raza, A., & Singhal, S. (2023). Deepfake Detection: A Multi-Algorithmic and Multi-Modal Approach for Robust Detection and Analysis. *2023 IEEE International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering, RMKMATE 2023*. doi:10.1109/RMKMATE59243.2023.10369155

Pienta, D., Thatcher, J. B., & Johnston, A. (2020). Protecting a whale in a sea of phish. *Journal of Information Technology*, *35*(3), 214–231. doi:10.1177/0268396220918594;WEBSITE:WEBSITE:SAGE;ISSUE:ISSUE:DOI

Rathore, P., & Kolhe, K. (2026). Integrating Automation and Orchestration in Security Incident Handling: A Review of SOAR Frameworks and Platforms. *Mechanisms and Machine Science*, *185*, 529–551. doi:10.1007/978-3-031-95963-9_38

Salloum, S., Gaber, T., Vadera, S., & Shaalan, K. (2022). A Systematic Literature Review on Phishing Email Detection Using Natural Language Processing Techniques. *IEEE Access*, *10*, 65703–65727. doi:10.1109/ACCESS.2022.3183083

Sarker, I. H. (2022). Machine Learning for Intelligent Data Analysis and Automation in Cybersecurity: Current and Future Prospects. *Annals of Data Science 2022 10:6*, *10*(6), 1473–1498. doi:10.1007/S40745-022-00444-2

Seo, B. S., Baek, J. M., & Ko, K. M. (2024). AI-Enabled Abnormal Behavior Detection and Visualization Technology on Blockchain Network. *2024 International Conference on Electronics, Information, and Communication, ICEIC 2024*. doi:10.1109/ICEIC61013.2024.10457211

Simone Raponi, B., & Di Pietro, R. (2021). AI-driven Detection of Cybersecurity-related Patterns.

Soe, Y. N., Feng, Y., Santosa, P. I., Hartanto, R., & Sakurai, K. (2019). Rule Generation for Signature Based Detection Systems of Cyber Attacks in IoT Environments. *Bulletin of Networking, Computing, Systems, and Software*, *8*(2), 93–97. Retrieved from http://ww.bncss.org/index.php/bncss/article/view/113

Sree, V. S., Koganti, C. S., Kalyana, S. K., & Anudeep, P. (2021). Artificial Intelligence Based Predictive Threat Hunting in the Field of Cyber Security. *2021 2nd Global Conference*

for Advancement in Technology, GCAT 2021. doi:10.1109/GCAT52182.2021.9587507

Subrahmanyam, S. (2025). Behavioral analysis for threat detection. *Handbook of AI-Driven Threat Detection and Prevention: A Holistic Approach to Security*, 95–115. doi:10.1201/9781003521020-6/BEHAVIORAL-ANALYSIS-THREAT-DETECTION-SATYA-SUBRAHMANYAM

Thomas, K., Grier, C., Ma, J., Paxson, V., & Song, D. (2011). Design and evaluation of a real-time URL spam filtering service. *Proceedings - IEEE Symposium on Security and Privacy*, 447–462. doi:10.1109/SP.2011.25

Tiwari, S. (2021). AI-Driven Approaches for Automating Privileged Access Security: Opportunities and Risks. *SSRN Electronic Journal*. doi:10.2139/SSRN.5259381

Tounsi, W., & Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & Security*, *72*, 212–233. doi:10.1016/J.COSE.2017.09.001

Vainio-Pekka, H., Agbese, M. O. O., Jantunen, M., Vakkuri, V., Mikkonen, T., Rousi, R., & Abrahamsson, P. (2023). The Role of Explainable AI in the Research Field of AI Ethics. *ACM Transactions on Interactive Intelligent Systems*, *13*(4), 26. doi:10.1145/3599974;WGROUP:STRING:ACM

Wang, Z., Ma, J., Wang, X., Hu, J., Qin, Z., & Ren, K. (2023). Threats to Training: A Survey of Poisoning Attacks and Defenses on Machine Learning Systems. *ACM Computing Surveys*, *55*(7). doi:10.1145/3538707;SERIALTOPIC:TOPIC:ACM-PUBTYPE

Zhao, J., Masood, R., & Seneviratne, S. (2021). A review of computer vision methods in network security. *IEEE Communications Surveys and Tutorials, 23*(3), 1838–1878. doi:10.1109/COMST.2021.3086475

NIFTY 50 ENDEKSİNDEKİ VERİLERİN ZAMAN SERİSİ MODELİ PROPHET İLE ANALİZİ

Bekir PARLAK[1]

1. GİRİŞ

Finansal zaman serisi analizleri, son zamanlarda hem akademi hem de finans alanında araştırmacılar için üst tercihlerden birisidir. [1]. Bu çalışma kapsamında, Hindistan'da bulunan borsa piyasasının endekslerinden birisi olan NIFTY50'de yer alan şirketlerin finansal verileri üzerinden zaman serisi analizi ile tahmin modelleri geliştirilmesi hedeflenmiştir. Zaman serisi analizi, finansal verilerde trend, mevsimsellik ve tahmin çalışmalarında çokça kullanılan bir teknik olmakla birlikte aralarında en iyisidir.[2]. Bu analiz belirli bir süre boyunca toplanan verilerin incelenip geleceğe dair bir tahmin yapmamızı sağlar. Zaman serisi analizinde sıklıkla ARIMA, Prophet, LSTM ve SARIMA gibi yöntemler kullanılmaktadır [3][4].

Bu çalışma kapsamında, Prophet zaman serisi analiz yöntemleri ile NIFTY50 endeksine ait bir şirketin kapanış fiyatları üzerinde tahmin çalışması gerçekleştirilecektir. Prophet modeli, mevsimsellik ve trend özelliklerini esnek bir şekilde modelleyebilmesi ile öne çıkmaktadır.[5] Trend, zaman serisinin genel yönelimini/eğilimini ifade eder. Mevsimsellik ise düzenli aralıklarla veya periyotlarla tekrar eden desenlerdir.[6] ARIMA

[1] Doç. Dr, Amasya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, ORCID: 0000-0001-8919-6481.

modeli ise zaman serisinin özüne uygun yapısal modeller geliştirerek tahminlerde bulunur [7].

2. YÖNTEMLER

2.1. Zaman Serisi Analizi

Zaman serisi analizi, bir dizi verinin zamana göre incelenerek gelecekteki trend ve olasılıkların durumunun tahmin edilmesini sağlar. Finansal piyasalarda, zaman serisi analizi tek değişken üzerinden mevsimselliği belirlemek, trend analizi yapmak ve uzun vadeli tahminler yapmak için kullanılır. Özellikle ARIMA ve Prophet modelleri, finansal verilerin karmaşık yapısını çözümleme konusunda daha başarılıdır [5][8].

2.2. Prophet Modeli

Prophet, Facebook tarafından geliştirilen açık kaynak kodlu bir model. Facebook'un veri bilimi takımı tarafından hazırlandığı için tahmin edeceğiniz üzere mevsimsellik ve trend analizlerini diğer modellerden daha iyi yapmaktadır.[9][10] Bu model, doğrusal veya lojistik bir trend yaklaşımıyla çalışabilir ve kullanıcı tarafından tanımlanan tatil etkilerini de hesaba katabilir. Finansal piyasaların dalgalanmasına ve kolay etkilenmesine rağmen, Prophet modelinin çok yönlülüğü, onu özellikle kısa ve orta vadeli tahminlerde daha yakın bir seçenek haline getirmiştir. Bu model, trend değişiklik noktalarını otomatik olarak belirleyebilir ve zaman serisinde belirgin olan haftalık veya yıllık mevsimselliği modelleyebilir [8][9][10][11].

3. VERİ SETİ

NIFTY50 Veri seti kaggle sitesinden alınmıştır. (date(tarih), symbol (sembol),series (güvenlik tipi),prev close (önceki kapanış), open, high, low, last, close, VWAP (Hacim

ağırlıklı ortalama)) değerlerini içermektedir. 15 sütundan, 3201 satırdan oluşmaktadır. 50 Firma için de aynı şekilde veriler mevcuttur.

Veri önişleme, her modelde olduğu gibi zaman serisi analizi sürecinde de kritik bir adımdır ve model performansının doğruluğu üzerine etkisi olabilir.[12] Bu çalışmada, NIFTY50 endeksine ait kapanış fiyatları verisi üzerinde aşağıdaki veri önişleme adımları uygulanmıştır:

- Eksik veri kontrolü ve işlenmesi: Kaggle'dan aldığımız NIFTY 50 veri setinde eksik veri kontrolü sonucunda eksik veri olmadığı görülmüştür.
- Zaman Sütunu değişimi: Verimizde 'Date' sütunu object türünde olduğu için model eğitiminde kullanılamıyordu. Bunu çözmek için DateTime'a çevrilmiştir..

df['ds'] = pd.to_datetime(df['ds'])

- Trend ve Mevsimsellik: Prophet modeli varsayılan olarak yıllık ve haftalık mevsimsellik bileşenlerini analiz eder. İlgili grafikleri ve örneklerini ilerde değinilecektir.
- Veri setinin hazırlanması: Veri, eğitim için ilk %80, test için ise ilk %20'lik kısım olacak şekilde bölünmüştür.

Veri önişleme adımları tamamlandıktan sonra , Prophet modelinin uygulanmasına geçilmiştir. Bu adımlar model performansını optimize etmek ve daha doğru tahminler için oldukça önemlidir[13].

Tablo 1. Bir firmanın örnek verileri

Date	Symbol	Series	Prev Clos	Open	High	Low	Last	Close
2008-05-26	BAJAJFINSV	EQ	2101.05	600	619	501	505.1	509.1
2008-05-27	BAJAJFINSV	EQ	509.1	505	610.95	491.1	564	554.65
2008-05-28	BAJAJFINSV	EQ	554.65	564	665.6	564	643	640.95
2008-05-29	BAJAJFINSV	EQ	640.95	656.65	703	608	634.5	632.4
2008-05-30	BAJAJFINSV	EQ	632.4	642.4	668	588.3	647	644
2008-06-02	BAJAJFINSV	EQ	644	650	699	622	687	686.95
2008-06-03	BAJAJFINSV	EQ	686.95	672	689.8	632.55	670.55	672.05
2008-06-04	BAJAJFINSV	EQ	672.05	674	674	566.6	595	598.95
2008-06-05	BAJAJFINSV	EQ	598.95	603	639.8	572.3	631.5	631.85

Bu haliyle ‘Date’ sütununun veri tipi object olduğu için Modelimize uygun bir şekilde tarihleri geçiremiyoruz. Çözüm olarak aşağıdaki satırlar ile çözüm ürettik.

```
df.columns = ['ds', 'y']  # Prophet formatına uygun hale getirme
df['ds'] = pd.to_datetime(df['ds'])
```

Tablo 2. Modele uygun tarih ve Close sütunu

Index	ds	y
0	2008-05-26 00:00:00	509.1
1	2008-05-27 00:00:00	554.65
2	2008-05-28 00:00:00	640.95
3	2008-05-29 00:00:00	632.4
4	2008-05-30 00:00:00	644
5	2008-06-02 00:00:00	686.95

Bu çalışmada kapanış değerleri üzerinden tahmin ve işlem yapacağımız için verinin son hali budur[4].

4. DENEYSEL ÇALIŞMA

Firmalardan bazıları Prophet modeliyle %80’i eğitim %20’si test şeklinde ayırdığımız veriyi girdi olarak verip eğitimi tamamladıktan sonra test verisindeki tarihlere karşılık gelen kapanış değerlerini modelimize tahmin ettirmeye çalışılmıştır. Bazı firmalar için model test edilmiştir. Firmaların iş alanları:

1. BAJAJFINSV: Bankacılık dışı finansal hizmetler
2. ADANIPORTS: Lojistik ve Liman operatörü
3. MARUTI: Otomotiv
4. RELIANCE: Enerji, petrokimya, tekstil, doğalgaz, eğlence, perakende, medya, eğlence vb.

Şekil 1. BAJAJFINSV firmasının tahmin ve gerçek değerleri

Tablolarda ; eğitim verisi mavi, gerçek değer turuncu, tahmin değerleri ise yeşil renkte gösterilecektir. Şekil 1 için, tahmin gerçek değerin genelde daha üstündedir.

Şekil 1'deki model tahmininin sonuçlarının performans ölçümleri şöyledir.

Accuracy: 70.74% (1-MAPE)

Mean Absolute Error (MAE): 1794.76

Mean Squared Error (MSE): 5849742.95

Mean Absolute Percentage Error (MAPE): 29.26%

5.1. Mean absolute error (MAE)

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| \qquad (1)$$

Burada N örnek sayısıdır, y_i gerçek değerdir, $\hat{y}_i$ tahmin edilen değerdir, e= y_i-$\hat{y}_i$ tahmin hatasıdır, üstü çizgili y gerçek değerlerin ortalaması, üstü çizgili e hataların ortalamasıdır. Denklem (1) ortalama mutlak hatayı gösterir. Tahmin hatalarının mutlak değerlerinin ortalaması olarak tanımlanır.[14]

5.2. Mean square error (MSE)

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2 \qquad (2)$$

(best value = 0; worst value = +∞)

Tespit edilmesi gereken aykırı değerler için kullanılır. Formülde kare olduğu için aykırı değerler hatayı daha da büyütür.[15]

5.3. Mean absolute percentage Error (MAPE)

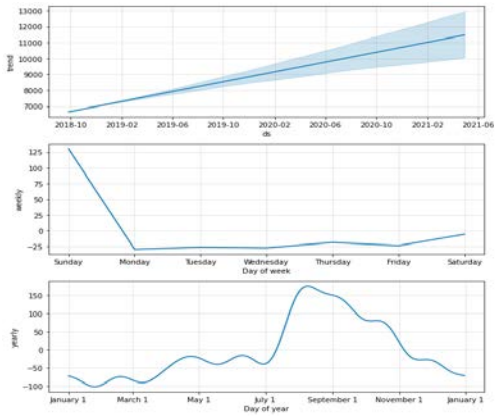
$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i - X_i}{Y_i}\right| \quad (3)$$

(best value = 0; worst value = +∞)

MAPE göreceli hata açısından çok sezgisel yorum yapan bir modeldir. Mutlak varyasyonlardan ziyade göreceli varyasyonlara duyarlı olmanın daha önemli olduğu görevlerde kullanılması önerilir.[15][16][17]

Arkaplandaki analizin çıktısı için components_fig = model.plot_components(forecast) değişkeninin görseli aşağıdadır.

Şekil 2. BAJAJFINSV trend, haftalık ve yıllık mevsimsellik değerleri



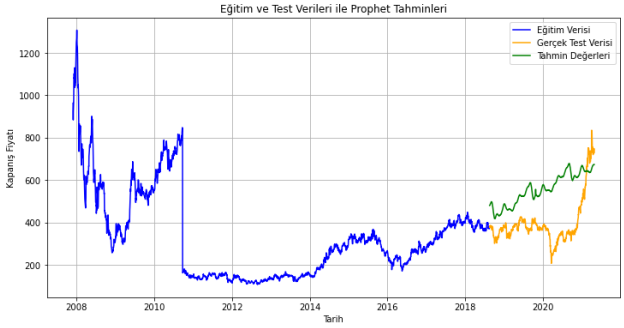
Görsel incelendiğinde, En üstteki grafik Trend grafiğidir. Zaman serisinin genel eğilimini gösterir. Grafik, zamanla kapanış fiyatlarında yükseliş trendi olduğunu göstermektedir. Mavi alan ise belirsizliği göstermektedir, zaman ilerledikçe belirsizlik artmaktadır.

Ortadaki grafik, haftanın günlerine bağlı fiyat değişikliklerini göstermektedir. Pazartesi günleri fiyatlar daha düşükken cumartesi ve pazar gününe doğru artış olduğu görülmektedir.

En alttaki grafik, yıldaki aylara göre fiyat değişikliklerini göstermektedir. Temmuz ve eylül aylarında artarken , kasım ve aralık aylarında düştüğü görülmektedir. [19]

Daha derin analizlerle bu farklılıkların neden kaynaklandığı araştırılıp daha doğru sonuçlar alınabilir. Örneğin firmanın alanına göre kışın daha fazla satış yapıp yazın daha az satış yapma durumlarına göre etkiler bulunabilir.

Şekil 3. ADANIPORTS firmasının gerçek ve tahmini değerleri



Accuracy: 51.84%

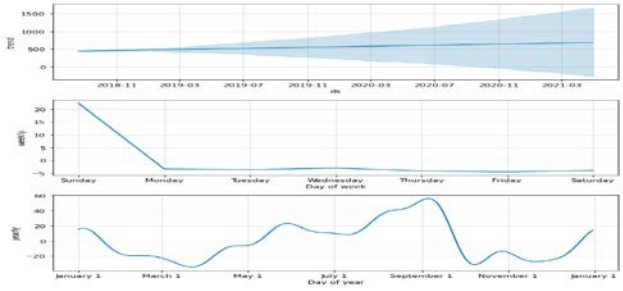
Mean Absolute Error (MAE): 171.33

Mean Squared Error (MSE): 36200.63

Mean Absolute Percentage Error (MAPE): 48.16%

Şekil 3 ve hata değerleri incelendiğinde önceki firmadan daha kötü bir tahmin yaptığı görülmektedir. Bu tahminlerde de gerçek değerden daha yüksek tahminler yapıldığı görülmektedir.

Şekil 4. ADANIPORTS firmasının trend, haftalık ve yıllık mevsimsellik değerleri



Trend grafiği, zaman geçtikçe artış çok az da olsa yatay trend göstermektedir ve belirsizlik önceki firmaya göre daha da fazla. Model uzun vadede fiyatların sabit kalacağını öngördü.

Haftalık mevsimsellik grafiği, Pazar günleri en yüksek ama pazartesi ve salı düşüş eğiliminde.

Yıllık mevsimsellik grafiği, Ekim-Kasım aylarında yükseliş görülmekte. Yani ekonomik hareketlilik veya yatırımcı durumunu yansıtabilir. Mayıs-Temmuz aylarındaki düşüş de mevsimsel bir faktöre bağlı olabilir.

Şekil 5. MARUTI firmasının gerçek ve tahmini değerleri

Accuracy: 58.24%

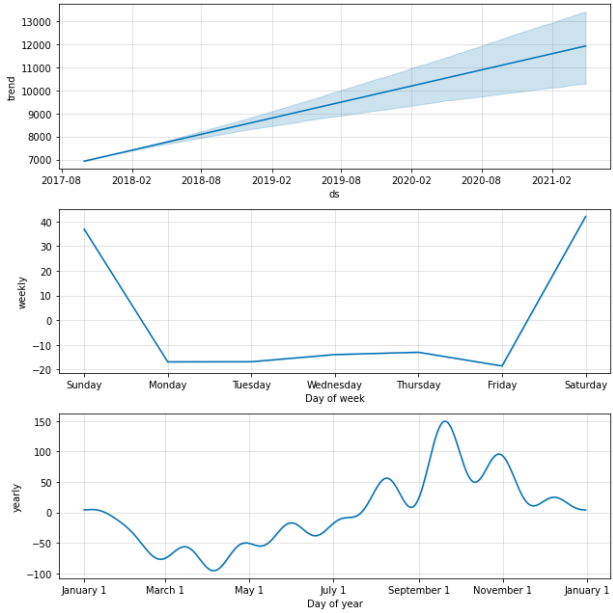
Mean Absolute Error (MAE): 2792.45

Mean Squared Error (MSE): 9964805.95

Mean Absolute Percentage Error (MAPE): 41.76%

Şekil 5 incelendiğinde, tahmin değerleri ilk safhada gerçek değerin altında olurken 2019'dan sonra gerçek değerin üstünde tahmin yaptığı görülmüştür.

Şekil 6. MARUTI firmasının trend ve mevsimsellik değerleri



Şekil 6'daki trend grafiği incelendiğinde, zaman geçtikçe artış olduğu görülmektedir. Belirsizlik bandı da zaman geçtikçe belirsizlik artıyor.

Haftalık mevsimsellik, Pazar ve cumartesi günleri fiyatlar daha yüksekken hafta içi fiyatlar daha düşük.

Yıllık mevsimsellik, eylül-kasım aralığında fiyatlar zirvede. Yıl sonuna doğru tekrar düşüşe geçtiği, yazın tekrar yükseldiği görülmektedir.

Şekil 7. RELIANCE firmasının gerçek ve tahmini değerleri

Accuracy: 80.20%

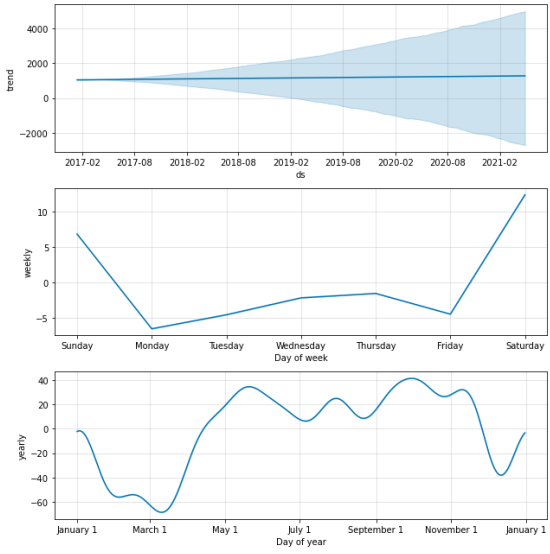
Mean Absolute Error (MAE): 309.55

Mean Squared Error (MSE): 170916.48

Mean Absolute Percentage Error (MAPE): 19.80%

Şekil 7 incelendiğinde, gerçek değerlerle içli dışlı ve yüksek doğrulukta bir tahmin yaptığı görülmektedir. Şu ana kadarki en yüksek doğruluktaki tahmin bu firmaya aittir.

Şekil 8. RELIANCE firmasının trend ve mevsimsellik değerleri



Şekil 8'deki trend grafiği incelendiğinde, az artışa sahip yatay bir trend görülmektedir. Belirsizlik de zaman ilerledikçe çok artış göstermekte.

Haftalık mevsimsellik incelendiğinde, en yüksek değer cumartesi günü. En düşük değer ise pazartesi ve Salı günü en düşük değerde.

Yıllık mevsimsellik, şubat ve mart aylarında en düşük değerde. Mayıs ile kasım arasında dalgalı artış ve iniş olsa da genel değer ortalamasının yüksek olduğu görülmektedir.

5. SONUÇLAR

Bu çalışmada, NIFTY50 endeksinde yer alan firmaların kapanış fiyatları PROPHET modeli kullanılarak tahmin edilmiştir. PROPHET modeli, trend ve mevsimsellikteki başarısından tercih edilmiştir. Veri setinde 4 firmanın verilerinden, kullanılmayan sütunları ve boş verileri doldurduktan sonra 80 eğitim 20 test olacak şekilde model eğitilmiştir. Tahmin edilen 20lik kısmında MAE, MAPE, MSE metrikleriyle modelin başarısı incelenmiştir. Firmalara göre tahmin başarı yüzdeleri incelendiğinde en iyi tahmin yaptığı firmanın RELIANCE olduğu görülmüştür. En kötü tahmin ettiği firma ise ADANIPORTS çıkmıştır. Her bir firma için trend, haftalık ve yıllık mevsimsellik grafikleri çıkartılıp yorumlanmıştır. Daha ileri analizlerde verideki diğer sütunlar için modeller üretilip hangisi için daha doğru tahminler yapılabileceği incelenebilir. Daha fazla firma için bu model denenebilir. Hindistan'daki borsa piyasasından eğittiğimiz model ile başka ülkelerdeki borsa piyasalarında tahmin yapılabilir.

KAYNAKÇA

[1] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied soft computing, 90, 106181.

[2] Devi, B. U., Sundar, D., & Alli, P. (2013). An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50. International Journal of Data Mining & Knowledge Management Process, 3(1), 65.

[3] Taylor, S. J., & Letham, B. (2018). Forecasting at scale. The American Statistician, 72(1), 37-45.

[4] Satrio, C. B. A., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. Procedia Computer Science, 179, 524-532.

[5] Yenidoğan, I., Çayir, A., Kozan, O., Dağ, T., & Arslan, Ç. (2018, September). Bitcoin forecasting using ARIMA and PROPHET. In 2018 3rd international conference on computer science and engineering (UBMK) (pp. 621-624). IEEE.

[6] Kitagawa, G., & Gersch, W. (1984). A smoothness priors–state space modeling of time series with trend and seasonality. Journal of the American Statistical Association, 79(386), 378-389.

[7] Hyndman, R. J. (2018). Forecasting: principles and practice. OTexts.

[8] Garlapati, A., Krishna, D. R., Garlapati, K., Rahul, U., & Narayanan, G. (2021, April). Stock price prediction using Facebook Prophet and Arima models. In 2021 6th

International Conference for Convergence in Technology (I2CT) (pp. 1-7). IEEE.

[9] Jha, B. K., & Pande, S. (2021, April). Time series forecasting model for supermarket sales using FB-prophet. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 547-554). IEEE.

[10] Khayyat, M., Laabidi, K., Almalki, N., & Al-Zahrani, M. (2021). Time Series Facebook Prophet Model and Python for COVID-19 Outbreak Prediction. Computers, Materials & Continua, 67(3).

[11] Ning, Y., Kazemi, H., & Tahmasebi, P. (2022). A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. Computers & Geosciences, 164, 105126

[12] García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. Big data analytics, 1, 1-22.

[13] Piramuthu, S. (2004). Evaluating feature selection methods for learning in data mining applications. European journal of operational research, 156(2), 483-494.

[14] González-Sopeña, J. M., Pakrashi, V., & Ghosh, B. (2021). An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renewable and Sustainable Energy Reviews, 138, 110515.

[15] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. Peerj computer science, 7, e623.

[16] De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. Neurocomputing, 192, 38-48.

[17] Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. International journal of forecasting, 8(1), 69-80.

[18] Gupta, G. K., & Ghosh, J. (2001, April). Detecting seasonal trends and cluster motion visualization for very high dimensional transactional data. In Proceedings of the 2001 SIAM International Conference on Data Mining (pp. 1-17). Society for Industrial and Applied Mathematics.

HABER VERİLERİNİN MAKİNE ÖĞRENME VE DERİN ÖĞRENME TEKNİKLERİ İLE SINIFLANDIRILMASI

Bekir PARLAK[1]

1. GİRİŞ

Günümüzde haber kaynaklarının artması ve bu kaynakların çokluğu, kişilerin bilgiye ulaşmasını zor hale getirmiştir. Bu sebeple, bilgiye vaktinde ve doğru bir şekilde erişmenin zorlaştığından, haberlerin kategorilere ayrılması ciddi bir çözüm olarak karşımıza çıkmaktadır [1].

Doğal dil işleme (NPL) teknikleri, haber sınıflandırma çalışmalarında ilerleme göstermiştir. Makine öğrenimi yöntemlerinden biri olan Destek Vektör Makineleri (SVM), özellikle çok boyutlu veri setleri üzerinde iyi performans gösterirken [2], Uzun Kısa Süreli Bellek (LSTM) modeli, sıralı veri analizinde üstün bir performans sergilemektedir [3]. Haber başlıklarının kategorilere ayrılmasında SVM ve LSTM modelleri karşılaştırılmaktadır.

Yapılan çalışmalar, haberlerin sınıflandırılmasında farklı yöntemlerin etkili bir şekilde kullanıldığını ortaya sunmuştur. Örneğin, metin madenciliğinde Destek Vektör Makineleri (SVM) ile Uzun Kısa Süreli Bellek (LSTM) modelinin karşılaştırmalı incelemeleri yapılmıştır [4]. Bu çalışmalar, NLP modelinin, haberlerin etkili bir şekilde sınıflandırılmasına nasıl katkıda bulunduğunu göstermektedir [5].

[1] Doç. Dr., Amasya Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, ORCID: 0000-0001-8919-6481.

Buna ek olarak, TF-IDF (Term Frequency-Inverse Document Frequency) yöntemi, metinlerde önemli kelimeleri tespit etmek için etkili bir yöntem olarak öne çıkar. Bu yöntem hem klasik sınıflandırma algoritmalarını hem de derin öğrenme modelleriyle uyumlu bir şekilde çalışabilir [6]. Özellikle LSTM gibi sıralı modeller, TF-IDF ile güçlendirerek daha komplike metin yapılarına karşı yüksek performans sergilemektedir [7].

Haberlerin otomatik olarak kategorize edilmesi, kişilerin bilgiye daha hızlı ulaşmasını sağlarken, aynı zamanda haber ajansları ve medya kuruluşlarının iş yükünü azaltan bir otomasyon aracı olarak da büyük bir önem taşımaktadır. Bu doğrultuda, haberlerin başlıklarına göre sınıflandırılmasında SVM ve LSTM modelleri kullanılarak haber başlıklarının etkin bir şekilde sınıflandırılması amaçlanmıştır.

2. YÖNTEM

Bu bölümde, veri setinin temizlenmesi ve işlenmesi, metin verilerinin TF-IDF kullanılarak sayısal hale getirilmesi, veri setinin test ve eğitim verisi ayrılması ve sınıflandırma modellerinin eğitilmesi anlatılmıştır.

2.1. Veri Seti ve Ön İşleme

Kaggle'dan alınan News Category Dataset adlı veri seti, haberlerin; başlık, link, kısa açıklama, anahtar kelime ve kategori bilgilerini içermektedir [8]. Eksik veriler temizlenmiş, başlık ve açıklama metinleri kelime köküne indirgenmiş ve durak kelimeleri ("the", "is", "in", "and", "to", "of", "a", "on", "for", "with", "as", "by", "an", "at", "this", "that", "it", "from", "or", "be", "are", "was", "were", "has", "have", "not", "but", "we", "they", "you", "he", "she", "which", "about") temizleme adımları uygulanmıştır. Metinlerin temizlenmesinde, NLTK kütüphanesi kullanılarak şu adımlar izlenmiştir:

- Noktalama işaretleri, sayılar ve özel karakterlerin temizlenmesi.
- Tüm metinlerin küçük harfe dönüştürülmesi.
- Durak kelimelerin (stopwords) kaldırılması.
- Kelimelerin köklerine indirgenmesi (lemmatization).

Bu adımlar, veri setinin makine öğrenimi modellerine uygun hale getirmiştir.

2.2. TF-IDF Vektörizasyonu

TF-IDF, metin madenciliğinde kelimelerin ne kadar önemli olduğunu ölçen bir yöntemdir. Bu yöntem, iki ana bileşenden oluşur: Bir kelimenin bir belgede ne kadar sık geçtiğini gösteren Term Frequency (TF) ve kelimenin tüm belgedeki nadirliğini ölçen Inverse Document Frequency (IDF). TF-IDF, bu sayede anlamlı ve dikkat çekici kelimeleri bulmamıza yardımcı olur.

Bu çalışmada en sık geçen 5000 kelime seçilmiştir.

2.3. Veri Setinin Eğitim ve Test Verisi Olarak Ayrılması

Modelleri Yorumlayabilmek için veri setini %80 eğitim %20 test olacak şekilde ayrılmıştır.

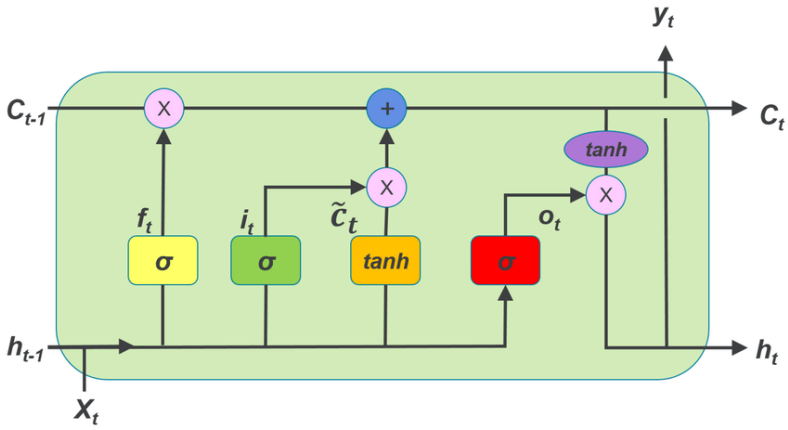
2.4. Modellerin Kullanılması

Destek Vektör Makineleri (SVM): İlk olarak, Vapnik ve Cortes (1995) tarafından, sınıflandırma ve örüntü problemlerine çözüm sunmak amacıyla geliştirilmiştir [9]. Denetimli öğrenme yöntemlerinden biri olup genellikle verilerin sınıflandırılması veya regresyon işlemleri için kullanılır. SVM, verileri ayıran bir sınır bulmayı amaçlar ve bu sınır, sınıflar arasındaki mesafeyi en geniş yapacak şekilde belirlenir [10]. Böylece modelin daha iyi genelleme yapması sağlanır. Eğer veriler doğrusal değilse, SVM

çekirdek (kernel) fonksiyonları kullanarak veriyi daha yüksek bir boyuta taşıyıp daha doğru sonuçlar elde etmeye çalışılır.

Uzun Kısa Süreli Bellek (LSTM): Hochreiter ve Schmidhuber tarafından 1997 yılında tanıtılan LSTM, derin öğrenmede kullanılan bir tür tekrarlayan sinir ağı (RNN) modelidir [11]. LSTM, özellikle sıralı ve zaman serisi verileri üzerinde geçmiş bilgileri uzun süre hatırlama yeteneğine sahiptir. Geleneksel RNN'lerdeki uzun vadeli bağımlılık sorununu çözmek için geliştirilmiştir. LSTM, hücre durumu ve kapılar (input, forget, output) kullanılarak bilgi akışını düzenler [12]. Gereksiz bilgileri unutur ve önemli olanları hatırlayarak daha verimli öğrenir. Bu özellikler sayesinde metin, ses ve video gibi sıralı veri türlerinde başarılı sonuçlar elde edilir.

Şekil 1. LSTM mimarisi (Ismail, Wood, & Bravo, 2018)



Şekil 1'de σ sigmoid fonksiyonu, ht gizli birim, it giriş kapısı, ot çıkış kapısı, ft unutma kapısı, ct ve c̃t bellek hücrelerini ifade etmektedir [13].

2.5. Veri Dağılımı Analizi

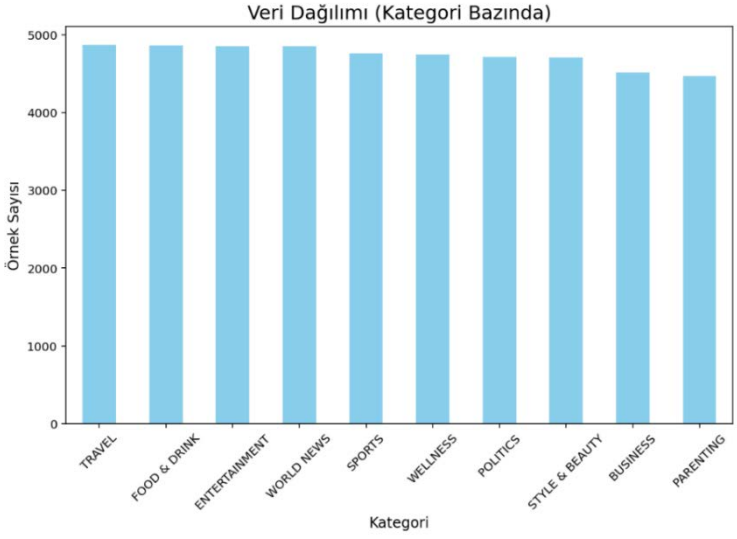
Veri setindeki haberlerin kategoriye göre dağılımı Tablo 1'de verilmiştir.

Tablo 1. Veri Dağılımı

Kategori	Haber Sayısı
TRAVEL	4865
FOOD & DRINK	4863
ENTERTAINMENT	4855
WORLD NEWS	4851
SPORTS	4759
WELLNES	4741
POLITICS	4712
STYLE & BEAUTY	4708
BUSINESS	4512
PARENTING	4466

Veri setinin genel olarak dengeli bir şekilde kategorilere ayrıldığı görülmektedir. Fakat bazı kategorilerdeki küçük farklar sınıflandırma performansını etkileyebilir. Aşağıda veri dağılımını görselleştiren bir grafik yer almaktadır:

Şekil 2. Haber veri setindeki kategori bazlı dağılım grafiği. Grafikte kategorilerin toplam haber sayıları gösterilmektedir.



Şekil 2. Haber veri setindeki kategori bazlı dağılım grafiği. Grafikte kategorilerin toplam haber sayıları gösterilmektedir.

2.6. Değerlendirme

Sınıflandırma işleminin değerlendirilmesinde, deney sonuçlarını analiz etmek için farklı ölçüm yöntemleri kullanılmaktadır [14]. Değerlendirme ölçütü, sınıflandırıcının performansını belirleyen bir araçtır. Farklı ölçüm metrikleri, sınıflandırıcıların çeşitli özelliklerini analiz etmeye yönelik kullanılır [15]. En etkili algoritmayı belirlemek için birçok farklı kriter kullanılmaktadır. Bu kriterler arasında doğru negatif (DN) oranı, doğru pozitif (DP) oranı, yanlış negatif (YN) oranı, yanlış pozitif (YP) oranı, geri çağırma (recall), doğruluk (accuracy), f1-skor, hassasiyet (precision), support (destek), makro ortalama (macro avg) ve ağırlıklı ortalama (weighted avg) yer almaktadır [16, 17].

Support: Bir sınıfın veri setindeki örnek sayısını ifade eder.

Macro Avg: Her bir sınıfın değerlendirmesi sonucu metriklerin eşit ağırlıkla dikkate alarak hesaplanmasıdır.

Weighted Avg: Sınıfların veri setindeki örnek sayısına göre ağırlıklandırılarak hesaplanan bir ortalamadır.

Precision: Bir sınıflandırma modelinin pozitif sınıf olarak tahmin ettiği ögelerin ne kadarını gerçekten doğru olduğunu gösteren bir ölçümdür. Yani modelin pozitif olarak işaretlediği ögelerin ne kadarının doğru tahmin olduğunu belirler.

Recall: Modelin gerçek pozitif örneklerini ne kadar doğru bulduğunu gösteren bir ölçümdür. Yani gerçekten pozitif olan verilerin ne kadarını doğru şekilde sınıflandırdığını ölçer.

F1-Score: Hassasiyet (precision) ve geri çağırma (recall) arasındaki dengeyi ölçen bir metriktir. İki değeri dikkate alarak, modelin performansını değerlendirmemizi sağlar.

Accuracy: Modelin doğru tahmin ettiği örneklerin, toplam tahmin edilen örneklere oranıdır. Yani modelin ne kadar doğru sonuç verdiğini gösteren bir ölçümdür.

SVM Modelinin Performansı: Modelin doğruluğu %78 olarak hesaplanmıştır. Bu da modelin genel anlamda başarılı olduğunu göstermektedir.

Tablo 2. SVM Sınıflandırma Raporu

	Precision	Recall	F1-score	Support
0	0.69	0.77	0.73	955
1	0.72	0.75	0.74	985
2	0.81	0.81	0.81	1021
3	0.74	0.75	0.74	1030
4	0.76	0.70	0.73	1034
5	0.87	0.87	0.87	995
6	0.83	0.81	0.82	986
7	0.79	0.77	0.78	1008
8	0.71	0.70	0.70	1009
9	0.76	0.75	0.75	977
Accuracy			0.78	1000
Marco avg	0.78	0.78	0.78	1000
Weighted avg	0.78	0.78	0.78	1000

LSTM Modelinin Performansı: Modelin doğruluğu %79 olarak hesaplanmıştır. Bu da modelin genel anlamda başarılı olduğunu göstermektedir.

Tablo 3. LSTM Sınıflandırma Raporu

	Precision	Recall	F1-score	Support
0	0.71	0.80	0.75	955
1	0.74	0.78	0.76	985
2	0.84	0.83	0.84	1021
3	0.76	0.76	0.76	1030
4	0.79	0.73	0.76	1034
5	0.89	0.89	0.89	995
6	0.86	0.84	0.85	986
7	0.82	0.79	0.80	1008
8	0.73	0.72	0.72	1009
9	0.79	0.77	0.78	977
Accuracy			0.79	1000
Marco avg	0.79	0.79	0.79	1000
Weighted avg	0.79	0.79	0.79	1000

3. SONUÇLAR

Yapılan çalışmada her iki model sonucunun yüksek doğruluk oranlarına sahip olsa da LSTM modeli, %79 doğruluk oranı ile SVM modelinden biraz daha iyi performans sergilemektedir.

Derin öğrenme alanındaki gelişmeler son zamanlarda oldukça hız kazanmış olup her geçen gün yeni ve yenilikçi mimariler ortaya çıkmaktadır. Sınıflandırma modelleri için bir alternatif olarak farklı derin öğrenme mimarilerinin araştırılması ve denenmesi çalışmayı bir adım daha ileriye taşıyabilir.

KAYNAKÇA

[1] Joachims, T. “Text categorization with Support Vector Machines: Learning with many relevant features”. Machine Learning, 1998.

[2] Zhang, Y., & Yang, Q. “A Survey on Multi-Task Learning”. IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 12, pp. 5586–5609, 2019.

[3] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning–Based Text Classification: A Comprehensive Review. ACM Computing Surveys (CSUR), 54(3), 1-40.

[4] Qiu, X., Sun, T., Xu, Y., Shao, Y., & Dai, N. (2020). Pre-trained Models for Natural Language Processing: A Survey. Science China Technological Sciences, 63(10), 1872-1897.

[5] Manning, C. D., Raghavan, P., & Schütze, H. “Introduction to Information Retrieval”. Cambridge University Press, 2008.

[6] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 328-339).

[7] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186).

[8] https://www.kaggle.com/datasets/setseries/news-category-dataset

[9] Cortes C., Vapnik V., Support-vector networks, Machine Learning, 1995,20(3), 273-297.

[10] Burges C.J., A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, 1998, 2(2), 121-167.

[11] Hochreiter S., Schmidhuber J., Long Short Term Memory, Neural Computation, 1997, 9(8), 1735–1780.

[12] Kocaeli Üniversitesi. "Türkiye'deki Havayolu Firmalarıyla İlgili Sosyal Medya Yorumlarının Makine Öğrenmesi Yöntemleriyle Sınıflandırılması". Kocaeli Üniversitesi, 2018, s. 15.

[13] Çetinkaya, E., & Akyokuş, S. "Türkçe Haber Başlıklarından Konu Tespiti: Topic Detection from Turkish News Texts". ResearchGate, 2022, s. 4.

[14] Kocaeli Üniversitesi. "Türkiye'deki Havayolu Firmalarıyla İlgili Sosyal Medya Yorumlarının Makine Öğrenmesi Yöntemleriyle Sınıflandırılması". Kocaeli Üniversitesi, 2018, s. 19.

[15] Hossin M., Sulaiman M. N., A review on evaluation metrics for data classification evaluations, International Journal of Data Mining & Knowledge Management Process, 2015, 5(2), 1.

[16] Filiz E., Makine Öğrenmesi Yöntemleri ve Eğitim Verisi Üzerine Bir Uygulama: Uluslararası Matematik ve Fen Eğilimleri Araştırması 2015 Türkiye Örneği, Doktora Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 2019, 598315

[17] Nergis P. Ankara Üniversitesi. “Sürdürülebilir Kalkınma ve Çevresel Etki: Türkiye Örneği”. Ankara Üniversitesi, 2022, s. 63.

A COMPREHENSIVE SURVEY OF RANSOMWARE ATTACKS: LIFECYCLE, ANALYSIS METHODS, AND DETECTION TOOLS

Sara NAGHIB ZADEH[1]

Emin Furkan DOĞAN[2]

Yaren DOĞAN[3]

Merve BERBEROĞLU[4]

1. INTRODUCTION

In recent years, with the rapid expansion of digital technologies and the growing dependence of organizations and individuals on information systems, cyber threats have become one of the major challenges in information security. Among these threats, ransomware has emerged as one of the most destructive and profitable forms of malware, causing extensive damage on a global scale. By encrypting critical data, restricting user access, and exerting psychological and financial pressure, ransomware attacks force victims to pay a ransom. The increasing sophistication of modern ransomware, such as the use of powerful cryptographic algorithms, obfuscation techniques, lateral

[1] Dr. Lecture, Halic University, Vocational School, Department of Computer Programming, ORCID: 0009-0005-6959-1165.

[2] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0000-0191-9412.

[3] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0004-0306-6424.

[4] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0003-8534-6288.

movement within networks, and double extortion strategies, has made mitigation and defense against these attacks increasingly difficult (Rajasekharaiah, Dule, & Sudarshan, 2020).

Large-scale incidents such as the WannaCry outbreak have demonstrated that ransomware can disrupt critical infrastructures and large organizations worldwide within a very short period of time. Furthermore, changes in working patterns, particularly during the COVID-19 pandemic, created favorable conditions for a significant rise in ransomware attacks. Under such circumstances, accurate identification of ransomware behavior and a thorough understanding of its lifecycle play a crucial role in designing effective preventive and defensive strategies (Pranggono & Arabo, 2021).

Therefore, ransomware analysis has become a fundamental pillar of cybersecurity research and practice. Static, dynamic, and hybrid analysis approaches, each with their own methodologies, advantages, and limitations, serve as essential tools for understanding ransomware behavior, identifying malicious patterns, and developing detection and mitigation systems (Kang & Gu, 2023). The objective of this paper is to provide a comprehensive overview of ransomware attack stages, their lifecycle, commonly used analysis methods, and the tools employed in this domain, thereby offering a coherent framework for researchers and cybersecurity professionals to address this complex and evolving threat.

2. RANSOMWARE AS A CRITICAL CYBERSECURITY THREAT

Ransomware has become one of the most destructive and profitable forms of cyber threats in the digital age. These malicious programs encrypt critical data, block user access, and force victims to pay attackers in order to recover their

information. Advanced ransomware variants often exhibit long periods of dormancy and stealth, embedding themselves within legitimate files and backup systems. As a result, backups, typically considered the last line of defense, can no longer be regarded as fully reliable. If the victim refuses to pay the ransom, the attackers may permanently delete the data, destroy the decryption keys, or even sell and publicly leak sensitive information as a means of extortion. Consequently, ransomware has evolved into a persistent and severe security challenge for organizations lacking strong defensive infrastructures (Jeelan et al., 2025).

The global WannaCry outbreak in 2017 marked a significant turning point in cybersecurity history, demonstrating the profitability, simplicity, and scalability of ransomware attacks and triggering a new wave of similar incidents worldwide. The security landscape worsened further during the COVID-19 pandemic, as widespread remote work and misconfigured systems created ideal conditions for exploitation. In the third quarter of 2020 alone, ransomware attacks rose by more than 50%, once again underscoring the magnitude of this threat (Temara, 2024).

3. GENERAL STAGES OF A RANSOMWARE ATTACK

A ransomware attack typically unfolds through several structured phases, each designed to maximize impact, maintain stealth, and ultimately coerce the victim into paying the ransom (Kara & Aydos, 2022).

1. Initial Entry and Activation

The attack begins when a malicious file reaches the victim’s system through phishing emails, infected attachments,

fraudulent websites, exploit kits, or vulnerabilities in outdated software. Once the user executes the malicious file, the ransomware immediately initiates its infection process.

2. **Initiation of Encryption**

Once established on the system, the ransomware begins encrypting targeted files using robust cryptographic algorithms such as AES or RSA. More advanced variants use hybrid encryption, combining symmetric and asymmetric methods, to ensure that recovering the data without the attacker's private key is virtually impossible.

3. **Identification and Selection of Valuable Files**

The ransomware selectively targets files with the highest operational or personal value to the victim, such as documents, images, videos, databases, and project or server files. This targeted approach increases both the emotional and operational pressure on the victim.

4. **Generation of Encryption Keys**

Each infected device typically receives a unique encryption key. The private key associated with this process is stored on the attacker's command-and-control server and is only offered to the victim upon payment, reinforcing complete dependency on the attacker.

5. **Displaying the Ransom Note**

Once encryption is complete, the ransomware presents a threatening message informing the user that their files have been taken hostage. This ransom note usually details the payment method—often involving cryptocurrencies—and may include modified wallpapers, text files, or automatically opened HTML pages.

6. Establishing Persistence

To ensure it runs automatically after every reboot, the ransomware employs persistence mechanisms such as registry modifications, creation of scheduled tasks, or installation of new system services. These techniques make removal extremely difficult without specialized intervention.

7. Stealth and Evasion Techniques

Ransomware uses a range of advanced evasion strategies to avoid detection, including code obfuscation, packing techniques, virtual machine or sandbox detection, and disabling antivirus tools. These mechanisms significantly reduce the likelihood of early detection.

8. Lateral Movement Across the Network

More sophisticated variants, such as WannaCry, possess self-propagation capabilities that allow them to spread rapidly across a network by exploiting vulnerabilities in file-sharing protocols and system services. This often results in large-scale organizational outbreaks.

9. Data Theft and Double Extortion

In many modern attacks, sensitive data is exfiltrated to the attacker’s server before encryption occurs. This enables a double extortion strategy: even if the victim restores files from a backup, the attacker can still threaten to publicly release the stolen information.

10. Ransom Payment and Potential Decryption

In the final phase, the attacker demands payment in exchange for the decryption key. However, there is no guarantee that paying the ransom will result in receiving a valid key or recovering the encrypted data. In many cases, files remain partially restored or permanently damaged(Kara & Aydos, 2022).

4. RANSOMWARE LIFECYCLE

The ransomware lifecycle is one of the most critical concepts in the technical and behavioral analysis of this malware family. Each stage of this lifecycle reveals specific points that can be leveraged for prevention, detection, or interruption of an attack. The lifecycle begins when the final version of the malicious code is prepared for deployment and ends when the ransom demand is presented to the victim. A clear understanding of these stages enables organizations to establish defensive measures before the ransomware reaches the encryption phase. As shown in Fig. 1, generally, the ransomware lifecycle consists of four main stages: preparation, distribution, phishing/initial intrusion, and the assault phase (Kara & Aydos, 2022).

4.1. Preparation Phase

In this stage, the ransomware developer designs and finalizes the core malware. Once development is complete, the final version is typically uploaded to anonymous platforms or darknet networks such as Tor for use by attackers. Key activities in this phase involve creating or modifying a ransomware variant tailored to the intended target, integrating obfuscation techniques to bypass antivirus mechanisms, and testing the malware's performance across different operating systems or environmental conditions. During this stage, attackers also prepare the necessary infrastructure required for the upcoming operation. The overall objective of this phase is to deliver a functional and stable ransomware build that is fully ready for deployment in real-world attacks (Sgandurra, Muñoz-González, Mohsen, & Lupu, 2016).

4.2. Distribution Phase

At this point, the attacker attempts to deliver the malicious payload to potential victims. The distribution process may be large-scale or highly targeted. In this phase, attackers employ multiple techniques to deliver the malicious payload to the victim.

These methods include sending infected email attachments in formats such as Word, PDF, ZIP, and other commonly used files; hosting the malware on compromised websites and leveraging drive-by downloads; deploying malicious online advertisements (malvertising); uploading the payload to breached servers; and exploiting hacked RDP access or existing network vulnerabilities. The primary objective of this stage is to place the malicious file in front of the user or ensure its execution on the target system, thereby enabling subsequent stages of the attack (Mbol, Robert, & Sadighian, 2016).

4.3. Phishing / Reconnaissance Phase

This stage aims at deceiving the victim or exploiting human vulnerabilities. The attacker attempts to trick the user into interacting with the malicious payload and enabling its execution. Common techniques in this phase involve identifying potential victims, employing various social engineering strategies, sending spoofed emails that impersonate banks or other trusted institutions, redirecting users to fraudulent banking or social media websites, and attempting to harvest sensitive information such as passwords, card numbers, or account credentials. The overarching aim of this stage is to secure initial access, thereby creating the necessary conditions for the ransomware to be executed later in the attack chain (Scaife, Carter, Traynor, & Butler, 2016).

4.4. Assault Phase

This is the core operational stage of the lifecycle and begins once the malicious file is executed on the victim’s system. After execution, a sequence of structured and coordinated activities takes place.

4.4.1. Initial execution and system configuration

During the initial execution and system configuration phase, the ransomware generates a unique identifier for the victim, disables antivirus tools or other security services to prevent detection, registers itself to ensure persistence after system reboot, and collects detailed information about the system, including the device configuration, IP address, and user credentials.

4.4.2. Connection to command-and-control (c2) server

During this step, the ransomware connects to a command-and-control server and retrieves the encryption key. The overall strength and security of the encryption mechanism largely depend on this stage.

4.4.3. File discovery and encryption

In this stage, the ransomware systematically searches for files with sensitive extensions such as pdf, doc, xls, pptx, and jpeg, after which it extracts or relocates them, encrypts their contents using hybrid cryptographic algorithms such as AES combined with RSA, renames the resulting encrypted files, and ultimately deletes the original unencrypted versions to prevent any possibility of recovery.

4.4.4. Displaying the ransom note

Finally, the victim is presented with a ransom note, usually delivered through a text file or a desktop message, that specifies the demanded payment amount, outlines the deadline for completing the transaction, provides the cryptocurrency wallet address for transferring the funds, and issues explicit threats to destroy the decryption key or publicly leak the stolen data in the case of double-extortion attacks (Al-rimy, Maarof, & Shaid, 2018).

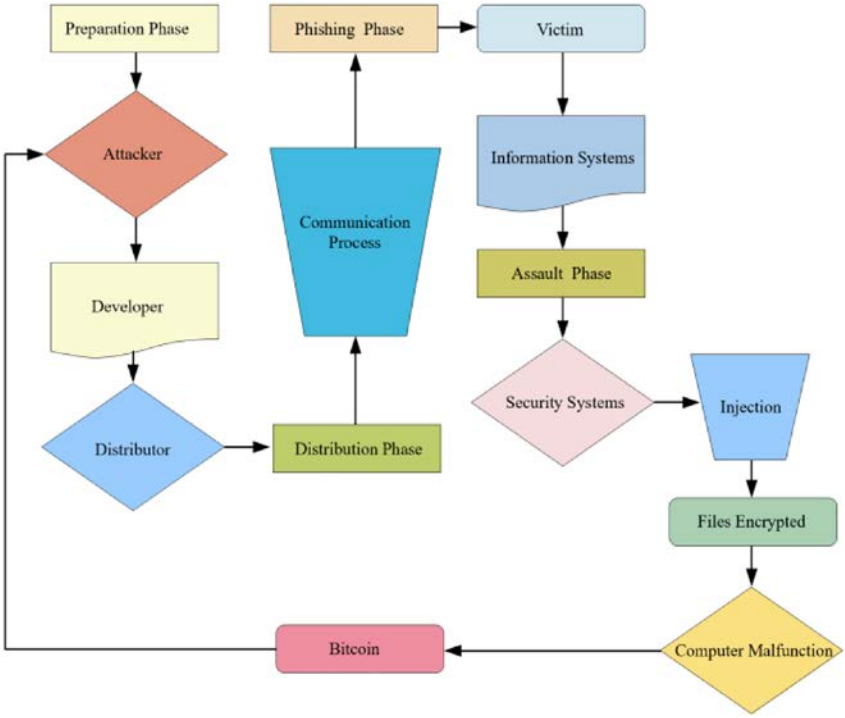


Fig. 1. Ransomware lifecycle (Kara & Aydos, 2022).

5. RANSOMWARE ANALYSIS METHODS

Ransomware analysis constitutes a fundamental component in the development of defensive strategies and in understanding the actual operational behavior of these threats. As illustrated in Fig. 2, this process is typically categorized into three primary approaches: static analysis, dynamic analysis, and hybrid analysis (Kara & Aydos, 2022).

5.1. Static Analysis

Static analysis is a non-executive approach in which the malware is examined without being run (Galal, Mahdy, & Atiea, 2015). In this method, the analyst inspects the file structure, internal code properties, and embedded patterns to obtain an

initial understanding of the malware's behavior without incurring the risks associated with actual execution. This approach is primarily used for safe identification, feature extraction, analysis of encryption mechanisms, and examination of control-flow structures, and it also underpins early-stage detection tools such as CryptoDrop (Scaife et al., 2016).

Several key techniques are commonly employed in static analysis. The first is Opcode analysis, one of the most widely used code-based methods, which can classify malware types before execution by examining instruction-level patterns. The second involves the use of open-source platforms such as VirusTotal, which enable researchers to locate similar samples, compare behavioral characteristics, and study the life cycles of related malware families. A third approach is the use of Control Flow Graphs (CFGs), which assist in categorizing and identifying distinct malware variants, although their reliability may diminish when dealing with techniques such as code packing. The fourth method, lexical analysis, focuses on evaluating code-level attributes or identifying malicious URLs; however, this technique can be circumvented through deliberate manipulation of URL structures or code packaging (Kara & Aydos, 2022).

Despite its utility, static analysis is constrained by several limitations. Key challenges include its inability to effectively detect packed malware, its susceptibility to evasion through obfuscation and polymorphism techniques, and the difficulty of accurately inferring real behavioral patterns without executing the malware. Consequently, static analysis is often combined with dynamic and hybrid methods to produce a more comprehensive understanding of ransomware behavior(Scaife et al., 2016).

5.2. Dynamic analysis

Dynamic analysis involves executing the ransomware within a controlled environment, typically a virtual machine or

sandbox, to observe its real behavior as it interacts with the operating system. This approach enables analysts to examine how the malware manipulates files, registry keys, network resources, and system memory, offering a more accurate picture of the attack sequence and operational logic(Madani et al., 2023).

A range of techniques support this form of analysis. Memory analysis focuses on extracting traces left by the malware during execution, taking advantage of the volatile nature of in-memory data. Registry analysis examines system configurations, services, and security-related entries that may reveal how the ransomware establishes persistence. Network traffic analysis inspects communication patterns, packet flows, and attempts to reach command-and-control servers or exchange suspicious data. Finally, code-level behavioral analysis evaluates the actual functions executed by the malware; however, this process can become significantly more challenging when obfuscation techniques are used to conceal malicious logic(Nauman, Azam, & Yao, 2016).

Overall, dynamic analysis offers a deeper, behavior-oriented understanding of ransomware, complementing static approaches and contributing to more effective detection and defense strategies (Madani et al., 2023).

5.3. Hybrid analysis

Hybrid analysis represents an integrated approach that combines the strengths of both static and dynamic techniques to provide a more comprehensive understanding of ransomware behavior. This method is designed to overcome the weaknesses inherent in each individual technique by correlating code-level insights with observations from real-time execution. As a result, hybrid analysis offers broader behavioral coverage, improves detection accuracy, and minimizes errors caused by obfuscation or delayed execution patterns (Alkhateeb, 2017).

Common practices in hybrid analysis include automated inspection, where sandbox environments are used to analyze the malware without manual intervention, though many ransomware families attempt to alter their behavior once they detect virtualization. Another component is anti-debugging, which involves mechanisms intended to prevent analysts from stepping through the malware's execution; attackers, in turn, often develop strategies to bypass such controls. A related technique is anti-emulation, in which the malware checks whether it is running inside an emulator or virtualized environment and suppresses its true behavior if such conditions are detected. Together, these methods illustrate the ongoing interplay between defensive analysis techniques and adversarial evasion strategies (Hull, John, & Arief, 2019).

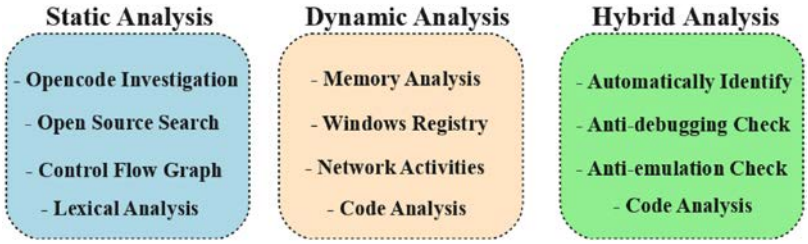


Fig 2. Commonly employed malware detection methods in the literature (Kara & Aydos, 2022).

6. TOOLS AVAILABLE FOR RANSOMWARE ANALYSIS

As presented in Table 1, various tools are available for ransomware analysis, encompassing static, dynamic, and hybrid approaches as reported by Talukder et al.(2020), the tools used for ransomware analysis fall into three overarching categories: static, dynamic, and hybrid analysis tools. As shown in Fig. 3, each category encompasses instruments designed to support different

stages of malware investigation, from initial detection to in-depth behavioral analysis (Talukder et al., 2020).

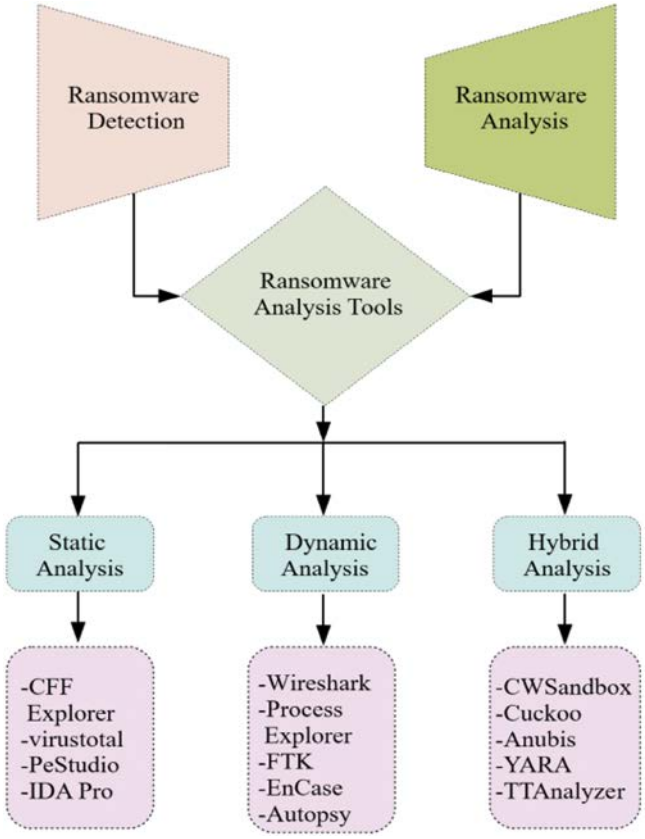


Fig 3. Ransomware detection and analysis tools(Kara & Aydos, 2022).

6.1. Static Analysis Tools

Static analysis utilities focus on examining executable files without running the malware. They analyze file structures, binary components, and inherent code characteristics to uncover suspicious patterns. Key tools in this group include:

- **CFF Explorer**, which enables detailed inspection of PE files and their internal architecture.

- **VirusTotal**, an online platform that evaluates suspicious samples across numerous antivirus engines.
- **PEStudio**, used to identify potential malicious indicators within executables prior to execution.
- **YARA**, a rule-based framework widely applied for malware research and pattern identification.
- **IDA Pro**, one of the most advanced disassemblers for exploring low-level instructions within malware binaries (Kara & Aydos, 2022).

6.2. Dynamic Analysis Tools

Dynamic analysis tools operate within controlled environments where the ransomware is executed so that its real behavior can be monitored. These tools capture modifications to the file system, registry, and network activity during runtime. Prominent dynamic analysis tools include:

- **Wireshark**, a powerful packet analysis tool that reveals suspicious or unauthorized network communications.
- **Process Explorer**, which monitors running processes, associated DLLs, and anomalous execution behavior.
- **FTK (AccessData)**, a digital forensics suite capable of examining live system activity and related artifacts.
- **EnCase**, a comprehensive forensic environment used for data investigation and evidence extraction.
- **Autopsy**, an open-source forensics platform able to observe and document malware actions during execution (Kara & Aydos, 2022).

6.3. Hybrid Analysis Tools

Hybrid tools integrate both static and dynamic methods, providing a unified framework for analyzing ransomware before

and during execution. These platforms often rely on automated sandbox environments to generate a holistic behavioral profile. Notable hybrid analysis tools include:

- **CWsandbox**, an automated sandbox system for malware examination, including ransomware families.
- **Cuckoo Sandbox**, the most widely adopted open-source malware analysis environment, offering blended static and dynamic capabilities.
- **TTAnalyzer**, an automated tool designed to detect and analyze ransomware samples (Kara & Aydos, 2022).

Table 1. Tools Available for Ransomware Analysis (Talukder et al., 2020)

Static Analysis Tools	Dynamic Analysis Tools	Hybrid Analysis Tools
CFF Explorer – Analysis of suspicious PE files	Wireshark – Packet and network traffic analysis	CWsandbox – Automated sandbox-based malware analysis
VirusTotal – Online multi-engine malware scanning	Process Explorer – Examination of processes and loaded DLLs	Cuckoo Sandbox – Open-source platform combining static and dynamic analysis
PEStudio – Detection of artifacts and indicators in executables	FTK (AccessData) – Dynamic analysis and digital forensics	TTAnalyzer – Automated ransomware identification and analysis
YARA – Rule-based malware identification and research	EnCase – Comprehensive forensic investigation of system data	—
IDA Pro – Disassembly and analysis of binary instructions	Autopsy – Free platform for dynamic analysis and digital forensics	—

7. CONCLUSION

Ransomware has become one of the most serious cybersecurity threats in recent years, targeting not only data but also the operational continuity of organizations and the trust of users. As discussed in this paper, ransomware attacks follow a multi-stage structure and a well-defined lifecycle that begins with

preparation and distribution and culminates in execution, data encryption, and financial extortion. A clear understanding of these stages enables the identification of attack weaknesses and facilitates effective intervention before the encryption phase is reached.

An examination of static, dynamic, and hybrid analysis approaches indicates that none of these methods alone is sufficient to fully address the complexity of modern ransomware. While static analysis offers a high level of safety, it is limited in the presence of techniques such as code obfuscation and packing. Dynamic analysis, although providing a more accurate behavioral perspective, faces challenges such as malware detection of virtualized environments. In this context, hybrid analysis, which integrates both approaches, provides a more comprehensive and effective solution for ransomware detection and investigation.

Furthermore, a wide range of ransomware analysis tools, from static and dynamic platforms to automated sandbox environments, play a crucial role in supporting both research and operational activities. However, the continuous evolution of evasion techniques demonstrates that combating ransomware requires constant tool updates, the adoption of intelligent analysis methods, and the development of advanced analytical frameworks. Ultimately, it can be concluded that combining knowledge of the ransomware lifecycle with multi-layered analysis approaches and specialized tools offers an effective pathway to enhancing system resilience against this rapidly evolving threat.

REFERENCES

Alkhateeb, Ehab Mufid Shafiq. 2017. “Dynamic Malware Detection Using API Similarity.” *IEEE CIT 2017 - 17th IEEE International Conference on Computer and Information Technology* 297–301. doi:10.1109/CIT.2017.14.

Al-rimy, Bander Ali Saleh, Mohd Aizaini Maarof, and Syed Zainudeen Mohd Shaid. 2018. “Ransomware Threat Success Factors, Taxonomy, and Countermeasures: A Survey and Research Directions.” *Computers & Security* 74:144–66. doi:10.1016/J.COSE.2018.01.001.

Galal, Hisham Shehata, Yousef Bassyouni Mahdy, and Mohammed Ali Atiea. 2015. “Behavior-Based Features Model for Malware Detection.” *Journal of Computer Virology and Hacking Techniques 2015 12:2* 12(2):59–67. doi:10.1007/S11416-015-0244-0.

Hull, Gavin, Henna John, and Budi Arief. 2019. “Ransomware Deployment Methods and Analysis: Views from a Predictive Model and Human Responses.” *Crime Science 2019 8:1* 8(1):2-. doi:10.1186/S40163-019-0097-9.

Jeelan, Peer Mohammed, Ramesh Saini, Sasmita Parida, Deepak Minhas, Manashree, and Ankita Agarwal. 2025. “The Threat Landscape of Ransomware in Critical Infrastructure: An Optimization Perspective.” *2025 International Conference on Automation and Computation, AUTOCOM 2025* 917–22. doi:10.1109/AUTOCOM64127.2025.10957218.

Kang, Qian, and Yuanyuan Gu. 2023. “A Survey on Ransomware Threats: Contrasting Static and Dynamic Analysis Methods.” doi:10.20944/preprints202311.0798.v1.

Kara, Ilker, and Murat Aydos. 2022. “The Rise of Ransomware: Forensic Analysis for Windows Based Ransomware Attacks.” *Expert Systems with Applications* 190. doi:10.1016/j.eswa.2021.116198.

Madani, Houria, Noura Ouerdi, and Abdelmalek Azizi. n.d. *Ransomware: Analysis of Encrypted Files*. Vol. 14. www.ijacsa.thesai.org.

Mbol, Faustin, Jean Marc Robert, and Alireza Sadighian. 2016. “An Efficient Approach to Detect TorrentLocker Ransomware in Computer Systems.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10052 LNCS:532–41. doi:10.1007/978-3-319-48965-0_32.

Nauman, Mohammad, Nouman Azam, and Jing Tao Yao. 2016. “A Three-Way Decision Making Approach to Malware Analysis Using Probabilistic Rough Sets.” *Information Sciences* 374:193–209. doi:10.1016/J.INS.2016.09.037.

Pranggono, Bernardi, and Abdullahi Arabo. 2021. “COVID-19 Pandemic Cybersecurity Issues.” *Internet Technology Letters* 4(2):e247. doi:10.1002/ITL2.247;SUBPAGE:STRING:FULL.

Rajasekharaiah, K. M., Chhaya S. Dule, and E. Sudarshan. 2020. “Cyber Security Challenges and Its Emerging Trends on Latest Technologies.” *IOP Conference Series: Materials Science and Engineering* 981(2):022062. doi:10.1088/1757-899X/981/2/022062.

Scaife, Nolen, Henry Carter, Patrick Traynor, and Kevin R. B. Butler. 2016. “CryptoLock (and Drop It): Stopping Ransomware Attacks on User Data.” *Proceedings - International Conference on Distributed Computing*

Systems 2016-August:303–12. doi:10.1109/ICDCS.2016.46.

Sgandurra, Daniele, Luis Muñoz-González, Rabih Mohsen, and Emil C. Lupu. 2016. “Automated Dynamic Analysis of Ransomware: Benefits, Limitations and Use for Detection.” https://arxiv.org/pdf/1609.03020.

Talukder, S., … Z. Talukder-Journal of Network Security &. Its, and undefined 2020. n.d. “A Survey on Malware Detection and Analysis Tools.” *Researchgate.NetS Talukder, Z TalukderInternational Journal of Network Security & Its Applications (IJNSA) Vol, 2020•researchgate.Net*. doi:10.5121/IJNSA10.5121/IJNSA:2020.1220010.5121/IJNSA:2020.12203.

Temara, Sheetal. 2024. “The Ransomware Epidemic: Recent Cybersecurity Incidents Demystified.” *Asian Journal of Advanced Research and Reports* 18(3):1–16. doi:10.9734/AJARR/2024/V18I3610.

BAKIR-ÇİNKO ALAŞIMINDA ÖZDİRENÇ TAHMİNİ: ÖZELLİK MÜHENDİSLİĞİ VE MODEL KARŞILAŞTIRMASI

Hasan GÜLER[1]

Rasim ÖZDEMİR[2]

1. GİRİŞ

Gelişmiş ve hedeflenmiş özelliklere sahip malzemelerin geliştirilmesi, modern mühendislik ve teknolojinin temel yapı taşlarından birini oluşturmaktadır. Bakır–çinko (CuZn) alaşımları, yaygın olarak pirinç olarak bilinmekte olup, mekanik dayanım, korozyon direnci ve elektriksel iletkenliğin dengeli bir kombinasyonunu sunmaları nedeniyle çok sayıda endüstriyel uygulamada tercih edilmektedir. Özellikle elektriksel özdirenç, konnektörler, terminaller ve iletken bileşenler gibi elektrik ve elektronik uygulamalardaki performansı doğrudan etkileyen kritik bir malzeme özelliğidir (Pollock, 2016).

CuZn alaşımlarının elektriksel özdirencinin doğru biçimde tahmin edilmesi, sentez sürecinin çok parametreli yapısı nedeniyle önemli bir zorluk teşkil etmektedir. Özdirenç; Cu/Zn oranı gibi kimyasal bileşim değişkenlerinin yanı sıra işlem sıcaklığı, elektrokimyasal potansiyel, tane boyutu ve elektrolit konsantrasyonu gibi birbiriyle ilişkili çok sayıda parametreden etkilenmektedir (Chandrasekar & Pushpavanam, 2008). Geleneksel ampirik yaklaşımlar ve analitik modeller, bu

[1] Dr. Öğr. Üyesi, Kilis 7 Aralık Üniversitesi, Teknik Bilimler MYO, Bilgisayar Teknolojileri Bölümü, ORCİD: 0000-0001-5312-171X.

[2] Prof. Dr., Kilis 7 Aralık Üniversitesi, Teknik Bilimler MYO, Elektrik ve Enerji Bölümü, ORCİD: 0000-0003-1439-0444.

parametreler arasındaki karmaşık ve doğrusal olmayan etkileşimleri tam anlamıyla temsil etmekte yetersiz kalmaktadır. Bu bağlamda, makine öğrenmesi (ML), malzeme biliminde çok boyutlu problemlerin ele alınmasında güçlü bir yaklaşım olarak öne çıkmaktadır (Jordan & Mitchell, 2015). ML algoritmaları, deneysel verilerden gizli örüntüleri ve doğrusal olmayan ilişkileri doğrudan öğrenebilme yeteneğine sahiptir. Ayrıca, gelişmiş yorumlanabilirlik teknikleri sayesinde eğitilmiş modellerden fiziksel olarak anlamlı içgörülerin çıkarılması mümkün olmaktadır (Lundberg et al., 2020).

Bu çalışma, elektrokimyasal olarak sentezlenen CuZn alaşımlarının elektriksel özdirencinin, süreç parametreleri ve bileşim oranları kullanılarak tahmin edilmesi problemine odaklanmaktadır. Bu amaçla, ağaç tabanlı topluluk yöntemlerinden Rastgele Orman (Random Forest-RF) ve Ekstra Ağaçlar (Extra Trees-ET) ile olasılıksal bir yaklaşım olan Gauss Süreç Regresyonu (Gaussian Process Regression-GPR) modelleri karşılaştırmalı olarak değerlendirilmiştir. Bu modeller, doğrusal olmayan ilişkileri yakalama yetenekleri ve yorumlanabilirlikleri açısından incelenmiştir (Breiman, 2001; Rasmussen & Williams, 2008).

Çalışmanın önemli katkılarından biri, açıklanabilir yapay zekâ (XAI) yaklaşımlarının kapsamlı biçimde uygulanmasıdır. SHAP yöntemi kullanılarak her bir girdinin tahminler üzerindeki katkısı nicel olarak ortaya konmuş, permütasyon önemi ile tamamlayıcı analizler gerçekleştirilmiştir (Lundberg et al., 2020). Buna ek olarak, fizik temelli özellik mühendisliği yaklaşımı ile; Cu/Zn oranları, sıcaklık dönüşümleri ve tane boyutu dönüşümleri gibi alan bilgisine dayalı türetilmiş özellikler modele dâhil edilmiştir. Bu sistematik yaklaşım, benzer malzeme özellik tahmin problemleri için model seçimine yönelik yol gösterici bir çerçeve sunabilir ve süreç optimizasyonu için uygulanabilir içgörüler sağlayabilir.

2. MATERYAL VE YÖNTEM

2.1. Deneysel Analiz ve Veriseti

Bu çalışmada kullanılan veri seti, kontrollü laboratuvar koşulları altında elektrokimyasal yöntemle sentezlenen CuZn alaşımlarına ait 127 deneysel ölçümden oluşmaktadır. Sentez işlemi, üç elektrotlu standart bir elektrokimyasal hücre konfigürasyonu kullanılarak gerçekleştirilmiştir. Hücre düzeninde alüminyum altlık çalışma elektrodu, platin karşı elektrot ve doygun kalomel elektrot referans elektrot olarak kullanılmıştır. Elektrolit çözeltisi, farklı Cu/Zn alaşım oranlarının elde edilmesi amacıyla değişken molar konsantrasyonlarda bakır sülfat ($CuSO_4 5H_2O$) ve çinko sülfat ($ZnSO_4 7H_2O$) içeren sulu çözeltilerden hazırlanmıştır. İşlem, potansiyostatik kontrol altında yürütülmüş olup sıcaklık, ±1 °C hassasiyetli termostatik su banyosu ile 100–407 K aralığında kontrol edilmiştir. Elde edilen CuZn alaşım filmleri X-ışını kırınımı (XRD) ile karakterize edilmiş ve tane boyutları Debye-Scherrer denklemi kullanılarak hesaplanmıştır (Patterson, 1939). Alaşımın elementel bileşimi, enerji dağılımlı X-ışını spektroskopisi (EDS) ile belirlenmiştir (Goldstein et al., 2018)(Goldstein et al., 2018). Elektriksel özdirenç ölçümleri, temas direnci etkilerini minimize etmek amacıyla dört nokta prob yöntemi ile gerçekleştirilmiştir (Van Der Pauw, 1991). Veri seti 8 girdi değişkeni ve 1 hedef değişkenden oluşmaktadır. Tablo 1'de değişkenlerin özeti sunulmaktadır.

Tablo 1. Veri seti değişkenleri ve değer aralıkları

Değişken	Sembol	Birim	Aralık	Açıklama
Sıcaklık	T	K	100–407	Sentez sıcaklığı
Alaşımdaki Cu	FCu	%	6.15–38.31	Filmdeki Cu yüzdesi
Alaşımdaki Zn	FZn	%	61.69–93.85	Filmdeki Zn yüzdesi
Elektrolit Konsantrasyonu	C	mol/L	0.3–0.9	Toplam metal iyon konsantrasyonu
Elektrolit Cu	ECu	%	22.8	Elektrolit Cu yüzdesi
Elektrolit Zn	EZn	%	77.2	Elektrolit Zn yüzdesi
Tane Boyutu	d	nm	28.0–32.4	XRD'den hesaplanan tane kristal boyutu
Potansiyel	V	V	1.517–1.549	Elektrodepolamanın başlama potansiyeli
Özdirenç	ρ	Ω·m	0.091–0.256	Elektriksel özdirenç

2.2. Fizik Temelli Özellik Mühendisliği

Özellik mühendisliği, makine öğrenmesi modellerinin performansını doğrudan etkileyen kritik bir aşamadır. Malzeme bilimi uygulamalarında etkin özellik mühendisliği, alan bilgisinin veri odaklı tekniklerle bütünleştirilmesini gerektirmektedir (Ward, Agrawal, Choudhary, & Wolverton, 2016). Bu çalışmada, metalik alaşımlarda elektriksel özdirenci yöneten fiziksel prensiplere dayalı türetilmiş özellikler tanımlanmıştır (Tablo 2).

Tablo 2. Fizik temelli türetilmiş özellikler ve fiziksel gerekçeleri

No	Özellik	Tanım	Fiziksel Motivasyon	İlgili Teori
f_1	Cu/Zn Oranı	FCu/(FZn+ε)	Faz bileşimini ve elektronik yapı	Cu–Zn faz diyagramı (Mizutani, 2016)
f_2	Zn Zenginleşme İndeksi	FZn – Ecu	Tercihli biriktirme ve stokiyometri sapması	Elektrokimyasal kinetik (Dini, 1997)
f_3	Ters Sıcaklık	1 / T	Termal aktivasyon	Arrhenius ilişkisi (Laidler, 1984)
f_4	Sıcaklığın Karesi	T^2	Doğrusal olmayan fonon katkıları	Bloch–Grüneisen teorisi (Grüneisen, 1933)
f_5	Sıcaklığın Karekökü	√T	Geçiş rejimi sıcaklık etkileri	Ampirik davranış (Kittel, 2005)
f_6	Ters Tane Boyutu	1 / (d + ε)	Tane sınırı saçılması	Mayadas–Shatzkes modeli (Mayadas & Shatzkes, 1970)
f_7	Logaritmik Tane Boyutu	ln(d + ε)	Güç yasası ilişkisi	Hall–Petch tipi modeller (Hall, 1951)
f_8	Cu–Sıcaklık Etkileşimi	FCu · T	Faza bağlı sıcaklık duyarlılığı	Bileşim–fonon etkileşimi (Gale & Totemeier, 2004)
f_9	Zn–Sıcaklık Etkileşimi	FZn · T	Tamamlayıcı bileşen etkisi	Faz-spesifik sıcaklık etkileri
f_{10}	Konsantrasyon–Sıcaklık	C · T	Biriktirme kinetiği ve kusur oluşumu	Elektrokimyasal kinetik (Schlesinger & Paunovic, 2010)
f_{11}	Potansiyel–Sıcaklık	V · T	Elektrokimyasal itici kuvvet	Mikro yapı evrimi (Newman & Thomas-Alyea, 2004)

Türetilen özellikler dört kategoride gruplandırılmıştır:

Bileşim Özellikleri (f_1–f_2): Alaşımın faz yapısı ve tercihli biriktirme davranışlarını yansıtarak, nominal besleme oranları ile gerçek bileşim sapmaları arasındaki ilişkiyi modele dahil etmektedir.

Sıcaklık Dönüşümleri (f_3–f_5): Fonon saçılması ve termal aktivasyon gibi farklı sıcaklık rejimlerinde baskın hâle gelen mekanizmaların doğrusal olmayan davranışlarını temsil etmektedir.

Mikroyapı Özellikleri (f_6–f_7): Tane sınırı saçılmasının özdirenç üzerindeki belirleyici rolünü, Mayadas–Shatzkes ve Hall–Petch tipi modellerle tutarlı biçimde modele entegre etmektedir.

Etkileşim Terimleri (f_8–f_{11}): Bileşim, sıcaklık ve elektrokimyasal parametrelerin sinerjik etkilerini yansıtarak hem tahmin performansını artırmakta hem de fiziksel yorumlanabilirliği kolaylaştırmaktadır.

Bu yapılandırılmış özellik seti, model karşılaştırmaları ve yorumlanabilirlik analizleri için fiziksel olarak tutarlı bir temel oluşturmaktadır.

2.3. Makine Öğrenmesi Modelleri

Bu çalışmada, elektriksel özdirenç tahmini için üç farklı makine öğrenmesi modeli kullanılmıştır: RF, ET ve GPR regresyon modeli.

Ağaç tabanlı modeller, özellik uzayını yinelemeli olarak bölerek karmaşık doğrusal olmayan örüntüleri açık özellik mühendisliği gerektirmeden modelleyebilme avantajı sunmaktadır (Loh, 2011).

RF: Önyüklemeli (bootstrap) örnekler üzerinde eğitilen çok sayıda karar ağacının ortalamasıyla tahmin üretmektedir:

$$\hat{y}_{RF} = \frac{1}{B}\sum_{b=1}^{B} \hat{y}_b$$

Burada B ağaç sayısını, $\widehat{y_b}$ ise b-inci ağacın tahminini ifade etmektedir. Bu yaklaşım, her düğümde rastgele seçilen özellik alt kümesi üzerinde bölünme yaparak modeller arası korelasyonu azaltmakta ve genelleme yeteneğini artırmaktadır. Ayrıca gürültüye ve aykırı değerlere karşı dayanıklı sonuçlar sunmaktadır (Cutler, Cutler, & Stevens, 2012).

ET: RF'ye benzer bir yapıdadır ancak iki temel farklılık içerir: (1) bootstrap örnekleme yerine tüm eğitim verisi kullanılır ve (2) bölünme eşikleri optimal değer aranmaksızın rastgele seçilir. Bu yaklaşım şu şekilde ifade edilebilir:

$$\hat{y}_{ET} = \frac{1}{B}\sum_{b=1}^{B} \hat{y}_b^{rand}$$

Rastgele eşik seçimi, modelin varyansını düşürmekte ve çoğu durumda daha hızlı eğitim süresi sağlamaktadır (Geurts, Ernst, & Wehenkel, 2006). Bu özellik, özellikle gürültülü veri setlerinde aşırı öğrenmeye karşı dayanaklıdır.

GPR tahmin belirsizliği sağlayabilen güçlü bir olasılıksal modeldir. GPR, girdiler üzerinde tanımlı bir fonksiyonlar dağılımı olarak ifade edilmektedir (Rasmussen & Williams, 2006):

$$f(x) \sim \mathcal{GP}\big(m(x), k(x, x')\big)$$

Burada $m(x)$ ortalama fonksiyonunu, $k(x, x')$ ise kovaryans (çekirdek) fonksiyonunu temsil etmektedir. Örneğin, radyal tabanlı fonksiyon (RBF) çekirdek şöyle ifade edilebilir:

$$k(x, x') = \sigma_f^2 \, exp\left(-\frac{|x - x'|^2}{2l^2}\right)$$

Burada σ_f^2 sinyal varyansını, l ise uzunluk ölçeğini ifade etmektedir. GPR'nin tahmin dağılımı şu şekilde elde edilmektedir:

$$p(y_*|x_*, X, y) = \mathcal{N}(\mu_*, \sigma_*^2)$$

$$\mu_* = k_*^T (K + \sigma_n^2 I)^{-1} y$$

$$\sigma_*^2 = k(x_*, x_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_*$$

GPR'ın avantajları şunlardır: (1) her tahmin için belirsizlik tahmini sağlaması, (2) küçük veri setlerinde etkili performans göstermesi, (3) hiperparametrelerin marjinal olabilirlik maksimizasyonu ile otomatik olarak öğrenilebilmesi ve (4) aşırı öğrenmeye karşı doğal düzenlileştirme sunmasıdır (Schulz, Speekenbrink, & Krause, 2018).

Bu çoklu model yaklaşımı; doğruluk, genelleme, yorumlanabilirlik ve güvenilirlik arasındaki ödünleşimlerin nicel olarak değerlendirilmesine olanak tanımaktadır.

2.4. Hiperparametre Optimizasyonu

Makine öğrenmesi modellerinin en iyi performansa ulaşabilmesi için hiperparametre optimizasyonu kritik bir aşamadır. Bu çalışmada, arama uzaylarında daha verimli sonuçlar sağladığı için Bayesyen optimizasyon (BO) yaklaşımı benimsenmiş ve Optuna çerçevesi içerisinde Tree-structured Parzen Estimator (TPE) algoritması kullanılmıştır (Akiba, Sano, Yanase, Ohta, & Koyama, 2019). TPE yaklaşımı, iyi ve kötü denemeler için hiperparametre dağılımlarını ayrı ayrı modelleyerek,

$$p(x \mid y) = \begin{cases} l(x), & y < y^* \\ g(x), & y \geq y^* \end{cases}$$

şeklinde tanımlanan koşullu olasılık fonksiyonlarını oluşturmakta ve Beklenen İyileşme (Expected Improvement, EI) ölçütünü maksimize eden parametreleri seçmektedir:

$$EI(x) \propto \frac{l(x)}{g(x)}$$

Bu adaptif strateji, olasılığı yüksek bölgelerde aramayı yoğunlaştırarak optimum hiperparametrelerin çok daha az değerlendirme ile bulunmasını sağlamaktadır. Modellere özgü hiperparametreler ve arama uzayları Tablo 3'te verilmiştir.

Tablo 3. Modellere özgü hiperparametreler ve arama uzayları

Model	Hiperparametre	Arama Uzayı	Açıklama
RF	n_estimators	[50, 500]	Ağaç sayısı
	max_depth	[3, 30]	Maksimum derinlik
	min_samples_split	[2, 20]	Bölünme için minimum örnek
ET	n_estimators	[50, 500]	Ağaç sayısı
	max_depth	[3, 30]	Maksimum derinlik
	min_samples_split	[2, 20]	Bölünme için minimum örnek
GPR	kernel	[RBF, Matérn, RationalQuadratic]	Çekirdek fonksiyonu
	alpha	[1e-10, 1e-2]	Gürültü seviyesi
	length_scale	[0.1, 10]	Uzunluk ölçeği

2.5. Çapraz Doğrulama Stratejisi

Hiperparametre ayarlama sürecinde doğrulama setine aşırı uyumu (overfitting) önlemek amacıyla, katmanlı çapraz doğrulama uygulanmıştır (Kohavi, 1995). Bu yaklaşımda:

- Veri seti %60 eğitim, %20 doğrulama ve %20 test olarak bölünmüştür
- Hiperparametre optimizasyonu için 5-katlı çapraz doğrulama kullanılmıştır
- Optuna TPE algoritması ile her model için 100 deneme gerçekleştirilmiştir

Böylece, hiperparametreler yalnızca eğitim verisi üzerinde optimize edilmekte ve final model performansı daha önce hiç görülmemiş test seti üzerinde değerlendirilmektedir. Bu strateji, yanlılıksız bir genelleme performansı tahmini sağlamaktadır.

2.6. Performans Metrikleri

Model performansının çok yönlü değerlendirilmesi amacıyla birden fazla tamamlayıcı hata metriği kullanılmıştır.

Kök Ortalama Kare Hata (RMSE): Birincil performans ölçütü olarak kullanılmış olup, büyük hataları daha güçlü biçimde cezalandırmaktadır:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Ortalama Mutlak Hata (MAE): Aykırı değerlere karşı daha dayanıklı bir hata ölçütüdür:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Ortalama Mutlak Yüzde Hata (MAPE): Göreli hata analizini mümkün kılmaktadır:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i|}$$

Belirleme Katsayısı (R^2): Açıklanan varyans oranını ifade etmektedir:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

Düzeltilmiş R^2 (R^2_adj): Model karmaşıklığını hesaba almaktadır:

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{(n - p - 1)}$$

Açıklanan Varyans Skoru (EV): Modelin veri dağılımını ne ölçüde yakaladığını değerlendirmektedir:

$$EV = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

Metrikleri, 5-katlı çapraz doğrulama ile hesaplanmış ve ortalama ± standart sapma şeklinde sunulmuştur.

2.7. Model Yorumlanabilirliği

Model çıktılarının yorumlanabilirliğini artırmak amacıyla çok katmanlı bir açıklanabilirlik çerçevesi sunan SHAP (SHapley Additive exPlanations) yöntemi kullanılmıştır. Kooperatif oyun teorisine dayanan SHAP, her bir özelliğin tahmin üzerindeki katkısını Shapley değerleri aracılığıyla nicel olarak ifade etmektedir (Shapley, 1953; Lundberg et al., 2020). Özellik i'nin Shapley değeri şu şekilde tanımlanır:

$$\phi_i = \sum_{S} \frac{|S|!\,(M - |S| - 1)!}{M!} [f(S \cup \{i\}) - f(S)]$$

Burada M tüm özellikler, S özelliklerin bir alt kümesi, $f(S)$ ise S özellikleri ile yapılan tahmindir. Böylece yerel doğruluk, eksiksizlik ve tutarlılık özellikleri sağlanır.

Permütasyon Önemi: Modelden bağımsız bir yöntem olarak, özellik değerleri rastgele karıştırıldığında performansta meydana gelen düşüşü ölçerek özellik önemini değerlendirmektedir.

3. SONUÇLAR VE TARTIŞMA

3.1. Model Performansları ve Hiperparametre Optimizasyonu

Seçilen üç regresyon modeli (RF, ET ve GPR) için Optuna kütüphanesi kullanılarak TPE tabanlı Bayesian optimizasyonu gerçekleştirilmiştir. Her model için 100 deneme ile 5-katlı çapraz doğrulama (5-fold CV) yapılarak hiperparametreler optimize

edilmiştir. Tablo 4'te modellerin temel (baseline) performansları ve Optuna TPE optimizasyonu sonrasındaki RMSE değerleri karşılaştırılmıştır.

Tablo 4. Optuna TPE Optimizasyonu ile Model Karşılaştırması

Model	Baseline RMSE	Optuna TPE RMSE	İyileşme (%)
Gaussian Process	0.002326	0.002087	10.26
Extra Trees	0.002481	0.002238	9.77
Random Forest	0.007672	0.004384	42.85

Optuna TPE optimizasyonu sonucunda en düşük çapraz doğrulama RMSE değeri 0.002087 ile GPR modelinde elde edilmiştir. GPR modelinin optimal hiperparametreleri şu şekildedir: kernel tipi olarak RationalQuadratic seçilmiş, length_scale 0.157, constant_value 0.961, gürültü seviyesi noise_level 0.0032 ve α parametresi 2.50 olarak belirlenmiştir. RationalQuadratic çekirdeğinin seçilmesi, veri setindeki farklı uzunluk ölçeklerindeki korelasyonları yakalama kapasitesini yansıtmaktadır (Rasmussen & Williams, 2008).

ET modeli 0.002238 RMSE değeri ile ikinci sırada yer almış olup, optimal parametreleri n_estimators=318, max_depth=10, min_samples_split=2, min_samples_leaf=1 ve max_features='log2' şeklindedir. Bootstrap örneklemenin kapalı tutulması, Geurts ve ark. (2006) tarafından önerilen orijinal algoritma ile tutarlıdır ve modelin varyansını düşürmektedir.

RF modeli ise baseline performansında en yüksek iyileşmeyi (%42.85) göstermiş, ancak mutlak RMSE değeri (0.004384) diğer modellerin gerisinde kalmıştır. Bu durum, RF'nin daha derin ağaçlara (max_depth=20) ihtiyaç duyması ve bootstrap örneklemenin bu veri seti için ek varyans getirmesi ile açıklanabilir (Breiman, 2001).

En iyi performansı sergileyen GPR modeli, eğitim sürecinde hiç kullanılmamış test veri seti üzerinde

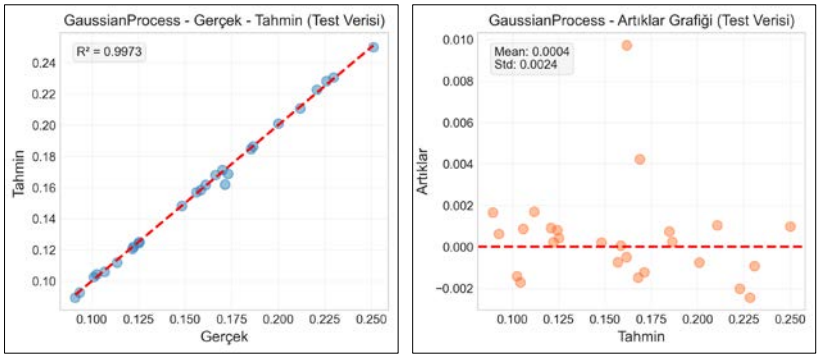
değerlendirilmiştir. Final model metrikleri Tablo 5'te sunulmaktadır.

Tablo 5. GPR Modelinin Test Seti Performans Metrikleri

Metrik	Değer
Test RMSE	0.002356
Test MAE	0.001448
MAPE	%0.95
R^2	0.9973
Düzeltilmiş R^2	0.9888
Açıklanan Varyans	0.9974

Çapraz doğrulama RMSE değeri (0.002087) ile test RMSE değeri (0.002356) arasındaki fark 0.000269 olup, bu durum modelin aşırı uyum (overfitting) göstermediğini ve genelleme kapasitesinin yüksek olduğunu göstermektedir (Hastie, Tibshirani, & Friedman, 2009). GPR modelinin R^2 değerinin 0.9973 olması, modelin CuZn alaşımlarının özdirenç değerlerindeki varyansın %99.73'ünü açıkladığını ifade etmektedir. Bu sonuç, malzeme özellik tahmininde GPR'nin etkinliğini gösteren önceki çalışmalarla uyumludur (Balachandran, Xue, Theiler, Hogden, & Lookman, 2016).

Şekil 1'de GPR modelinin test seti üzerindeki tahmin performansı görselleştirilmiştir. Sol panelde gerçek değerler ile tahmin edilen değerler arasındaki ilişki, sağ panelde ise artık (residual) dağılımı sunulmaktadır. Artıkların sıfır etrafında rastgele dağılımı, modelin sistematik bir hata içermediğini doğrulamaktadır.



a) b)

Şekil 1. GPR modelinin test seti performansı: (a) Gerçek-Tahmin karşılaştırması, (b) Artık analizi

3.2. SHAP Analizi

GPR modeli için hesaplanan SHAP değerleri, her bir özniteliğin tahminler üzerindeki marjinal katkısını nicel olarak ortaya koymaktadır. Tablo 6'da ortalama mutlak SHAP değerlerine göre sıralanan en etkili beş öznitelik sunulmaktadır.

Tablo 6. SHAP Analizi Sonuçları- En Etkili Öznitelikler

Sıra	Öznitelik	Ortalama SHAP Etkisi
1	FZn_Temperature	0.062
2	T_squared	0.051
3	Mol_Temperature	0.032
4	FCu_Temperature	0.016
5	Potential_Temperature	0.014

SHAP analizi sonuçları fiziksel olarak anlamlı örüntüler ortaya koymaktadır. FZn_Temperature (Çinko oranı × Sıcaklık etkileşimi) en güçlü tahmin edici öznitelik olarak belirlenmiştir. Bu bulgu, alaşım özdirencinin bileşim ve sıcaklığın birlikte belirlediği karmaşık bir fonksiyon olduğunu doğrulamaktadır (Rossiter, 1987). T_squared (sıcaklığın karesi) özniteliğinin ikinci sırada yer alması, özdirenç-sıcaklık ilişkisinin doğrusal olmadığını ve Bloch-Grüneisen formülasyonu ile uyumlu kuadratik bir bileşen içerdiğini göstermektedir (Grüneisen, 1933).

İlk 5 özniteliğin tamamının sıcaklık ile etkileşim içeren türetilmiş öznitelikler olması, fizik temelli özellik mühendisliğinin model performansına olan kritik katkısını vurgulamaktadır (Ward et al., 2016).

3.3. Permütasyon Önemi Analizi

Permütasyon önemi, modelden bağımsız bir yorumlanabilirlik yöntemi olarak SHAP sonuçlarını doğrulamak amacıyla uygulanmıştır. Tablo 7'de permütasyon önemi sonuçları sunulmaktadır.

Tablo 7. Permütasyon Önemi Analizi Sonuçları

Sıra	Öznitelik	Önem Skoru	Std
1	FZn_Temperature	21.44	±9.96
2	T_squared	16.75	±2.49
3	Mol_Temperature	1.80	±0.56
4	FCu_Temperature	0.50	±0.17
5	Potential_Temperature	0.38	±0.10

Permütasyon önemi sıralamasının SHAP sonuçları ile tam örtüşmesi, elde edilen bulguların güvenilirliğini artırmaktadır. Her iki yöntemde de FZn_Temperature ve T_squared özniteliklerinin baskın konumu, bu değişkenlerin özdirenç tahmini için kritik önemini teyit etmektedir.

3.4. Fiziksel Yorumlama

Açıklanabilir yapay zekâ analizlerinden elde edilen bulgular, metalik alaşımlarda elektriksel iletkenliği yöneten temel fiziksel mekanizmalarla tutarlıdır. Fizik tabanlı türetilmiş özniteliklerin (özellikle sıcaklık etkileşimleri) ham özniteliklere göre çok daha güçlü tahmin gücü sağlaması, Matthiessen kuralı ve alaşım saçılma teorisi ile uyumludur (Rossiter, 1987). Bu kural, toplam özdirencin farklı saçılma mekanizmalarının (fonon, kusur, tane sınırı) bağımsız katkılarının toplamı olduğunu ifade etmektedir. T_squared özniteliğinin yüksek önemi, yüksek sıcaklıklarda fonon-elektron saçılmasının özdirence kuadratik bir katkı sağladığını gösteren Bloch-Grüneisen teorisi ile

örtüşmektedir (Grüneisen, 1933). Bu bulgu, basit doğrusal modellerin metalik alaşımlarda yetersiz kalacağını ve fizik temelli dönüşümlerin gerekliliğini vurgulamaktadır. FZn_Temperature etkileşiminin en yüksek SHAP değerine sahip olması, CuZn alaşımlarında çinko konsantrasyonunun sıcaklık bağımlılığı üzerindeki kritik rolünü ortaya koymaktadır. Bu durum, Cu-Zn faz diyagramındaki α ve β fazları arasındaki yapı farklılıkları ile açıklanabilir (Mizutani, 2016).

Gaussian Process Regression modelinin ağaç tabanlı topluluk yöntemlerine göre daha iyi performans göstermesi, birkaç faktörle açıklanabilir: (i) GPR'nin sürekli ve türevlenebilir bir fonksiyon uzayında çalışması, fiziksel süreçlerin doğasına daha uygun bir yapı sunmaktadır; (ii) RationalQuadratic çekirdeğinin farklı uzunluk ölçeklerini otomatik olarak öğrenebilmesi, veri setindeki çok ölçekli korelasyonları yakalamaktadır; (iii) Bayesian çerçevenin doğal düzenlileştirme sağlaması, küçük veri setlerinde aşırı öğrenmeyi önlemektedir (Rasmussen & Williams, 2008; Schulz et al., 2018).

KAYNAKÇA

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. Anchorage AK USA: ACM. https://doi.org/10.1145/3292500.3330701

Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., & Lookman, T. (2016). Adaptive Strategies for Materials Design using Uncertainties. *Scientific Reports*, *6*(1), 19660. https://doi.org/10.1038/srep19660

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chandrasekar, M. S., & Pushpavanam, M. (2008). Pulse and pulse reverse plating—Conceptual, advantages and applications. *Electrochimica Acta*, *53*(8), 3313–3322. https://doi.org/10.1016/j.electacta.2007.11.054

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random Forests. In C. Zhang & Y. Ma (Eds.), *Ensemble Machine Learning* (pp. 157–175). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4419-9326-7_5

Dini, J. W. (1997). Properties of Coatings: Comparisons of Electroplated, Physical Vapor Deposited, Chemical Vapor Deposited, and Plasma Sprayed Coatings. *Materials and Manufacturing Processes*, *12*(3), 437–472. https://doi.org/10.1080/10426919708935157

Gale, W. F., & Totemeier, T. C. (2004). *Smithells metals reference book* (8th ed). Oxford: Elsevier Butterworth-Heinemann.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), 3–42. https://doi.org/10.1007/s10994-006-6226-1

Goldstein, J. I., Newbury, D. E., Michael, J. R., Ritchie, N. W. M., Scott, J. H. J., & Joy, D. C. (2018). *Scanning Electron Microscopy and X-Ray Microanalysis*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-6676-9

Grüneisen, E. (1933). Die Abhängigkeit des elektrischen Widerstandes reiner Metalle von der Temperatur. *Annalen Der Physik*, *408*(5), 530–540. https://doi.org/10.1002/andp.19334080504

Hall, E. O. (1951). The Deformation and Ageing of Mild Steel: III Discussion of Results. *Proceedings of the Physical Society. Section B*, *64*(9), 747–753. https://doi.org/10.1088/0370-1301/64/9/303

Hastie, T. J., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed). New York: Springer.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Kittel, C. (2005). *Introduction to solid state physics* (8th ed). Hoboken, NJ: Wiley.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, 1137–1143. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Laidler, K. J. (1984). The development of the Arrhenius equation. *Journal of Chemical Education*, *61*(6), 494. https://doi.org/10.1021/ed061p494

Loh, W. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, *1*(1), 14–23. https://doi.org/10.1002/widm.8

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Mayadas, A. F., & Shatzkes, M. (1970). Electrical-Resistivity Model for Polycrystalline Films: The Case of Arbitrary Reflection at External Surfaces. *Physical Review B*, *1*(4), 1382–1389. https://doi.org/10.1103/PhysRevB.1.1382

Mizutani, U. (2016). *Hume-Rothery Rules for Structurally Complex Alloy Phases* (0 ed.). CRC Press. https://doi.org/10.1201/b10324

Newman, J., & Thomas-Alyea, K. E. (2004). *Electrochemical systems* (3. ed). Hoboken, NJ: Wiley-Interscience.

Patterson, A. L. (1939). The Scherrer Formula for X-Ray Particle Size Determination. *Physical Review*, *56*(10), 978–982. https://doi.org/10.1103/PhysRev.56.978

Pollock, T. M. (2016). Alloy design for aircraft engines. *Nature Materials*, *15*(8), 809–815. https://doi.org/10.1038/nmat4709

Rasmussen, C. E., & Williams, C. K. I. (2008). *Gaussian processes for machine learning* (3. print). Cambridge, Mass.: MIT Press.

Rossiter, P. L. (1987). *The Electrical Resistivity of Metals and Alloys* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511600289

Schlesinger, M., & Paunovic, M. (Eds.). (2010). *Modern Electroplating* (1st ed.). Wiley. https://doi.org/10.1002/9780470602638

Schulz, E., Speekenbrink, M., & Krause, A. (2018). A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. *Journal of Mathematical Psychology*, *85*, 1–16. https://doi.org/10.1016/j.jmp.2018.03.001

Shapley, L. S. (1953). A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Van Der Pauw, L. J. (1991). A Method of Measuring Specific Resistivity and Hall Effect of Discs of Arbitrary Shape. In S. M. Sze, *Semiconductor Devices: Pioneering Papers* (pp. 174–182). WORLD SCIENTIFIC. https://doi.org/10.1142/9789814503464_0017

Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. *Npj Computational Materials*, *2*(1), 16028. https://doi.org/10.1038/npjcompumats.2016.28

EARLY DETECTION AND PREVENTION OF CYBER ATTACKS IN MOBILE BANKING USING MACHINE LEARNING: A MACHINE LEARNING–BASED REVIEW

Beytullah OKTAY[1]

Sara NAGHIB ZADEH[2]

Mustafa CİCİDURUCU[3]

1. INTRODUCTION

With the expansion of digital transformation processes and the widespread global use of smartphones, mobile banking has become an integral component of financial transactions. The ability for users to access their bank accounts anytime and anywhere has significantly improved the convenience of banking services; however, it has also attracted the attention of cybercriminals. In recent years, a large portion of banking operations has shifted from physical branches to desktop computers and, more recently, to mobile devices. While this technological advancement offers substantial convenience, it has simultaneously introduced new security vulnerabilities (Dasgupta et al., 2023).

The cost of cybercrime targeting the financial sector continues to increase each year. According to a report published

[1] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0000-9042-8913.

[2] Dr. Lecture, Halic University, Vocational School, Department of Computer Programming, ORCID: 0009-0005-6959-1165.

[3] Halic University, Vocational School, Department of Information Security, ORCID: 0009-0009-2491-4011.

by Accenture, the total financial losses incurred by the banking industry due to cybercrime between 2020 and 2025 are expected to approach USD 350 billion (Cele, 2023). These losses are not limited to large-scale data breaches but also reflect the cumulative impact of numerous small yet frequent phishing and fraud attacks. Fraudulent activities conducted through mobile banking channels impose significant operational and financial burdens on banks. For instance, it has been reported that 33% of fraud-related costs incurred by banks in the United States are directly attributed to mobile banking, compared to 26% in 2020, indicating a rapid growth of this threat (Valiquette L'Heureux, 2022).

As the number of mobile banking users increases, cyberattacks targeting this domain have also become more diverse and sophisticated. During the COVID-19 pandemic and the rapid, mandatory shift toward digital banking, the Federal Bureau of Investigation (FBI) warned of a substantial rise in malware attacks, particularly banking Trojans and fake applications targeting mobile banking platforms (Tsantikidou & Sklavos, 2024) . Banking Trojans, in particular, operate by concealing themselves on users' devices under the guise of legitimate applications and stealing sensitive financial information. These malicious programs are activated when a mobile banking application is launched and deploy a fraudulent overlay that imitates the bank's login screen, thereby capturing users' authentication credentials (Daninga, Losioki, Kitali, Adili, & Bulengela, 2022).

Currently, banks attempt to protect customer accounts through various security measures, including firewalls, antivirus software, and multi-factor authentication (MFA) based on SMS or biometric verification (Karim et al., 2024). However, traditional security solutions are largely reactive in nature and rely on known threat signatures. When attackers develop entirely new malware or attack techniques, such as zero-day exploits,

systems may remain vulnerable until these threats are identified and added to signature databases. As emphasized by Santeri Kangas in a World Economic Forum report, given the rapid evolution of cyber threats, reliance solely on human intervention and rule-based security systems is no longer sufficient (Pureti, 2022).

In response to constantly evolving and unpredictable attack methods, there is a growing need for a more proactive and preventive defense approach. At this point, techniques based on artificial intelligence (AI) and machine learning (ML) become increasingly important. Due to their strong capability to identify patterns and anomalies within large-scale datasets, machine learning methods can detect emerging threats at a speed that is difficult for human analysts to achieve and can enable automated responses (Muppalaneni, Inaganti, & Ravichandran, 2024).

The objective of this study is to examine contemporary attack vectors threatening the mobile banking ecosystem and to demonstrate how machine learning techniques can detect and prevent these attacks, even at early stages of execution. Based on recent data and case studies from the existing literature, this research discusses the effectiveness of AI-driven security systems in enhancing the resilience of mobile banking platforms.

2. TYPES OF CYBER ATTACKS AND VULNERABILITIES IN MOBILE BANKING

Cyber attacks targeting mobile banking systems can be classified into different categories based on the platform used, the techniques employed, and the specific vulnerability being exploited (Abdullah, Ahmed, & Ameen, 2018). Attackers typically target human errors, software weaknesses in applications, or network-level vulnerabilities.

2.1. Malicious Mobile Applications (Trojans and Fake Applications)

Malware developed by cybercriminals to target mobile devices poses a particularly serious threat, especially within the Android ecosystem. Malicious software commonly referred to as banking Trojans disguises itself as legitimate applications installed on users' devices and aims to steal login credentials, SMS-based verification codes, and credit card information (Reijonen, 2024).

According to a report published by Kaspersky in 2024, attempts to steal banking data from smartphones increased by 196% compared to the previous year (Deshpande, 2025). In 2024 alone, more than 33 million attacks targeting mobile users were detected worldwide. These malicious applications are often presented to users as games, messaging apps, flashlight applications, or useful tools (such as PDF readers); however, in the background, they actively collect financial information. Many banking Trojans are triggered when the victim launches a mobile banking application and deceive the user by overlaying a fake login screen on top of the legitimate application without the user's awareness (Abiola, 2023).

In addition, fully fake mobile banking applications represent a significant security threat. A study conducted in 2018 revealed that even official application stores contained tens of thousands of fake banking applications that were able to bypass security filters. If users enter their authentication credentials into these deceptive applications, attackers can gain direct access to their bank accounts (Muhammad et al., 2023).

2.2. Phishing and Social Engineering

Phishing attacks carried out via mobile devices are among the most common methods used to steal banking customers' account information. The small screen size of mobile devices and

the limited visibility of full Uniform Resource Locators (URLs) make it more difficult for users to distinguish fraudulent websites from legitimate ones (Panda, n.d.).

Attackers may deceive users through SMS messages (smishing), emails, or instant messaging platforms such as WhatsApp, redirecting them to fake banking login pages. For example, messages containing urgency or incentives, such as "Your account has been suspended—click to reactivate" or "Click here to receive a credit card fee refund," are commonly used to lure victims (Pinjarkar et al., 2024). According to data from a security company, approximately 53% of all detected phishing attacks target financial services. These attacks, often combined with social engineering techniques, do not attempt to bypass technical security controls directly; instead, they exploit human psychology and user inattention (Alkhalil, Hewage, Nawaf, & Khan, 2021).

2.3. Device Vulnerabilities and Network Attacks

Security vulnerabilities in smartphone operating systems (Android or iOS), as well as unauthorized modifications performed by users on their devices (rooting or jailbreaking), significantly expand the attack surface. On a rooted device, it becomes considerably easier to access the protected memory areas of banking applications and to steal stored credentials (Cortes, 2024).

In addition, when users perform banking transactions over unsecured public Wi-Fi networks (e.g., cafés or airports), the risk of Man-in-the-Middle (MitM) attacks increases. Attackers can intercept and read data packets by eavesdropping on network traffic or by connecting users to malicious or spoofed wireless networks. Although modern mobile banking applications employ end-to-end encryption (SSL/TLS), advanced techniques such as *SSL stripping* or the installation of forged digital certificates are

used in attempts to bypass these protections (Cekerevac, Cekerevac, Prigoda, Journal, & 2025, 2025).

2.4. Unauthorized Access and Account Takeover

In some cases, attackers do not exploit technical vulnerabilities directly but instead rely on weak user passwords or credentials leaked from other platforms. In a technique known as credential stuffing, stolen username–password pairs obtained from compromised e-commerce or online services are automatically tested against banking applications using bots (Ahuja et al., 2025). Users' tendency to reuse the same passwords across multiple platforms significantly increases the success rate of this type of attack. Once an account is compromised, attackers rapidly transfer funds to so-called money mule accounts to launder the stolen assets (Esoimeme, 2021).

2.5. Notable Mobile Banking Attack Cases Worldwide and in Türkiye

Beyond the theoretical aspects of cyber attacks, examining real-world large-scale incidents is essential for understanding the severity of the threat. In recent years, several critical attacks targeting the financial sector have been reported:

2.5.1. The Carbanak (Anunak) Attack

Carbanak, one of the largest cyber heists targeting mobile banking and financial networks, was carried out by an Advanced Persistent Threat (APT) group . The attackers infiltrated banking systems by infecting bank employees' computers and mobile devices with malware. By observing and learning system operations over several months, often described as a malicious use of techniques similar to machine learning, they were able to withdraw cash directly from ATMs and transfer funds to fraudulent accounts, stealing approximately USD 1 billion worldwide. This incident demonstrated that compromised

internal devices, whether mobile or desktop, can pose a serious threat, not only external attacks (Bhardwaj, 2021).

2.5.2. The FluBot Malware

FluBot, which spread rapidly across Europe, Türkiye, and other regions in 2021 and 2022, is a banking Trojan targeting Android-based devices. It was distributed through SMS-based phishing messages (smishing) claiming “package delivery tracking” or “new voicemail notifications,” which persuaded users to install a malicious application. Once installed, the malware propagated like a worm by sending the same messages to all contacts in the victim’s address book and stole credentials from banking applications on the device. Such fast-spreading malware, which is often difficult for traditional antivirus solutions to detect, highlights the necessity of machine learning–based behavioral analysis systems (Ashawa, 2021).

2.5.3. Phishing Trends in Türkiye

According to data from USOM (National Cyber Incident Response Center) and banking sector reports, fake social media advertisements are among the most common attack vectors targeting mobile banking users in Türkiye. Attackers lure users to fraudulent banking websites with promises such as “credit card fee refunds” or “low-interest loans,” and then empty accounts by capturing SMS-based verification codes. Because these attacks primarily exploit human vulnerabilities rather than technical flaws, banks increasingly rely on artificial intelligence–based systems that analyze user behavior, for example, by comparing the user’s real-time location with the location from which a transaction is initiated, to detect and prevent fraud (Sermet, 2021).

2.6. Comparison of Traditional Methods and Machine Learning–Based Approaches

In mobile banking security, there are fundamental differences between traditional (rule-based) methods and modern approaches based on machine learning. As presented in Table 1, the following comparison clearly illustrates why a transition toward artificial intelligence–driven solutions has become necessary (Islam, Haque, Naser, & Karim, 2024).

Table 1. Comparison of Traditional Security and Machine Learning–Based Security(Choudhury & Paul, 2025)

Feature	Traditional Security Methods	Machine Learning (Artificial Intelligence)–Based Methods
Detection Mechanism	Signature-based (identifies known threats)	Behavior-based (detects anomalies and deviations)
Unknown Threats	Unable to detect zero-day attacks	Capable of detecting new and unknown (zero-day) attacks
Speed	Requires frequent updates and has longer response times	Provides real-time analysis and response (within milliseconds)
False Positives	Usually low but rigid and inflexible	May be high initially, but decreases over time through learning
Data Processing	Operates on limited datasets	Becomes more effective as it is trained on large-scale (Big Data)
Example Scenario	“Lock the account if the password is entered incorrectly three times.”	“Even if the password is correct, lock the account if abnormal behavior (e.g., robotic typing speed) is detected.”

As shown in the table 1, machine learning moves beyond static rules to establish a dynamic and intelligent defense layer that can adapt to emerging attack patterns.

3. THE ROLE OF MACHINE LEARNING IN CYBERSECURITY

Machine learning (ML) is a subfield of artificial intelligence based on statistical techniques that enables computer systems to learn from data without explicit programming. In

recent years, the application of machine learning in the field of cybersecurity has attracted significant attention in both academia and industry. The primary reason for this interest is the strong capability of machine learning in data-driven attack detection, allowing systems to perform faster, continuous, and large-scale analyses compared to human experts (Apruzzese et al., 2023).

The roles and methods through which machine learning contributes to attack detection and prevention can be examined under the following aspects:

3.1. Pattern Recognition and Anomaly Detection

One of the most powerful capabilities of machine learning is its ability to discover complex patterns within large-scale datasets. By leveraging statistical methods, these technologies can identify weak signals and detect deviations from normal system behavior, commonly referred to as anomalies (Gudala et al., 2019).

As noted by an artificial intelligence expert, the most prominent strength of AI lies in its ability to perform pattern recognition much faster and more accurately than humans (Korteling, van de Boer-Visschedijk, Blankendaal, Boonekamp, & Eikelboom, 2021). This capability enables banking systems to instantly identify suspicious transactions or user behaviors among millions of incoming operations. For example, if a customer initiates a high-value transfer at an unusual time (e.g., 3:00 a.m.) from an unexpected location (e.g., an overseas IP address), such anomalous behavior can be detected by machine learning models within milliseconds. The system can classify this transaction as a potential threat and immediately flag it for further investigation or block it altogether (Shabbir, Shabir, Javed, Chakraborty, & Rizwan, 2022).

In the case of Commonwealth Bank, the AI-based system developed by the bank generates a “baseline behavioral profile”

for each customer using behavioral biometric data, such as keyboard and mouse usage patterns. Subsequent transactions are compared against this reference profile. If abnormal behavior is detected, for instance, robotic typing speed instead of natural, hesitation-based input, a real-time, two-way alert is sent to the customer via the mobile application (Aljamal et al., 2025).

3.2. Detection of Unknown Threats Using Unsupervised Learning

Cyberattacks do not always have predefined signatures. To detect previously unseen threats, such as zero-day attacks, unsupervised learning techniques are employed. These methods are particularly suitable for discovering unusual patterns in unlabeled data, where the nature of the data is not specified in advance (Zoppi, Ceccarelli, Capecchi, & Bondavalli, 2021).

Clustering Algorithms: In this context, the k-means algorithm is widely used to group transactions with similar characteristics into clusters. Data points that lie far from dense clusters formed by millions of normal transactions are marked as outliers and may indicate potential anomalies or attack attempts (Datta, Dasgupta, Dasgupta, & Reddy, 2021).

Autoencoder Networks: Autoencoders are a deep learning architecture that is especially effective in financial anomaly detection. They are trained on normal transaction data and learn to generate a compressed representation of such data. When evaluating a new transaction, if the model produces a high reconstruction error, this suggests that the transaction significantly deviates from normal patterns and is likely indicative of a fraudulent attempt (Wei, Chow, & Yiu, 2020).

Unsupervised learning also plays a critical role in network traffic monitoring solutions for banks. For example, CyGlass, a “Network Defense as a Service” solution, employs an AI engine that models normal behavior in banking networks and

continuously learns from data. By progressively refining its understanding of what constitutes “normal” network activity, the system can detect even subtle deviations in traffic patterns, such as a server transmitting 5% more data than usual (Wei, Chow, & Yiu, 2020).

4. EARLY DETECTION AND PROACTIVE PREVENTION USING MACHINE LEARNING

Machine learning–based systems create a fundamental advantage over traditional approaches by enabling the identification of attacks at very early stages, thereby preventing damage before it occurs. The importance of early detection becomes particularly evident in the financial sector, given the potentially severe financial losses and reputational damage that may result from a successful attack (Liu & Lang, 2019).

4.1. Automated Response and Risk Mitigation

Once a threat is detected, machine learning systems can automatically generate alerts and even execute preventive actions without waiting for human intervention. This level of automation is essential to counter the speed at which attackers operate , often at machine scale (ArnaldoIgnacio & VeeramachaneniKalyan, 2019).

For example, the Commonwealth Bank’s fraud detection platform immediately alerts users and suspends transactions when unusual activity is detected in an account (Cugnasco, 2023). If the system identifies a login attempt from an unfamiliar device or location, it sends a two-way notification to the customer asking, “Is this login yours?” As soon as the user indicates that the activity is unauthorized, the platform automatically terminates the session, secures the account information, and prompts the user to reset their password. The bank has reported that this proactive

approach prevented fraud attempts amounting to USD 100 million for its customers over a specific period (Barker, 2020).

4.2. Prevention at the Network and Infrastructure Level

Machine learning is not limited to the user interface layer (mobile applications); it also contributes to early detection and prevention within a bank's server infrastructure (Al Hwaitat et al., 2024). Network-based intrusion detection and prevention systems (IDS/IPS), when enhanced with machine learning, can learn patterns of normal network traffic and identify anomalies, stopping attacks before they penetrate deeper layers of the network (Jayalaxmi, Saha, Kumar, Conti, & Kim, 2022).

In one case study conducted by CyGlass, a cloud-based security service for small banks, it was reported that AI-supported 24/7 network monitoring enabled the security team to gain clear visibility into network activity and respond rapidly. By continuously scanning incoming network data, the system identifies, prioritizes, and reports high-risk activities (Boamah, Asante, Timean, & Okai, 2025).

4.3. Management of False Positives and Human–Machine Collaboration

Relying on machine learning model decisions for automated blocking can introduce the risk of false positives. In such cases, legitimate activities (for example, customer spending while traveling) may be mistakenly flagged as malicious, potentially leading to customer dissatisfaction (Van Rooy, 2024).

To address this issue, many modern systems adopt a hybrid approach. Alerts generated by machine learning are evaluated in conjunction with rule sets or expert validation. For instance, anomalies detected by AI are cross-checked using a rule engine. If the anomaly matches known Indicators of Compromise

(IoCs), automated intervention is triggered; otherwise, the system generates an informational warning only (Noor, Imoize, Li, & Weng, 2025).

4.4. Trend Analysis and Future Prediction

Another key advantage of machine learning–based early warning systems is their ability to perform trend analysis. Models that continuously process data streams gradually learn patterns of attack behavior. For example, periodic increases in specific types of fraud or a rise in anomalous requests from a particular geographic region can be detected by the model, allowing relevant units to be alerted in advance. As a result, banks can proactively prepare for attack waves that have not yet directly affected them but are beginning to spread across the industry (Muralitharan et al., 2021).

5. CONCLUSION

As the number, diversity, and sophistication of cyberattacks in mobile banking continue to grow, the shift in financial institutions' security strategies from a reactive to a proactive approach has become inevitable. This study highlights the critical role of machine learning in mobile banking security and the tangible benefits it provides.

The findings and reviewed case studies indicate that:

- **Speed Factor:** Attacks that occur at speeds beyond human detection capabilities can be identified within milliseconds through machine learning algorithms.
- **Detection of the Unknown:** Zero-day attacks and emerging malware that evade traditional signature-based systems can be detected through behavioral analysis performed by AI models.

- **Cost Advantage:** Preventing attacks before they occur protects banks from multi-million-dollar fraud losses and, more importantly, from reputational damage.

In conclusion, machine learning for the early detection and proactive prevention of cyberattacks in mobile banking is no longer a luxury but has become an essential component of cybersecurity. With future advancements, cyber defense systems are expected to become increasingly autonomous, enabling real-time threat intelligence sharing whereby insights gained by one bank can instantly protect others, forming a collective cyber defense ecosystem. As the banking industry continues to invest in these intelligent security technologies, the operational space for cybercriminals will narrow, and the digital financial ecosystem will become more secure.

REFRENCES

Abdullah, S. M., Ahmed, B., & Ameen, M. (2018). A New Taxonomy of Mobile Banking Threats, Attacks and User Vulnerabilities. *Eurasian Journal Of Science And Engineering*, *3*(3), 12–20. doi:10.23918/EAJSE.V3I3P12

Abiola, O. B. (2023). Exploring Mobile Banking App Security from User's Perspectives. doi:10.20533/ijisr.2042.4639.2023.0122

Ahuja, B., Prabha, C., & Garg, G. (1AD). Emerging Technologies in E-Commerce Security. *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/979-8-3693-5718-7.Ch010*, 235–260. doi:10.4018/979-8-3693-5718-7.CH010

Al Hwaitat, A. K., Fakhouri, H. N., Alawida, M., Atoum, M. S., Abu-Salih, B., Salah, I. K. M., … Alassaf, N. (2024). Overview of Mobile Attack Detection and Prevention Techniques Using Machine Learning. *International Journal of Interactive Mobile Technologies*, *18*(10), 125. doi:10.3991/IJIM.V18I10.46485

Aljamal, M., Alsarhan, A., Aljaidi, M., Mughaid, A., Al-Jamal, W. Q., & Twairesh, A. E. (2025). Securing the Mobile Future: An Extensive Analysis of the Threat Landscape from Mobile Devices User Perspectives. *International Journal of Interactive Mobile Technologies*, *19*(8), 140. doi:10.3991/IJIM.V19I08.51959

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science*, *3*, 563060. doi:10.3389/FCOMP.2021.563060/BIBTEX

Apruzzese, G., Laskov, P., Montes De Oca, E., Mallouli, W., Brdalo Rapa, L., Grammatopoulos, A. V., & Di Franco, F.

(2023). The Role of Machine Learning in Cybersecurity. *Digital Threats: Research and Practice*, *4*(1). doi:10.1145/3545574;Issue:Issue:Doi

ArnaldoIgnacio, & VeeramachaneniKalyan. (2019). The Holy Grail of 'Systems for Machine Learning'. *ACM SIGKDD Explorations Newsletter*, *21*(2), 39–47. doi:10.1145/3373464.3373472

Ashawa, M. A. (2021). The detection and prevention of Malware attacks on android mobile through the application of artificial intelligence techniques.

Barker, R. (2020). The use of proactive communication through knowledge management to create awareness and educate clients on e-banking fraud prevention. *South African Journal of Business Management*, *51*(1).

Bhardwaj, A. (2021). Cybersecurity Incident Response Against Advanced Persistent Threats (APTs). *EAI/Springer Innovations in Communication and Computing*, 189–209. doi:10.1007/978-3-030-69174-5_9

Boamah, K., Asante, A., Timean, A., & Okai, K. (2025). Artificial intelligence integration in cyber incident response teams to enable faster containment, forensic accuracy, and resilient business continuity.

Cekerevac, Z., Cekerevac, P., Prigoda, L., Journal, F. A.-N.-M., & 2025, undefined. (2025). Security Risks From The Modern Man-In-The-Middle Attacks. *Meste.Org*, *13*(1), 34–51. doi:10.12709/mest.13.13.01.04

Cele, S. (2023). The Disruptive Impact of Financial Technologies on the Financial Services Industry in South Africa.

Choudhury, N. R., & Paul, S. (1AD). Comparative Analysis of Traditional vs. AI-Driven Network Security.

Cortes, I. (2024). Security Vulnerabilities in Mobile Operating Systems Used in IoT Devices: An Examination of Current Challenges and Countermeasures.

Cugnasco, F. (2023). Analysis of a multi-channel anti-fraud platform in banking.

Daninga, P., Losioki, B., Kitali, L., Adili, Z., & Bulengela, G. (2022). The Mwalimu Nyerere Memorial Academy Directorate Of Research, Consultancy And Publication Proceedings of the I st Academic Conference in Commemoration of the Late Mwalimu Julius Kambarage Nyerere, the First President of United Republic of Tanzania and Fa….

Dasgupta, S., Yelikar, B. V., Ramnarayan, Naredla, S., Ibrahim, R. K., & Alazzam, M. B. (2023). AI-Powered Cybersecurity: Identifying Threats in Digital Banking. *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2023*, 2614–2619. doi:10.1109/ICACITE57410.2023.10182479

Datta, J., Dasgupta, R., Dasgupta, S., & Reddy, K. R. (2021). Real-Time Threat Detection in UEBA using Unsupervised Learning Algorithms. *2021 5th International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2021*. doi:10.1109/IEMENTECH53263.2021.9614848

Deshpande, A. V. (2025). Securing Mobile Banking: Global Trends, Threats, and a Conceptual Security Framework. *Authorea Preprints*. doi:10.22541/AU.176252964.45219761/V1

Esoimeme, E. E. (2021). Identifying and reducing the money laundering risks posed by individuals who have been

unknowingly recruited as money mules. *Journal of Money Laundering Control*, *24*(1), 201–212. doi:10.1108/JMLC-05-2020-0053

Gudala, L., Shaik, M., … S. V.-… L. and B., & 2019, undefined. (n.d.). Leveraging artificial intelligence for enhanced threat detection, response, and anomaly identification in resource-constrained iot networks.

Islam, S., Haque, M. M., Naser, A., & Karim, M. R. (2024). A rule-based machine learning model for financial fraud detection. *International Journal of Electrical and Computer Engineering (IJECE)*, *14*(1), 759–771. doi:10.11591/ijece.v14i1.pp759-771

Jayalaxmi, P. L. S., Saha, R., Kumar, G., Conti, M., & Kim, T. H. (2022). Machine and Deep Learning Solutions for Intrusion Detection and Prevention in IoTs: A Survey. *IEEE Access*, *10*, 121173–121192. doi:10.1109/ACCESS.2022.3220622

Karim, N. A., Khashan, O. A., Kanaker, H., Abdulraheem, W. K., Alshinwan, M., & Al-Banna, A. K. (2024). Online Banking User Authentication Methods: A Systematic Literature Review. *IEEE Access*, *12*, 741–757. doi:10.1109/ACCESS.2023.3346045

Korteling, J. E. (Hans), van de Boer-Visschedijk, G. C., Blankendaal, R. A. M., Boonekamp, R. C., & Eikelboom, A. R. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, *4*, 622364. doi:10.3389/FRAI.2021.622364/BIBTEX

Liu, H., & Lang, B. (2019). Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. *Applied Sciences 2019, Vol. 9, Page 4396*, *9*(20), 4396. doi:10.3390/APP9204396

Muhammad, Z., Anwar, Z., Javed, A. R., Saleem, B., Abbas, S., & Gadekallu, T. R. (2023). Smartphone Security and Privacy: A Survey on APTs, Sensor-Based Attacks, Side-Channel Attacks, Google Play Attacks, and Defenses. *Technologies 2023, Vol. 11, Page 76*, *11*(3), 76. doi:10.3390/TECHNOLOGIES11030076

Muppalaneni, R., Inaganti, A. C., & Ravichandran, N. (2024). AI-Driven Threat Intelligence: Enhancing Cyber Defense with Machine Learning. *Journal of Computing Innovations and Applications*, *2*(1), 1–11. doi:10.63575/

Muralitharan, S., Nelson, W., Di, S., McGillion, M., Devereaux, P. J., Barr, N. G., & Petch, J. (2021). Machine learning–Based early warning systems for clinical deterioration: Systematic scoping review. *Journal of Medical Internet Research*, *23*(2), e25187. doi:10.2196/25187

Noor, K., Imoize, A. L., Li, C. T., & Weng, C. Y. (2025). A Review of Machine Learning and Transfer Learning Strategies for Intrusion Detection Systems in 5G and Beyond. *Mathematics 2025, Vol. 13, Page 1088*, *13*(7), 1088. doi:10.3390/MATH13071088

Panda, S. P. (n.d.). The Evolution and Defense Against Social Engineering and Phishing Attacks. *International Journal of Science and Research*. doi:10.21275/SR25504223645

Pinjarkar, L., Hete, P. R., Mattada, M., Nejakar, S., Agrawal, P., & Kaur, G. (2024). An Examination of Prevalent Online Scams: Phishing Attacks, Banking Frauds, and E-Commerce Deceptions. *2nd IEEE International Conference on Advances in Information Technology, ICAIT 2024 - Proceedings*. doi:10.1109/ICAIT61638.2024.10690377

Pureti, N. (Nagaraju). (2022). Zero-Day Exploits: Understanding the Most Dangerous Cyber Threats. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(2), 70–97. Retrieved from https://www.neliti.com/publications/603775/

Reijonen, A. (2024). The Evolution of Mobile Malware. Retrieved from http://www.theseus.fi/handle/10024/851969

Sermet, B. (2021). Detection and remediation of Turkish phishing websites. Retrieved from https://risc01.sabanciuniv.edu/record=b2586113_(Table of contents)

Shabbir, A., Shabir, M., Javed, A. R., Chakraborty, C., & Rizwan, M. (2022). Suspicious transaction detection in banking cyber–physical systems. *Computers & Electrical Engineering*, *97*, 107596. doi:10.1016/J.COMPELECENG.2021.107596

Tsantikidou, K., & Sklavos, N. (2024). Threats, Attacks, and Cryptography Frameworks of Cybersecurity in Critical Infrastructures. *Cryptography*, *8*(1). doi:10.3390/CRYPTOGRAPHY8010007

Valiquette L'Heureux, A. (2022). The Case Study of Los Angeles City & County Fraud, Embezzlement and Corruption Safeguards during times of pandemic. *Public Organization Review*, *22*(3), 593–610. doi:10.1007/S11115-022-00641-W

Van Rooy, D. (2024). Human–machine collaboration for enhanced decision-making in governance. *Data & Policy*, *6*, e60. doi:10.1017/DAP.2024.72

Wei, Y., Chow, K. P., & Yiu, S. M. (2020a). Insider Threat Detection Using Multi-autoencoder Filtering and

Unsupervised Learning. *IFIP Advances in Information and Communication Technology*, *589 IFIP*, 273–290. doi:10.1007/978-3-030-56223-6_15

Wei, Y., Chow, K. P., & Yiu, S. M. (2020b). Insider Threat Detection Using Multi-autoencoder Filtering and Unsupervised Learning. *IFIP Advances in Information and Communication Technology*, *589 IFIP*, 273–290. doi:10.1007/978-3-030-56223-6_15

Zoppi, T., Ceccarelli, A., Capecchi, T., & Bondavalli, A. (2021). Unsupervised Anomaly Detectors to Detect Intrusions in the Current Threat Landscape. *ACM/IMS Transactions on Data Science*, *2*(2), 1–26. doi:10.1145/3441140;TAXONOMY:TAXONOMY:ACM-PUBTYPE;PAGEGROUP:STRING:PUBLICATION

PREDICTION OF ALZHEIMER DISEASE WITH CNN MODEL

Rıfat AŞLIYAN[1]

1. INTRODUCTION

Alzheimer's disease (AD) is one of the vital neurological problems resulting from damage to brain cells. Mostly affecting the elderly, this disease is irreversible under current conditions, and its progression can only be slowed through medication. Accounting for 60-80% of dementia, the AD symptoms include an inability to maintain daily life activities, cognitive impairment, and memory loss. Nowadays, there is no definitive cure; however, if it is diagnosed early, the progression of the disease can be slowed by therapeutic and pharmacological interventions. Therefore, making an accurate diagnosis during the initial stages of the disease is of great importance, as it enables the earlier initiation of treatment. Consequently, numerous studies are being conducted with the aim of AD's early diagnosis (Al Shehri, 2022; Bhade & Bamnote, 2024; Boyapati et al., 2023; Lien et al., 2023).

The diagnosis of AD is achieved through neuropsychological tests and the examination of MRI images by radiologists. Detecting microscopic differences in brain tissue is both a time-consuming process and prone to conflicting interpretations by experts. Consequently, the rate of accurate early diagnosis for AD can be low. For these reasons, deep learning models, a specialized subfield of artificial intelligence, have gained prominence in the analysis of medical images. These

[1] Dr. Öğr. Üyesi, Aydın Adnan Menderes Üniversitesi, Fen Fakültesi, Matematik, ORCID: 0000-0003-1495-713X.

models are capable of processing large datasets and detecting complex structures. Convolutional Neural Network (CNN) models have been shown to be highly successful in image classification and segmentation in recent years,. These systems provide significant support to radiologists, thereby enabling earlier and more accurate diagnoses (Mridha et al., 2025; Priyadharshini et al., 2025; Shastry et al., 2022; Singh et al., 2024; Soujanya et al., 2023; Yeasir et al., 2025).

Arpitha (2024) discusses deep learning techniques for AD predictions, different works about MRI images and text data. This paper reviews many CNN, DNN models including enhancement methods, performances of models and extractions of features.

Rathod and Degadwala (2024) mention different ML and DL techniques to categorize the disease of Alzheimer. They have assessed these methods as RNN and CNN comparing the classification metrics for enhancing the accuracy.

Gagneja et al. (2024) proposed a model for AD with CNN. The model classifies AD based on different criteria. The model has achieved test accuracy of 86.91% by addressing the data in medical imaging.

Saraswathi et al. (2025) compared the CNNs, VGG16, and ResNet to detect AD phases with brain MRI. Whereas CNNs have been used, VGG16 has been superior of them in the accuracy of categorization. The model has distinguished among the four impairment categories by the capabilities of strong feature extraction.

In the study of De Silva & Kunz (2023), a model of CNN has been used to detect AD from MRI, achieving a Correlation Coefficient of Matthew of 0.77, AUC of 0.92, F1-score of 0.89, and accuracy of 0.89. This system has indicated great diagnostic capability.

In this study, a model of CNN has been proposed to diagnose early AD. The brain MRI dataset consists of four classes as Healthy, Moderate, Mild and Very Mild AD classes. The purpose of the CNN model is to learn the microscopic structures of the brain and to detect the differences among the categories. The dataset has been divided into test set and train set. Specifically, 5120 images have been utilized for training, whereas 1280 images have been used for testing. Models have been constructed using different optimization techniques and learning rates. Then the models' performances have been compared with each other. The of the models has been evaluated using F1-Score, Accuracy, and ROC curves.

2. MATERIALS AND METHODS

In this study, a CNN model has been developed to determine the stage of AD. The system analyzes brain MRI to detect whether the disease is present and, if so, at what stage. This section presents details about the Alzheimer's MRI dataset and the constructed CNN model.

2.1. Alzheimer Disease MRI Dataset

The AD Magnetic Resonance Imaging (MRI) dataset (Salieh, 2023) is a dataset mostly utilized to train and test deep learning techniques for classification and diagnosis of AD. The dataset includes MRI images labeled with the medical experts by the examination the conditions of many patients' cognition. The dataset has specific four categories of AD as "Non-Demented", "Mild Demented", "Moderate Demented" and "Very Mild Demented". The category of "Non-Demented" states healthy MRI image. The classes of "Mild Demented", "Moderate Demented" and "Very Mild Demented" indicate mild, very mild and moderate dementia respectively.

Table 1 shows the number of images and memory capacities of the Alzheimer's disease test and training datasets in "parquet" format. Figures 1 and 2 graphically indicate the number of images in the training and test datasets.

"Mild Demented" has been encoded with 0, "Moderate Demented" with 1, "Non-Demented" with 2, and "Very Mild Demented" with 3 in this AD dataset. The dataset provides machine learning techniques to decide both an individual's dementia and the disease's severity. In general, this dataset comprises of standardized 2D MRI scans on the basis of the OASIS database. One major problem in multi-class datasets is to be unbalanced. Namely, the sizes of the classes are not equal each other. Especially the number of images labeled as "Moderate Demented" is quite lower than healthy images. The performance of the model has been improved with data augmentation techniques. Therefore, the model can train efficiently every category in unbalanced dataset.

Table 1. Some features of Alzheimer Disease MRI Dataset.

Features	Train Dataset	Test Dataset
Sample Size	5120	1280
Number of Bytes	22560791	5637447

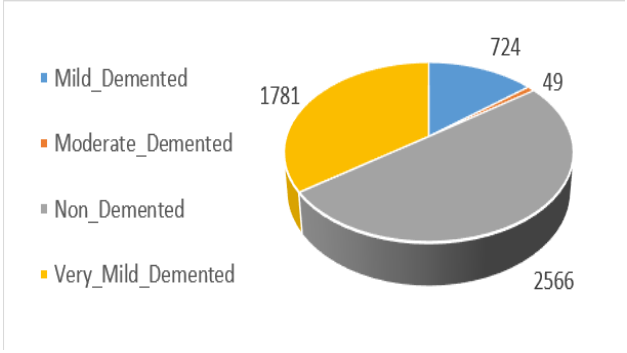


Figure 1. Class distribution of the train dataset.

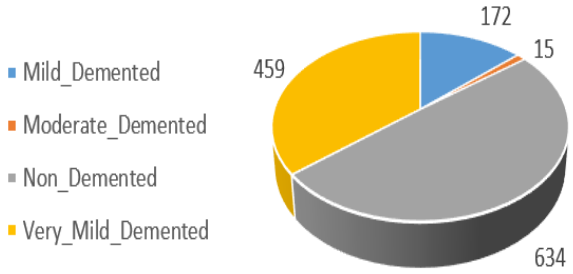


Figure 2. Class distribution of the test dataset.

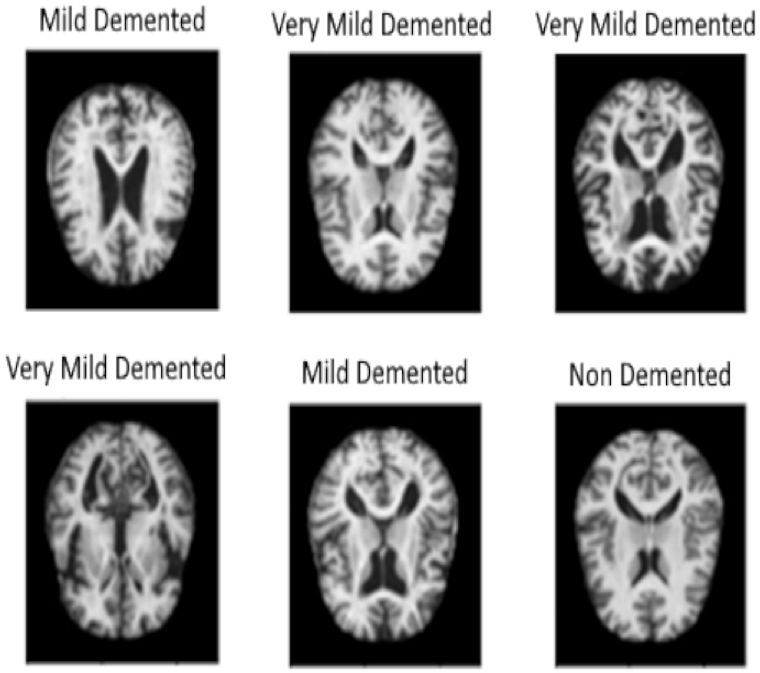


Figure 3. Some Alzheimer MRI images of the dataset.

2.2. Convolutional Neural Network

Convolutional Neural Network (CNN), a subset of deep learning, is an architecture specifically designed to process images. Their most distinguishing feature is the ability to extract image features from the dataset automatically. The foundation of these networks rests on the experiments conducted by Hubel and Wiesel on cats. Additionally, Yann LeCun's work on CNNs in the late 1980s made significant contributions to the area.

Multi-Layer Perceptrons, a type of traditional Artificial Neural Network, perform calculations by converting image data into one-dimensional arrays. Thus, the spatial relationships between pixels within the image are disregarded, which leads to a reduction in system performance. Furthermore, in such networks, the computational load increases drastically as the number of layers grows, particularly due to the full connectivity between layers.

In CNN models, however, the number of parameters can be optimized using local connectivity and weight sharing approaches. Images are processed as matrices rather than one-dimensional arrays, ensuring that spatial features are preserved. Generally, CNN architectures are divided into two main sections: "Feature Extraction" and "Classification". In feature extraction stage, image features are extracted using filters automatically. These features are then processed in the classification stage. During training, the initial layers of the CNN learn important features as corners and edges, while next layers integrate the features to recognize larger and more complex patterns.

Convolutional Neural Networks are built upon four structures arranged in a hierarchical and sequential manner. At the core of these networks lies the kernel, which processes the input image. There is a Convolutional Layer where the kernel detects edge, corner, and texture features by sliding over the

image. Activation Functions are included to these obtained linear maps in order to learn complex structures. Then, a Pooling Layer is added to reduce dimensionality and overfitting. Finally, a Fully Connected Layer is included at final layer of this architecture.

The convolution layer is where features are extracted from images. The convolution process occurs by sliding matrices called kernels over the image. Filters create feature maps containing details such as edges, corners, or textures by performing matrix multiplication with the image. We determine how many pixels the kernel matrix will skip using the Stride value. Data loss occurs at the edge pixels during the sliding of the kernels. The Padding method is used to preserve the dimensions of the output matrix. By adding appropriate values (such as 0) to the image's border, it ensures that the filter processes all pixels with equal weight.

Activation functions can produce solutions using non-linearity for complex problems that cannot be linearly separated. The most frequently used activation functions are Sigmoid and Tanh functions. While the Sigmoid function produces output in the 0-1 range, the Tanh function yields results in the range of 1 to -1. However, in multi-layered deep networks, they cause the vanishing gradient problem as they hinder weight updates. Therefore, as a solution to this problem, the Rectified Linear Unit function (ReLU), $f(x)=\max(0,x)$, is utilized for CNNs in general.

The Pooling layer, which follows the convolution layers, is a critical building block used to reduce computational costs and prevent the network from memorizing, overfitting, by reducing the dimensional size of feature maps, down sampling. One of the two fundamental methods applied in this process, Max Pooling, selects the highest pixel value in the relevant area, ensuring that

the most dominant features as textures, edges, and corners are transferred to the subsequent layer.

The Fully Connected Layer is the decision mechanism where the feature extraction phase of the CNN model concludes and the classification process takes place. Multi-dimensional feature maps obtained in the convolution and pooling layers, which contain the hierarchical features of the image, are converted into a one-dimensional vector via the "Flattening" process and transmitted to this layer.

One of the most fundamental problems faced by deep learning models is overfitting, which causes the model to make incorrect predictions by memorizing noise or specific examples in the training data. The Dropout method, which prevents the network from becoming overly dependent on specific connections by randomly disabling neurons with a certain probability during training, and the Batch Normalization technique, which both accelerates training and provides stability by standardizing the data distribution between layers, are widely used.

Dropout is a widely accepted regularization method in the literature developed to prevent overfitting, which is one of the most critical problems encountered during the training process of deep neural network.

The "Internal Covariate Shift" problem, defined as the continuous change in the distribution of data arriving at each layer as the network's parameters are updated during training, slows down the model's learning speed and makes training difficult. Batch Normalization minimizes this distribution shift by normalizing the incoming data at each training step so that it has a mean of 0 and a variance of 1.

2.3. Proposed CNN Model

As seen in Figure 4, the model used for the detection of AD in this study has been constructed. The CNN model proposed here consists of 14 layers. This model includes 3 convolutional, 4 ReLU, 1 Batch Normalization, 3 max pooling, 1 flatten, and 2 fully connected layers.

Figure 4. The architecture of the proposed CNN model.

3. RESULTS AND DISCUSSION

The proposed CNN model's performance for the detection of Alzheimer's disease has been evaluated using Precision, Recall, ROC curve, AUC, and Accuracy metrics in this work. Additionally, the models have been compared using different parameters.

The ratio of correctly categorized samples determines the robust metrics. As shown in Equation 1-4, the study utilizes the metrics: Recall (Re.), Precision (Pr.), F1-Score, and Accuracy (Ac.).

$$\text{Re.} = \text{TP}/(\text{TP} + \text{FN}) \quad (1)$$

$$\text{Pr.} = \text{TP}/(\text{TP} + \text{FP}) \quad (2)$$

$$\text{Ac.} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (3)$$

$$\text{F1} - \text{Score} = (2 \text{ x Pr. x Re. l})/(\text{ Pr.} + \text{Re.}) \quad (4)$$

The Receiver Operating Characteristic (ROC) curve serves as a visual graph to evaluate model performances, whereas the Area Under the Curve (AUC) is used to compare the effectiveness of different systems. All implementation and testing have been conducted using the Python programming language.

The CNN models have been trained and tested on a workstation equipped as Intel Core i7 2.30 GHz processor, 40 GB of RAM, and the Windows 11 operating system.

3.1. CNN Model Results According To Learning Rates

In this study, every model was run 5 times. Table 2 displays the CNN models' results executed 5 times according to different learning rates. When the learning rate was 0.1 and 0.01, F1-Score, Recall, Precision, AUC, and Accuracy values could not

exceed 50%. Thus, the model was very unsuccessful at these learning rates. However, the highest success values occurred when the learning rate was 0.0001, with Accuracy, Precision, Recall, and F1-Score reaching 98.5%. The AUC value was obtained as 99.9% in the most successful model.

In Table 3, the means and standard deviations (std) of the metrics according to learning rates are shown when a batch size of 8 and Adam optimizer have been used. The best average Accuracy, F1-Score, Precision, and Recall obtained from the 5 runs were 97.7%, and the average AUC was 97.7% when the learning rate was 0.0001. The standard deviation was found to be 0.008 for Accuracy, F1-Score, Precision, and Recall, and 0.002 for AUC.

Table 2. The results of Accuracy, Precision, Recall, F1-Score and AUC according to learning rates when the optimizer is Adam.

Learning Rate	Run	Accuracy	Precision	Recall	F1-Score	AUC
0.1	1	0.495	0.245	0.495	0.328	0.5
0.1	2	0.495	0.245	0.495	0.328	0.5
0.1	3	0.495	0.245	0.495	0.328	0.5
0.1	4	0.495	0.245	0.495	0.328	0.5
0.1	5	0.495	0.245	0.495	0.328	0.5
0.01	1	0.495	0.245	0.495	0.328	0.5
0.01	2	0.495	0.245	0.495	0.328	0.5
0.01	3	0.495	0.245	0.495	0.328	0.5
0.01	4	0.495	0.245	0.495	0.328	0.5
0.01	5	0.495	0.245	0.495	0.328	0.5
0.001	1	0.923	0.924	0.923	0.923	0.985
0.001	2	0.963	0.963	0.963	0.962	0.995
0.001	3	0.952	0.952	0.952	0.952	0.992
0.001	4	0.967	0.967	0.967	0.967	0.996
0.001	5	0.874	0.874	0.874	0.873	0.967
0.0001	1	**0.985**	**0.985**	**0.985**	**0.985**	**0.999**
0.0001	2	0.981	0.982	0.981	0.981	0.998
0.0001	3	0.976	0.976	0.976	0.976	0.997
0.0001	4	0.964	0.964	0.964	0.964	0.994
0.0001	5	0.979	0.979	0.979	0.979	0.998
0.00001	1	0.859	0.861	0.859	0.857	0.96
0.00001	2	0.869	0.875	0.869	0.868	0.961
0.00001	3	0.8	0.821	0.8	0.796	0.939
0.00001	4	0.876	0.877	0.876	0.875	0.968
0.00001	5	0.838	0.842	0.838	0.838	0.942

Table 3. The mean and standard deviation results of Accuracy, Precision, Recall, F1-Score and AUC according to learning rates when the optimizer is Adam.

Learning Rate	Accuracy		Precision		Recall		F1-Score		AUC	
	mean	std	mean	std	mean	std	mean	std	mean	std
0.00001	0.848	0.031	0.855	0.024	0.848	0.031	0.847	0.032	0.954	0.013
0.0001	**0.977**	0.008	**0.977**	0.008	**0.977**	0.008	**0.977**	0.008	**0.997**	0.002
0.001	0.936	0.038	0.936	0.039	0.936	0.038	0.935	0.039	0.987	0.012
0.01	0.495	0	0.245	0	0.495	0	0.328	0	0.5	0
0.1	0.495	0	0.245	0	0.495	0	0.328	0	0.5	0

3.2. CNN Model Results According To Optimizers

Table 4 presents the results of the CNN models executed five times using different optimizers. In terms of Accuracy, Precision, Recall, AUC, and F1-Score, Adagrad proved to be the least effective, whereas RMSprop, Adam, Radam, and AdamW were the most successful optimizers. The highest performance values were achieved with RMSprop, reaching an Accuracy of 98.8% and an F1-Score of 98.7%.

Table 5 displays the means and standard deviations (std) of the metrics across different optimizers, with a fixed learning rate of 0.0001 and a batch size of 8. Following the five runs, the best average Accuracy and F1-Score were recorded at 98.3%, along with an average AUC of 99.9%.

Table 4. The results of Accuracy, Precision, Recall, F1-Score and AUC according to optimizers when learning rate is 0.0001.

Optimizer	Run	Accuracy	Precision	Recall	F1-Score	AUC
Adam	1	0.978	0.978	0.978	0.978	0.997
Adam	2	0.981	0.981	0.981	0.981	0.999
Adam	3	0.984	0.985	0.984	0.984	0.997
Adam	4	0.984	0.984	0.984	0.983	0.999
Adam	5	0.976	0.976	0.976	0.976	0.998
AdamW	1	0.981	0.981	0.981	0.981	0.998
AdamW	2	0.973	0.974	0.973	0.973	0.998
AdamW	3	0.976	0.976	0.976	0.976	0.997
AdamW	4	0.976	0.976	0.976	0.976	0.998
AdamW	5	0.968	0.969	0.968	0.968	0.997
RMSprop	1	0.979	0.979	0.979	0.979	0.998
RMSprop	2	0.985	0.985	0.985	0.985	0.998

RMSprop	3	0.984	0.985	0.984	0.984	0.998
RMSprop	4	**0.988**	**0.988**	**0.988**	**0.987**	**0.999**
RMSprop	5	0.981	0.981	0.981	0.981	0.999
RAdam	1	0.98	0.981	0.98	0.98	0.998
RAdam	2	0.979	0.979	0.979	0.979	0.998
RAdam	3	0.98	0.98	0.98	0.98	0.998
RAdam	4	0.975	0.975	0.975	0.975	0.996
RAdam	5	0.98	0.981	0.98	0.98	0.998
Adagrad	1	0.645	0.64	0.645	0.63	0.811
Adagrad	2	0.648	0.637	0.648	0.627	0.807
Adagrad	3	0.659	0.645	0.659	0.639	0.818
Adagrad	4	0.634	0.614	0.634	0.607	0.807
Adagrad	5	0.67	0.679	0.67	0.654	0.827

Table 5. The mean and standard deviation results of Accuracy, Precision, Recall, F1-Score and AUC according to optimizers when learning rate is 0.0001.

Optimizer	Accuracy		Precision		Recall		F1-Score		AUC	
	mean	std	mean	std	mean	std	mean	std	mean	std
Adagrad	0.651	0.014	0.643	0.023	0.651	0.014	0.632	0.017	0.814	0.009
Adam	0.981	0.004	0.981	0.004	0.981	0.004	0.981	0.004	0.998	0.001
AdamW	0.975	0.005	0.975	0.005	0.975	0.005	0.975	0.005	0.998	0.001
RAdam	0.979	0.002	0.979	0.002	0.979	0.002	0.979	0.002	0.998	0.001
RMSprop	**0.983**	0.003	**0.984**	0.003	**0.983**	0.003	**0.983**	0.003	**0.999**	0.001

3.3. The Best CNN Model Result Confusion Matrix, ROC Curve and Training Accuracy Graph

Figures 5, 6, and 7 display the confusion matrix, ROC curve, and training accuracy graph, respectively, obtained using the RMSprop optimizer with a learning rate of 0.0001 (Run 4), which given the best results. As observed in the confusion matrix, all 634 images in the “Healthy” class were correctly predicted, whereas 14 out of the 15 images in the “Moderate Dementia” class were classified correctly. Only one image from the “Moderate Dementia” class was misclassified as “Very Mild Dementia”. The ROC curve in Figure 2 demonstrates the high performance of the model, with AUC values for each class measured as 1.0. Finally, Figure 3 presents the graph of accuracy

values against batch iterations, indicating that the model demonstrated excellent learning capability.

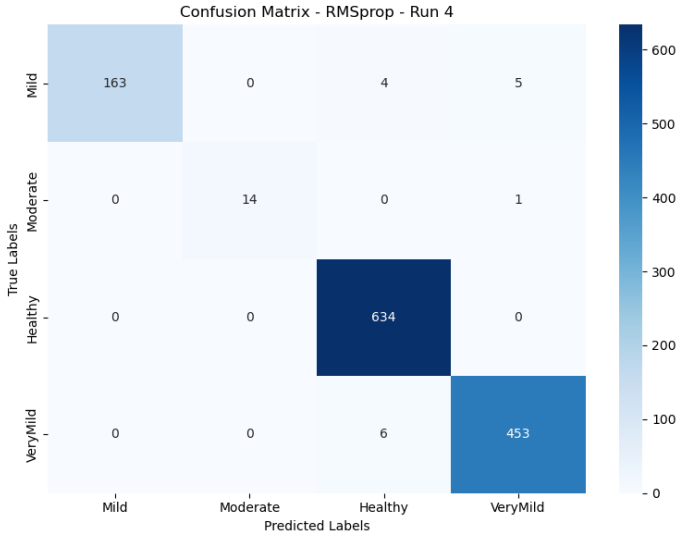


Figure 5. The confusion matrix of the best CNN model (Optimizer=RMSprop, Learning rate=0.001 at run 4).

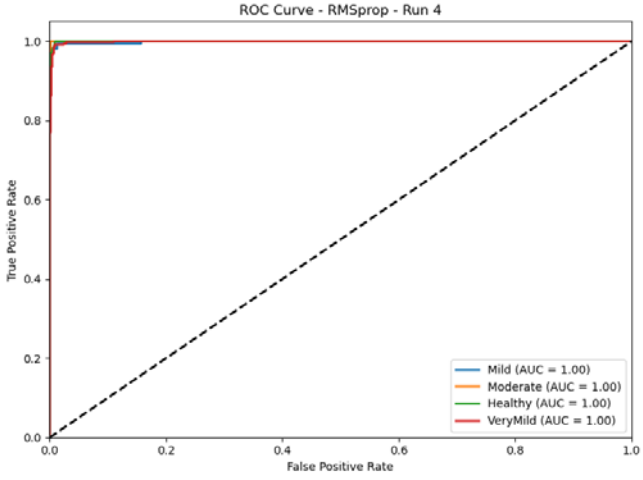


Figure 6. The ROC curve of the best CNN model (Optimizer=RMSprop, Learning rate=0.001 at run 4).

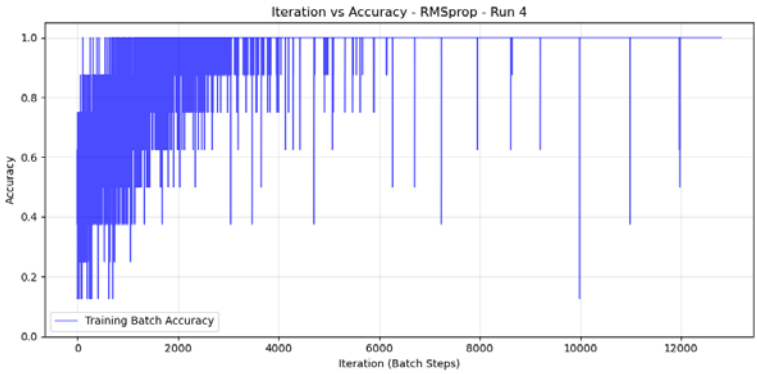


Figure 7. The training accuracy graph of the best CNN model (Optimizer=RMSprop, Learning rate=0.001 at run 4).

4. CONCLUSION

In this study, a CNN model has been proposed to diagnose early Alzheimer's disease with brain MRI dataset which consists of four classes as Healthy, Moderate, Mild and Very Mild AD classes. The proposed models with different parameters as learning rates and optimizers have been trained with 5,120 images. And they have been tested 1,280 testing images. The performance of the models has been evaluated using Accuracy, Precision, Recall, AUC, F1-Score, and ROC curves. The best average Accuracy and F1-Score have obtained at 98.3%, along with an average AUC of 99.9% when learning rate, optimizer and batch size are 0.0001, 8 and RMSprop respectively.

For future research, different CNN models can be developed using various parameters. Additionally, the performance of the models could be further enhanced by employing ensemble techniques.

REFERENCES

Al Shehri, W. (2022). Alzheimer's disease diagnosis and classification using deep learning techniques. *PeerJ Computer Science*, 8, e1177.

Arpitha, V. (2024). Unraveling Alzheimer's Disease Classification with Deep Learning: A Comprehensive Review. *International Journal For Science Technology And Engineering*, 12(8), 1048-1052.

Aşlıyan, R. (2025). Categorization of Tea Leaf Disease With Darknet53 CNN Model. *Bilgisayar Bilimleri ve Mühendisliği Değerlendirmeleri*, Yaz Yayınları, 978-625-5838-79-7, 1, 116-134.

Aşlıyan, R., & Cemek, B. (2025). Determining Categories of Leaf Images Using Transfer Learning of VGGNet. *International Journal of Advanced Natural Sciences and Engineering Researches*, 9(6), 298-310.

Aşlıyan, R. (2024). Classification of Leaf Images with CNN and RF. *International Journal of Advanced Natural Sciences and Engineering Researches*, 4(8), 473-480.

Bhade, A. W., & Bamnote, G. R. (2024). Prediction of Alzheimer's disease using stacked ensemble transfer neural network model. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3), 3407–3415.

Boyapati, N., Tej, M. B., Darshitha, M., & Veluppal, A. (2023). Alzheimer's disease prediction using convolutional neural network (CNN) with generative adversarial network (GAN). *2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, 1-6.

De Silva, K., & Kunz, H. (2023). Prediction of Alzheimer's disease from magnetic resonance imaging using a

convolutional neural network. *Intelligence-Based Medicine*, 7, 100091.

Gagneja, S., Kaur, B., Gupta, R., & Kumar, V. (2024). Deep Learning and AI for Alzheimer's Disease Prediction. 1–6.

Lien, W.-C., Yeh, C.-H., Chang, C.-Y., Chang, C.-H., Wang, W.-M., Chen, C.-H., & Lin, Y.-C. (2023). Convolutional neural networks to classify Alzheimer's disease severity based on SPECT images: A comparative study. *Journal of Clinical Medicine*, 12(6), 2218.

Mridha, M. M. R., Imtiaz, S., Nakib, F. N., & Tasnim, Z. (2025). Revolutionizing Alzheimer's: A CNN model for early detection and accurate prediction. *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1-6.

Priyadharshini, G., Ponni, P., & Subramanian, S. (2025). Improving Alzheimer's diagnosis with deep learning and pre-trained CNN models. *2025 11th International Conference on Communication and Signal Processing (ICCSP)*, 1-5.

Rathod, P. S., & Degadwala, S. (2024). A Review on Alzheimer Disease Classification using different ML and DL Models. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(3), 412-423.

Salieh , F.G. (2023). Alzheimer MRI Dataset, Hugging Face, version 1.0. https://huggingface.co/datasets/Falah/Alzheimer_MRI.

Saraswathi, K., Harita, T., Nandal, N., & Vibinikha, E. M. (2025). Alzheimer Disease Prediction using Deep Learning. 708–716.

Shastry, K. A., Vijayakumar, V., Manoj Kumar, M. V., Manjunatha, B. A., & Chandrashekhar, B. N. (2022). Deep learning techniques for the effective prediction of Alzheimer's disease: A comprehensive review. *Healthcare*, 10(10), 1842.

Singh, G., Guleria, K., & Sharma, S. (2024). A deep learning-based convolutional neural network model for Alzheimer's disease detection. *2024 Second International Conference on Computational and Characterization Techniques in Engineering & Sciences (IC3TES)*, 944-948.

Soujanya, S., Anuradha, K., Srilakshmi, V., & Adilakshmi, K. (2023). Comparative analysis of CNN algorithms for Alzheimer's detection. *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 753-758.

Yeasir, M. R., Nakib, F. N., Mizan, M. A. I., Mridha, M. M. R., Shakil, M. S. A., & Imtiaz, S. (2025). Innovative lightweight CNN model for stage-wise Alzheimer's disease detection. *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, 1-6.

MAKİNE ÖĞRENMESİ TABANLI ŞİRKET DEĞERLEME MODELLERİ: YÖNTEMLER VE EĞİLİMLER[1]

Yağmur ATEŞ[2]

Ülviye HACIZADE[3]

1. GİRİŞ

Şirket değerlemesi, yatırım kararlarının alınması, birleşme ve devralma süreçlerinin yürütülmesi, sermaye maliyetinin belirlenmesi ve kurumsal performansın değerlendirilmesi açısından finans literatürünün temel araştırma alanlarından biridir. Bir işletmenin ekonomik değerinin güvenilir biçimde tahmin edilmesi, yalnızca yatırımcılar için değil; finansal kurumlar, yöneticiler ve düzenleyici otoriteler açısından da stratejik öneme sahiptir (Damodaran, 2012). Bu nedenle şirket değerlemesi, hem teorik hem de uygulamalı finans çalışmalarında uzun süredir ele alınan bir konu olmuştur.

Geleneksel şirket değerleme yaklaşımları genel olarak gelir, piyasa ve maliyet temelli yöntemler etrafında şekillenmektedir. İndirgenmiş nakit akımları, çarpan analizi ve benzer şirket karşılaştırmaları gibi yöntemler, belirli varsayımlar altında analitik tutarlılık sağlamakla birlikte, çoğunlukla doğrusal ilişkiler ve durağan parametreler üzerine kuruludur (Koller et al., 2010; Penman, 2013). Finansal piyasaların zaman içinde artan oynaklığı, şirket değerini etkileyen değişkenlerin çeşitlenmesi ve

---

[1] Yüksek Lisans Tezinden üretilmiştir

[2] Haliç Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği (Tezli), ORCID: 0000-0002-3637-0586,

[3] Doç. Dr., Haliç Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, ORCID: 0000-0002-0073-996X

veri yapılarının daha karmaşık hâle gelmesi, bu yaklaşımların bazı bağlamlarda açıklayıcılığını ve tahmin gücünü sınırlayabilmektedir.

Son yıllarda finansal raporlama süreçlerindeki dijitalleşme, büyük veri teknolojilerinin yaygınlaşması ve hesaplama kapasitesindeki artış, şirket değerlemesinde veri odaklı yöntemlere olan ilgiyi artırmıştır. Bu gelişmelerle birlikte makine öğrenmesi algoritmaları, çok boyutlu ve doğrusal olmayan veri yapılarının modellenmesine imkân tanıyan esnek araçlar olarak finans literatüründe giderek daha fazla yer bulmaya başlamıştır (Vayas-Ortega et al., 2020). Makine öğrenmesi yöntemleri, önceden tanımlanmış fonksiyonel formlara bağlı kalmaksızın verinin içsel örüntülerini öğrenebilme yetenekleri sayesinde, şirket değerlemesini bir tahmin problemi olarak ele alma potansiyeli sunmaktadır.

Makine öğrenmesinin finans alanındaki ilk uygulamaları ağırlıklı olarak hisse senedi fiyat tahmini, kredi riski değerlendirmesi ve finansal başarısızlık öngörüsü gibi alanlarda yoğunlaşmıştır (Olson & Mossman, 2003). Zaman içinde bu yaklaşımlar, şirket değeri ve yatırım değeri tahmini gibi daha karmaşık problemlere doğru genişlemiştir. Literatürde yer alan bazı çalışmalar, belirli örneklemler ve piyasa koşulları altında makine öğrenmesi tabanlı modellerin geleneksel regresyon yaklaşımlarına kıyasla daha düşük hata oranları veya daha yüksek açıklayıcılık sağlayabildiğini rapor etmektedir. Bununla birlikte, bu sonuçların veri setinin yapısı, sektör özellikleri ve modelleme tercihleriyle yakından ilişkili olduğu da vurgulanmaktadır (Hu et al., 2022).

Mevcut çalışmalar incelendiğinde, makine öğrenmesi tabanlı şirket değerleme modellerinin tek tip bir yaklaşım sunmadığı; aksine farklı algoritmaların, farklı veri yapıları ve bağlamlar altında öne çıktığı görülmektedir. Yapay sinir ağları,

ağaç tabanlı modeller, topluluk yöntemleri ve hibrit yaklaşımlar literatürde eş zamanlı olarak yer almakta; bu durum, şirket değerlemesinde makine öğrenmesinin bütüncül ve yöntemsel olarak değerlendirilmesini gerekli kılmaktadır. Bu bağlamda, makine öğrenmesi yöntemlerinin geleneksel değerleme yaklaşımlarının doğrudan alternatifi olmaktan ziyade, onları tamamlayıcı analitik araçlar olarak konumlandığı yönünde ortak bir eğilim dikkat çekmektedir (Penman, 2013; Vayas-Ortega et al., 2020).

Şirket değerlemesi literatüründe son yıllarda artan veri hacmi ve yöntemsel çeşitlilik, makine öğrenmesi tabanlı modelleme yaklaşımlarına olan ilgiyi belirgin biçimde artırmıştır. Bu çerçevede, şirket değerleme problemlerinde kullanılan başlıca makine öğrenmesi modelleri ve bu modellerin hangi bağlamlarda öne çıktığı literatürde giderek daha görünür hâle gelmektedir. Farklı algoritmaların kullanım eğilimleri ile çalışmalarda raporlanan ortak bulgular, makine öğrenmesi yöntemlerinin şirket değerlemesindeki konumunun yöntemsel bir bakış açısıyla değerlendirilmesini gerekli kılmaktadır. Bu doğrultuda, ilgili literatürde öne çıkan yaklaşımlar ve güncel eğilimler bütüncül bir çerçevede ele alınmaktadır.

2. ŞİRKET DEĞERLEMESİNDE MAKİNE ÖĞRENMESİ YAKLAŞIMI

Şirket değerlemesi, temel olarak bir işletmenin ekonomik değerinin belirli bir zaman diliminde tahmin edilmesine yönelik analitik bir problem olarak ele alınabilir. Bu yönüyle değerleme, finansal oranlar, muhasebe verileri, makroekonomik göstergeler ve piyasa bilgileri gibi çok sayıda değişkenin birlikte değerlendirilmesini gerektiren çok boyutlu bir tahmin sürecini ifade etmektedir. Makine öğrenmesi yaklaşımları, bu karmaşık veri yapısını önceden tanımlanmış doğrusal varsayımlara bağlı

kalmaksızın modelleyebilme potansiyeli nedeniyle şirket değerleme literatüründe giderek daha fazla kullanılmaktadır (Vayas-Ortega et al., 2020).

Makine öğrenmesi perspektifinden bakıldığında şirket değerlemesi, genellikle denetimli öğrenme çerçevesinde ele alınmaktadır. Bu bağlamda modelin amacı; geçmiş dönemlere ait finansal ve finansal olmayan değişkenleri girdi olarak kullanarak, piyasa değeri, firma değeri veya Tobin's Q gibi hedef değişkenleri tahmin etmektir (Hu et al., 2022). Finansal oranlar, nakit akışı göstergeleri, sermaye yapısı değişkenleri ve büyüme ölçütleri, değerleme modellerinde en sık kullanılan girdiler arasında yer almaktadır. Bunun yanı sıra bazı çalışmalarda makroekonomik göstergeler ve zaman serisi bileşenleri de modele dâhil edilerek tahmin performansının artırılması hedeflenmektedir.

Makine öğrenmesi yöntemlerinin şirket değerlemesine sunduğu temel katkılardan biri, değişkenler arasındaki doğrusal olmayan ilişkileri yakalayabilme kapasitesidir. Geleneksel regresyon temelli modeller, çoğu zaman sabit parametre varsayımlarına dayanırken; makine öğrenmesi algoritmaları veriye dayalı olarak esnek karar sınırları oluşturabilmektedir (Penman, 2013). Bu özellik, özellikle finansal verilerin gürültülü ve heterojen yapısı dikkate alındığında, değerleme problemleri açısından önemli bir avantaj olarak değerlendirilmektedir.

Bununla birlikte, makine öğrenmesi yaklaşımlarının şirket değerlemesinde kullanımı bazı metodolojik sınırlılıkları da beraberinde getirmektedir. Model sonuçlarının yorumlanabilirliği, veri kalitesine duyarlılık ve örneklem bağımlılığı, literatürde sıklıkla vurgulanan konular arasında yer almaktadır. Bu nedenle makine öğrenmesi tabanlı değerleme modelleri, çoğu çalışmada geleneksel değerleme yaklaşımlarının yerine geçmekten ziyade, onları tamamlayıcı analitik araçlar

olarak konumlandırılmaktadır (Koller et al., 2010; Vayas-Ortega et al., 2020).

Sonuç olarak, makine öğrenmesi yaklaşımı şirket değerlemesini, sabit varsayımlara dayalı tek yönlü bir hesaplama süreci yerine, veri odaklı ve uyarlanabilir bir tahmin problemi olarak ele almaktadır. Bu yaklaşım, değerleme literatüründe yöntemsel çeşitliliği artırmakta ve farklı piyasa koşullarında alternatif analiz imkânları sunmaktadır.

3. MAKİNE ÖĞRENMESİ TABANLI ŞİRKET DEĞERLEME MODELLERİ

Makine öğrenmesi tabanlı şirket değerleme modelleri, finansal değerleme problemini önceden tanımlanmış doğrusal varsayımlar yerine, veriye dayalı öğrenme süreçleri aracılığıyla ele almaktadır. Literatürde bu modellerin temel amacı, şirket değerini temsil eden bir hedef değişkenin (piyasa değeri, firma değeri veya Tobin's Q gibi) çok sayıda finansal ve finansal olmayan değişken kullanılarak tahmin edilmesidir. Bu yaklaşım, şirket değerini etkileyen faktörler arasındaki karmaşık ve doğrusal olmayan ilişkilerin daha esnek biçimde modellenmesine imkân tanımaktadır (Hu et al., 2022).

Şirket değerlemesinde kullanılan makine öğrenmesi modelleri, yöntemsel açıdan tek tip bir yapı sunmamaktadır. Aksine, kullanılan algoritmalar veri setinin yapısına, örneklem büyüklüğüne, hedef değişkenin tanımına ve analiz amacına bağlı olarak farklı avantajlar ve sınırlılıklar ortaya koymaktadır. Bu nedenle literatürde, belirli bir algoritmanın tüm değerleme problemleri için evrensel olarak üstün olduğu yönünde bir uzlaşı bulunmamaktadır. Bunun yerine, farklı koşullar altında öne çıkan model grupları söz konusudur.

Tablo 1. Şirket Değerlemesinde Kullanılan Makine Öğrenmesi Modellerinin Genel Özellikleri

Model Türü	Örnek Algoritmalar	Kullanım Amacı	Başlıca Avantajlar	Temel Sınırlılıklar
Yapay Sinir Ağları	ANN, MLP	Firma değeri tahmini	Doğrusal olmayan ilişkiler	Düşük yorumlanabilirlik
Destek Vektör Modelleri	SVM, SVR	Küçük veri setleri	Gürültüye dayanıklılık	Ölçeklenebilirlik
Ağaç Tabanlı Modeller	CART, RF	Heterojen veri	Kararlılık	Model karmaşıklığı
Boosting Yöntemleri	XGBoost, LightGBM	Büyük veri setleri	Yüksek doğruluk	Parametre hassasiyeti
Hibrit / Derin Öğrenme	ANFIS, LSTM	Zaman serisi	Dinamik yapı	Yüksek maliyet

Kaynak: Yazarın derlemesi (2025)

Tablo 1, şirket değerlemesinde kullanılan makine öğrenmesi modellerinin yöntemsel çeşitliliğini ve her bir model grubunun farklı veri yapıları ve analiz amaçları açısından sunduğu temel özellikleri özetlemektedir. Tablo, tek bir modelin tüm değerleme problemleri için evrensel bir çözüm sunmadığını; aksine model seçiminin veri setinin yapısı, hedef değişkenin tanımı ve araştırma bağlamına bağlı olarak şekillendiğini ortaya koymaktadır. Bu çerçeve doğrultusunda izleyen alt bölümlerde, tabloda yer alan model grupları yöntemsel yaklaşımları ve uygulama bağlamları açısından ayrıntılı biçimde ele alınmaktadır.

3.1. Doğrusal Olmayan Öğrenme Yaklaşımları: Yapay Sinir Ağları ve Destek Vektör Modelleri

Makine öğrenmesi tabanlı şirket değerleme literatüründe kullanılan ilk sistematik yaklaşımlar, doğrusal olmayan öğrenme modellerine dayanmaktadır. Bu bağlamda yapay sinir ağları ve destek vektör makineleri, finansal değişkenler ile şirket değeri arasındaki karmaşık ilişkilerin modellenmesinde öncü yöntemler olarak öne çıkmıştır. Söz konusu modeller, geleneksel regresyon

yaklaşımlarının doğrusal varsayımlarına alternatif olarak, veriye dayalı ve esnek bir öğrenme yapısı sunmaktadır.

Yapay sinir ağları (Artificial Neural Networks – ANN), biyolojik sinir sistemlerinden esinlenen çok katmanlı yapıları sayesinde, finansal oranlar, muhasebe kalemleri ve piyasa göstergeleri ile şirket değeri arasındaki doğrusal olmayan ilişkileri öğrenebilme kapasitesine sahiptir. Şirket değerlemesi uygulamalarında ANN modelleri genellikle finansal oranlar, nakit akışı göstergeleri, kârlılık ve büyüme ölçütleri gibi değişkenleri girdi olarak almakta; çıktı katmanında ise piyasa değeri veya firma değeri gibi hedef değişkenleri tahmin etmektedir. Bu yapı, önceden tanımlanmış bir fonksiyonel form gerektirmemesi nedeniyle, karmaşık ve gürültülü finansal verilerle çalışmada önemli bir avantaj sağlamaktadır (Olson & Mossman, 2003).

Literatürde yapay sinir ağlarının şirket değerlemesinde kullanımı, özellikle küçük ve orta ölçekli veri setlerinde esnek tahminler sunabilmesi nedeniyle yaygınlık kazanmıştır. Bununla birlikte, ağ mimarisinin belirlenmesi, gizli katman sayısı ve öğrenme parametrelerinin seçimi büyük ölçüde araştırmacı tercihine bağlıdır. Ayrıca ANN modellerinin "kara kutu" niteliği, model çıktılarının ekonomik olarak yorumlanmasını zorlaştırmakta ve bu durum değerleme gibi açıklanabilirliğin önem taşıdığı alanlarda yöntemsel bir sınırlılık olarak değerlendirilmektedir (Penman, 2013). Bu nedenle literatürde ANN tabanlı modeller çoğunlukla karşılaştırmalı analizlerde veya tamamlayıcı araçlar olarak konumlandırılmaktadır.

Destek vektör makineleri (Support Vector Machines – SVM), doğrusal olmayan öğrenme yaklaşımları içerisinde farklı bir yöntemsel çerçeve sunmaktadır. SVM modelleri, veri noktaları arasında optimum ayırma düzlemleri oluşturarak tahmin yapmayı amaçlamakta; şirket değerlemesi uygulamalarında ise çoğunlukla destek vektör regresyonu (Support Vector Regression

– SVR) biçiminde kullanılmaktadır. Bu yaklaşım, özellikle sınırlı örneklem büyüklüğüne sahip veri setlerinde ve yüksek boyutlu değişken uzaylarında kararlı sonuçlar üretebilmesi nedeniyle tercih edilmektedir.

SVM tabanlı modellerin şirket değerlemesindeki en önemli avantajlarından biri, yapısal risk minimizasyonu ilkesine dayanmaları sayesinde aşırı öğrenme riskini görece sınırlı tutabilmeleridir. Bununla birlikte, veri seti büyüdükçe hesaplama maliyetinin artması ve modelin ölçeklenebilirliğinin sınırlı olması, bu yaklaşımın büyük veri setlerinde kullanımını kısıtlamaktadır (Milosevic, 2016). Bu nedenle destek vektör modelleri, şirket değerleme literatüründe çoğunlukla yapay sinir ağları, rastgele orman veya boosting tabanlı yöntemlerle karşılaştırmalı biçimde ele alınmakta ya da hibrit model yapıları içerisinde değerlendirilmektedir.

Genel olarak değerlendirildiğinde, yapay sinir ağları ve destek vektör modelleri, makine öğrenmesi tabanlı şirket değerleme çalışmalarının kavramsal ve yöntemsel temelini oluşturmaktadır. Bu modeller, şirket değerinin doğrusal olmayan ve çok boyutlu yapısını modelleyebilme potansiyeli sunmakla birlikte, yorumlanabilirlik ve ölçeklenebilirlik gibi konularda belirli sınırlılıklar taşımaktadır. Bu sınırlılıklar, literatürde daha sonraki dönemlerde ağaç tabanlı ve topluluk öğrenme modellerine yönelimin artmasına zemin hazırlamıştır.

Yapay sinir ağları ve destek vektör modelleri, şirket değerlemesinde makine öğrenmesi yaklaşımlarının kavramsal temelini oluşturmakla birlikte, bu modellerin yorumlanabilirlik ve ölçeklenebilirlik açısından taşıdığı sınırlılıklar literatürde yeni yöntem arayışlarını beraberinde getirmiştir. Özellikle daha büyük ve heterojen veri setlerinin analizinde, daha kararlı ve genellenebilir sonuçlar sunabilen model yapılarına olan ihtiyaç artmıştır. Bu bağlamda, ağaç tabanlı ve topluluk (ensemble)

öğrenme yaklaşımları, şirket değerleme literatüründe öne çıkan alternatifler olarak dikkat çekmektedir.

3.2.Ağaç Tabanlı ve Topluluk (Ensemble) Öğrenme Modelleri

Ağaç tabanlı öğrenme modelleri, şirket değerleme literatüründe makine öğrenmesi yöntemlerinin daha olgun ve uygulamaya dönük bir evresini temsil etmektedir. Bu modeller, finansal değişkenler ile şirket değeri arasındaki ilişkileri kural tabanlı bir yapı üzerinden ele alarak, doğrusal olmayan etkileşimleri modelleyebilme kapasitesi sunmaktadır. Aynı zamanda karar süreçlerinin hiyerarşik biçimde izlenebilmesine imkân tanıması, ağaç tabanlı modelleri özellikle finans alanında tercih edilir kılmaktadır.

Karar ağaçları (Classification and Regression Trees – CART), şirket değerlemesinde kullanılan en temel ağaç tabanlı modeller arasında yer almaktadır. Bu modeller, değişken uzayını belirli eşik değerler üzerinden bölerek, şirket değerini etkileyen faktörleri ardışık karar kuralları çerçevesinde analiz etmektedir. Karar ağaçlarının en önemli avantajı, model çıktılarının açık ve sezgisel biçimde yorumlanabilmesidir. Ancak tekil karar ağaçlarının örnekleme duyarlılığı yüksek olup, aşırı uyum (overfitting) eğilimi göstermeleri, genellenebilirlik açısından önemli bir sınırlılık oluşturmaktadır (Milosevic, 2016).

Bu sınırlılıkların giderilmesi amacıyla geliştirilen topluluk (ensemble) öğrenme yaklaşımları, şirket değerleme literatüründe giderek daha yaygın biçimde kullanılmaktadır. Topluluk modelleri, birden fazla zayıf öğrenicinin bir araya getirilmesiyle daha güçlü ve kararlı tahminler üretmeyi amaçlamaktadır. Bu yaklaşım, özellikle finansal veri setlerinde sıkça karşılaşılan gürültü, eksik gözlem ve çoklu doğrusal bağlantı gibi sorunların etkisini azaltma potansiyeli sunmaktadır.

Rastgele orman (Random Forest – RF) algoritması, şirket değerlemesinde en sık başvurulan topluluk öğrenme yöntemlerinden biridir. RF, farklı örneklem alt kümeleri ve rastgele seçilmiş değişken setleri üzerinden oluşturulan çok sayıda karar ağacının çıktısını birleştirerek tahmin üretmektedir. Bu yapı, tekil karar ağaçlarının kararsızlığını önemli ölçüde azaltmakta ve modelin genellenebilirliğini artırmaktadır. Literatürde RF tabanlı değerleme modellerinin, özellikle heterojen ve yüksek boyutlu finansal veri setlerinde istikrarlı sonuçlar sunduğu vurgulanmaktadır (Milosevic, 2016).

Rastgele orman modellerinin şirket değerlemesi açısından dikkat çeken bir diğer özelliği, değişken önem sıralaması yapabilme yeteneğidir. Bu özellik, şirket değerini etkileyen finansal oranların ve muhasebe kalemlerinin göreli katkılarının analiz edilmesine olanak tanımakta; böylece model çıktılarının ekonomik açıdan anlamlandırılmasını kısmen kolaylaştırmaktadır. Bu yönüyle RF, tahmin performansı ile yorumlanabilirlik arasında görece dengeli bir çözüm sunmaktadır.

Topluluk öğrenme yaklaşımları içerisinde boosting tabanlı algoritmalar, literatürde özellikle son yıllarda öne çıkan yöntemler arasında yer almaktadır. Gradient Boosting, XGBoost ve LightGBM gibi algoritmalar, zayıf öğrenicileri ardışık biçimde eğiterek önceki modellerin hatalarına odaklanmakta ve bu sayede toplam hata oranını azaltmayı hedeflemektedir. Şirket değerlemesi bağlamında boosting yöntemleri, büyük örneklem büyüklüğüne sahip veri setlerinde ve karmaşık değişken etkileşimlerinin bulunduğu durumlarda sıklıkla tercih edilmektedir (Hu et al., 2022).

Boosting tabanlı modellerin en önemli avantajı, yüksek tahmin performansı sağlayabilmeleridir. Bununla birlikte, bu modeller hiperparametre ayarlarına karşı duyarlı olup dikkatli bir modelleme süreci gerektirmektedir. Ayrıca model

karmaşıklığının artması, yorumlanabilirlik açısından belirli sınırlılıkları beraberinde getirmektedir. Bu nedenle boosting yöntemleri, şirket değerlemesi literatüründe çoğunlukla performans odaklı analizlerde veya diğer yöntemlerle birlikte kullanılan tamamlayıcı araçlar olarak konumlandırılmaktadır.

Genel olarak değerlendirildiğinde, ağaç tabanlı ve topluluk öğrenme modelleri, şirket değerlemesinde doğrusal olmayan ilişkileri modelleme kapasitesi ile genellenebilirlik arasında daha dengeli bir yapı sunmaktadır. Bu modeller, yapay sinir ağları ve destek vektör makinelerine kıyasla daha istikrarlı sonuçlar üretebilmekte; bu yönüyle makine öğrenmesi tabanlı şirket değerleme literatüründe önemli bir yöntemsel aşamayı temsil etmektedir. Bununla birlikte, artan model karmaşıklığı ve hesaplama maliyetleri, literatürde daha gelişmiş hibrit ve derin öğrenme yaklaşımlarına yönelimin de zeminini oluşturmaktadır.

3.3.Hibrit ve Derin Öğrenme Yaklaşımları

Makine öğrenmesi tabanlı şirket değerleme literatüründe son dönemde öne çıkan eğilimlerden biri, farklı modelleme yaklaşımlarının bir arada kullanıldığı hibrit yapılar ile derin öğrenme temelli modellerin artan kullanım alanıdır. Bu yönelim, ağaç tabanlı ve topluluk öğrenme modellerinin sunduğu genellenebilirlik ve performans avantajlarının, zaman boyutu ve karmaşık veri yapıları açısından daha da geliştirilmesi ihtiyacından kaynaklanmaktadır. Özellikle finansal verilerin zamana bağlı dinamik özellikleri dikkate alındığında, statik model yapılarına kıyasla daha esnek ve uyarlanabilir yöntemlere olan ilgi belirgin biçimde artmıştır.

Hibrit modeller, klasik makine öğrenmesi algoritmalarını farklı yapay zekâ teknikleri veya geleneksel finansal yaklaşımlarla birleştiren yöntemler olarak tanımlanabilir. Şirket değerlemesi bağlamında bu modeller, genellikle yapay sinir ağları ile bulanık mantık, regresyon temelli yöntemler veya finansal

değerleme bileşenlerinin birlikte kullanıldığı yapılar şeklinde karşımıza çıkmaktadır. Bu yaklaşımın temel amacı, tekil modellerin güçlü yönlerini bir araya getirerek daha dengeli ve yorumlanabilir tahminler elde etmektir.

Bu çerçevede literatürde sıkça karşılaşılan hibrit modellerden biri, adaptif sinirsel bulanık çıkarım sistemleridir (Adaptive Neuro-Fuzzy Inference System – ANFIS). ANFIS modelleri, yapay sinir ağlarının öğrenme kapasitesi ile bulanık mantığın kural tabanlı yapısını birleştirerek, şirket değerini etkileyen değişkenler arasındaki ilişkilerin hem öğrenilmesini hem de kısmen yorumlanabilir biçimde ifade edilmesini sağlamaktadır. Özellikle finansal oranlar ile firma değeri arasındaki ilişkilerin belirsizlik içermesi, bu tür hibrit yaklaşımları değerleme literatürü açısından cazip kılmaktadır (Ekşi et al., 2014).

Derin öğrenme tabanlı modeller ise, çok katmanlı sinir ağları aracılığıyla yüksek boyutlu ve karmaşık veri yapılarının modellenmesine odaklanmaktadır. Şirket değerlemesi literatüründe derin öğrenme yaklaşımları, özellikle zaman serisi özellikleri güçlü olan finansal verilerin analizinde kullanılmaktadır. Uzun kısa süreli bellek ağları (Long Short-Term Memory – LSTM), geçmiş dönem bilgilerini belirli bir hafıza mekanizması aracılığıyla taşıyabilme kapasitesi sayesinde, şirket değerinin zaman içerisindeki dinamik değişimini modelleme potansiyeli sunmaktadır. Bu özellik, klasik makine öğrenmesi modellerine kıyasla önemli bir yöntemsel farklılık oluşturmaktadır.

Derin öğrenme modellerinin şirket değerlemesinde kullanımına ilişkin literatürde raporlanan bulgular, bu yaklaşımların belirli koşullar altında yüksek tahmin performansı sağlayabildiğini göstermektedir. Bununla birlikte, bu modellerin yüksek veri gereksinimi, hesaplama maliyetleri ve sınırlı

yorumlanabilirliği, yöntemsel açıdan dikkatle ele alınması gereken hususlar arasında yer almaktadır (Vayas-Ortega et al., 2020). Özellikle finans alanında model çıktılarının ekonomik anlamlandırılması beklentisi, derin öğrenme yaklaşımlarının tek başına değil, çoğu zaman diğer yöntemlerle birlikte değerlendirilmesine yol açmaktadır.

Literatürde hibrit ve derin öğrenme tabanlı modellerin çoğunlukla karşılaştırmalı analizler çerçevesinde ele alındığı görülmektedir. Bu çalışmalar, söz konusu yaklaşımların her veri seti ve piyasa koşulu için üstün sonuçlar sunmadığını; aksine performansın büyük ölçüde veri yapısı, örneklem büyüklüğü ve modelleme tercihlerine bağlı olduğunu ortaya koymaktadır. Bu bulgu, makine öğrenmesi tabanlı şirket değerlemesinde "tek bir en iyi model" yaklaşımının geçerliliğini sınırlamakta ve yöntemsel çeşitliliğin önemini vurgulamaktadır (Hu et al., 2022).

Genel olarak değerlendirildiğinde, hibrit ve derin öğrenme yaklaşımları, makine öğrenmesi tabanlı şirket değerleme literatüründe yöntemsel genişlemenin ve olgunlaşmanın bir göstergesi olarak değerlendirilebilir. Bu modeller, şirket değerinin dinamik ve çok boyutlu yapısını daha ayrıntılı biçimde ele alma potansiyeli sunmakla birlikte, uygulama maliyetleri ve yorumlanabilirlik sınırlılıkları nedeniyle dikkatli ve bağlama özgü bir kullanım gerektirmektedir. Bu durum, makine öğrenmesi yöntemlerinin şirket değerlemesinde geleneksel yaklaşımların yerini alan araçlar olmaktan ziyade, onları tamamlayan ve destekleyen analitik çerçeveler olarak konumlandığı yönündeki genel literatür eğilimiyle örtüşmektedir.

3.4. Makine Öğrenmesi Tabanlı Şirket Değerleme Modellerinin Karşılaştırmalı Değerlendirilmesi

Bu bölümde ele alınan makine öğrenmesi tabanlı şirket değerleme modelleri, yöntemsel yaklaşımları, veri gereksinimleri ve uygulama bağlamları açısından önemli farklılıklar

göstermektedir. Yapay sinir ağları ve destek vektör modelleri, doğrusal olmayan ilişkileri modelleme kapasitesiyle erken dönem çalışmalarda öne çıkarken; ağaç tabanlı ve topluluk öğrenme yaklaşımları daha büyük ve heterojen veri setlerinde genellenebilirlik açısından avantaj sağlamaktadır. Hibrit ve derin öğrenme modelleri ise zaman boyutunu ve karmaşık veri yapısını dikkate alarak literatürde yöntemsel genişlemeyi temsil etmektedir.

Bu çerçevede, tek bir makine öğrenmesi modelinin tüm şirket değerleme problemleri için evrensel olarak üstün olduğu yönünde bir yaklaşım benimsenmemektedir. Aksine, model performansının veri setinin yapısı, örneklem büyüklüğü, hedef değişkenin tanımı ve analiz amacıyla yakından ilişkili olduğu görülmektedir. Dolayısıyla şirket değerlemesinde makine öğrenmesi uygulamalarının, bağlama özgü ve karşılaştırmalı bir değerlendirme çerçevesinde ele alınması gerekmektedir.

Bu değerlendirmeyi sistematik hâle getirmek amacıyla Tablo 2, bu bölümde ele alınan başlıca makine öğrenmesi model gruplarını yöntemsel özellikleri açısından karşılaştırmalı olarak sunmaktadır.

Tablo 2. Makine Öğrenmesi Tabanlı Şirket Değerleme Modellerinin Karşılaştırmalı Özeti

Model Grubu	Öğrenme Yapısı	Veri Gereksinimi	Yorumlanabilirlik	Şirket Değerlemesindeki Rolü
Yapay Sinir Ağları	Çok katmanlı öğrenme	Orta	Düşük	Doğrusal olmayan ilişki modelleme
Destek Vektör Modelleri	Marjinal optimizasyon	Düşük–Orta	Orta	Kararlı karşılaştırmalı tahmin
Ağaç Tabanlı Modeller	Kural tabanlı bölme	Orta	Yüksek	Yapısal ilişki analizi
Topluluk Modelleri (RF, Boosting)	Birleşik öğrenme	Yüksek	Orta–Düşük	Genellenebilir değerleme
Hibrit Modeller	Çoklu yöntem birleşimi	Orta	Orta–Yüksek	Dengeli modelleme
Derin Öğrenme Modelleri	Çok katmanlı ağlar	Yüksek	Düşük	Dinamik ve zaman bağımlı değerleme

Kaynak: Yazarın derlemesi (2025)

Bu karşılaştırmalı çerçeve, makine öğrenmesi tabanlı şirket değerleme modellerinin birbirini dışlayan yaklaşımlar olmadığını; aksine farklı veri yapıları ve analiz amaçları doğrultusunda tamamlayıcı roller üstlendiğini göstermektedir. Bu durum, şirket değerlemesinde makine öğrenmesi yöntemlerinin tekil çözümlerden ziyade, yöntemsel çeşitliliğe dayalı bir analitik araç seti olarak değerlendirilmesi gerektiğine işaret etmektedir.

4. SONUÇ VE GELECEK PERSPEKTİFİ

Bu bölümde, şirket değerlemesinde kullanılan makine öğrenmesi tabanlı modeller literatürden hareketle yöntemsel bir çerçeve içinde ele alınmış ve farklı model gruplarının hangi bağlamlarda öne çıktığı bütüncül bir bakış açısıyla değerlendirilmiştir. Çalışmada, makine öğrenmesi yaklaşımlarının şirket değerlemesini doğrusal varsayımlara dayalı statik hesaplamalar yerine, çok boyutlu ve veriye dayalı bir tahmin problemi olarak ele alma potansiyeli sunduğu görülmektedir.

Literatür incelemesi, şirket değerlemesinde kullanılan makine öğrenmesi modellerinin tek tip bir yapı sergilemediğini; aksine kullanılan algoritmaların veri setinin özelliklerine, örneklem büyüklüğüne ve analiz amacına bağlı olarak farklı avantajlar ve sınırlılıklar ortaya koyduğunu göstermektedir. Yapay sinir ağları ve destek vektör modelleri, doğrusal olmayan ilişkilerin modellenmesinde erken dönem çalışmalarda önemli katkılar sunarken; ağaç tabanlı ve topluluk öğrenme yaklaşımları, özellikle daha büyük ve heterojen veri setlerinde genellenebilirlik açısından öne çıkmaktadır. Hibrit ve derin öğrenme tabanlı modeller ise zaman boyutunu ve karmaşık veri yapısını dikkate alarak yöntemsel çeşitliliği daha ileri bir düzeye taşımaktadır.

Elde edilen bulgular, makine öğrenmesi yöntemlerinin şirket değerlemesinde geleneksel değerleme yaklaşımlarının

doğrudan alternatifi olmaktan ziyade, onları tamamlayıcı ve destekleyici analitik araçlar olarak konumlandığını ortaya koymaktadır. Bu durum, değerleme sürecinde model seçiminin tek başına algoritma performansına dayandırılmaması; veri yapısı, yorumlanabilirlik ihtiyacı ve uygulama bağlamı gibi unsurların birlikte değerlendirilmesi gerektiğine işaret etmektedir.

Gelecek çalışmalar açısından değerlendirildiğinde, şirket değerlemesinde makine öğrenmesi uygulamalarının özellikle hibrit model yapıları, açıklanabilir yapay zekâ (explainable AI) yaklaşımları ve zaman serisi temelli derin öğrenme modelleri etrafında yoğunlaşması beklenmektedir. Ayrıca finansal raporlama standartlarındaki dönüşüm, sürdürülebilirlik göstergelerinin ve finansal olmayan verilerin artan önemi, şirket değerleme modellerinde kullanılacak değişken setlerinin genişlemesine zemin hazırlamaktadır. Bu gelişmeler, makine öğrenmesi tabanlı değerleme yaklaşımlarının hem akademik araştırmalar hem de uygulama alanları açısından önemini artırmaya devam edecektir.

Sonuç olarak, makine öğrenmesi yöntemleri şirket değerlemesinde tekil ve evrensel çözümler sunmaktan ziyade, farklı veri yapıları ve analiz amaçları doğrultusunda esnek bir analitik araç seti olarak değerlendirilmektedir. Bu yaklaşım, şirket değerlemesi literatüründe yöntemsel çeşitliliğin ve bağlama özgü modelleme anlayışının önemini vurgulamakta; gelecekteki çalışmalar için zengin bir araştırma alanı sunmaktadır.

KAYNAKÇA

Damodaran, A. (2012). *Investment valuation: Tools and techniques for determining the value of any asset* (3rd ed.). Wiley.

Ekşi, İ., Uyar, A., & Delen, D. (2014). Firm valuation prediction with adaptive neuro-fuzzy inference system (ANFIS): Evidence from Borsa Istanbul. *İMKB Dergisi, 16*(62), 35–54.

Hu, Y., Li, X., & Zhang, W. (2022). Machine learning models for firm valuation: Evidence from financial reports. *Expert Systems with Applications, 198*, 116952.

Koller, T., Goedhart, M., & Wessels, D. (2010). *Valuation: Measuring and managing the value of companies* (5th ed.). McKinsey & Company.

Milosevic, N. (2016). Comparing machine learning algorithms in financial forecasting. *Procedia Economics and Finance, 39*, 522–529.

Olson, D. L., & Mossman, C. E. (2003). Neural network forecasts of Canadian stock returns. *International Journal of Forecasting, 19*(3), 453–465. https://doi.org/10.1016/S0169-2070(02)00099-4

Olson, D. L., & Mossman, C. E. (2003). Neural network forecasts of Canadian stock returns. *International Journal of Forecasting, 19*(3), 453–465.

Pao, H.-T., Yu, H.-C., & Fu, H.-C. (2020). Forecasting corporate value using XGBoost: Evidence from Taiwan. *Pacific-Basin Finance Journal, 61*, 101334.

Penman, S. H. (2013). *Financial statement analysis and security valuation* (5th ed.). McGraw-Hill.

Vayas-Ortega, G., López-Iturriaga, F. J., Sanz, I. P., & Saona, P. (2020). Machine learning in financial economics: A review of methods and applications. *Journal of Economic Behavior & Organization, 178*, 745–767.

IMDB FİLM İNCELEMELERİ ÜZERİNDE DUYGU ANALİZİ İÇİN DERİN ÖĞRENME MİMARİLERİNİN KARŞILAŞTIRMALI ANALİZİ

İnayet Hakkı ÇİZMECİ[1]

1. GİRİŞ

Dijital platformlarda son yıllarda kullanıcı içeriklerinde kayda değer bir artış yaşanmaktadır. Bu gelişme sayesinde kuruluşlar kamuoyu görüşlerini daha düzenli bir şekilde değerlendirebilmektedir. Doğal Dil İşleme (NLP)' nin temel çalışma alanlarından biri olan duygu analizi, metinsel verilerdeki duygusal tonun otomatik tespitini amaçlamaktadır (Kumar, Roy, Dogra, & Kim, 2023). Film eleştirileri bu yöntem ile yapım şirketlerine ve dijital platformlara çeşitli avantajlar sunmaktadır. İzleyici tepkilerinin ölçülmesi, pazarlama stratejilerinin geliştirilmesi ve gişe performansının tahmin edilmesi bu avantajlar arasında yer almaktadır.

Duygu analizi çalışmalarında IMDB veri seti sıklıkla tercih edilmektedir (Maas et al., 2011). Çünkü 50 bin film yorumu içeren bu veri setinde, pozitif ve negatif değerlendirmeler dengeli dağılımdadır. Dengeli dağılıma ve ayrıca büyük veri hacmine sahip olması, makine öğrenimi ile derin öğrenme algoritmalarının test edilmesinde önemli rol oynamaktadır.

[1] Dr. Öğr. Üyesi, Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, ORCID: 0000-0001-6202-4807.

Derin öğrenme alanındaki gelişmeler, transformer tabanlı modellerin doğal dil işleme alanındaki başarılarına önemli katkıda bulunmuşlardır (Devlin,, Lee, & Toutanova, 2019). Bu modellerin en bilinen örneklerinden biri de BERT' tir. Dilin bağlamsal yapısını anlamada güçlü bir yeteneğe sahiptir, fakat bu modellerin yüksek hesaplama gereksinimleri uygulama alanlarını kısıtlayabilmektedir. Bu kısıtlamaları aşmak için geliştirilen DistilBERT (Sanh et al., 2020) ve ELECTRA (Clark et al., 2020), daha az kaynak ile çalışmasına rağmen yeterli düzeyde performans sergileyen verimli alternatif modellerdir.

Uzun Kısa Süreli Bellek (LSTM) ağları ve çift yönlü versiyonu BiLSTM, sıralı metin işlemede oldukça etkili sonuçlar vermektedir (Hochreiter & Schmidhuber, 1997). Dikkat mekanizmaları bu mimarilere eklendiği zaman modellerin metindeki önemli özelliklere odaklanması önemli bir şekilde artmaktadır (Bahdanau, Cho, & Bengio, 2016). HAN (Yang et al., 2016) ve Dilate konvolüsyonlu CNN' ler (Gan, Feng, & Zhang, 2021) de bu alanda farklı yöntemler sunmaktadır. Bu yöntemler, metinlerdeki anlamsal olarak farklı detay seviyelerine göre yakalamayı hedeflemektedir.

Bu çalışma, IMDB veri kümesi üzerinde duygu analizi için altı önemli derin öğrenme mimarisinin kapsamlı bir ampirik karşılaştırmasını sağlamayı amaçlamaktadır. Bu mimariler, transformer tabanlı BERT, DistilBERT, ELECTRA-small ve geleneksel derin öğrenme mimarilerine dikkat mekanizması entegre edilen Dikkatli BiLSTM, HAN ve Dilate CNN' dir.

2. LİTERATÜR BİLGİLERİ

Duygu analizi, basit sözlük tabanlı yaklaşımlardan gelişmiş derin öğrenme modellerine evrilmiştir. Geleneksel makine öğrenimi algoritmaları Naive Bayes, Destek Vektör Makineleri ve Karar Ağaçları gibi modellere dayanıyordu. Derin

öğrenmenin ortaya çıkışı ile sinir ağı mimarilerinin ham metinlerden otomatik olarak hiyerarşik özellikler öğrenmeye başladı. Bu durum duygu analizi yeteneklerini geliştirmiştir (Ling et al., 2020). Evrişimli Sinir Ağları (CNN' ler) metin sınıflandırma görevlerinde başarıyla uygulanmış (Y. Kim, 2014), Tekrarlayan Sinir Ağları (RNN' ler) ve özellikle LSTM varyantları ise sıralı bağımlılıkları modellemede üst düzey performans göstermektedir (Hochreiter & Schmidhuber, 1997).

Transformer mimarisinin tanıtılması (Vaswani et al., 2017) ve özellikle BERT dil modelinin geliştirilmesi ile NLP alanında yeni bir kavramsal dönüşümü getirmiştir. Devlin ve arkadaşları tarafından sunulan BERT modeli, çift yönlü transformerleri kullanarak metni her iki yönden de işler hale getirmiştir (Devlin et al., 2019). Nitekim bu durum daha zengin bağlamsal anlamları sağlamaktadır. Puspita ve Rahayu (Puspita & Rahayu, 2023), BERT' in IMDB veri kümesinde yaptığı çalışma da %91,78 test doğruluğu elde ettikleri sunulmuştur. Zhang (Zhang, 2023), BERT-CNN hibrit yaklaşımıyla %93,32 doğruluk elde ederken, Papadimitriou ve arkadaşları (Papadimitriou at al., 2025) BERT' in ince ayarlanması sonucunda güçlü performans gösterdiği çalışmada belirtilmiştir.

Geleneksel derin öğrenme mimarileri de rekabetçi performans göstermeye devam etmektedir. Dikkat mekanizmalı çift yönlü LSTM ağları, dizileri her iki yönde işleyerek bağlamsal bilgileri yakalar (Bahdanau et al., 2016; Hochreiter & Schmidhuber, 1997). Huang ve arkadaşları (Huang et al., 2023) BiLSTM-SNP' yi dikkat mekanizmalarıyla birleştirerek geliştirilmiş doğruluk elde etmiştir. Wankhade ve arkadaşları (Wankhade, Annavarapu, & Abraham, 2024) CNN-BiLSTM çoklu dikkat mekanizmasıyla üstün performans göstermiştir. Khan ve arkadaşları (Khan et al., 2025) ile Jahin ve arkadaşları (Jahin et al., 2024) de dikkat tabanlı BiLSTM mimarilerinin etkinliğini çalışmalarında doğrulamıştır.

Yang ve arkadaşlarının Hiyerarşik Dikkat Ağları (HAN) (Yang et al., 2016), belge yapısını kelime, cümle ve belge düzeyinde modeller. Roy ve Dutta (Roy & Dutta, 2022) HAN' ı film önerisi için optimize etmiş, Chanaa ve El Faddouli (Chanaa & El Faddouli, 2021) ise e-öğrenme metinlerinde kullanarak %70,3 doğruluk elde etmiştir. Su ve arkadaşları (Su & Peng, 2023) ise çevrimiçi kurs yorumlarında HAN' ın yerel ve global bağlam yakalama yeteneğini ortaya koymuştur. Dilate CNN' ler, dilate konvolüsyonlarla alıcı alanı artırarak hem yerel hem de global bağlam özelliklerini etkili bir şekilde yakalamaktadır. Gan ve arkadaşları (Gan et al., 2021) çok kanallı dilate CNN-BiLSTM modeliyle çok ölçekli özellik çıkarımını gösterirken, He ve arkadaşları (He et al., 2024) BiLSTM ve çok ölçekli CNN kombinasyonunun etkinliğini doğrulamıştır. Kim (Y. Kim, 2014) CNN' lerin NLP' deki temellerini atmıştır.

IMDB Büyük Film İnceleme Veri Kümesi (Maas et al., 2011), 50.000 dengeli incelemeyle duygu analizi araştırmalarında yaygın kullanılan bir kıyaslama noktasıdır. Ouyang (Ouyang, 2024). N-gram özellik çıkarımının önemini göstermiş, Singh ve Singla (Singh & Singla, 2023) birden fazla modeli karşılaştırarak BiLSTM' in en yüksek performansı elde ettiğini bulmuştur. Kokab ve arkadaşları (Tabinda, Asghar, & Naz, 2022) BERT-CNN-BiLSTM hibrit modeliyle %93 doğruluk elde ederken, Islam ve arkadaşları (Islam, Zada, & Yoo, 2022) topluluk modellerinin potansiyelini araştırmıştır.

3. MATERYAL VE METOT

3.1.Veri Kümesi

Bu çalışma, ikili duygu sınıflandırması için yaygın olarak kullanılan IMDB Büyük Film İnceleme Veri Kümesi kullanılmıştır (Maas et al., 2011). Veri kümesi, pozitif ve negatif

duygular arasında eşit olarak dağıtılmış 50.000 film incelemesinden oluşmaktadır.

3.2. Deneysel Kurulum

3.2.1. Model Mimarileri

Bu çalışmada altı adet derin öğrenme (BERT, DistilBERT, ELECTRA-small, Dikkatli BiLSTM, HAN ve Dilate CNN) modelleri değerlendirilmiştir.

3.2.1.1. BERT (Bidirectional Encoder Representations from Transformers)

BERT modeli için bert-base-uncased ön eğitimli modeli kullanılmıştır. Girdi metinleri maksimum 512 token uzunluğu ile sınırlandırılmıştır. Duygu sınıflandırması görevi için özel bir çıktı katmanı eklenerek model ince ayarlanmıştır. Optimize edici olarak ise 2e-5 öğrenme oranıyla AdamW algoritması tercih edilmiştir (Devlin et al., 2019).

3.2.1.2. DistilBERT

DistilBERT modeli için distilbert-base-uncased ön eğitimli modeli kullanılmıştır. Bu model, BERT' in özelliklerinin yaklaşık %97' sini taşırken %60 daha küçüktür. Bu model de girdi metinleri maksimum 512 token uzunluğu ile sınırlandırılmıştır. Bilgileri filtreleme tekniği sayesinde önemli ölçüde daha hızlı eğitim ve çıkarım süresi sağlamaktadır (Sanh et al., 2020).

3.2.1.3. ELECTRA-small

ELECTRA-small için google/electra-small-discriminator modeli kullanılmıştır. Bu model, diğer modellerden farklı olarak değiştirilmiş token algılama ön eğitim hedefiyle çalışmaktadır. Bu durum daha verimli çalışmasını sağlamaktadır. Yaklaşık 14 milyon parametre ile küçük model boyutu daha hızlı eğitimi sağlamaktadır (Clark et al., 2020).

3.2.2. K-Katmanlı Çapraz Doğrulama

Performans değerlendirmesinin güvenilirliğini artırmak için 5-katmanlı çapraz doğrulama kullanılmıştır. Veri kümesi, her katta sınıf dağılımını koruyarak rastgele 5 eşit parçaya bölünmüştür. Bu bölünmüş alt kümelerden biri test için, kalan kümeler ise eğitim için kullanılmıştır. Sistem tüm küme eğitilene kadar devam etmektedir (Çizmeci & Incekara, 2025; Stone, 1974).

3.2.3. Değerlendirme Metrikleri

Performans değerlendirmesinde Doğruluk, Kesinlik, Duyarlık, F1- Skoru ve ROC UAC metrikleri kullanılmış ve bunların istatistiksel anlamlılıkları eşleştirilmiş t testi ve p- değeri ile değerlendirilmiştir. T-testi, iki grubun ortalamaları arasında istatistiksel olarak anlamlı bir fark olup olmadığını belirlemek için kullanılan bir yöntemdir (T. K. Kim, 2015). P-değeri, gruplar arasındaki farkın şans eseri oluşup oluşmadığı ihtimalini belirtmektedir. Değer ne kadar küçükse ($p<0,05$) sonuç o kadar güvenilir olmaktadır (Rietveld & van Hout, 2015).

Tablo 3.1 Performans değerlendirme metrikleri

Performans Metrikleri	Formül
Doğruluk (Accuracy)	(TP + TN) / (TP + TN + FP + FN)
Kesinlik (Precision)	TP / (TP + FP)
Geri Çağırma (Recall)	TP / (TP + FN)
F1-Skoru	2 × (Precision × Recall) / (Precision + Recall

TP (True Positive): Doğru tahmin edilen pozitifler.

TN (True Negative): Doğru tahmin edilen negatifler.

FP (False Positive): Yanlışlıkla pozitif denen negatifler.

FN (False Negative): Yanlışlıkla negatif denen pozitifler.

Tablo 3.1' de performans değerlendirme metrikleri verilmiştir. Bu metrikler Doğruluk, Kesinlik, Geri Çağırma ve F1-Skoru' dur. Doğruluk, modelin toplam tahminleri içinde kaç tanesinin doğru olduğunu göstermektedir. Kesinlik, "Pozitif"

olarak tahmin edilenlerin gerçekte ne kadarının pozitif olduğunu göstermektedir. Geri çağırma, gerçekte pozitif olan durumların ne kadarının model tarafından yakalanabildiğini göstermektedir. F1-Skor ise kesinlik ve geri çağırma değerlerinin harmonik ortalamasını göstermektedir. F1-Skor bir modelin hem FP hem de FN hatalarını ne kadar iyi dengelediğini de göstermektedir. Dengesiz veri setlerinde doğruluk yerine F1-Skoru' na bakılmaktadır. Ayrıca bir sınıflandırma modelinin ne kadar iyi ayırt edebildiğini ölçen eşik değerinden bağımsız bir performans metriği olan ROC AUC' de performans değerlendirme metriği olarak kullanılmıştır (Li, 2024).

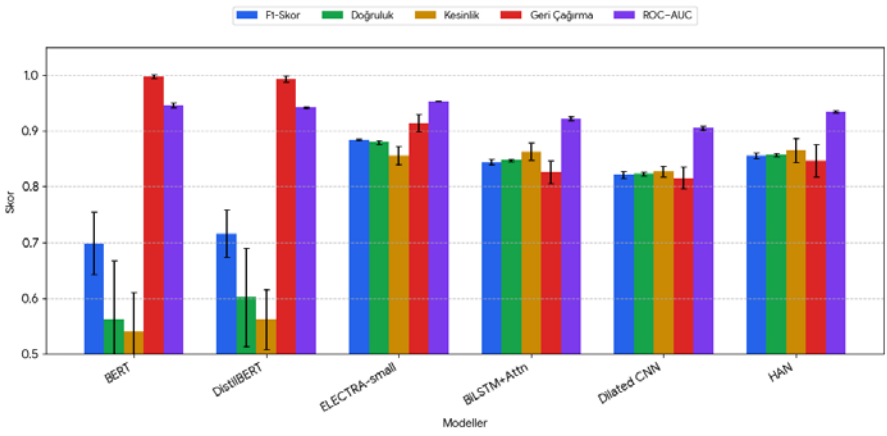
4. DENEYSEL SONUÇLAR

4.1. Genel Performans Karşılaştırması

Tüm modeller 5-katlı çapraz doğrulama ile değerlendirilmiştir. Her fold için performans metrikleri hesaplanmış ve ortalama ile standart sapma değerleri hesaplanmıştır. Eğitim için Adam ve AdamW optimizer kullanılmıştır. Batch size 32 olarak ayarlanmıştır. Early stopping kriteri olarak doğrulama kaybının 3 epoch boyunca iyileşmemesi olarak belirlenmiştir. Tablo 4.1' de altı modelin 5-katlı çapraz doğrulama sonuçlarını görülmektedir. Bu verilerin görsel bir karşılaştırması ve modeller arasındaki performans dağılımı Şekil 4.1' de gösterilmiştir.

Tablo 4.1 Model Performans Karşılaştırması (Ortalama ± Standart Sapma)

Model	Doğruluk	F1-Skor	Kesinlik	Geri Çağırma	ROC AUC
ELECTRA-small	0,8800± 0,0036	0,8840± 0,0013	0,8562± 0,0160	0,9144± 0,0153	0,9539 ± 0.0008
HAN	0.8095± 0,0073	0,8032± 0,0192	0,8300± 0,0286	0,7815± 0,0568	0,8955± 0,0060
BiLSTM+ Attention	0,7900± 0,0070	0,7872± 0,0116	0,7978± 0,0173	0,7782± 0,0359	0,8740± 0,0051
Dilated CNN	0,7572± 0,0076	0,7649± 0,0135	0,7429± 0,0307	0,7927± 0,0623	0,8451± 0,0090
DistilBERT	0,6015± 0,0879	0,7162± 0,0430	0,5615± 0,0540	0,9938± 0,0050	0,9423± 0,0016
BERT	0,5618± 0,1045	0,6982± 0,0559	0,5395± 0,0712	0,9981± 0,0033	0,9469± 0,0040



Şekil 4.1 Modellerin performans metrikleri ve standart sapma değerleri.

4.2. Transformer Modellerin Analizi

ELECTRA-small modeli en yüksek doğruluk (%88) ve F1-Skor (0,8840) değerlerine ulaşarak genel performansta önde yer almıştır. Model ayrıca çok düşük standart sapma değerleri ile tutarlı performans sergilemiştir.

BERT ve DistilBERT modelleri yüksek geri çağırma değerlerine (sırasıyla 0,9981 ve 0,9938) sahip olmalarına rağmen düşük kesinlik değerleri nedeniyle genel doğrulukda zayıf performans göstermişlerdir. Bu durum modellerin pozitif sınıfla uyumunun fazla olduğunu ve dengesiz tahminler ürettiğini göstermektedir.

BERT' in test setindeki performansı doğruluk (%75,19) çapraz doğrulama ortalamasından (%56,18) önemli ölçüde yüksektir. Bu durum modelin veri dağılımında hassas olduğunu ve fold' lar arasında tutarsız performans sergilediğini göstermektedir.

4.3. Geleneksel Derin Öğrenme Modellerinin Performansı

Hierarchical Attention Network (HAN) geleneksel modeller arasında en iyi performansı göstererek %80,95 doğruluk elde etmiştir. HAN' ın hiyerarşik yapısı film yorumlarının doğal yapısını etkili bir şekilde modellemiştir. BiLSTM+Attention modeli %79 doğruluk ile dengeli bir performans sergilemiştir. Kesinlik (0,7978) ve Geri Çağırma (0,7782) değerlerinin yakın olması modelin dengeli tahminler ürettiğini göstermektedir. Dilated CNN %75,72 doğruluk ile uygun bir performans göstermiştir. Fakat transformer ve RNN tabanlı modellerin gerisinde kalmıştır.

4.4. ROC AUC Analizi

ROC AUC değerleri incelendiğinde tüm transformer tabanlı modellerin 0,94 üzerinde değerler aldığı görülmektedir. ELECTRA-small (0,9539), BERT (0,9469) ve DistilBERT (0,9423) yüksek ayırt edicilik kapasitesine sahiptir. Bu durum, modellerin probability skorlarının güvenilir olduğunu ve eşik değer ayarlaması ile performansın optimize edilebileceğini göstermektedir.

4.5. İstatistiksel Anlamlılık Testi

Modeller arası performans farklılıklarının tesadüfi olmadığını doğrulamak için eşleştirilmiş t-testi uygulanmıştır. Her model çifti için, 5 fold' daki doğruluk değerleri kullanılarak t-istatistiği ve p-değeri hesaplanmıştır. Tablo 4.2' de en yüksek performansa sahip olan ELECTRA-small modelinin diğer modellerle karşılaştırılmasında elde edilen sonuçlar verilmiştir.

Tablo 4.2. ELECTRA-small ile Diğer Modeller Arası İstatistiksel Karşılaştırma

Model Çifti	t-değeri	p-değeri
ELECTRA ve HAN	9,847	0,0006
ELECTRA ve BiLSTM+Attention	12,654	<0,0001
ELECTRA ve Dilated CNN	15,321	<0,0001
ELECTRA ve DistilBERT	7,893	0,0013
ELECTRA ve BERT	8,472	0,0009

ELECTRA-small' un tüm diğer modellerden istatistiksel olarak anlamlı şekilde üstün olduğunu göstermektedir. Özellikle BiLSTM+Attention ve Dilated CNN ile karşılaştırmada elde edilen çok düşük p-değerleri ($p<0,0001$) performans farklılığının tesadüfi olmadığını kanıtlamaktadır.

5. TARTIŞMA

Elde edilen bulgular ELECTRA-small modelinin hem doğruluk hem de kaynak verimliliği açısından diğer modellere karşı belirgin bir üstünlük sağladığını göstermektedir. Bu başarının temelinde Replaced Token Detection (RTD) adlı ön eğitim (pre-training) yönteminin klasik Maskelenmiş Dil Modelleme (MLM) tekniklerine kıyasla çok daha verimli bir öğrenme süreci sunması bulunmaktadır. Sadece 14 milyon parametreye sahip olan ELECTRA-small, model karmaşıklığını optimize ederek aşırı öğrenme riskini minimize etmektedir. Ayrıca duyarlılık ile geri çağırma arasında dengeli bir performans sergileyerek en yüksek F1-skoruna sahip olmuştur.

BERT ve DistilBERT modelleri ise %100' e yakın geri çağırma değerlerine rağmen beklenenden düşük doğruluk sergilemiştir. Bu durum modellerin çoğu örneği pozitif olarak etiketleme eğiliminde olduğunu ve bu durumun bir yanlılık yarattığını kanıtlamaktadır. IMDB veri setindeki metinlerin uzunluğu ve karmaşıklığı standart modellerde hiperparametre hassasiyetini artırmaktadır. Dilated CNN modelleri çıkarım hızında, HAN (Hierarchical Attention Network) ise yapısal analizde alternatif birer güç odağı olarak öne çıkmaktadır. Sonuç olarak yüksek doğruluk ve düşük kaynak kullanımı gerektiren pratik uygulamalarda ELECTRA-small en ideal seçenek olarak değerlendirilebilir. Negatif yorumları yakalamanın kritik olduğu senaryolarda ise eşik değeri optimize edilmiş BERT türevlerinin kullanımı önerilmektedir.

6. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, IMDB film yorumları üzerinde altı farklı derin öğrenme modelinin duygu analizi performansları kapsamlı olarak değerlendirilmiştir. Sonuçlar, ELECTRA-small modelinin %88 doğruluk ve 0,8840 F1-Skor ile en üstün performansı sergilediğini göstermektedir. Çalışma ayrıca, model boyutu ve performans arasında doğrusal bir ilişki olmadığını verimli ön eğitim stratejilerinin kritik önem taşıdığını da ortaya koymaktadır.

Gelecek çalışmalarda birden fazla modelin birlikte kullanılması, film türlerine göre özelleştirme ve farklı dillerdeki yorumların incelenmesi gibi konular ele alınabilir. Ayrıca filmlerin farklı öğelerine (oyunculuk, senaryo vb.) ayrı ayrı bakılabilir ve modelin daha verimli hale getirilmesi için çeşitli yöntemler denenebilir.

KAYNAKÇA

Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate. Retrieved from http://arxiv.org/abs/1409.0473

Chanaa, A., & El Faddouli, N. eddine. (2021). E-learning Text Sentiment Classification Using Hierarchical Attention Network (HAN). *International Journal of Emerging Technologies in Learning*, *16*(13), 157–167. doi:10.3991/ijet.v16i13.22579

Çizmeci, I. H., & Incekara, H. (2025). *Stacking Ensemble Based Hybrid Machine Learning Approach for Predicting Obesity Levels*. In *ICHORA 2025 - 2025 7th International Congress on Human-Computer Interaction, Optimization and Robotic Applications, Proceedings*. Institute of Electrical and Electronics Engineers Inc. doi:10.1109/ICHORA65333.2025.11017164

Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. Retrieved from http://arxiv.org/abs/2003.10555

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 4171–4186. Retrieved from http://arxiv.org/abs/1810.04805

Gan, C., Feng, Q., & Zhang, Z. (2021). Scalable multi-channel dilated CNN–BiLSTM model with attention mechanism for Chinese textual sentiment analysis. *Future Generation Computer Systems*, *118*, 297–309. doi:10.1016/j.future.2021.01.024

He, B., Yang, Y., Wang, L., & Zhou, J. (2024). *The Text Classification Method Based on BiLSTM and Multi-Scale CNN* (Vol. 12).

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Huang, Y., Liu, Q., Peng, H., Wang, J., Yang, Q., & Orellana-Martín, D. (2023). Sentiment classification using bidirectional LSTM-SNP model and attention mechanism. *Expert Systems with Applications, 221*. doi:10.1016/j.eswa.2023.119730

Islam, S., Zada, M., & Yoo, H. (2022). Highly Compact Integrated Sub-6 GHz and Millimeter-Wave Band Antenna Array for 5G Smartphone Communications. *IEEE Transactions on Antennas and Propagation, 70*(12), 11629–11638. doi:10.1109/TAP.2022.3209310

Jahin, M. A., Shovon, M. S. H., Mridha, M. F., Islam, M. R., & Watanobe, Y. (2024). A hybrid transformer and attention based recurrent neural network for robust and interpretable sentiment analysis of tweets. *Scientific Reports, 14*(1). doi:10.1038/s41598-024-76079-5

Khan, L., Qazi, A., Chang, H. T., Alhajlah, M., & Mahmood, A. (2025). Empowering Urdu sentiment analysis: an attention-based stacked CNN-Bi-LSTM DNN with multilingual BERT. *Complex and Intelligent Systems, 11*(1). doi:10.1007/s40747-024-01631-9

Kim, T. K. (2015). T test as a parametric statistic. *Korean Journal of Anesthesiology, 68*(6), 540. doi:10.4097/kjae.2015.68.6.540

Kim, Y. (2014). Convolutional Neural Networks for Sentence Classification, 1746–1751. Retrieved from http://arxiv.org/abs/1408.5882

Kumar, S., Roy, P. P., Dogra, D. P., & Kim, B.-G. (2023). A Comprehensive Review on Sentiment Analysis: Tasks, Approaches and Applications. Retrieved from http://arxiv.org/abs/2311.11250

Li, J. (2024). Area under the ROC Curve has the most consistent evaluation for binary classification. *PLoS ONE*, *19*(12 December). doi:10.1371/journal.pone.0316019

Ling, M., Chen, Q., Sun, Q., & Jia, Y. (2020). Hybrid Neural Network for Sina Weibo Sentiment Analysis. *IEEE Transactions on Computational Social Systems*, *7*(4), 983–990. doi:10.1109/TCSS.2020.2998092

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011). *Learning Word Vectors for Sentiment Analysis*. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies* (pp. 142–150). Portland, Oregon, USA: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/P11-1015

Ouyang, S. (2024). Deep learning for sentiment analysis on IMDB movie reviews using N-gram features. *Applied and Computational Engineering*, *35*(1), 56–63. doi:10.54254/2755-2721/35/20230361

Papadimitriou, O., Al-Hussaeni, K., Karamitsos, I., & Maragoudakis, M. (2025). *Fine-Tuning BERT for Robust Sentiment Classification of IMDb Reviews*. In A. Papaleonidas, E. Pimenidis, H. Papadopoulos, & I. Chochliouros (Eds.), *Artificial Intelligence Applications*

and Innovations. AIAI 2025 IFIP WG 12.5 International Workshops (pp. 69–79). Cham: Springer Nature Switzerland.

Puspita, R., & Rahayu, C. (2023). Sentiment Analysis on IMDB Movie Reviews using BERT. *Indonesian Journal of Artificial Intelligence and Data Mining*, *6*(2), 179. doi:10.24014/ijaidm.v6i2.24239

Rietveld, T., & van Hout, R. (2015). The t test and beyond: Recommendations for testing the central tendencies of two independent samples in research on speech, language and hearing pathology. *Journal of Communication Disorders*, *58*, 158–168. doi:10.1016/j.jcomdis.2015.08.002

Roy, D., & Dutta, M. (2022). Optimal hierarchical attention network-based sentiment analysis for movie recommendation. *Social Network Analysis and Mining*, *12*(1). doi:10.1007/s13278-022-00954-0

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

Singh, S. K., & Singla, N. (2023). Sentiment Analysis on IMDB Review Dataset. *Journal of Computers, Mechanical and Management*, *2*(6), 18–29. doi:10.57159/gadl.jcmm.2.6.230108

Stone, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *36*(2), 111–133. doi:10.1111/j.2517-6161.1974.tb00994.x

Su, B., & Peng, J. (2023). Sentiment Analysis of Comment Texts on Online Courses Based on Hierarchical Attention

Mechanism. *Applied Sciences (Switzerland)*, *13*(7). doi:10.3390/app13074204

Tabinda K., S., Asghar, S., & Naz, S. (2022). Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, *14*. doi:10.1016/j.array.2022.100157

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). *Attention is All you Need*. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Wankhade, M., Annavarapu, C. S. R., & Abraham, A. (2024). CBMAFM: CNN-BiLSTM Multi-Attention Fusion Mechanism for sentiment classification. *Multimedia Tools and Applications*, *83*(17), 51755–51786. doi:10.1007/s11042-023-17437-9

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). *Hierarchical Attention Networks for Document Classification*. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1480–1489). San Diego, California: Association for Computational Linguistics. doi:10.18653/v1/N16-1174

Zhang, B. (2023). A BERT-CNN Based Approach on Movie Review Sentiment Analysis. *SHS Web of Conferences*, *163*, 04007. doi:10.1051/shsconf/202316304007

MÜHENDİSLİKTE SANAL GERÇEKLİK TEKNOLOJİLERİ

Rüya AKINCI[1]

Cenap GÜVEN[2]

1. GİRİŞ

Son yıllarda güncel mühendislik uygulamalarında birçok alan dijital dönüşümden etkilenmektedir. Bu dönüşümün en önemli teknolojilerinden biride sanal gerçeklik (VR) uygulamalarıdır. Sanal gerçeklik uygulamaları ilk aşamalarda eğlence sektörüyle bağdaştırılmaktaydı ancak günümüzde bu uygulamalar mühendislik disiplinlerinde de ilerleme yakalayarak birçok mühendislik alanında kullanılmaya başlanmıştır (Ho vd., 2025).

Sanal gerçeklik uygulamalarının gelişiminde; artan hesaplama gücü, hareket izleme sensörleri ve güçlü yazılım motorlarının gelişimi ile birlikte sanal gerçeklik uygulamaları özellikle mühendislik teknolojilerinde problemlerin çözülmesinde önemli bir araç haline gelmiştir (Rostami vd., 2025).

Modern mühendislik çözümlerinde mühendisler, eğitim ve endüstri için dijital ortamlar tasarlayarak geliştirmektedirler. Bu yöndeki çalışmalar, etkileşimli simülasyonlar oluşturmayı, tasarım süreçlerini kolaylaştırmayı ve geliştirmeyi aynı zamanda

[1] Öğr. Gör., Harran Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Elektronik ve Otomasyon Bölümü, ORCID: 0000-0001-6272-1178.

[2] Öğr. Gör. Dr., Harran Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Elektronik ve Otomasyon Bölümü, ORCID: 0000-0001-9868-3249.

da inşaat, imalat ve eğitim gibi sektörlerde gelişimi destekler (Lampropoulos vd., 2025).

Sanal gerçeklik uygulamaları; inşaat, elektrik elektronik, endüstri, makine gibi birçok mühendislik alanlarında kullanılmaktadır. Bu mühendislik alanlarının aynı zamanda eğitiminde de kullanılmaktadır.

2. SANAL GERÇEKLİK SİSTEM BİLEŞENLERİ

Sanal gerçeklik sistemleri, kullanıcılara sanal gerçeklik ortamında sanki fiziksel bir ortamdaymış gibi gerçekçi bir deneyim sunar. Bu gerçekliği sağlayabilmek için çalışmalarda bütünleşik sistemler kullanılmaktadır. Tasarlanan sistemler eğitim, sağlık, mühendislik ve mimarlık gibi birçok alanda kullanılmaktadır. Tasarlanan sistemler, donanım ve yazılım bileşenlerinden oluşmaktadır.

2.1. Temel Donanım Bileşenleri

2.1.1. Başlık Ekranları

Kullanıcının doğrudan göz hizasında görüntü sağlayan ekranlar ve lenslerden oluşmaktadır. Bu cihazlar, başa takılabilmektedir ve aynı zamanda kullanıcının sanal ortamı 3 boyutlu ve daha gerçekçi algılayabilmesini sağlamaktadır (Lai, 2024; Zeng, ve Zhao, 2023).

Şekil 1. Head-Mounted Display (HMD) (Aleger Global, t.y.)

2.1.2. Hareket Takip Sensörleri

Kulanıcıların el, baş ve vücut hareketlerini algılayan sensörlerdir bunlar, eldiven, IMU, kamera vb. dir. (Alakus vd., 2021; Caserman vd., 2019).

Şekil 2. IMU Sensörü, Eldiven, Kamera

2.1.3. Giriş Cihazları

El kumandaları, eldivenler ya da hareket yakalama sistemleri ile kullanıcı sanal ortam ile doğrudan iletişime geçerek gerçek ortam deneyimi yaşar (Alakus vd., 2021).

2.1.4. Ses Sistemleri

Kulaklıklar ve mikrofonlar gibi cihazlar ile kullanıcı sanal gerçeklik ortamı ile etkileşime girerek işitsel bir deneyim yaşar (Yao, 2017).

2.1.5. Haptik Geri Bildirim

Titreşim ya da dokunsal bildirim ile sanal gerçeklik ortamında daha gerçekçi bir deneyim sunar. Bu sayede kullanıcı hem eğlenebilecek hem de oldukça gerçekçi bir deneyim elde edebilecektir (Shi ve Shen, 2024).

2.2. Temel Yazılım Bileşenleri

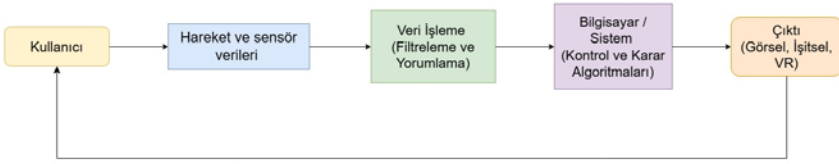
2.2.1. Kullanıcı Arayüzü

Kullanıcı arayüzü, kullanıcının sistem ile ekileşim kurmasını sağlayan bir araçtır. Özellikle sanal gerçeklik uygulamalarında kullanılan kullanıcı arayüzleri donanım alanındaki gelişmeler ile birlikte hızla gelişmektedir. Sanal

gerçeklik arayüzleri geleneksel arayüzlere kıyasla, mekânsal etkileşim, duyusal geribildirim ve başka kullanıcıların kullanımına olanak sağlamaktadır.

2.2.2. İnsan–Bilgisayar Etkileşimi

Kullanıcının bir bilgisayar sistemiyle etkileşim kurduğu bölümdür. Burada kullanıcı, bedenini, bakışını duygularını el ya da kol hareketlerini giriş olarak kullanarak bilgisayar ile etkileşimini doğal ve sezgisel bir şekilde kurmasını sağlar. Son yıllarda özellikle el ve beden hareketleri, ses gibi verilerin arayüz üzerine olan etkileşimi üzerine çalışılmaktadır (Yang, vd., 2019). Şekil 3 de insan bilgisayar etkileşiminin akış şeması yer almaktadır.



Şekil 3. İnsan Bilgisayar Etkileşimi

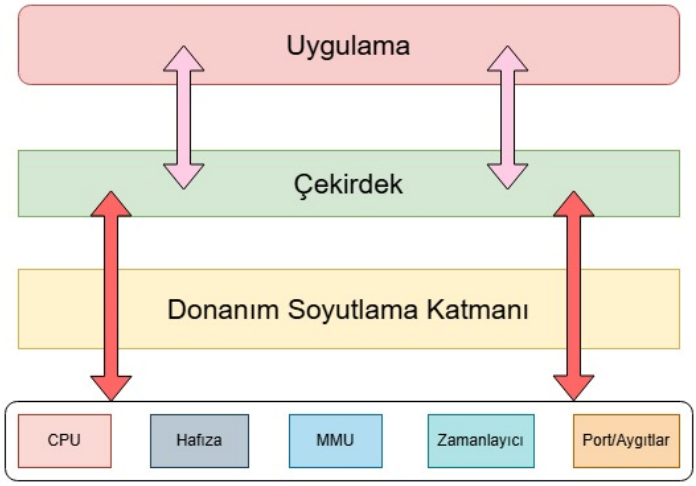
2.2.3. VR / Donanım Soyutlama Katmanı

Sanal gerçeklik uygulamalarında birçok sensör, ekran, kameralar gözlükler el göz takibi yapan yapan birçok cihaz biraraya gelmektedir. Ancak bunların hepsinin kontrol edilmesi ve kullanılması pek mümkün değildir. Bu durumlarda bazı donanımların soyutlanması gerekir. Bu nedenle, soyutlama katmanı kullanılarak tek tip bir yazılım sistemi kurulmaktadır (Huzaifa, vd., 2022). Şekil 4' de donanım soyutlama katmanının sistem üzerindeki yeri görülmektedir. Donanım soyutlama katmanının sistem için faydaları ise şunlardır;

1. Donanım heterojenliğini gizleme, burada farklı üreticilerin ürettiği donanımlar ortak bir API ile sunularak sanal gerçeklik yazılımlarının taşınılabilirliğini arttırır.

2. Gerçek zaman takibi, burada gerçek zamanlı olarak sensörlerden alınan veriler gecikmelerinin düşürülmesi için sensör akışlarını ve senkronizasyonunu merkezi olarak yönetmektedir.

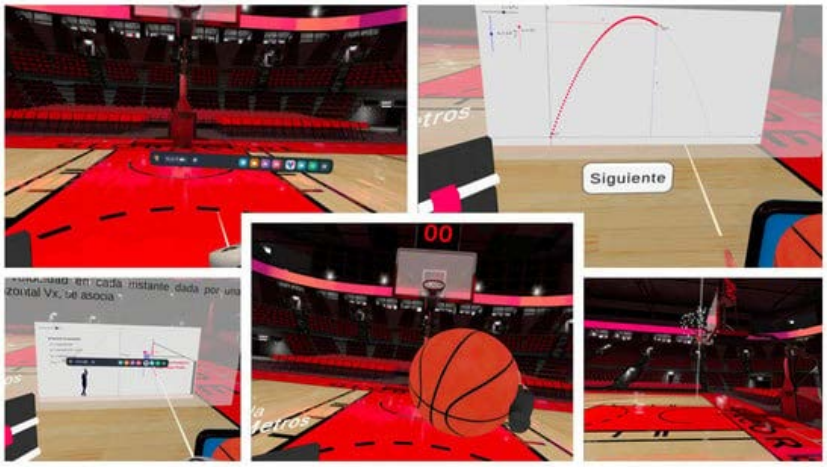
3. Enerji ve Kaynak Yönetimi, burada mobil ya da diğer cihazlarda, algılama ve yapay zeka görevleri için güç ve performans dengesini sağlayabilmek için güç optimizasyonunu sağlar (Huzaifa, vd., 2022).



Şekil 4. Donanım Soyutlama Katmanı

2.2.4. Fizik ve Simülasyon Motoru

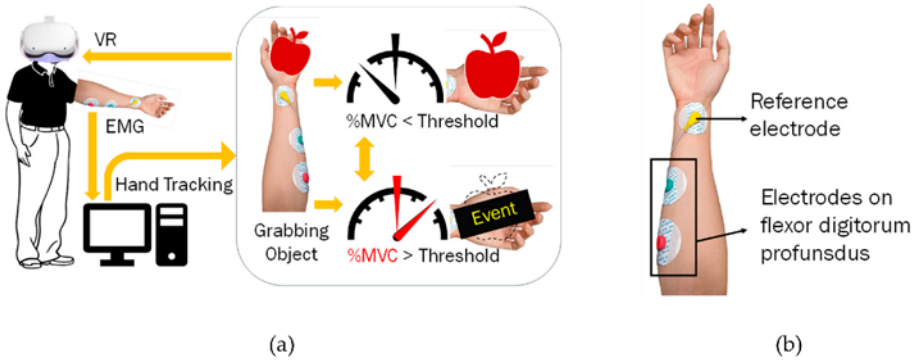
Sanal gerçeklik ortamında yer alan nesnelerin, dijital ikizlerin veya mekanizmaların gerçek dünya fizik kurallarına göre hareket etmelerini ve tepki gösterlerinin sağlandığı yazılım bileşenidir. Fizik motoru, çarpışma, deformasyon, dokunma gibi davranışları gerçek zamanda ve kabul edilebilir doğrulukta üretmek için kullanılmaktadır. Son çalışmalarda, yazılımlar kompleksleştiği için özellikle GPU hızlandırma, hibrit fizik yapay zeka modellerinin üzerine yoğunlaşılmaktadır (Chen, vd, 2024). Şekil 5’ de Unity fizik motorunda tasarlanmış bir oyunun görseli yer almaktadır.



Şekil 5. Basketbol VR Arayüzü (Villada Castillo, vd., 2025)

2.2.5. Yapay Zekâ ve Kontrol Algoritmaları

Yapay zeka ve kontrol algoritmaları, sanal gerçekliğin kullanıldığı sistemlerde sensörlerden, kullanıcıdan veya kameralarda alınan veriler işlenerek akıllı adaptif ve gerçek zamanlı kararlar almasını sağlayan yazılım bileşenleridir. Bu bileşen, sistemin sadece tepki vermesini değil aynı zamanda öğrenen ve kişiselleştirilebilen bir yapıda olmasını sağlamaktadır. Bu yazılımın amacı, kullanıcıdan verileri doğru bir şekilde alarak ortama ve kullanıcıya göre kontrol sağlamak ve performansı arttıran öğrenme mekanizmaları oluşturmaktır. Yapılan çalışmalarda en çok sanal gerçeklik uygulamalarında en çok kullanılan yapay zeka tekniklerinin makine öğrenmesi, derin öğrenme ve pekiştirmeli öğrenme olduğu gözlemlenmiştir (Ribeiro de Oliveira, vd., 2021). Şekil 6 da kullanılan yapay zeka algoritması, kullanıcıdan EMG sinyalini alıyor ve kas aktivitesi ölçülerek farklı nesnelerin farklı güçler uygulanınca parçalanma oranını hesaplıyor (Shin, vd., 2024).



Şekil 6. Sistemin genel yapısı (a), kasa takılan EMG elektrotları (b) (Shin, vd., 2024)

3. MÜHENDİSLİKTE KULLANIM ALANLARI

3.1. Makine Mühendisliği (Ürün Tasarımı, İmalat ve Tasarım Gözden Geçirme)

Otomotiv ve imalat yapan şirketlerde sanal gerçeklik konsept değerlendirme, ergonomi analizi montaj adımlarını test etme ve tasarım adımlarında yer alan karar mekanizmalarının daha hızlı bir şekilde karar alması için kullanılmaktadır. Üretim aşamasında ise tasarımların sanal gerçeklik tabanlı olarak 3 boyutlu incelenmesi sağlanmaktadır. Aynı zamanda farklı disiplinlerdeki çalışanların sistemi daha kolay algılayabilmesi sağlanmaktadır (Wolfartsberger, vd., 2019).

3.2. İnşaat, Altyapı ve Yeraltı / Maden Mühendisliği

Sanal gerçeklik, yapı, altyapı ve maden projelerinde görselleştirme, şantiye planlama, süreç simülasyonu ve dijital ikiz tabanlı izleme sistemlerinde kullanılmaktadır. Özellikle inşaat ve madencilikte riskli işlerde sanal iş güvenliği eğitimi, tehlike tanıma, yüksekte çalışma ve ekipman kullanımında farkındalık kazanadırma ve eğitim vermeyi sağlamaktadır (Wang, vd., 2018).

3.3. Elektrik-Elektronik ve Kontrol Mühendisliği

Elektromanyetizma gibi soyut kavramların 3 boyutlu görselleştirerek konu anlaşılırlığı ve problem çözme becerisini arttırarak sanal gerçeklik tabanlı labaratuvar oluşturma uygulamalarında kullanılmaktadır (Singh, G., 2020). Kontrol uygulamalarında scada benzeri arayüzler sanal gerçeklik ortamında oluşturularak operatörlerin sistemi sanal ortamda izleyerek kumanda edebilmesi sağlanmaktadır. Bu uygulama operatörlerin daha heyecanlı bir şekilde çalışma yapmasını sağlamaktadır (Kizilov, vd., 2023).

3.4. Mekatronik / Robotik ve Otomasyon Mühendisliği

Mekatronik ve robotik otomasyon alanında özellikle robot teleoperasyonu, endüstriyel robot eğitimi, insan-robot işbirliği ve tehlikeli ortamlarda görev alması için kullanımaktadır. Çalışmalar, sanal gerçekliğin hem güvenliği hem de kullanıcı performansını arttırdığını göstermektedir. Yapılan çalışmalar, askeri mobil robot TAROS için bir sanal gerçeklik ortamı tasarlanmıştır. Bu ortama mobil robotun bir dijital ikizi yerleştirilerek robotun sezgisel kontrolü ve daha iyi durum farkındalığı kazanması sağlanmıştır. Özellikle manipülatör için ekstra ani çarpışma sistemide entegre edilerek robotun kontrolünün daha verimli bir şekilde uygulanması sağlanmıştır (Kot, vd., 2018). Başka bir çalışmada sanal gerçeklik ve dijital ikiz ile yol bakım gibi tehlikeli sahalarda dinamik sahnenin yeniden oluşturulmasıyla teleoperasyon ile operatörün uzaktan çarpışmadan kaçınarak kontrol sağlamasına olanak sunuyor (Bavelos, vd., 2024).

3.5. Biyomedikal Mühendisliği

Biyomedikal mühendisliğinde sanal gerçeklik özellikle nörolojik rehabilitasyon, sanal cerrahi ve tıp eğitimi bilişsel rehabilitasyon ve biyomedikal veri görselleştirme alanarında hızla yaygınlaşmaktadır. Özellikle rehabilitasyon çalışmalarında

son yıllarda sanal gerçeklik uygulamaları yaygınlaşmıştır. Sanal gerçekliğin üst ya da alt ekstremite motor fonksiyonlarını, denge, yürüme ve günlük yaşam aktivitelerini önemli ölçüde iyileştirdiği ve terapi seanslarını daha eğlenceli hale getirerek hastaya psikolojik olarak iyi geldiği anlaşılmıştır (Zhang, vd., 2021; Sveistrup, vd., 2004). Aynı zamanda sanal gerçeklik cerrahi simülatörlerde de kullanılmaktadır. Bu simülatörler artroskopi gibi alanlarda teknik beceriyi, hız ve doğruluğu arttırarak gerçek ameliyathane performansına katkı sağlamaktadır (Dhillon, vd., 2024).

3.6. Havacılık ve Uzay Mühendisliği

Sanal gerçeklik, havacılık ve uzay mühendisliğinde uçuş/astronot eğitimi ve tasarım, analiz ve planlama için kullanılmaktadır. Pilot eğitiminde kullanılan sanal gerçeklik sistemi hem düşük maliye hem de kullanıcıya durumsal bir farkındalık katmaktadır (Cross, vd., 2022). Uçak ve roket tasarımı çalışmalarında sanal gerçeklik, kompresör kanadı, uçak/roket alt sistemlerinin 3 boyutlu olarak incelenmesini sağlamaktadır. Ayrıca aero tasarım anlayışını geliştirmekte ve hızlı iterasyon ve veri analitiğini sağlamaktadır (Tadeja, vd., 2020).

3.7. Yazılım ve Bilgisayar Mühendisliği

Sanal gerçeklik, yazılım ve bilgisayar mühendisliğinde hem öğretim ortamı oluşturmak hem de arge platformu olarak kullanılabilmektedir. Yazılım geliştirme uygulamalarında sanal gerçeklik ortamında yazılım geliştiriciyi 3 boyutlu bir proje ortamına yerleştirerek kod yapılarını, görev yönetimini ve etkip etkileşimini sağlamaktadır (Güleç, vd., 2018).

4. SONUÇ

Bu çalışmada sanal gerçeklik teknolojilerinin mühendislik alanlarındaki kullanımı incelenmiştir. Çalışmada sanal gerçeklik bileşenleri açıklanmış ve bunların tümleşik bir şekilde kullanılmasının sanal gerçeklik ortamının daha gerçekçi, etkileşimli ve daha güvenilir bir şekilde oluşturmasını mümkün kılmaktadır.

Mühendislik uygulamalarında, sanal gerçeklik; tasarım, eğitim, analiz ve simülasyon çalışmalarında oldukça iyi performans gösterdiği anlaşılmıştır. Fizik ve simülasyon motorları sayesinde gerçek dünyadaki fiziksel performans sanal gerçeklik ortamınada aktarılabilmektedir. Aynı zamanda yapay zeka ve öğrenme tabanlı kontrol algoritmaları sayesinde sistemler adaptif ve gerçek zamanlı olarak kontrol edilebilmektedir. Bu sayede sistemler kişiye özel olarak kontrol edilebilmektedir.

Sonuç olarak sanal gerçeklik teknolojileri mühendislik çalışmalarında sadece görselleştirme amacıyla değil; aynı zamanda karar verme sistemi kontrol etme gibi bütün bu parametreleri bütüncül bir şekilde birleştiren bir yapıyı sunmaktadır. Gelişen donanım ve yazılım teknojileri ile birlikte önümüzdeki yıllarda daha yaygınlaşarak akıllı ve disiplinler arası çözümler sunması beklenmektedir.

KAYNAKÇA

Aleger Global. (t.y.). *Head-Mounted Displays*. Aleger Global. https://alegerglobal.com/en/augmented-reality/smart-glasses/head-mounted-displays/

Alakus, F., Isık, A. H., & Eskicioğlu, Ö. C. (2021). Hareket Yakalama ve Sanal Gerçeklik Teknolojileri Kullanarak Oyun Tabanlı Rehabilitasyon. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *9*(6), 269-279.

Bavelos, A., Anastasiou, E., Dimitropoulos, N., Michalos, G., & Makris, S. (2024). Virtual reality-based dynamic scene recreation and robot teleoperation for hazardous environments. *Computer-Aided Civil and Infrastructure Engineering*, 40, 392 - 408. https://doi.org/10.1111/mice.13337.

Caserman, P., Garcia-Agundez, A., Konrad, R., Göbel, S., & Steinmetz, R. (2019). Real-time body tracking in virtual reality using a Vive tracker. *Virtual Reality*, *23*(2), 155-168.

Chen, D., Youssef, A., Pendse, R., Schleife, A., Clark, B. K., Hamann, H., ... & Nagpurkar, P. (2024). Transforming the hybrid cloud for emerging AI workloads. *arXiv preprint arXiv:2411.13239*.

Cross, J., Boag-Hodgson, C., Ryley, T., Mavin, T., & Potter, L. (2022). Using Extended Reality in Flight Simulators: A Literature Review. *IEEE Transactions on Visualization and Computer Graphics*, 29, 3961-3975. https://doi.org/10.1109/tvcg.2022.3173921.

Dhillon, J., Tanguilig, G., & Kraeutler, M. (2024). Virtual and Augmented Reality Simulators Show Intraoperative, Surgical Training, and Athletic Training Applications: A Scoping Review.. *Arthroscopy : the journal of arthroscopic*

& related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association. https://doi.org/10.1016/j.arthro.2024.02.011.

Kizilov, S., Nikitenko, M., Khudonogov, D., Neogi, B., & Verkhovtsev, D. (2023). Technological equipment control system using virtual reality devices. *Mining Industry Journal (Gornay Promishlennost)*. https://doi.org/10.30686/1609-9192-2023-5-124-129.

Kot, T., & Novák, P. (2018). Application of virtual reality in teleoperation of the military mobile robotic system TAROS. *International Journal of Advanced Robotic Systems*, 15. https://doi.org/10.1177/1729881417751545.

Lai, Z. (2024). Diving into the virtual realm: Exploring the mechanics of virtual reality. *Applied and Computational Engineering*, *31*(1), 268-273.

Lampropoulos, G., Fernández-Arias, P., de Bosque, A., & Vergara, D. (2025). Virtual Reality in Engineering Education: A Scoping Review. *Education Sciences*, *15*(8), 1027.

Güleç, U. (2018). *A virtual reality-based training environment designed for hands-on experience of software development* (Doctoral dissertation, Middle East Technical University (Turkey)).

Ho, T. L. H., & Thanh, N. C. (2025). Virtual Reality and Augmented Reality Technology in Bridge Engineering: A Review. *Engineering, Technology & Applied Science Research*, *15*(2), 20729-20736.

Huzaifa, M., Desai, R., Grayson, S., Jiang, X., Jing, Y., Lee, J., ... & Adve, S. V. (2022). Illixr: An open testbed to enable

extended reality systems research. *IEEE Micro, 42*(4), 97-106.

Ribeiro de Oliveira, T., Moura da Silva, M., Nepomuceno Spinasse, R. A., Giesen Ludke, G., Ruy Soares Gaudio, M., Iglesias Rocha Gomes, G., ... & Mestria, M. (2021, October). Systematic review of virtual reality solutions employing artificial intelligence methods. In *Proceedings of the 23rd Symposium on Virtual and Augmented Reality* (pp. 42-55).

Rostami, M., Pradhan, P., Karki, N., Omorodion, J., Milani, P., Kamoonpuri, J., ... & Chung, J. (2025). A comprehensive review of extended reality and its application in aerospace engineering. *Progress in Aerospace Sciences*, 101118.

Sveistrup, H. (2004). Motor rehabilitation using virtual reality. *Journal of NeuroEngineering and Rehabilitation*, 1, 10 - 10. https://doi.org/10.1186/1743-0003-1-10.

Singh, G., Mantri, A., Sharma, O., & Kaur, R. (2020). Virtual reality learning environment for enhancing electronics engineering laboratory experience. *Computer Applications in Engineering Education*, 29, 229 - 243. https://doi.org/10.1002/cae.22333.

Shi, Y., & Shen, G. (2024). Haptic sensing and feedback techniques toward virtual reality. *Research, 7*, 0333.

Shin, Y., & Lee, M. (2024). Development of Sensory Virtual Reality Interface Using EMG Signal-Based Grip Strength Reflection System. *Applied Sciences, 14*(11), 4415.

Tadeja, S., Seshadri, P., & Kristensson, P. (2020). AeroVR: An immersive visualisation system for aerospace design and digital twinning in virtual reality. *The Aeronautical Journal*, 124, 1615 - 1635. https://doi.org/10.1017/aer.2020.49.

Villada Castillo, J. F., Bohorquez Santiago, L., & Martínez García, S. (2025). Optimization of Physics Learning Through Immersive Virtual Reality: A Study on the Efficacy of Serious Games. *Applied Sciences*, *15*(6), 3405.

Yang, L. I., Huang, J., Feng, T. I. A. N., Hong-An, W. A. N. G., & Guo-Zhong, D. A. I. (2019). Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware*, *1*(1), 84-112.

Yao, S. N. (2017). Headphone-based immersive audio for virtual reality headsets. *IEEE Transactions on Consumer Electronics*, *63*(3), 300-308.

Wang, P., Wu, P., Wang, J., Chi, H. L., & Wang, X. (2018). A critical review of the use of virtual reality in construction engineering education and training. *International journal of environmental research and public health*, *15*(6), 1204.

Wolfartsberger, J. (2019). Analyzing the potential of Virtual Reality for engineering design review. *Automation in construction*, *104*, 27-37.

Zeng, H., & Zhao, R. (2023). Perceptually-guided dual-mode virtual reality system for motion-adaptive display. *IEEE Transactions on Visualization and Computer Graphics*, *29*(5), 2249-2257.

Zhang, B., Li, D., Liu, Y., Wang, J., & Xiao, Q. (2021). Virtual reality for limb motor function, balance, gait, cognition and daily function of stroke patients: A systematic review and meta-analysis.. *Journal of advanced nursing*. https://doi.org/10.1111/jan.14800.

TÜRKİYE BAĞLAMINDA BÜYÜK DİL MODELLERİNDE RİSKLER, GÜVENLİK VE UYUM YAKLAŞIMLARI

Fesih KESKIN[1]

Gulser OZ[2]

1. GİRİŞ

Doğal Dil İşleme (NLP, Natural Language Processing) alanı, ölçeklenen veri ve hesaplama gücü sayesinde Büyük Dil Modellerinin (LLM, Large Language Models) hızla olgunlaşmasıyla yeni bir döneme girmiştir. Transformer tabanlı mimarilerle eğitilen bu modeller, metin üretimi, muhakeme, özetleme ve kod üretimi gibi yetenekleriyle kamu ve özel sektörde kritik süreçlere entegre edilmektedir (Abdali, Anarfi, Barberan, He, & Shayegani, 2024). Bununla birlikte, LLM'lerin kritik altyapılara ve hassas karar süreçlerine yerleşme hızı, güvenlik ve emniyet mekanizmalarının aynı hızla olgunlaşmasını her zaman mümkün kılmamıştır. İnternetten derlenen büyük ölçekli ve çoğu zaman denetimsiz veri üzerinde ön eğitim yapan modeller, bu verinin taşıdığı önyargıları, toksik örüntüleri ve güvenlik açıklarını da içselleştirebilmekte; bu durum hem toplumsal zarar riskini hem de kötüye kullanım ihtimalini artırmaktadır (Weidinger, ve diğerleri, 2021).

Bu noktada, Yapay zekâ emniyeti ile Yapay zekâ güvenliği arasındaki ayrım işlevseldir. Yapay zekâ emniyeti,

[1] Dr. Öğr. Üyesi, Iğdır Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği, ORCID: 0000-0002-3798-2912.

[2] Arş. Gör, Iğdır Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği, ORCID: 0000-0002-9620-4598.

modelin amaç fonksiyonundan, eğitim verisinden veya tasarım tercihlerinden kaynaklanan istenmeyen zararları önlemeyi hedefler; toksik çıktı, ayrımcı ifade ve halüsinasyon buna dahildir (Shi, ve diğerleri, 2024). Yapay zekâ güvenliği ise kasıtlı saldırılara, dış tehdit aktörlerine ve sistem bütünlüğünü ihlal eden davranışlara odaklanır; jailbreak, veri zehirleme ve istem yönlendirme bu gruptadır (Akiri, Simpson, Aryal, Khanna, & Gupta, 2025; Abdali, Anarfi, Barberan, He, & Shayegani, 2024). Türkiye'de kurumsal uygulamalar açısından bu ayrım, bir yandan uyum ve etik gereklilikleri (özellikle Kişisel Verilerin Korunması Kanunu, KVKK, Law on the Protection of Personal Data) diğer yandan siber güvenlik kontrolleri ve saldırı yüzeyi yönetimi arasındaki görev paylaşımını netleştirmeye yardımcı olur.

LLM risklerini değerlendirmek için bu bölüm üç ana risk alanını esas alır: ayrımcılık ve toksisite, bilgi tehlikeleri (mahremiyet ve veri sızıntısı), kötüye kullanım ve teknik zafiyetler (Weidinger, ve diğerleri, 2021). Türkiye bağlamı, bu alanların her birinde yerelleştirme ihtiyacını güçlendirir. Örneğin Türkçe, dilbilgisel olarak cinsiyetsiz bir üçüncü tekil şahıs zamirine sahiptir; "o" hem he hem she hem it karşılığıdır. Buna rağmen LLM'ler, özellikle çeviri veya çıkarım üretirken toplumsal cinsiyet kalıp yargılarını içeren varsayımlar ekleyebilmektedir (Caglidil, Ostendorff, & Rehm, 2024). Benzer şekilde, KVKK'nın özel nitelikli veri rejimi, LLM'lerin eğitim ve çıktı aşamalarında kişisel veriyi "hatırlama ve yeniden üretme" riskine karşı daha sıkı bir uyum ihtiyacı doğurur (Kişisel Verileri Koruma Kurumu (KVKK), 2025). Teknik tarafta ise Türkçenin eklemeli yapısı, tokenizasyon kalitesi üzerinden hem performansı hem de emniyet filtrelerinin etkinliğini etkileyen bir risk kaynağıdır (Bayram, ve diğerleri, 2025).

Aşağıdaki tablo, risk sınıflarını küresel örneklerle birlikte Türkiye'ye özgü görünümleriyle özetler.

Tablo 1. Büyük Dil Modellerinde temel risk alanları ve Türkiye bağlamındaki yansımaları

Risk alanı	Küresel görünüm	Türkiye bağlamında tipik yansıma
Ayrımcılık ve toksisite	Nefret söylemi, stereotipleştirme, dışlayıcı dil ve toksik içerik üretimi (Weidinger, ve diğerleri, 2021)	"O" zamirinin cinsiyetsiz yapısına rağmen çeviri ve bağlamsal çıkarımda cinsiyet varsayımı; örtük hakaret ve ironiye dayalı saldırgan dilin yakalanmasında zorluk (Caglidil, Ostendorff, & Rehm, 2024)
Bilgi tehlikeleri	Kişisel verinin modele sızması, çıktıda PII üretimi (PII, Personally Identifiable Information) ve gizli bilgi ifşası (Abdali, Anarfi, Barberan, He, & Shayegani, 2024)	KVKK'nın açık rıza ve özel nitelikli veri koşulları; silme hakkı ve modelden veri "unutturma" gerilimi (Kişisel Verileri Koruma Kurumu (KVKK), 2025; Vassilev, 2025)
Kötüye kullanım ve teknik zafiyet	Kandırma, dolandırıcılık, propaganda; istem yönlendirme ve veri zehirleme gibi saldırılar (Abdali, Anarfi, Barberan, He, & Shayegani, 2024)	Derin sahte içeriklerle seçim güvenliği ve itibar saldırıları; Türkçe token parçalanması nedeniyle filtre zayıflaması ve bağlam kaybı (Belada, 2024; Bayram, ve diğerleri, 2025)

Bu risk alanları, küresel çerçevelerle uyumlu şekilde ele alınmalıdır. Örneğin NIST Yapay Zekâ Risk Yönetimi Çerçevesi (Tabassi, 2023) kurumsal risk yönetişimi için temel ilkeleri sunar; ancak Türkiye'de uygulama, Türkçenin dilsel özellikleri ile KVKK ve Türk Ceza Kanunu (TCK, Turkish Penal Code) gibi düzenlemelerin gerekleri üzerinden yerelleştirilmek zorundadır (Akbulut, 2023; STB & CDDO, Ulusal Yapay Zekâ Stratejisi 2021-2025, 2021).

2. TÜRKÇE LLM'LERDE DİLBİLİMSEL VE TEKNİK RİSKLER

Türkiye'de LLM tabanlı uygulamaların kalitesi ve emniyeti, yalnızca modelin büyüklüğüyle değil, Türkçeye uyumlu bileşenlerin tasarımıyla belirlenir. Bu noktada tokenizasyon, önyargı üretimi ve toksik içerik tespiti öne çıkar.

Türkçe eklemeli bir dildir. Kök üzerine çok sayıda ek eklenerek tek kelimede yoğun anlam taşınabilir. Tokenleştirme algoritmaları, özellikle İngilizce merkezli alt birim yöntemleriyle (örneğin BPE) eğitilmişse, Türkçe kelimeleri anlamlı morfem sınırlarından koparak parçalayabilir. "Evlerimizden" gibi bir kelimenin "ev-ler-imiz-den" biçiminde anlamlı bileşenlerle temsil edilmesi idealken, anlamsal olarak zayıf parçacıklara bölünmesi bağlam kaybına ve daha fazla token tüketimine yol açar (Bayram, ve diğerleri, 2025). Bu durum yalnızca verimlilik ve gecikme maliyeti yaratmaz; aynı zamanda emniyet filtrelerinin belirli kalıpları yakalama kapasitesini de düşürebilir. Filtreler belirli sözcük biçimlerine veya n-gram örüntülerine dayanıyorsa, parçalanma bu örüntüleri görünmez hale getirebilir. Dolayısıyla Türkçe için token saflığını artıran, morfoloji farkındalığı yüksek tokenizasyon tasarımları, hem performans hem de emniyet açısından kritik bir hafifletme alanıdır (Bayram, ve diğerleri, 2025).

İkinci risk alanı kültürel ve toplumsal önyargılardır. Türkçe dilbilgisel olarak cinsiyet belirtmediği halde, LLM'ler eğitim verisindeki dağılımları yansıtarak cinsiyetle ilişkilendirilmiş meslek kalıplarını üretime taşıyabilir. Çeviri senaryolarında "o bir doktor" ifadesinin erkek zamiriyle, "o bir hemşire" ifadesinin kadın zamiriyle verilmesi buna tipik bir örnektir (Caglidil, Ostendorff, & Rehm, 2024). Bu tür önyargılar, işe alım, performans değerlendirme, müşteri temsil süreçleri veya eğitimde yönlendirme gibi alanlarda adaletsiz sonuçlara

dönüşebilir. Bu nedenle Türkiye'de LLM değerlendirmeleri, sadece genel doğruluk ölçütleriyle değil, Türkçe bağlamlı ayrımcılık ve temsil zararlarını ölçen test setleriyle de desteklenmelidir.

Üçüncü olarak, Türkçede toksik içerik tespiti çoğu zaman örtük anlam, ironi ve bağlama dayalı saldırganlık üzerinden ilerler. İngilizce için iyi çalışan normalizasyon ve anahtar kelime tabanlı yaklaşımlar Türkçede benzer performansı göstermeyebilir; ayrıca eklemeli yapı, saldırgan kelimelerin farklı eklerle çoğaltılmasını kolaylaştırır. Bu durum, bağlam farkındalığı yüksek sınıflandırıcıların ve Türkçe odaklı veri artırma tekniklerinin gerekliliğini ortaya koyar (Tanyel, ALKURDI, & AYVAZ, 2024). Kurumsal uygulamalarda, özellikle kullanıcıyla doğrudan etkileşen sohbet botlarında, toksisite tespiti sadece model içi politika katmanına bırakılmamalı; bağımsız denetleyici sınıflandırıcılar ve insan denetimli süreçler ile tamamlanmalıdır.

3. TÜRKİYE'DE HUKUKİ SİSTEM VE DÜZENLEMELER AÇISINDAN RİSKLER

LLM'lerin Türkiye'de kullanımı, teknik risklerin ötesinde KVKK uyumu, telif ve fikri mülkiyet sorunları ile ceza sorumluluğu tartışmaları nedeniyle çok katmanlı bir hukuki değerlendirme gerektirir.

KVKK açısından en kritik konu, eğitim verisinin temini ve model çıktılarının kişisel veri işleme sayılıp sayılmadığıdır. Büyük veri kümelerinin web kazıma yoluyla toplanması yaygındır; ancak KVKK rehber yaklaşımı, kişisel verinin alenileştirilmiş olmasının sınırsız işleme serbestisi doğurmadığını vurgular. Kişinin veriyi kamuya açık hale getirme iradesi belirli bir amaçla sınırlı olabilir; bu amaç dışı model eğitimi, amaçla sınırlılık ilkesini ihlal edebilir (Kişisel Verileri Koruma Kurumu

(KVKK), 2025). Bu nedenle LLM geliştiren veya kullanan aktörlerin, hangi veri kategorilerini hangi hukuki dayanakla işlediğini açıkça belgelemesi, veri minimizasyonu uygulaması ve mümkün olduğunda özel nitelikli veriyi eğitim sürecinden dışlaması gerekir.

KVKK'nın veri sahibi haklarıyla LLM'lerin teknik doğası arasında belirgin bir gerilim vardır. Silme ve yok etme talepleri, klasik veri tabanlarında doğrudan uygulanabilirken, LLM'lerde kişisel verinin ağırlıklarda temsili nedeniyle "unutma" talebi teknik açıdan zorlaşır. Makine unutturma (machine unlearning) yaklaşımları gelişmekle birlikte, maliyetli olabilir ve her zaman kesin garanti vermeyebilir (Vassilev, 2025; Shi, ve diğerleri, 2024). Bu gerçek, kurumların uyum riskini artırır. Uygulamada daha güvenli bir yaklaşım, kişisel verinin modele girmesini azaltan önlemler, eğitim verisinde otomatik PII temizleme, eğitim sırasında mahremiyet artırıcı teknolojiler (PET, Privacy Enhancing Technologies) ve çıktıda hassas veri sızıntısını tespit eden denetim katmanlarının birlikte kullanılmasını gerektirir (Kişisel Verileri Koruma Kurumu (KVKK), 2025).

Fikri mülkiyet boyutunda, Fikir ve Sanat Eserleri Kanunu (FSEK, Law on Intellectual and Artistic Works) çerçevesinde eser sahibinin gerçek kişi olması ilkesi önemlidir. Bu nedenle mevcut yaklaşım, LLM'nin eser sahibi olarak kabul edilmemesi yönündedir; üretimin telif statüsü ise insan katkısının düzeyine göre değerlendirilebilir (Kızrak, ve diğerleri, 2019; Dericioglu Egemen, ve diğerleri, 2024). Bunun yanında, telifli eserlerin eğitim verisinde izinsiz kullanımı ve model çıktısında telifli metinlere aşırı benzerlik üretilmesi, hem eser sahipliği tartışması hem de ihlal iddiaları doğurabilir. Türkiye'de bu alanın yargı içtihadı ile netleşmesi zaman alabileceği için, kurumsal risk yönetimi lisanslı veri setleri, açık lisanslı kaynaklar ve telifli içerik filtreleme yaklaşımlarıyla desteklenmelidir.

Ceza hukuku bakımından, Türkiye'de LLM'nin fail olarak kabul edilmesi mümkün değildir; ceza sorumluluğu şahsidir ve yapay sistemlerin kast veya taksirle hareket ettiğinden söz edilemez (Kızrak, ve diğerleri, 2019). Buna rağmen LLM'ler, dolandırıcılık, sahtecilik, kişilik haklarına saldırı, özel hayatın gizliliğini ihlal gibi suçlarda araç olarak kullanılabilir. Örneğin bir kişinin LLM ile kimlik avı metni üretip bunu bilişim sistemleri üzerinden kullanması halinde, suçun faili kullanan kişidir ve TCK hükümleri uygulanır (Akbulut, 2023). Daha zor alan, geliştirici veya dağıtıcıların sorumluluğudur. Öngörülebilir risklere rağmen yeterli önlemi almamak, belirli koşullarda taksir tartışmasını gündeme getirebilir; ancak illiyet bağı ve özen yükümlülüğünün ihlali teknik ve hukuki değerlendirme gerektirir (Kızrak, ve diğerleri, 2019). Bu belirsizlik, kurumların güvenlik testlerini, kullanım kısıtlarını, kayıt ve izleme süreçlerini belgelemesini ve "özen standardı" açısından savunulabilir bir uyum dosyası oluşturmasını pratikte daha önemli hale getirir.

4. TOPLUMSAL ZARARLAR VE GÜVENLİK TEHDİTLERİ

Türkiye'de LLM'lerin toplumsal etkisi, özellikle dezenformasyon, siber güvenlik ve akademik etik alanlarında belirginleşir. Üretken yapay zekâ sistemleri, ikna edici metinleri ve sentetik medyayı düşük maliyetle ölçekleyebildiği için bilgi ekosisteminin güvenilirliğini zorlar.

Derin sahte içerikler (deepfake) bu bağlamda en kritik risklerden biridir. Derin sahte, biyometrik veriyi (ses, yüz, mimik) taklit eden ve gerçeğe çok yakın görsel veya işitsel içerik üretebilen yöntemlerle üretilir (Belada, 2024). Seçim dönemleri gibi yüksek hassasiyetli süreçlerde, sahte içeriklerin yayılması kadar, bu teknolojilerin varlığının yarattığı "gerçeği inkâr avantajı" da önem kazanır. Kişiler, gerçek kayıtları dahi "sahte

olabilir" iddiasıyla tartışmalı hale getirerek kamu güvenini aşındırabilir (OpenAI, ve diğerleri, 2023). Buna ek olarak, hedefli itibar saldırıları ve rıza dışı sentetik içerikler, bireysel düzeyde ağır psikolojik ve sosyal zararlar doğurur (Belada, 2024). Türkiye'de bu risk, hızlı yayılım gösteren sosyal ağ pratikleriyle birleştiğinde, hukuk yolları devreye girse bile zararın geri döndürülmesini güçleştirebilir.

Siber güvenlik açısından, LLM'lerin uygulama içine gömülmesi yeni saldırı yüzeyleri yaratır. İstem yönlendirme (prompt injection) saldırıları, sistem talimatları ile kullanıcı girdisinin birleştirilme biçimini kötüye kullanarak modele yetkisiz hedefler benimsetebilir (Abdali, Anarfi, Barberan, He, & Shayegani, 2024). Dolaylı istem yönlendirme (indirect prompt injection) özellikle belge işleyen veya web içeriği okuyan sistemlerde tehlikelidir; modele dış kaynaktan gelen metnin içine gizlenmiş talimatlar yerleştirilebilir (Wu, Zhang, Jha, McDaniel, & Xiao, 2024). Jailbreak teknikleri ise, modelin politika filtrelerini aşarak yasaklı çıktılar üretmesini hedefler ve mevcut hizalama yöntemlerinin kırılganlığını ortaya koyar (Pathade, 2025). Veri zehirleme (data poisoning) ve arka kapı (backdoor) senaryoları da hem eğitim hem de kurum içi bilgi tabanına dayalı sistemlerde (özellikle RAG mimarilerinde) ciddi risk oluşturur (Fendley, ve diğerleri, 2025; Lin, Wang, Chen, & Mao, 2025).

Akademik etik boyutunda, Yükseköğretim Kurulu (YÖK, Council of Higher Education) üretken yapay zekânın araştırma süreçlerine yardımcı olabileceğini, ancak bir yazar gibi konumlandırılamayacağını belirtir (YÖK, 2024). LLM çıktılarının doğrulanmadan kullanılması, uydurma kaynak, hatalı alıntı ve yanlı yorum risklerini artırır. Türkiye'de akademik bütünlüğü korumak için, üretken yapay zekâ kullanımının yöntem bölümünde şeffaf biçimde beyan edilmesi ve nihai sorumluluğun insan araştırmacıda kalması yönündeki yaklaşım,

hem bilimsel güvenilirliği hem de mesleki etik standartları destekler (YÖK, 2024).

5. HAFİFLETME VE UYUM STRATEJİLERİ

LLM risklerini hafifletme, tek bir teknik önlemle sağlanamaz. Türkiye bağlamında etkili yaklaşım, model yaşam döngüsünün tamamına yayılan çok katmanlı bir tasarım ve yönetişim anlayışıdır. Bu bölümde kültürel uyumlu hizalama, güvenilir bilgiye dayalı mimari tercihleri, yerelleştirilmiş adversaryal test ve kurumsal yönetişim eksenleri öne çıkar.

İlk eksen, insan geri bildirimiyle pekiştirmeli öğrenme (RLHF, Reinforcement Learning from Human Feedback) ve Türkçe odaklı ince ayar süreçleridir. İngilizce merkezli geri bildirim setleri, Türkçenin kültürel bağlamını ve söylem nüanslarını yakalamakta yetersiz kalabilir. Bu nedenle, Türkçe değerlendirici havuzlarıyla ödül modeli ve politika modeli eğitiminin yerelleştirilmesi, hem toksik içerik tespiti hem de cinsiyet ve kültürel önyargıların azaltılması açısından önemlidir (Caglidil, Ostendorff, & Rehm, 2024; Tanyel, ALKURDI, & AYVAZ, 2024). Benzer biçimde, Türkçe morfolojiye duyarlı değerlendirme ölçütleri, modelin emniyet davranışını gerçek kullanım bağlamına yaklaştırır. Kurumsal uygulamalarda, sadece eğitim sonrası denetim değil, üretim aşamasında da içerik denetleyicileri ve kayıt temelli izleme mekanizmaları gerekir.

İkinci eksen, geri getirim destekli üretim (RAG, Retrieval-Augmented Generation) mimarisidir. Halüsinasyon riskini azaltmak için LLM'nin parametre içi "bilgi" yerine doğrulanmış kaynaklardan geri getirilen belge parçalarıyla yanıt üretmesi, özellikle hukuk ve kamu hizmetlerinde güvenilirliği artırır (Huang, ve diğerleri, 2023; Roy, ve diğerleri, 2024). Türkiye'de bu yaklaşım, Resmî Gazete, mevzuat veri tabanları ve yüksek yargı kararları gibi yetkili kaynaklara bağlanan kurumsal bilgi

tabanlarıyla uygulandığında, modelin güncel hukuku yanlış üretmesi riskini anlamlı ölçüde düşürebilir (Ayaz, 2025). Ancak RAG, bilgi tabanı zehirlenmesi riskini beraberinde getirir. Bu nedenle belge bütünlüğü, erişim kontrolü, değişiklik kaydı ve vektör veri tabanına yönelik yetkilendirme, RAG mimarisinin ayrılmaz güvenlik katmanları olarak tasarlanmalıdır (Yao, ve diğerleri, 2025).

Üçüncü eksen, adversaryal test ve kırmızı takım çalışmalarıdır. NIST'in değerlendirme yaklaşımı, LLM'lerin yayına alınmadan önce sistematik saldırgan testlere tabi tutulmasını önerir (Vassilev, 2025). Türkiye'de etkili kırmızı takım, yerel gündem, sosyo politik hassasiyetler, Türkçe argo ve dolaylı anlatım biçimleri gibi yerel saldırı vektörlerini içermelidir. Bu testler, yalnızca yasaklı içerik üretimi değil, aynı zamanda yanıltıcı güven iddiaları, araç kullanımını gizleme gibi şeffaflık ihlalleri ve iş akışı manipülasyonu gibi davranışsal riskleri de kapsamalıdır (Liu, Cui, & Zhang, 2025; Shi, ve diğerleri, 2024).

Dördüncü eksen, yönetişim ve uyum mimarisidir. Türkiye Ulusal Yapay Zekâ Stratejisi (STB & CDDO, 2021) güvenilir yapay zekâ hedefi doğrultusunda, denetim, şeffaflık ve hesap verebilirlik ilkelerini vurgular. Kurumların iç politika setleri, KVKK rol tanımları, veri işleme envanteri, tedarik zinciri riskleri ve model güncelleme süreçlerini kapsamalıdır. Kamu tarafında, kamu görevlilerinin yapay zekâ kullanımına ilişkin etik ilkeler, tarafsızlık, temel haklara saygı ve şeffaflık gibi çerçevelerle uygulanmalıdır (T.C. Kamu Görevlileri Etik Kurulu, 2024). Sertifikasyon tarafında ise Türk Standardları Enstitüsü (TSE, Turkish Standards Institution) tarafından geliştirilen güvenilir yapay zekâ damgası ve ilgili yönetim sistemi yaklaşımları, kurumsal kontrol setlerinin standartlaşmasını destekleyebilir (STB & CDDO, 2024; TSE GLOBAL, 2025). Bu tür çerçeveler, hem mühendislik ekiplerine somut kontrol listeleri sağlar hem de

hukuk ve uyum ekiplerinin denetlenebilir süreçler inşa etmesini kolaylaştırır.

6. SONUÇ VE GELECEK PERSPEKTİFİ

Büyük Dil Modelleri (LLM, Large Language Models) Türkiye için stratejik bir fırsat alanı sunarken, eş zamanlı olarak mahremiyet, güvenlik, ayrımcılık ve dezenformasyon risklerini de büyütmektedir. Bu bölümde, Türkçenin eklemeli yapısının tokenizasyon üzerinden performans ve emniyeti etkilediği, cinsiyetsiz dil yapısına rağmen modellerin cinsiyet önyargılarını üretebildiği ve KVKK'nın kişisel veri koruma rejiminin LLM yaşam döngüsüne doğrudan yansıdığı gösterilmiştir. TCK açısından ise LLM'lerin fail değil araç olarak değerlendirilmesi, sorumluluğun insan aktörlerde toplandığı bir çerçeve yaratırken, geliştirici ve dağıtıcıların özen standardına ilişkin belirsizlikler, proaktif risk yönetimini daha önemli kılar.

Gelecek dönemde Türkiye'nin güvenilir yapay zekâ hedefi, iki yönlü ilerlemeyi gerektirir. Birincisi, teknik tarafta Türkçe odaklı hizalama, güvenilir kaynaklara dayalı RAG mimarileri ve yerel kırmızı takım testleriyle emniyet ve güvenlik katmanlarının olgunlaştırılmasıdır. İkincisi, yönetişim ve uyum tarafında KVKK ile uyumlu veri yönetimi, tedarik zinciri denetimi, sertifikasyon ve düzenleyici deney alanları gibi mekanizmaların işletilmesidir. Bu iki eksen birlikte yürütüldüğünde, Türkiye yalnızca küresel teknolojiyi tüketen değil, kendi dilsel ve hukuki gerçekliklerine göre güvenli ve hesap verebilir üretken yapay zekâ sistemleri geliştiren bir ekosisteme dönüşebilir.

KAYNAKÇA

Abdali, S., Anarfi, R., Barberan, C. J., He, J., & Shayegani, E. (2024). Securing Large Language Models: Threats, Vulnerabilities and Responsible Practices. *arXiv.org*. doi:10.48550/ARXIV.2403.12503

Akbulut, B. (2023, October). YAPAY ZEKA VE CEZA HUKUKU SORUMLULUĞU. *Ankara Hacı Bayram Veli Üniversitesi Hukuk Fakültesi Dergisi, 27*, 267–319. doi:10.34246/ahbvuhfd.1339596

Akiri, C., Simpson, H., Aryal, K., Khanna, A., & Gupta, M. (2025). Safety and Security Analysis of Large Language Models: Benchmarking Risk Profile and Harm Potential. *arXiv.org*. doi:10.48550/ARXIV.2509.10655

Ayaz, S. (2025, July). Yapay Zekânın Hukuk Alanındaki Uygulamaları ve Hukuk Alanında Kullanılmasının Olası Sonuçları. *Sakarya Üniversitesi Hukuk Fakültesi Dergisi, 13*, 80–108. doi:10.56701/shd.1566441

Bayram, M. A., Fincan, A. A., Gümüş, A. S., Karakaş, S., Diri, B., & Yıldırım, S. (2025). Tokenization Standards for Linguistic Integrity: Turkish as a Benchmark. *Tokenization Standards for Linguistic Integrity: Turkish as a Benchmark*. arXiv. doi:10.48550/ARXIV.2502.07057

Belada, N. E. (2024, June). DEEPFAKE DEZENFORMASYONU. *Bilişim Hukuku Dergisi, 6*, 321–358. doi:10.55009/bilisimhukukudergisi.1387306

Caglidil, O. M., Ostendorff, M., & Rehm, G. (2024). Investigating Gender Bias in Turkish Language Models. *Investigating Gender Bias in Turkish Language Models*. arXiv. doi:10.48550/ARXIV.2404.11726

Dericioglu Egemen, A., Uyan, B., Küzeci, E., Bensason, İ., Sezgin, Ö., & Kaspı, R. (2024). *Yapay Zekâ Etik İlkeleri ve Hukuki Düzenlemeler Raporu.* Tech. rep., Türkiye Yapay Zekâ İnisiyatifi (TRAI), İstanbul, Turkey. https://turkiye.ai/wp-content/uploads/2024/06/TRAI-Yapay-Zeka-Etik-Ilkeleri-ve-Hukuki-Duzenlemeler-Raporu-Mayis-2024-5.pdf adresinden alındı

Fendley, N., Staley, E. W., Carney, J., Redman, W., Chau, M., & Drenkow, N. (2025). A Systematic Review of Poisoning Attacks Against Large Language Models. *arXiv.org*. doi:10.48550/ARXIV.2506.06518

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., . . . Liu, T. (2023). A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. doi:10.48550/ARXIV.2311.05232

Kişisel Verileri Koruma Kurumu (KVKK). (2025). *Üretken Yapay Zekâ ve Kişisel Verilerin Korunması Rehberi (15 Soruda).* Tech. rep., Kişisel Verileri Koruma Kurumu, Ankara. https://www.kvkk.gov.tr/SharedFolderServer/CMSFiles/MTY5MjNmNmIwZWY3YTE.pdf adresinden alındı

Kızrak, M. A., Buluz, B., "Ozparlak, B. O., "Unsal, B., G"urzumar, D. D., Atalar, G. D., . . . Arslan, S. (2019). *Yapay Zek[a] Çağında Hukuk: .Istanbul, Ankara ve .Izmir Baroları Çalıştay Raporu.* Tech. rep., .Istanbul Barosu, Ankara Barosu, .Izmir Barosu, .Istanbul, Turkey. https://www.istanbulbarosu.org.tr/files/docs/yapay_zeka_caginda_hukuk2019.pdf adresinden alındı

Lin, B., Wang, S., Chen, L., & Mao, X. (2025). Exploring the Security Threats of Knowledge Base Poisoning in Retrieval-Augmented Code Generation. *Exploring the Security Threats of Knowledge Base Poisoning in*

Retrieval-Augmented Code Generation. arXiv. doi:10.48550/ARXIV.2502.03233

Liu, Y., Cui, Y., & Zhang, H. (2025). RRTL: Red Teaming Reasoning Large Language Models in Tool Learning. *RRTL: Red Teaming Reasoning Large Language Models in Tool Learning*. arXiv. doi:10.48550/ARXIV.2505.17106

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., . . . Zoph, B. (2023). GPT-4 Technical Report. *GPT-4 Technical Report*. arXiv. doi:10.48550/ARXIV.2303.08774

Pathade, C. (2025). Red Teaming the Mind of the Machine: A Systematic Evaluation of Prompt Injection and Jailbreak Vulnerabilities in LLMs. *Red Teaming the Mind of the Machine: A Systematic Evaluation of Prompt Injection and Jailbreak Vulnerabilities in LLMs*. arXiv. doi:10.48550/ARXIV.2505.04806

Roy, D., Zhang, X., Bhave, R., Bansal, C., Las-Casas, P., Fonseca, R., & Rajmohan, S. (2024). Exploring LLM-based Agents for Root Cause Analysis. *Exploring LLM-based Agents for Root Cause Analysis*. arXiv. doi:10.48550/ARXIV.2403.04123

Shi, D., Shen, T., Huang, Y., Li, Z., Leng, Y., Jin, R., . . . Xiong, D. (2024). Large Language Model Safety: A Holistic Survey. *arXiv.org*. doi:10.48550/ARXIV.2412.17686

STB, & CDDO. (2021). *Ulusal Yapay Zekâ Stratejisi 2021-2025*. T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. Ankara: T.C. Sanayi ve Teknoloji Bakanlığı ve T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. https://cbddo.gov.tr/SharedFolderServer/Genel/File/TR-UlusalYZStratejisi2021-2025.pdf adresinden alındı

STB, & CDDO. (2024). *Ulusal Yapay Zekâ Stratejisi 2024-2025 Eylem Planı.* T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. Ankara: T.C. Sanayi ve Teknoloji Bakanlığı and T.C. Cumhurbaşkanlığı Dijital Dönüşüm Ofisi. https://cbddo.gov.tr/SharedFolderServer/Genel/File/UlusalYapayZekaStratejisi2024-2025EylemPlani.pdf adresinden alındı

T.C. Kamu Görevlileri Etik Kurulu. (2024). 2024/108 Sayılı İlke Kararı (Yapay Zekâ Sistemlerinin Kullanımında Kamu Görevlilerinin Uyması Gereken Etik Davranış İlkeleri). *2024/108 Sayılı İlke Kararı (Yapay Zekâ Sistemlerinin Kullanımında Kamu Görevlilerinin Uyması Gereken Etik Davranış İlkeleri)*. Ankara, Turkey. 2025 tarihinde https://www.etik.gov.tr/icerikler/2024-108-sayili-ilke-karari-yapay-zeka-sistemlerinin-kullaniminda-kamu-gorevlilerinin-uymasi-gereken-etik-davranis-ilkeleri/ adresinden alındı

Tabassi, E. (2023, January). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* National Institute of Standards and Technology (U.S.). doi:10.6028/nist.ai.100-1

Tanyel, T., ALKURDI, B. E., & AYVAZ, S. E. (2024, November). Developing Linguistic Patterns to Mitigate Inherent Human Bias in Offensive Language Detection. *Turkish Journal of Electrical Engineering and Computer Sciences, 32*, 829–848. doi:10.55730/1300-0632.4105

TSE GLOBAL. (2025, December). *ISO 42001 Yapay Zekâ Yönetim Sistemi Belgelendirmesi*. ISO 42001 Yapay Zekâ Yönetim Sistemi Belgelendirmesi: https://www.tseglobal.com.tr/sistem-belgelendirme?num=38?lg=tr adresinden alındı

Vassilev, A. (2025). *Adversarial Machine Learning:: A Taxonomy and Terminology of Attacks and Mitigations.* National Institute of Standards and Technology. doi:10.6028/nist.ai.100-2e2025

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., . . . Gabriel, I. (2021). Ethical and social risks of harm from Language Models. *Ethical and social risks of harm from Language Models*. arXiv. doi:10.48550/ARXIV.2112.04359

Wu, F., Zhang, N., Jha, S., McDaniel, P., & Xiao, C. (2024). A New Era in LLM Security: Exploring Security Concerns in Real-World LLM-based Systems. *A New Era in LLM Security: Exploring Security Concerns in Real-World LLM-based Systems*. arXiv. doi:10.48550/ARXIV.2402.18649

Yao, H., Shi, H., Chen, Y., Jiang, Y., Wang, C., & Qin, Z. (2025). ControlNET: A Firewall for RAG-based LLM System. *ControlNET: A Firewall for RAG-based LLM System*. arXiv. doi:10.48550/ARXIV.2504.09593

YÖK, Y. K. (2024, May). *Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Faaliyetlerinde Üretken Yapay Zekâ Kullanımına Dair Etik Rehber.* Tech. rep., Yükseköğretim Kurulu, Ankara. https://proje.yok.gov.tr/documentFiles/17539645334.Y%C3%BCksek%C3%B6%C4%9Fretimde%20%C3%BCretken%20yapay%20zeka%20kullan%C4%B1m%C4%B1-tr.pdf adresinden alındı

MAPPING THE RESEARCH LANDSCAPE OF DIGITAL TWIN AND CYBERSECURITY (2019–2025)

Özgür TONKAL[1]

1. INTRODUCTION

Rapid advances in digital technologies have led to the widespread adoption of the Digital Twin concept, a virtual representation of physical systems, across many sectors. Used in diverse fields such as industrial manufacturing, healthcare, and energy management, digital twin technology enables real-time monitoring, analysis, and optimization of systems. However, the constant data exchange between digital twins and physical systems makes them vulnerable to cyber threats and increases their security vulnerabilities (Alcaraz & Lopez, 2022).

The need to ensure the security of digital twin systems has led to a rapid increase in research conducted at the intersection of this technology and cybersecurity. However, the literature demonstrates a limited number of analyses systematically addressing the structural trends, academic collaboration networks, and thematic developments of studies conducted at the intersection of these two fields.

Digital twin technologies, which serve as real-time virtual representations of physical systems, are being increasingly utilised in critical infrastructures, industrial automation systems, and manufacturing processes. The constant data flow, cloud-

[1] Asst. Prof., Samsun University, Faculty of Engineering and Natural Sciences, Department of Software Engineering, ORCID: 0000-0001-7219-9053.

based connectivity, and direct interaction of these technologies with sensitive operational information make them vulnerable to cyberattacks. This has increased academic interest in the security and resilience of digital twin systems, particularly in the context of cybersecurity, and has accelerated efforts to protect these systems (Alcaraz & Lopez, 2022; Sun, Zhang, & Zhu, 2024).

In recent years, studies on the integration of digital twin technologies with cybersecurity have been primarily related to areas such as Industry 4.0, smart manufacturing, the Internet of Things (IoT), and blockchain Technologies (S. A. et al., 2023). Balta et al. (2024) addressed cybersecurity risks in smart manufacturing (SM) systems in their study. They proposed a DT framework to detect cyberattacks on cyber-physical manufacturing systems (CPMS) and expected anomalies in physical processes. Böhm et al. (2021), on the other hand, emphasize the growing industrial use of the Internet of Things and Cyber-Physical Systems, emphasizing the importance of augmented reality (AR) and DT solutions against cybersecurity risks. Furthermore, Mohsin et al. (2023) discussed how digital twins can be integrated into security operations using “shift left” and “shift right” approaches.

There has been a limited number of systematic literature studies on cybersecurity and DT. Alhamam et al. (2025) comprehensively address the cybersecurity aspects of DT technologies in their study, examining the current status, challenges, and future research directions. A systematic literature review using the PRISMA method was conducted to analyze 74 selected studies. Digital twins were examined at the component, process, asset, system, and network levels, and cyber threats encountered at each level were identified. The findings indicate that DTs contribute to security, resilience, and operational efficiency in various sectors, but are also exposed to serious threats such as unauthorized access, data breaches, and DDoS.

Lampropoulos et al. (2024) used a literature review, bibliometric analysis, and scientific mapping methods to examine the use of digital twin technology in critical infrastructures. A total of 3,414 documents obtained from the Scopus and Web of Science (WoS) databases were analyzed. The findings demonstrate that digital twins play a significant role in improving the security, resilience, reliability, maintenance, continuity, and operation of critical infrastructures across all sectors. Additionally, it has been demonstrated that digital twins can provide advanced capabilities such as intelligent and autonomous decision-making, process optimization, enhanced traceability, interactive visualization, real-time monitoring, analysis, and prediction. Furthermore, it has been demonstrated that digital twins can serve as a bridge between physical and virtual environments, contributing to the improvement of cybersecurity in critical infrastructure through continuous, real-time monitoring.

However, these technically focused approaches in the literature are often limited to specific solution proposals and do not include higher-level structural analyses such as academic collaborations, citation structures, and thematic clusters. For example, in a study by Homaei et al. (2024) on AI-enabled digital twin applications in the context of cybersecurity, although technical classifications are included, there is no mapping or structural analysis at the bibliometric level. Similarly, Empl et al. (2025) addressed the role of digital twins in cybersecurity operations at a conceptual level but did not analyze publication trends and scientific clusters in the literature.

In this context, it is clear that systematic bibliometric analyses examining academic production at the intersection of digital twin and cybersecurity concepts, in light of parameters such as publication trends by year, keyword clusters, co-citation structures, and international academic collaborations, would fill a

significant gap. Indeed, even the comprehensive 2022 review by Alcaraz & Lopez (2022) offers an analysis framework that lacks detailed bibliometric structures. A similar lack of structural orientation is evident in studies such as Homaei et al. (2024) and Otoom (2025).

Therefore, this study aims to systematically analyze publications at the academic intersection of digital twin and cybersecurity, both filling the gap in the literature and providing guiding findings for researchers and decision-makers.

This study analyzes 360 academic publications on the themes of “Digital Twin” and “Cybersecurity” published in the Web of Science database between 2019 and 2025 using bibliometric methods. The analysis includes: The aim was to visualize publication trends, author collaborations, keyword clusters, citation relationships, and thematic structures in the literature. The findings not only reveal the current state of the research field but also provide guidance for future academic studies.

2. MATERIALS AND METHODS

2.1. Data Analysis Used

The bibliometric dataset used in this study was obtained through a custom query conducted through the Web of Science Core Collection database. The query expression was structured to include both “Digital Twin” and “Cybersecurity” fields, and the query used was as follows:

TS = (“Digital Twin” OR “Digital Twins”) AND TS = (“Cybersecurity” OR “Cyber Security” OR “Information Security” OR “Network Security” OR “Attack Detection”)

This query only considered journal articles published between 2019 and 2025, yielding a total of 360 scientific publications.

Fig.1. Searched journals list

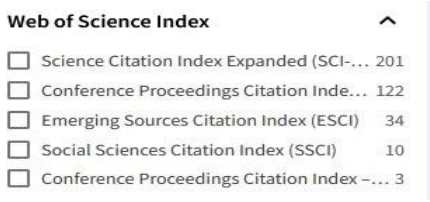
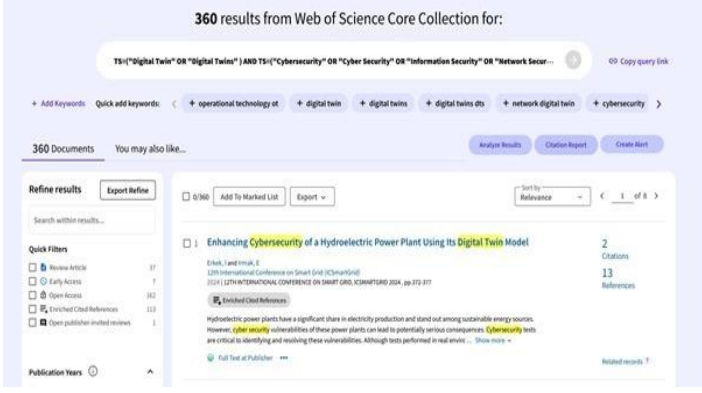


Fig. 2. Post-scan image



(Source: https://www.webofscience.com/wos/woscc/basic-search) (assessed on 15th May 2025)

The resulting publications were exported in two different formats to ensure robust bibliometric analysis. First, data preprocessing, filtering, and graphical analyses were performed using Microsoft Excel (.xlsx). Secondly, to enable detailed bibliometric analyses with VOSviewer software, the data were exported in Tab Delimited (.txt) format. This enabled the following analyses:

- Publication Trend Analysis By Year
- Journal (Source) Analysis

- Author Productivity And Citation Performance
- Co-Authorship Network Analysis
- Keyword Clustering Analysis
- Co-Citation Analysis
- Bibliographic Coupling Analysis

These analyses provided a comprehensive data base for revealing the structural characteristics of the research field and identifying future scientific trends.

2.2. Data Fields Used with VOSviewer

The VOSviewer software used in this study is a competent tool for visualizing structural relationships in the literatüre (van Eck & Waltman, 2010). The main data fields used in the analyses are as follows:

Table I. Vosviewer Data Fields and Their Purposes

Data Field	Purpose of Use
Authors / Author Full Names	co-authorship analysis to identify collaboration between authors.
Author Keywords / Keywords Plus	co-word analysis to determine thematic focuses and research trends.
Source Title	analyze the most productive journals and their citation connections.
Affiliations / Addresses	map institutional and international collaboration networks.
Cited References	co-citation and bibliographic coupling analysis to reveal knowledge structures between publications.
Times Cited	evaluate the impact level of the most cited publications.
Publication Year	analyze the temporal distribution of publication trends between 2019 and 2025.
DOI	a unique identifier to ensure reference accuracy and accessibility.

Through these fields, the study established multifaceted bibliometric relationships, and scientific structures were clearly analyzed using visual network maps provided by VOSviewer.

2.3. Bibliometric Analysis Tool and Data Source

In this study, VOSviewer software was chosen to effectively conduct bibliometric analyses. This choice was driven by VOSviewer's powerful analysis and visualization capabilities, such as visualizing structural relationships in the literature, mapping conceptual clusters, and uncovering collaboration networks. The software enables multidimensional and interactive analysis of research developments, collaborations among authors, and relationships between key concepts.

The Web of Science (WoS) Core Collection database was used as the data source. With its advanced search capabilities, detailed filtering options, and high publication standards, Web of Science offers a reliable and comprehensive resource for bibliometric studies. Furthermore, its broad interdisciplinary literature coverage provides a suitable environment for compiling data on multidisciplinary topics such as Digital Twins and Cybersecurity.

A total of 360 scientific publications obtained through the searches constituted the core dataset for the bibliometric analyses conducted in this study. These publications were examined in detail, including:

- Conceptual networks,
- Author, institutional, and country collaborations,
- Keyword clusters,
- Citation and source links,
- Research trends,

and a structural map of the field was presented.

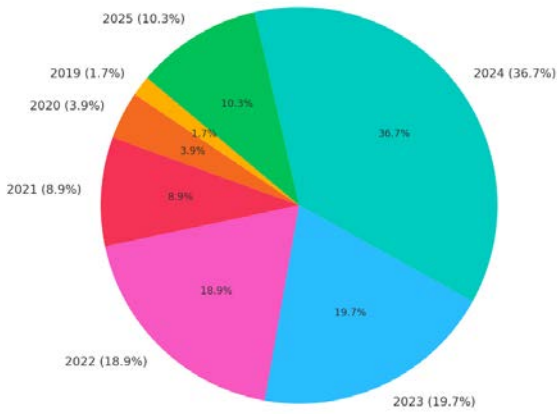
3. RESULTS AND DISCUSSION

The findings obtained in the study are given under this heading.

3.1. Number of Publications by Year

Examining the number of publications by year is crucial for revealing the dynamics of development in a research field over time and trends of increasing or decreasing scientific interest. Such analyses help understand the periods in which academic productivity in a given field is concentrated and the research activity that parallels technological advancements. Figure 3 below shows the percentage distribution of publications on the themes of "Digital Twin" and "Cybersecurity" by year between 2019 and 2025.

Fig. 3. The distribution of publication numbers by year



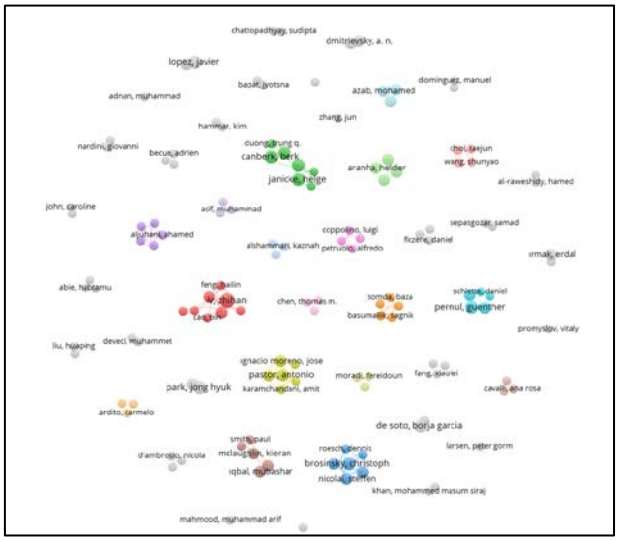
Interest in the field, which was limited to only 1.7% of publications in 2019, has shown a steady increase in the following years. Significant increases were particularly seen in 2022 (18.9%) and 2023 (19.7%), reaching a peak of 36.7% in 2024. This demonstrates that the subject is gaining increasing

importance in academic circles and has gained a central place on the research agenda.

Although the 2025 data remained at 10.3% because it was collected before the end of the year, even this rate demonstrates that interest is sustainable. The results demonstrate that the Digital Twin and Cybersecurity fields are beginning to be evaluated in an integrated manner, and that productivity in the literature continues to increase in this context. The increase in publications over the last three years (2022–2024) indicates that this research theme has become a rising field in the literature, and more academic contributions are expected in the future. This trend holds significant potential for the emergence of new research topics and the promotion of interdisciplinary research.

3.2. Author Productivity Analysis

Fig. 4. Author productivity analysis map



(Source: https://www.vosviewer.com)

In the collaboration network Figure 4., each node represents an author, and the lines between the nodes represent co-publication relationships between authors. Clusters of the

same color indicate groups of authors who frequently collaborate. Regions with dense clusters on the map reflect stronger academic collaborations.

From the analysis, Lv Zhihan emerges as the most productive author, with a total of five publications (Lv, 2020; Lv et al., 2022; Lv et al., 2023a; Lv et al., 2023b; Feng et al., 2024). He is followed by Christoph Brosinsky and Antonio Pastor, each with four publications (Brosinsky et al., 2019, 2020, 2021, 2024; Pastor et al., 2022, 2024). However, author impact should not be assessed solely based on publication count; the number of citations received by these publications is also a critical indicator of scholarly influence.

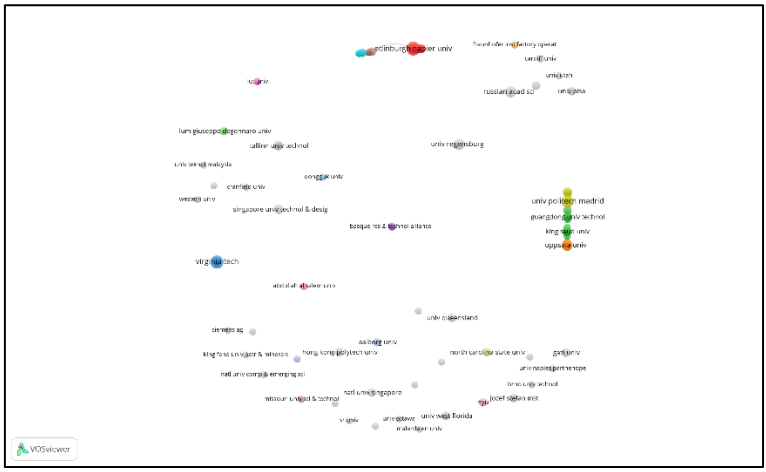
In this context, Lopez and Javier are the most highly cited authors, receiving 171 citations with only 4 publications. Similarly, Alcaraz and Cristina have achieved 171 citations with 3 publications, making a significant impact. Furthermore, authors like Teisserenc, Benjamin, Sepasgozar, and Samad, who have received 127 citations with 2 publications, have also gained significant visibility in the literature despite their small number of publications. These findings demonstrate the position of authors who are both productive and high-impact in the research field. The work of highly cited authors, particularly those with a small number of publications, is among the primary sources cited in the literature and guides the development of the field.

3.3. Country and Institution Analysis

The network map in Figure 5 visualizes scientific collaboration relationships between institutions working in the fields of digital twins and cybersecurity. Each node represents an institution, and connections between nodes indicate collaboration based on shared publications. Colored clusters represent groups of universities working together.

According to the analysis findings, Edinburgh Napier University and Virginia Tech are among the most productive institutions, with 7 publications each. Universidad Politécnica de Madrid and Queens University Belfast follow with 6 each. Another notable institution is Politecnico di Milano, with only 4 publications, yet demonstrating a high level of impact with 140 citations.

Fig. 5. Institution analysis map

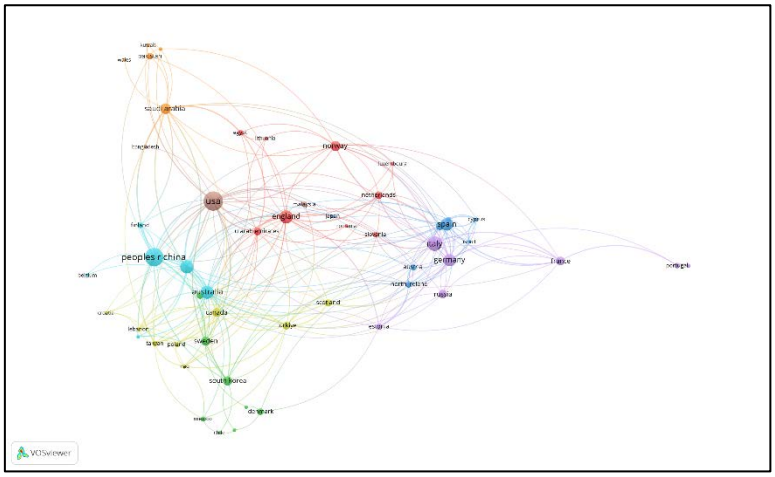


Furthermore, some institutions, such as Lebanese American University (163 citations) and Hong Kong Polytechnic University (123 citations), have achieved high citation counts, yet they appear to have limited connections within their collaboration networks. This suggests that these institutions produce impactful work individually, but their institutional collaborations are less frequent. When assessed in terms of network connectivity, Edinburgh Napier University, with a total network connectivity of 11, is central to the network and demonstrates its active collaboration with many institutions. Collaborations are generally observed to occur more frequently among institutions within the same cluster. This analysis contributes to the evaluation of

research activities conducted at the institutional level in terms of both productivity and interaction.

An examination of the distribution of publications by country, as shown in Figure 6, reveals that the United States produces the most publications. The United States tops the list with 59 publications, followed by China with 53. These two countries stand out significantly in terms of research productivity. Italy ranks third with 34 publications, followed by Australia (29), India (28), the United Kingdom (27), Spain (25), and Germany (23).

Fig. 6. Country analysis map



These data suggest that research activities are largely centered in North America, Asia, and Europe. Overall, the contribution of these countries to scientific publications drives the development of the field and reflects a significant global distribution in academic productivity.

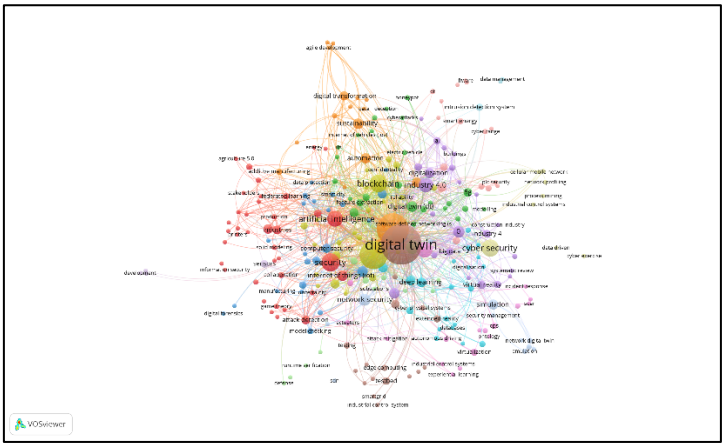
3.4. Keyword Analysis

Figure 7 shows a map of the keywords used in the studies. When we look at the prominent concepts in the keyword analysis, we see that “digital twin” is by far the most frequently repeated

term (152 times). The concepts surrounding this term reveal that the study generally focuses on digital twin technology.

Terms such as “cybersecurity” (95) and “digital twins” (74) are among the notable topics in the literature focusing on the security aspects and plural uses of digital twins. Furthermore, terms such as “security,” “blockchain,” “internet of things” (IoT), and “artificial intelligence” are also prominent with high repetitions. This demonstrates that digital twin technology is closely related to fields such as cybersecurity, artificial intelligence, and the internet of things.

Fig. 7. Keyword analysis map



The keyword map generally indicates that the concept of digital twin is used in a multidisciplinary context: This reflects the research conducted across a wide range of applications, from security and artificial intelligence to Industry 4.0 and cyber-physical systems. This diversity reveals the technical depth and breadth of scope of the work in the field.

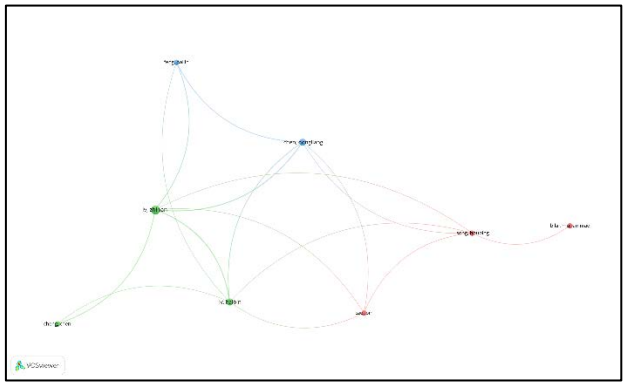
3.5. Co-Authorship Analysis

The co-authorship network visualizes the collaborative relationships between authors publishing together on digital twins and cybersecurity (Figure 8). Each node represents an author, and

the connecting lines represent joint publications. The colored clusters on the map represent groups of authors working together.

When examining the publication productivity and impact among authors, the most cited authors are Lopez, Javier, and Alcaraz, Cristina. Despite each having 4 and 3 publications, respectively, they have achieved significant visibility in the literature with 171 citations. This demonstrates that their publications are heavily cited and represent strong contributions to the field. Furthermore, authors with high impact include Sepasgozar, Samad, and Teisserenc, Benjamin, who have received 127 citations with only 2 publications. These examples demonstrate that academic impact is not only related to the number of publications but also to the quality of the publications and their resonance within the field.

Fig. 8. Co-Authorship map

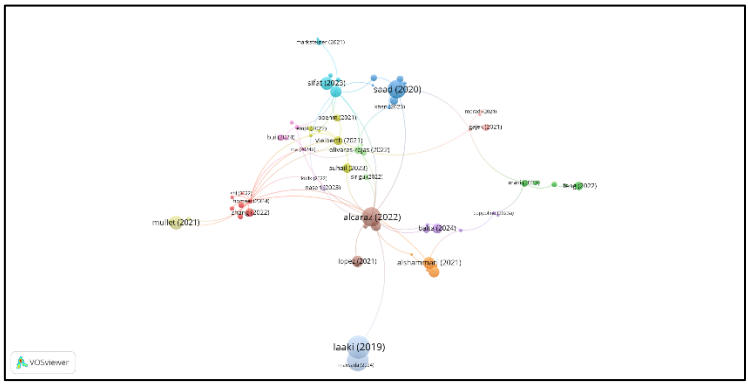


In terms of publication count, Lv, Zhihan stands out as the most productive author with 5 publications. He is followed by authors such as Park, Jong Hyuk, Pernul, Guenther and De Soto, and Borja Garcia, each with four publications. Overall, this network structure reveals the positions of both productive and influential authors in the field's literature, while also visualizing how highly cited publications interact with collaborative relationships.

3.6. Citation Analysis

The network map in Figure 9 visualizes the citation relationships established between publications on digital twins and cybersecurity. Each node represents a publication, and the lines between the nodes indicate the citations these publications make to each other. Centrally located publications are those that are more frequently cited in the literature and establish strong connections to other studies.

Fig. 9. Most cited publications map

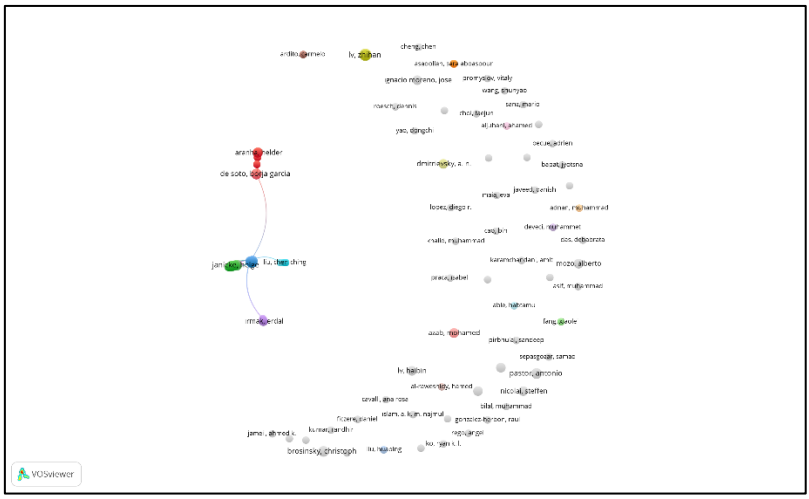


According to the analysis results, the most centrally located study in the network is Alcaraz (2022), which has high visibility in the literature due to its numerous citations from various publications. This demonstrates that Alcaraz's work stands out as a reference source in the research field and plays a significant role in the flow of knowledge.

Other important publications such as Laaki (2019), Saad (2020), and Sifat (2023) also have extensive connection networks. These publications are linked to various working groups through citations and contribute to key discussion points in the literature. Furthermore, publications such as Shi (2022), Zhang (2022), and Homayoun (2024) appear to be grouped within specific subclusters and represent current research focuses. This suggests that these studies focus on more niche topics but still

hold influential positions in the literature. Overall, the network structure reveals that the flow of scientific knowledge generated through citation is largely centered around the Alcaraz (2022) publication, which establishes connections with both previous and subsequent studies. These findings are important for identifying where knowledge centers are formed in the literature.

Fig. 10. Most cited authors and connections map



As shown in Figure 10, researchers Alcaraz, Cristina, and Lopez Javier are among the most cited authors in the literature with 171 citations. Both authors, with 29 connections, appear to hold a central position in the network.

Authors like Teisserenc, Benjamin, and Sepasgozar, Samad, on the other hand, achieved a remarkable level of influence, receiving 127 citations with just two publications. This demonstrates that even a small number of publications can have a high impact.

While some authors are positioned independently of their collaborations, concentrations and strong co-citation relationships are observed in certain groups. Overall, it appears that there may not be a direct correlation between citation count

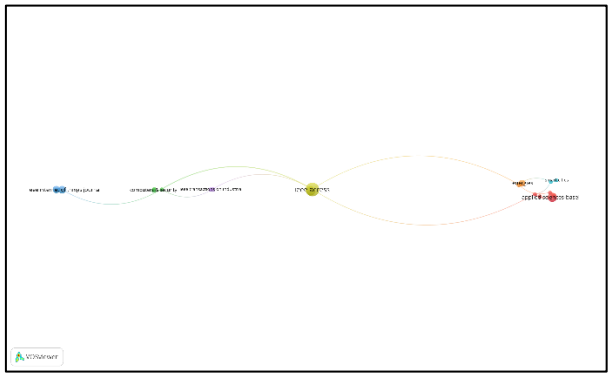
and collaboration network position, but both dimensions are important for academic impact.

3.7. Citation Analysis

In terms of publication count, IEEE Access stands out as the journal hosting the largest number of publications, with 21. Applied Sciences – Basel and Sensors follow with 10 each.

As Figure 11 shows, IEEE Access is at the center of the network not only in terms of publication count but also in terms of its strong content and citation connections with other journals. It particularly interacts heavily with journals such as Energies, Computers & Security, and IEEE Internet of Things Journal.

Fig. 11. Journal analysis map



These findings indicate that publications are largely clustered in electronics, energy, and cybersecurity-themed journals, and that IEEE Access plays a central role in this field.

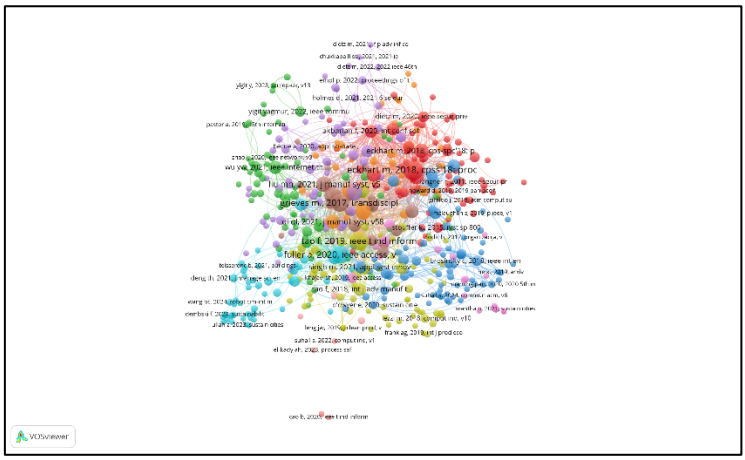
3.8. Co-Citation Analysis

Figure 12 shows the Co-Citation Analysis (References) map. According to the co-citation analysis, the most frequently cited study in the literature is Eckhart, M. (2018), with 33 co-citations. Tao, F. (2019) and Fuller, A. (2020) follow with 31 and 27 citations, respectively. These publications, frequently cited

together by other studies, have become essential resources in the field.

Figure 12 shows that these publications are positioned at the center of the network, acting as bridges between different clusters. Authors such as Grieves, M. (2017) and Qi, Q. (2021) are particularly notable for their high citation count and multiple affiliations.

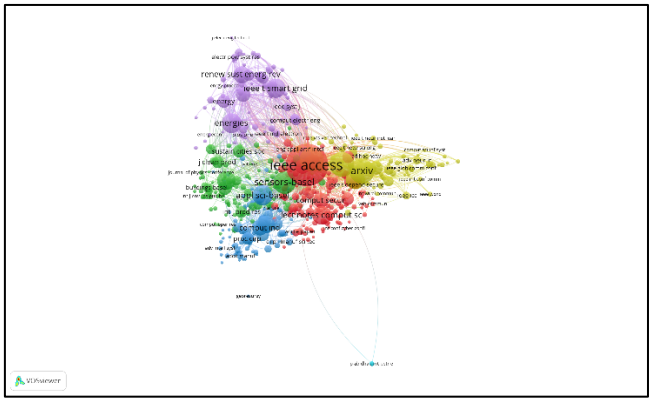
Fig. 12. Co-Citation analysis (references) map



In general, this structure shows that studies on digital twin, industry 4.0, production systems and cybersecurity are concentrated around certain references and these publications have high visibility in the interdisciplinary literature.

According to the co-citation analysis presented in Figure 13, IEEE Access is by far the most frequently cited source in the literature, sitting at the center of the network with 862 citations. This journal is followed by IEEE Internet of Things Journal (303), arXiv (297), and IEEE Transactions on Industrial Informatics (291).

Figure 13. Co-Citation analysis (source)map

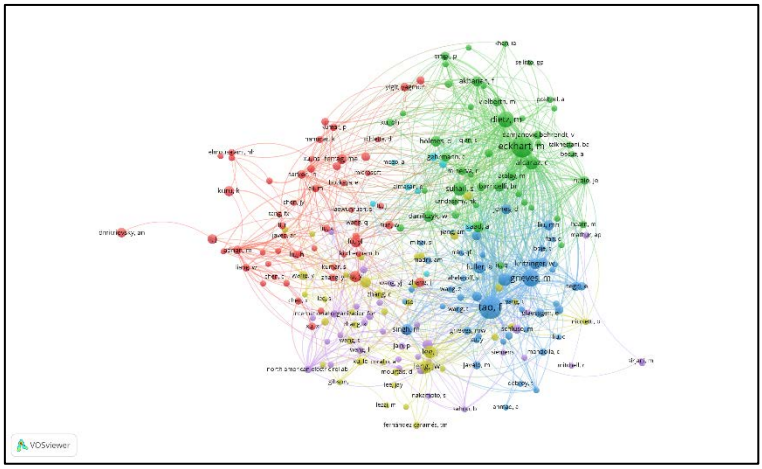


The network visualization reveals that IEEE Access connects to numerous journals clustered in different colors, acting as a bridge between research areas. Other frequently cited key publications include Sensors-Basel, Applied Sciences-Basel, Computers & Security, Lecture Notes in Computer Science, and Energies.

Generally, cited sources are concentrated in disciplines such as electronics, energy, information security, and manufacturing systems. This suggests that the knowledge bases of interdisciplinary studies are converging in specific high-impact journals.

According to the co-citation network among authors shown in Figure 14, the most frequently cited authors are Tao, F. (104 citations) and Eckhart, M. (92 citations). These two authors are at the center of the network and have published work on digital twins, manufacturing systems, and Industry 4.0. They are followed by researchers such as Grieves, M. (57) and Dietz, M. (51). Furthermore, authors such as Alcaraz, C., Suhail, S., Qi, Q., Lee, J., and Leng, J.W. also demonstrate significant influence, receiving over 30 citations.

Fig. 14. Co-Citation analysis (authorship) map



The network structure shows that these authors are largely located in dense clusters and are cited together in similar thematic areas. The colored clusters reveal that digital twin research is grouped by subfields such as security, manufacturing systems, and smart factories.

This analysis shows that knowledge structures in the literature are shaped around certain sets of authors and that these authors play an important role in determining the direction of the field.

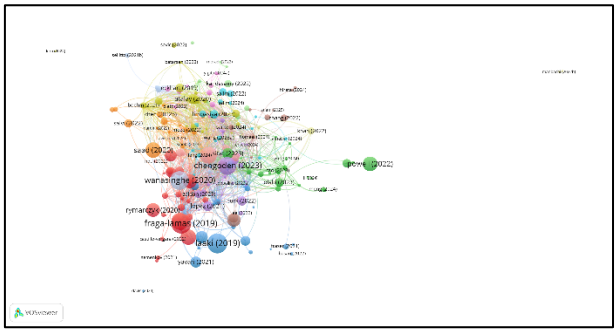
3.10. Bibliographic Coupling

According to the bibliographic matching analysis, strong relationships were observed between publications citing similar sources (Figure 15). Studies such as Fraga-Lamas (2019), Laaki (2019), Chengoden (2023), and Powell (2022), in particular, are positioned at the center of the network, sharing references with numerous publications.

The clusters formed in the network indicate a clustering of publications with similar content. This structure suggests that the studies are based on common knowledge bases and that

researchers frequently draw on the same sources. Overall, this analysis allows us to identify theoretical affinities among publications in the literature and visualize which central publications shape research topics.

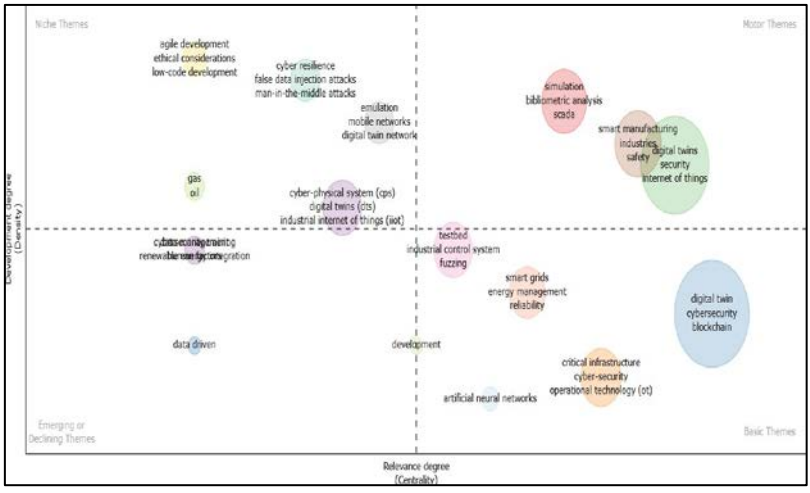
Fig. 15. Inter-Publication bibliographic matching network map



3.11. Conceptual Map (Thematic Map)

The map shows that the strongest areas for digital twins and cybersecurity are smart manufacturing and security applications. Furthermore, blockchain, the internet of things, and data-driven solutions also present significant opportunities.

Fig. 16. Inter-Publication bibliographic matching network map



In the thematic map presented in Figure 16, studies in the Digital Twin and Cybersecurity field are divided into four regions according to their intensity and centrality levels:

Motor Themes (Top Right): Among the most intensively studied topics in the literature are smart manufacturing, digital twins, security, and the internet of things. These themes constitute the driving force of the research field.

Core Themes (Bottom Right): Digital twin, cybersecurity, and blockchain are themes considered highly central and fundamental, serving as the starting point for many studies.

Niche Themes (Top Left): While topics such as agile development, cyber resilience, and low-code development have received limited research, they stand out as deepened and focused themes in specific subfields.

Emerging or Declining Themes (Bottom Left): Topics such as data-driven and renewable energy integration represent areas of either emerging interest or declining interest.

This map sheds light on the current state of the field as well as future research opportunities, demonstrating that smart manufacturing and cybersecurity applications are particularly central.

4. CONCLUSIONS AND RECOMMENDATIONS

Within the scope of this study, academic publications on the themes of “Digital Twin” and “Cybersecurity” between 2019 and 2025 were analyzed using bibliometric methods. A dataset of 360 publications obtained from the Web of Science database was processed with VOSviewer software to visualize structural relationships, thematic clusters, and collaboration networks in the literature.

The main findings can be summarized as follows:

- An increasing publication trend was observed over the years, with a significant density in the literature, particularly between 2022 and 2024.
- Authors such as Lv Zhihan (number of publications) and Lopez Javier (number of citations) stood out.
- The US and China led the way in publication production, while Edinburgh Napier University and Virginia Tech attracted attention at the institutional level.
- IEEE Access played a central role in both the number of publications and co-citation links.
- Among the keywords, “digital twin,” “cybersecurity,” “IoT,” and “blockchain” were the most frequently encountered.
- Co-citation and bibliographic matching analyses revealed key publications and clusters with content similarities in the literature.
- The thematic map was effective in distinguishing emerging, basic, and niche research areas within the literature.

Based on the analysis results, it is possible to make some recommendations for future research. First, specific themes such as agile development and low-code platforms, which have been the subject of limited studies in the literature, have the potential to yield new contributions through in-depth investigation. Furthermore, international collaborations with countries with high academic productivity, such as the US, China, and Germany, can increase the visibility and citation impact of these studies. Current themes such as smart manufacturing, cybersecurity, and IoT, in particular, provide a suitable basis for multidisciplinary approaches, offering researchers a wide range of application areas. Furthermore, closely monitoring emerging themes such as data-

driven systems is crucial to identifying new research opportunities that may emerge at the intersection of digital twins and cybersecurity. Studies developed within this framework will both contribute to the academic literature and yield tangible outcomes for industrial applications.

As a result, this study aims to provide guiding insights to researchers and policymakers by comprehensively analyzing the academic production in the fields of Digital Twin and Cybersecurity.

REFERENCES

Alcaraz, C., & Lopez, J. (2022). Digital twin: A comprehensive survey of security threats. *IEEE Communications Surveys & Tutorials, 24*(3), 1475–1503. doi:10.1109/COMST.2022.3171465

Alhamam, N., Hafizur Rahman, M. M., & Aljughaiman, A. (2025). A comprehensive review on cybersecurity of digital twins: Issues, challenges, and future research directions. *IEEE Access, 13*, 45106–45124. doi:10.1109/ACCESS.2025.3545004

Balta, E. C., Pease, M., Moyne, J., Barton, K., & Tilbury, D. M. (2024). Digital twin-based cyber-attack detection framework for cyber-physical manufacturing systems. *IEEE Transactions on Automation Science and Engineering, 21*(2), 1695–1712. doi:10.1109/TASE.2023.3243147

Böhm, F., Dietz, M., Preindl, T., & Pernul, G. (2021). Augmented reality and the digital twin: State-of-the-art and perspectives for cybersecurity. *Journal of Cybersecurity and Privacy, 1*(3), 519–538. doi:10.3390/jcp1030026

Brosinsky, C., Naglič, M., Lehnhoff, S., Krebs, R., & Westermann, D. (2024). A fortunate decision that you can trust: Digital twins as enablers for the next generation of energy management systems and sophisticated operator assistance systems. *IEEE Power & Energy Magazine, 22*(1), 24–34. doi:10.1109/MPE.2023.3330120

Empl, P., Koch, D., Dietz, M., & Pernul, G. (2025). Digital twins in security operations: State of the art and future perspectives. *ACM Computing Surveys*. Advance online publication. doi:10.1145/3746279

Feng, H., Chen, D., & Lv, Z. (2022). Blockchain in digital twins-based vehicle management in VANETs. *IEEE Transactions on Intelligent Transportation Systems, 23*(10), 19613–19623. doi:10.1109/TITS.2022.3202439

Feng, H., Chen, D., Lv, H., & Lv, Z. (2024). Game theory in network security for digital twins in industry. *Digital Communications and Networks, 10*(4), 1068–1078. doi:10.1016/j.dcan.2023.01.004

Homaei, M., Mogollón-Gutiérrez, Ó., Sancho, J. C., Ávila, M., & Caro, A. (2024). A review of digital twins and their application in cybersecurity based on artificial intelligence. *Artificial Intelligence Review, 57*(8), 201. doi:10.1007/s10462-024-10805-3

Karamchandani, A., et al. (2024). Design of an AI-driven network digital twin for advanced 5G–6G network management. In *Proceedings of the IEEE Network Operations and Management Symposium (NOMS 2024)* (pp. 1–7). doi:10.1109/NOMS59830.2024.10575106

Kummerow, A., Nicolai, S., Brosinsky, C., Westermann, D., Naumann, A., & Richter, M. (2020). Digital-twin based services for advanced monitoring and control of future power systems. In *Proceedings of the IEEE Power & Energy Society General Meeting* (pp. 1–5). doi:10.1109/PESGM41954.2020.9354468

Kummerow, A., Rosch, D., Nicolai, S., Brosinsky, C., Westermann, D., & Naumann, A. (2021). Attacking dynamic power system control centers: A cyber-physical threat analysis. In *Proceedings of the IEEE Innovative Smart Grid Technologies Conference* (pp. 1–5). doi:10.1109/ISGT49243.2021.9372285

Kummerow, A., et al. (2019). Challenges and opportunities for phasor data based event detection in transmission control centers under cyber security constraints. In *Proceedings of the IEEE Milan PowerTech* (pp. 1–6). doi:10.1109/PTC.2019.8810711

Lampropoulos, G., Larrucea, X., & Colomo-Palacios, R. (2024). Digital twins in critical infrastructure. *Information, 15*(8), 454. doi:10.3390/info15080454

López-Brea, J. T., et al. (2024). Dynamic deployment and security assessment of resilient services over digital twins. In *Proceedings of the Joint European Conference on Networks and Communications & 6G Summit* (pp. 836–841). doi:10.1109/EuCNC/6GSummit60053.2024.10597070

Lv, Z. (2020). Interaction of edge-cloud computing based on SDN and NFV for next generation IoT. *IEEE Access, 7*(7), 5706–5712.

Lv, Z., Cheng, C., & Lv, H. (2023). Blockchain-based decentralized learning for security in digital twins. *IEEE Internet of Things Journal, 10*(24), 21479–21488. doi:10.1109/JIOT.2023.3295499

Lv, Z., Cheng, C., Guerrieri, A., & Fortino, G. (2023). Behavioral modeling and prediction in social perception and computing: A survey. *IEEE Transactions on Computational Social Systems, 10*(4), 2008–2021. doi:10.1109/TCSS.2022.3230211

Mohsin, A., Janicke, H., Nepal, S., & Holmes, D. (2023). Digital twins and the future of their use enabling shift left and shift right cybersecurity operations. *arXiv preprint*. http://arxiv.org/abs/2309.13612

Mozo, A., Karamchandani, A., Gómez-Canaval, S., Sanz, M., Moreno, J. I., & Pastor, A. (2022). B5GEMINI: AI-driven network digital twin. *Sensors, 22*(11), 4106. doi:10.3390/s22114106

Otoom, S. (2025). Risk auditing for digital twins in cyber physical systems: A systematic review. *Journal of Cyber Security Risk Audit, 2025*(1), 22–35. doi:10.63180/jcsra.thestap.2025.1.3

Rebecchi, F., et al. (2022). A digital twin for the 5G era: The SPIDER cyber range. In *Proceedings of the IEEE WoWMoM* (pp. 567–572). doi:10.1109/WoWMoM54355.2022.00088

Sun, Z., Zhang, R., & Zhu, X. (2024). The progress and trend of digital twin research over the last 20 years: A bibliometrics-based visualization analysis. *Journal of Manufacturing Systems, 74*, 1–15. doi:10.1016/j.jmsy.2024.02.016

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. doi:10.1007/s11192-009-0146-3

YAPAY ZEKÂ DESTEKLİ SOHBET ROBOTLU REZERVASYON SİSTEMİ TASARIMI: KUAFÖRLERE ÖZGÜ UYGULAMA

Aslı SAĞLAM[1]

Onur GÖNÜLLÜ[2]

Durmuş Özkan ŞAHİN[3]

1. GİRİŞ

Bu çalışmanın amacı, geleneksel rezervasyon sistemlerinin aksine yapay zekâ desteği sunan sohbet robotu (chatbot) ile yenilikçi rezervasyon sistemi geliştirmektir. Bu sistem yapay zekâ desteği ile kullanıcıların sorduğu soruları doğru şekilde anlayıp anlamlı ve doğru yanıtlar vermeyi amaçlamaktadır. Ayrıca kullanıcıların, kendi isteklerine uygun, belirledikleri konumda veya almak istedikleri hizmetlere göre çıktılar oluşturulmaktadır. Böylece oluşturulan chatbot ile kullanıcılar sorularına cevap alabilmesinin yanında öneriler alıp kayıtlı işletmeler hakkında bilgi edinebileceklerdir. Bu doğrultuda rezervasyon sistemi üzerine bu chatbotun entegre edildiği bir web sitesi geliştirilmiştir. Web sitesi kuaförler üzerine odaklı olup yapılan chatbot kuaför rezervasyonlarına sorulabilecek sorular ile ince ayar yapılmıştır.

[1] Bilgisayar Mühendisi, Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, ORCID: 0009-0002-4017-2213.

[2] Bilgisayar Mühendisi, Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, ORCID: 0009-0005-6829-9808.

[3] Dr. Öğr. Üyesi, Ondokuz Mayıs Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, ORCID: 0000-0002-0831-7825.

Çalışma kapsamında, yalnızca yapay zekâ destekli bir rezervasyon sistemi geliştirilmemiş, aynı zamanda bu alana çeşitli özgün katkılar da sunulmuştur. Çalışmada, doğal dil işleme sürecinde Türkçe için özel olarak eğitilmiş BERTurk modeli kullanılarak Adlandırılmış varlık tanıma (Named Entity Recognition - NER) işlemleri gerçekleştirilmiş ve kullanıcıdan gelen verilerin daha doğru analiz edilmesi sağlanmıştır. Ayrıca, sistemin başarısını artırmak amacıyla, kuaför rezervasyonlarına özgü, manuel olarak oluşturulan özgün bir veri kümesi kullanılmıştır. Bu veri kümesi, gerçek kullanıcı senaryoları göz önünde bulundurularak hazırlanmış ve modelin alan odaklı çalışmasına olanak tanımıştır.

2. LİTERATÜR ARAŞTIRMASI

Chatbot, soruları cevaplayarak kullanıcılarla iletişim kuran, görevleri yerine getiren yazılımlardır. Joseph Weizenbaum tarafından 1966 yılında geliştirilen ELIZA, ilk chatbot örneği olarak kabul edilmektedir (Gökkaya ve İşcan, 2025). Chatbotlar günümüzde sesli iletişim, görsel analiz ve bağlamsal anlayış gibi karmaşık işlemleri yerine getirebilmektedir. Bu gelişimde Büyük Dil Modellerinin (Large Language Models - LLM) ve Transformer tabanlı modellerin katkısı oldukça büyüktür. Literatürde chatbotlar, müşteri hizmetleri, eğitim ve sağlık gibi birçok alanda kullanılmaktadır (Okonkwo ve Ade-Ibijola, 2021; Nirala vd., 2022). Chatbot üzerine yapılan bazı çalışmalar şöyle özetlenmiştir:

(Oliveira ve Matos, 2023) yaptıkları çalışmada yükseköğretim kurumunun web portalında öğrencilerle etkileşimi artırmayı ve gerektiğinde onlara bilgi vermeyi hedefleyen bir chatbot uygulamasını tanıtmaktadırlar. LLM'ler ve vektör veritabanları kullanarak veri yapıları işlemlerinde büyük bir başarı elde etmişlerdir. Bu çalışmada, web tabanlı uygulamalar

geliştirmek için kullanılan Streamlit, dil modellerini kullanarak uygulamalar geliştirmeye olanak sağlayan Langchain, vektör veritabanı hizmeti veren Pinecone, yapay zekâ ve dil modelleme için kullanılan OpenAI'nin GPT-3.5 (Generative Pre-trained Transformer) dil modeli, BERT, RoBERTa (Robustly Optimized BERT Approach) gibi transformer modellerini kullanarak metinleri vektör temsillerine dönüştüren SentenceTransformers, veri işleme ve yapısal olmayan verilerle çalışmak için kullanılan Unstructured teknolojileri kullanılmıştır.

(Toprak ve Rasheed, 2022) tarafından yapılan çalışmada kullanıcıların İstanbul'un belirli mekanları hakkında detaylı bilgiye, menülerine ve fotoğraflarına ulaşabildiği, gurmelerin bu restoranlar hakkında yazdıkları blog yazılarını okuyabildiği, yakındaki mekanları anlık konumlarına göre gösterebildiği ve kullanıcıların isteklerine göre en uygun mekanları bulabilen veya gidilecek mekan önerilerini sunabilen bir chatbot geliştirilmiştir. Doğal dil anlama (Natural Language Understanding - NLU) ve konuşma yönetimi bileşenlerini oluşturmak için Dialogflow çerçevesi kullanılmıştır. Mobil uygulama geliştirmek için Flutter/Dart, uygulamanın veritabanı ve bulut depolama çözümü olarak Firebase kullanılmıştır. Chatbot ile kullanıcının etkileşim kurması için NLP teknolojisi, bu etkileşimlerin daha akıllı hale gelmesi için ise makine öğrenmesi teknolojisinden yararlanılmıştır. Görev odaklı chatbot kullanılması önerilen chatbot destekli rezervasyon sistemi ile benzer özelliklere sahiptir.

(Albayrak vd., 2018) çalışmalarında yapay zeka tabanlı sohbet botlarının çalışma prensiplerini, kullanım alanları açıklamışlardır. Bunlara ek olarak ayrıca telekomünikasyon sektöründe bağış servisi için bir chatbot geliştirmişlerdir. Bu uygulama bağış işlemlerini kolaylaştırmak amacıyla Facebook Messenger platformu üzerinden gerçeklenmiştir. Chatbot'un kullanıcıyı en iyi şekilde anlaması ve yanıt vermesi için NLP,

NLU, Doğal dil üretimi (Natural Language Generation - NLG) ve makine öğrenimi gibi teknolojiler bir arada kullanılmıştır.

(Bratić vd., 2024) çalışmalarında öğrenim aşamasında materyal arama ve bulma zorluklarını göz önünde bulundurarak bir sistem önerisinde bulunmuştur. Bu zorluklara bir son vermek amacıyla tüm öğretim materyallerine kolay, hızlı ve güvenilir bir şekilde erişilmesini sağlayan mevcut bir LLM modeline dayalı bir chatbot geliştirilmiştir. Yapay zeka bir sorunun yanıtından emin olmadığında önceden belirlenmiş komutlara uyan, karmaşık sorgular için makine öğrenimini kullanan hibrit model tasarlanmıştır. Bu chatbot mimarisinde DialogFlow platformu ve veri tabanı işlemleri için Firebase teknolojileri kullanılmıştır. Bunlara ek olarak Transformer kütüphanesi, NLP, LLM ve Python programlama dili önerilen chatbotun temelini oluşturmaktadır. Kullanıcı dostu bir arayüz için çeşitli çerçeveler ve Tkinter teknolojisi bir araya getirilmiştir. Belirli testler sonucu chatbotun gerçek doğruluk oranı değerinin % 91.5 olduğu tespit edilmiştir.

(Işık ve Yağcı, 2020) tarafından yapılan bir çalışmada Uzun Kısa Süreli Bellek (Long Short-Term Memory - LSTM) ve Sıradan Sıraya (Sequence-to- Sequence - Seq2Seq) modelleri kullanarak bir Telegram chatbot uygulaması geliştirilmiştir. Çalışmanın asıl amacı, kullanıcıların sordukları sorulara doğru ve bağlama uygun yanıtlar üretebilen bir sistem tasarlamaktır. LSTM ve Seq2seq Modeli, konuşma geçmişini kullanarak gelecek yanıtları tahmin etmektedir. Telegram Bot, Uygulama Programlama Arayüzü (Application Programming Interface - API) ile kullanıcı arayüzü oluşturulmuştur. Çalışmada, veri kümesi olarak 1984 soru-cevap çiftinden oluşan “chatterbot” veri kümesi kullanılmıştır. Elde edilen doğruluk oranı %79’dur ve verilerin eğitim sayısı arttıkça doğruluk oranından artacağı belirtilmiştir. Kullanıcıların sorularını doğru anlamlandırmak için kelime gömme (embedding) yöntemleri kullanılmıştır.

(Kesarwani ve Juneja, 2023) öğrencilerin akademik ve idari sorularını hızlıca çözebilen bir chatbot önermektedir. Chatbotlar, öğrencilere ders, etkinlik, sınav hazırlığı gibi konularda bilgi sunarken, idari çalışanların iş yükünün azaltılması hedeflenmiştir. Önerilen modelde chatbot kullanıcının sorusuna yanlış cevap verirse kullanıcı "yanlış cevap" etiketi taşıyan bir düğmeyi seçerek yöneticiye geri bildirim gönderebilmektedir. Bu çalışmada chatbot geliştirme süreci için ChatterBot kütüphanesi, kullanıcı sorularına yanıt verebilmek için bulut tabanlı bir veri tabanı, kullanıcı etkileşimlerinden öğrenerek zamanla daha doğru yanıtlar verme yeteneği için makine öğrenmesi teknolojileri birlikte kullanılmıştır. Önerilen chatbot ile duygu analizi de yapılmaktadır. Bu sayede kullanıcı soruları olumlu, olumsuz veya nötr olarak başarıyla ayırt edilebilmektedir. Botun farklı türde sorulara doğru şekilde yanıt verme özelliği aktif öğrenme yoluyla geliştirilmektedir.

(Gemirter ve Goularas, 2021) tarafından yapılan çalışmada Türkçe dilinde bankacılık sektörüne özgü soru cevap chatbotu uygulaması sunulmuştur. Türkçe dilinin karmaşıklığından kaynaklanan sorunları çözmek amacıyla İngilizce'den Türkçe'ye çevrilmiş veri setleri kullanılarak BERT algoritması eğitilerek bir model hazırlanmıştır. Bu model eğitilirken ön eğitim aşamasında Wikipedia Türkçe kümeleri ve haber kümeleri kullanılmıştır. Ayrıca bankacılık sektörüne özgü dokümanlar kullanılarak model tarafından bankacılık özelindeki cümlelerin daha kolay anlaşılması sağlanmıştır. Oluşturulan bu model Açık Alan (Open Domain) soruları üzerinde test edildiğinde F-Score metriğine göre %67.07 oranında başarım göstermektedir. Buna ek olarak Kapalı Alan (Closed Domain - Bankacılık) sorularına cevap verme başarımı ise F-Score metriğine göre %79.01 olarak ölçülmüştür.

(Li vd., 2019) tarafından yapılan çalışmada otel rezervasyonlarında kullanılmak üzere gerçek dünya diyalogları

üreten yapay zeka tabanlı chatbot geliştirilmiştir. Bu bot Whatsapp Telegram gibi platformlara entegre edilebilir biçimdedir. Bot kullanıcıya tercih ettiği özellikleri sırasıyla doğal konuşma dilinde sorar ve gerekli cevapları veri tabanındaki kayıtlarla eşleştirir. Bu chatbot eğitilirken 300000 otel, 100000 şehiri kapsayan veri seti kullanılmış olup BERT algoritması ince ayar yapılarak model oluşturulmuştur. Mesajlardan otel ve şehir adları gibi bilgileri çıkarması için NER tekniği kullanılmış olup; SpaCy çerçevesi ile 21000 etiketlenmiş mesaj üzerinden özelleştirilmiş bir model ortaya çıkarılmıştır.

(Kanodia vd., 2021) tarafından yapılan çalışmada bir soru cevaplama modeli geliştirilmiştir. Bu model geliştirilirken BERT modeli ve Google Dialogflow entegrasyonu kullanılmıştır. Kullanıcı sorusunu arayüze girer daha sonra bu sorular Dialogflow yardımı ile niyetler ile eşleştirilir. Ardından BERT modeline gönderir. BERT modeli cevabı üreterek kullanıcıya API yardımı ile gönderir. BERT modelinin ince eğitimi için Conversational Question Answering (CoQA) veri kümesi kullanılmıştır.

(Tanwar vd., 2023) sağlık alanında kullanılmak üzere yapay zeka tabanlı bir chatbot geliştirmiştir. Geliştirilen chatbotta herhangi GPT-3, BERT gibi modeller kullanılmamış olup model eğitimi için Dialogflow ve makine öğrenmesi teknikleri kullanılmıştır. Sınıflandırma algoritmaları kullanılarak soruların kategorilere ayrılması, Terim Frekansı Ters Doküman Frekansı (Term Frequency-Inverse Document Frequency - TFIDF) gibi algoritmalar kullanılarak ise verinin özniteliklerinin çıkarılması hususunda çalışmalar yapılmıştır. Bu modeli eğitirken kullanılan veriler forum siteleri, sosyal medya gibi yerlerden toplanmıştır. Yapılan chatbotta 100 girdi veri setinden 20 tanesi test amacıyla kullanılmış olup botun yüksek oranda doğru cevaplar verdiği gözlemlenmiştir.

(Arısoy, 2024) yaptığı çalışmada NLP'de kullanılan algoritmaları konu almış ve bu algoritmalar arasında performans karşılaştırması yapmıştır. Çalışma kapsamında metin sınıflandırma, NER ve soru cevaplama problemleri ele alınmıştır. F1-skor metriklerine göre; metin sınıflandırma probleminde algoritmaların başarım sıralaması en iyiden en kötüye doğru BERT, destek vektör makinesi (Support Vector Machine - SVM), Naive Bayes şeklindedir. NER probleminde de BERT modeli en başarılı model olmuştur. Soru cevaplama probleminde ise GPT-3 modeli en başarılı model olarak ön plana çıkmaktadır. BERT modeli ise GPT-3 modelinin hemen arkasında yer almaktadır.

(Khin ve Soe, 2020) yaptığı çalışmada Seq2Seq yapısını kullanarak üniversitelerde kullanılabilecek bir chatbot yapmayı amaçlamışlardır. Bu çalışma Mynmar dili için oluşturulan ilk sinir ağı tabanlı chatbot yapımını temsil etmektedir. Tekrarlayan Sinir Ağı (Recurrent Neural Network - RNN) tabanlı Seq2Seq modelini temel alarak bir encoder-decoder mimarisi kullanmıştır. Mynmar dilinin az kullanılması sebebiyle veri seti hazırlamada zorluklar yaşayıp 5000 adet soru cevap çiftleri ile model eğitilmiştir. Bu modele ek olarak uzun cümlelerde daha verimli çıkarım yapması için Dikkat mekanizması (Attention mechanisim) eklenmiş olup Bilingual Evaluation Understudy (BLEU) metriğine göre model dikkat mekanizması olmadan 0.35 oranda başarım verirken dikkat mekanizması eklendiğinde bu başarım 0.41 oranına çıkmıştır.

(Acharya vd., 2022) çalışmasında hem bireysel hem akademik ihtiyaçlarda kullanılabilecek bir chatbot geliştirmiştir. Geliştirilen chatbot NLP ve BERT tabanlı bir kapalı alan soru cevap (Closed-Domain Question Answering - CDQA) chatbotudur. Eğitim verileri için kullanıcılardan gelen mesajları kullanmışlardır. Ek olarak SpaCy kütüphanesi desteği ile NER algoritmalarını kullanarak kullanıcılardan gelen önemli yerleri kişi adı, yer adı gibi bu verileri veri setine dahil edip BERT

modelini hazırladıkları veri seti dahilinde eğitmişlerdir. CDQA yanıtı üretmek için Retriever ve Reader algoritmaları kullanarak veri tabanından gerekli verilerin seçilmesini sağlamışlardır. Reader algoritmaları ise seçilen veriler ile anlamlı yanıtlar üreten algoritmadır. BERT ise sorunun bağlamını anlamak için kullanılmıştır. Gerekli testler sonucunda %80 oranında başarıma ulaşılmıştır.

(Banu ve Patil, 2020) yaptıkları çalışmada hazır platformlar kullanarak bir chatbot geliştirmeyi hedeflemişlerdir. Chatbotu geliştirirken kullanıcıları anlamak ve niyetlerini sınıflandırmak için Microsoft Dil Anlama Akıllı Hizmeti (Language Understanding Intelligent Service - LUIS) kullanmışlardır. LUIS Microsoftun doğal dil işleme ve makine öğrenmesi sağlayan hazır bulut hizmetidir.

(İşeri vd., 2021) tarafından yapılan çalışmada asıl amaç kullanıcıların karşılarında bir çalışana ihtiyaç duymadan firma veya ürünler hakkında hızlı ve doğru bilgiler almasını sağlayan yani bir çağrı merkezi çalışanın yerine geçen bir chatbot tasarlamaktır. Çalışmada veri seti olarak bir firmanın Sıkça Sorulan Sorular bölümünde bulunan sorular alınmıştır. İlk olarak BERT modelinde kullanılacak veriler hazırlanmıştır. Daha sonra BERT modelinin eğitilmesi ve eğitilen modelin WhatsApp üzerinden şirket web sayfasına bağlanması gerçekleştirilmiştir. Model başarımlarının günlük olarak takibi yapılmış olup veri seti 32 gün boyunca eğitilmiştir. Selenium kütüphanesi ile WhatsApp uygulamasına bağlantı gerçekleşmiştir. BERT sınıflandırma ile chatbot projesinde Google Colaboratory, Pycharm gibi ortamlardan yararlanılmıştır. Ayrıca Tensorflow, Keras, Numpy, Matplotlib, Pandas, Torch, Transformers, Selenium, Pyodbc gibi Python kütüphaneleri ve teknolojileri kullanılmıştır. Tüm bunların sonucunda % 77 oranında bir doğruluk elde edilmiştir.

(Kesgin vd., 2023) tarafından yapılan çalışmada çağrı merkezi maliyetlerini azaltmak ve müşteri deneyimini arttırmak için Ensemble chatbot modeli geliştirilmiştir. Büyük telekomünikasyon şirketlerinin sıkça sorulan sorularından oluşan 1887 soru-cevap çiftini içeren bir veri seti kullanılarak FastText yöntemiyle kelime vektörleri oluşturulmuştur. Chatbotun en doğru şekilde çalışabilmesi için, ortalama ve ağırlıklı ortalama cümle vektörlerinin kosinüs benzerliği, n-gram benzerliği ve kelime taşıma mesafesi gibi yöntemler bir ensemble modelde birleştirilmiştir. Model, küçük veri setlerinde test edildiğinde bile yüksek oranda doğruluk oranlarıyla üstün performans sergilemiştir. Gelecekte, BERT gibi gelişmiş dil modelleme tekniklerinin entegrasyonu ile modelin performansının artırılabileceği öngörülmektedir.

(Akarsu ve Er, 2023) tarafından yapılan çalışmada, e-sağlık sistemine entegre edilmiş yapay zeka tabanlı bir chatbot geliştirilmesi ele alınmaktadır. Temel amaç, NLP ve derin öğrenme yöntemlerini kullanarak, kullanıcıların sağlıkla ilgili sorularına doğru ve zamanında yanıtlar sunan ayrıca doktorların yükünü hafifletmek ve hastaların memnuniyetini artırmak için bir chatbot tasarlamaktır. Doğal dil işleme ve anlamlandırma yetenekleri için BERT modeli kullanılmıştır. Çalışma, chatbotun geliştirme süreci, kullanılan veri kümeleri ve performans değerlendirme kriterlerini detaylı bir şekilde açıklamaktadır.

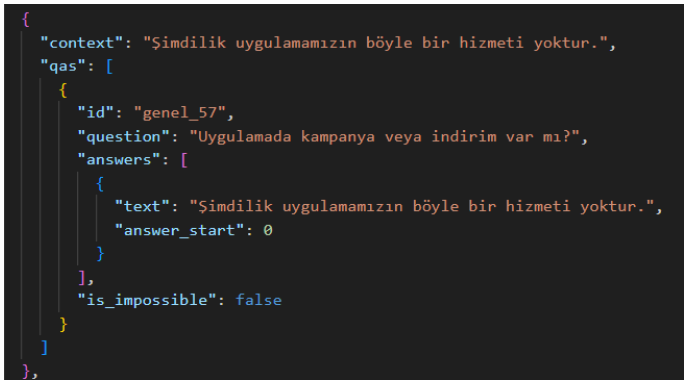
(Nsaif vd., 2024) çalışmalarında chatbot tasarım sürecinde kullanılan çerçeve ve platformların seçiminde dikkate alınması gerekenler ve alandaki bilimsel araştırmaları ele almaktadır. Makalenin amacı, chatbot geliştirme sürecinde karşılaşılan zorlukları ve bu zorlukların çözümüne yönelik en iyi durumları sunarak kullanıcılara rehberlik etmektedir. Chatbot performansını değerlendirmek için kullanılan ölçütler belirtilmiştir. F1 Skoru, Recall-Oriented Understudy for Gisting Evaluation (ROUGE), BLEU (Bilingual Evaluation Understudy) gibi ölçütler yer

almıştır. Çalışmada, Chatbot tasarımı ve geliştirilmesinde kullanılan çerçeve ve platformların uygulamaya olan etkisi üzerinde durulmuştur.

3. MATERYAL

3.1. Kullanılan Veri Kümesi

Veri kümesi, oluşturulan chatbot modeli için temel yapı taşı olmuştur. Bu veriler BERT modelinin öğrenmesinde doğru yanıtlar üretmesinde kaynak olup kullanıcıların niyetlerine (intent) göre modelin cevap vermesi sağlanmıştır. Bu veriler temelde üç temel unsura dayanır. Bunlar; bağlam (context), soru (question) ve cevaplardır (answers). Bu unsurlar göz önünde bulundurulduğunda niyetlerin vektörel soru kısmı ile asıl sorunun soru, oluşacak cevabın ise cevaplar ile eşleştirildiğinden söz edilebilmektedir. Oluşturulan chatbot kuaförlere odaklı olsa da bu bir rezervasyon sistemi için geliştirilen chatbottur. Bu sebeple var olan rezervasyon sitelerinde bulunan sık sorulan sorular toplanıp düzenlenlendi ve buna ek olarak doğal dilde geçebilecek insan kaynaklı tümceler oluşturulmuş ardından bu sorular ile benzer sorular hazırlanmıştır. Ek olarak veri seti hazırlama algoritmasında veri tabanı verilerinin çekilip, bu verilerle örneğin şehirler ile birçok yeni soru veri setine katılmıştır. Örneğin veri tabanında bulunan "şehirler" tablosundan şehir bilgileri seçilip her şehir ile yeni soru oluşturulmuştur. JavaScript Nesne Notasyonu (JavaScript Object Notation - JSON) formatında depolanarak model eğitimi için kullanıldı. Şekil 1’de küçük bir kesiti görünen yapılandırılmış JSON veri seti, 1400’den fazla soru cevap objesinden oluşmaktadır.



Şekil 1. Veri Kümesinden Bir Kesit

Bu JSON veri yapısında her kayıt, bir bağlam, bu bağlamla ilişkili bir veya birden fazla soru ve bu soruların bağlamdan alınmış yanıtlarını içermektedir. Her cevap, bağlam içerisindeki başlangıç konumu ile birlikte verilmektedir. Bu sayede chatbot, kullanıcıdan gelen doğal dildeki sorulara uygun ve doğru yanıtlar verebilecek şekilde eğitilebilmektedir.

3.2. Kullanılan Teknolojiler

Guido van Rossum tarafından 1991 yılında geliştirilen Python programlama dili, kolay öğrenilebilirliği ve geniş kütüphane desteği sayesinde yapay zekâ, web geliştirme, otomasyon gibi birçok alanda kullanılabildiğinden bu çalışmada tercih edilmiştir. Chatbot geliştirme sürecinde kullanılan çeşitli Python kütüphaneleri, NLP ve derin öğrenme gibi süreçleri önemli ölçüde kolaylaştırmaktadır. Güçlü bir model oluşturmayı sağlayanlara TensorFlow ve PyTorch örnek gösterilirken önceden eğitilmiş dil modelleriyle hızlı bir uygulama geliştirmeyi sağlayana ise Hugging Face Transformers örnek verilmektedir. Ek olarak Bilimsel hesaplamalar için geliştirilmiş temel bir araç olan NumPy, çok boyutlu diziler üzerinde hızlı bir şekilde işlem yapılmasını sağlamaktadır. Özellikle metin verisinin sayısal temsillere (embedding, vektör) dönüştürülmesinde ve bu temsillerin matris işlemlerine tabi tutulmasında kullanılmaktadır.

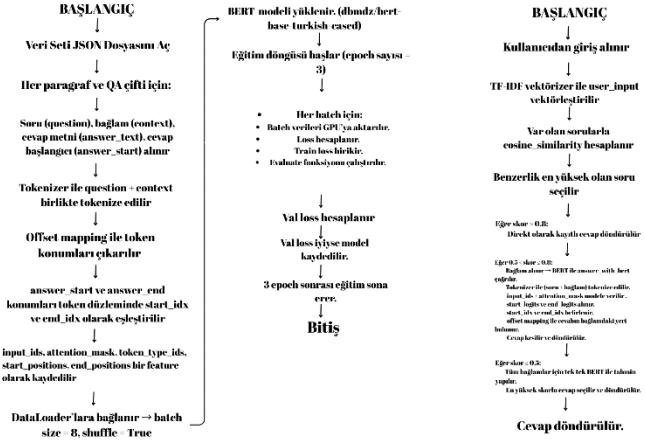
Ayrıca TensorFlow ve PyTorch gibi kütüphanelerin altyapısında da yaygın olarak kullanılmaktadır. PyTorch, derin öğrenme modellerinin geliştirilmesinde öne çıkan kütüphanelerden biridir. Gerek klasik yapay sinir ağları gerekse transformer tabanlı dil modelleri bu kütüphane üzerinde kolayca inşa edilebilir. GPU desteği sayesinde büyük veri kümeleri üzerinde verimli eğitim sağlamaktadır. Bu da PyTorch'u, özellikle doğal dil üretimi ve diyaloğa dayalı chatbot sistemleri için tercih edilen bir araç haline getirmektedir. Makine öğrenmesi algoritmalarını hızlı bir şekilde uygulamayı sağlayan Scikit-learn (sklearn), veri madenciliği ve modelleme süreçlerinde kullanılan bir kütüphanedir. Sınıflandırma, regresyon, kümeleme gibi temel algoritmaların yanı sıra veri ön işleme, model seçimi ve değerlendirme gibi görevleri yerine getirir. Chatbot geliştirme süreçlerinde genellikle modelden önceki veri hazırlık aşamasında tercih edilmektedir. Son olarak, Pydantic, Python veri yapılarını doğrulama ve otomatik tür dönüşümleri yapmak amacıyla kullanılan bir kütüphanedir. Genellikle API geliştirme süreçlerinde, dış kaynaklardan gelen verilerin yapılandırılması ve doğrulanmasında kullanılmaktadır. Chatbot uygulamalarında ise kullanıcıdan gelen verilerin beklenen biçimde olup olmadığını kontrol etmek, hatalı veri girişlerini tespit etmek açısından büyük kolaylık sağlamaktadır. FastAPI gibi modern web çatılarıyla doğal uyumu sayesinde, chatbot'un API tabanlı sistemlerle entegrasyonunu da oldukça pratik hale getirmektedir. Transformer kütüphanesi, açık kaynaklı bir Python kütüphanesi olmakla birlikte NLP ve derin öğrenme modellerinin basit bir şekilde uygulanmasını sağlamaktadır. BERT, GPT ve RoBERTa gibi önceden eğitilmiş modellerle geniş bir kullanım alanını kapsayan kütüphane, Hugging Face tarafından sunulmuştur. FastAPI, Python programlama dili tabanlı hızlı ve verimli API'ler geliştiren bir web framework'üdür. Otomatik belge üretimi sunar ve OpenAPI ve JSON Schema standartlarını yerine getirerek Temsili Durum Transferi (Representational State Transfer -

RESTful) API'ler oluşturmaktadır. FastAPI, asenkron programlama ile yüksek performans sağlayarak oldukça tercih edilmektedir.

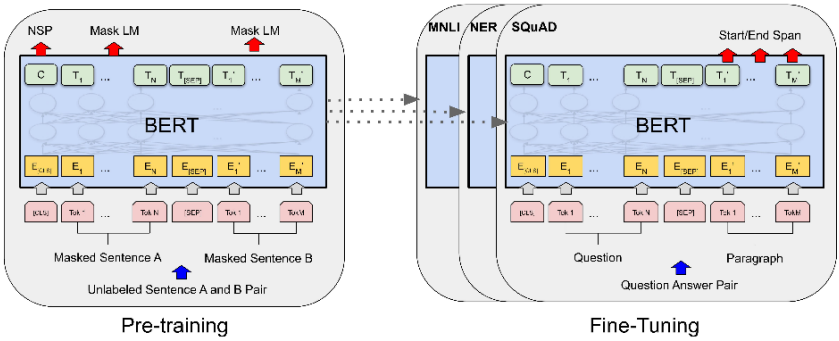
JavaScript, web geliştirme için kullanılan, dinamik ve çok yönlü bir programlama dilidir. Özellikle web tarayıcıları üzerinde çalışan etkileşimli öğeler oluşturmak için kullanılmaktadır. Web uygulama geliştirmenin olmazsa olmazı diyebileceğimiz Javascript programlama dili ile yüzlerce arayüz oluşturulmuştur. Bunlardan Express ve React kullanılan arayüzler arasındadır. Javascript söz dizimine sahip NodeJS Javascript programlama dilinin sunucu taraflı çalışan çalışma ortamıdır. Express ise açık kaynaklı yazılım olarak yayınlanan NodeJS ile RESTful API'ler oluşturmaya yönelik bir arka uç çerçevesidir. Kullanımı kolay, yüksek hız ve en büyük açık kaynak topluluğu Node Paket Yöneticisi (Node Package Manageri - NPM) barındırmasından dolayı bu çalışmanın web tarafında tercih edilmiştir. React, Meta şirketi tarafından 2013 yılında geliştirilen açık kaynak Javascript kütüphanesidir. React bileşen tabanlı bir yapıya sahiptir. Her bileşen ayrı ayrı derlenmektedir. React'ın Virtual Belge Nesne Modeli (Document Object Model - DOM) kullanımı ile değişiklikleri rakiplerine göre çok hızlı algılayıp yeniden derlenmesi performans ve hız açısından büyük avantaj sağlamaktadır. Bunlara ek olarak yine NPM kullanılır ve React kullanırken ihtiyaç duyduğunuz her şey için çeşitli kütüphane ve araçlar NPM'de bulunabilmektedir. Bu bağlamda bir Yapılandırılmış Sorgu Dili (Structured Query Language - SQL) tabanlı mimariye sahip PostgreSQL uygulaması tercih edilmiştir. SQL yapı veri yapısı tutarlılığı sebebiyle Sadece SQL Değil (Not Only SQL - NoSQL) yerine tercih edilmiştir. PostgreSQL kullanacak olmamızın sebebi ise; MySQL'deki gibi sadece ilişkisel veri tabanı özellikleri sağlamasının yanı sıra nesne ilişkisel model özelliklerini de sağlamaktadır. Veri tabanında nesneler, anahtar değer çiftleri gibi veriler tutulabilmektedir.

4. YÖNTEM

Bu çalışmada, derin öğrenme modeli olarak BERTürk kullanılmıştır. İlerleyen alt bölümlerde BERT modelinin genel yapısı, BERTürk modelinin detayları, ince ayar yöntemi, tokenleştirme, kullanıcı tarafından girilen soruların ve veri setindeki bağlamların vektörleştirilmesi açıklanmaktadır. Modelin eğitimi sürecini daha iyi açıklayabilmek adına kullanılan doğal dil işleme mimarisinin genel yapısını özetleyen bir şema Şekil 2'de sunulmuştur. Bu şema, verinin ham metin formatından başlayarak son çıktıya kadar geçtiği temel aşamaları görselleştirmektedir. Şemada gösterilen sıralamaya uygun olarak, modelin eğitim süreci aşağıda adım adım açıklanmıştır.



Şekil 2. Sistemin Genel Akış Diyagramı



Şekil 3. BERT Modeli (Devlin vd., 2019)

Açık kaynak kodlu ve önceden eğitilmiş bir doğal dil işleme modeli olan BERT, Google'ın yapay zeka araştırma ekibi tarafından geliştirilmiştir (Devlin vd., 2019). BERT, verilen metni çift yönlü olarak incelediği için kelimenin sağındaki ve solundaki kelimelerle olan ilişkisini iyi bir şekilde anlamaktadır ve kullandığı MLM ve NSP algoritmaları ile metni derinlemesine incelemektedir. Şekil 3'te BERT modelinin çalışma mimarisi gösterilmiştir. BERT'in eğitimi iki aşamalı olarak gerçekleştirilmiştir: 1. Ön Eğitim (Pre-training): Bu aşamada model çok büyük bir genel veri setleri ile eğitilmektedir. 2. İnce Ayar (Fine-tuning): Ön eğitim sürecinde eğitilen model spesifik bir görev için verilen veri kümesi ile yeniden eğitilmektedir.

Tokenizasyon, metni anlamlı parçalara (token) ayırma işlemidir. BERT modelinde kullanılan tokenizasyon yöntemi metini boşluk ve noktalama işaretleri gibi ayırıcılar kullanarak bölmekteedir. Ardından metni noktalama işaretleri, gereksiz kelimeler gibi etmenlerden arındırmaktadır. Şekil 4'te basit bir tokenizasyon işlemi örneği gösterilmiştir.

Şekil 4. Basit Bir Tokenizasyon Örneği

Türkçe BERT modeli Türkçe diline ait büyük dil modelidir. Türkçe kaynaklardan oluşan büyük bir veri seti ile eğitilen bu model Türkçe metinleri daha iyi anlamak ve daha doğru tahminler yapmak için tasarlanmıştır. Bu çalışmada Hugging Face platformundan alınan model kullanılmıştır.

BERTürk modeli hazırlanan veri kümesi ile eğitilerek kullanıma hazır hale getirilmiştir. İnce ayar adımları şu şekildedir: 1. Veri Yükleme: Eğitim ve doğrulama verileri DataLoader yapısıyla, küçük gruplar halinde modele aktarılmıştır. 2. İleri Geçiş (Forward Pass): Her batch modele verilmiş, model tarafından sorulara karşılık tahminler üretilmiştir. 3. Kayıp Hesaplama: Tahminler ile gerçek cevaplar karşılaştırılarak bir kayıp (loss) değeri hesaplanmıştır. 4. Geri Yayılım (Backpropagation): Kayıp değeri kullanılarak modelin ağırlıkları türevler yardımıyla güncellenmiştir. 5. Ağırlık Güncelleme: optimizer ile ağırlıklar güncellenmiştir. 6. Doğrulama: Her epoch sonunda model, öğrenme süreci durdurularak doğrulama verisi üzerinde performans değerlendirmesine tabi tutulmuştur. 7.Modelin Kaydedilmesi: Eğer doğrulama kaybı önceki en düşük değerden daha azsa, model ve tokenizer diske kaydedilmiştir. Modelin başarısı, eğitim ve doğrulama kayıpları üzerinden izlenmiştir.

- Eğitim Kaybı modelin eğitim verisi üzerindeki performansını gösterir. Ortalama eğitim kaybı şu formülle hesaplanmıştır:

Ortalama Eğitim Kaybı = Toplam Kayıp / Adım Sayısı

- Doğrulama Kaybı ise modelin daha önce görmediği veriler üzerindeki performansını ölçer. Aşırı öğrenme (overfitting) olup olmadığını tespit etmek için kullanılmaktadır.
- En İyi Kayıp: Doğrulama kaybı eğer şimdiye kadarki en düşük değer ise modelin güncel hali kaydedilmiştir. Bu, en iyi genel performansa sahip modelin korunmasını sağlamaktadır.

Modelin eğitimi kapsamında kullanılan parametre ayarları ise şu şekildedir: Eğitim sürecinde epoch sayısı 5 olarak belirlenmiştir; bu doğrultuda model, eğitim verisi üzerinde beş kez tekrar eden bir öğrenme sürecinden geçirilmiştir. Önceki denemelerde üç epoch ile tutarsız çıktılar üreten modelin, beş epoch ile daha istikrarlı ve güvenilir yanıtlar verdiği gözlemlenmiştir. Eğitim sırasında batch boyutu 8 olarak seçilmiştir; bu, modelin verileri sekizli mini kümeler şeklinde işleyerek öğrenmesini sağlamıştır. Öğrenme oranı (learning rate) ise 0.005 olarak belirlenmiştir. Bu düşük öğrenme oranı, modelin ağırlıklarını her güncelleme adımında daha küçük değişikliklerle optimize etmesine olanak tanıyarak, eğitim sürecinde istikrarı artırmıştır. Son olarak, optimizasyon sürecinde AdamW algoritması tercih edilmiştir. Transformer tabanlı mimarilerle uyumlu olan bu algoritma, ağırlıkların düzenlenmesini sağlayarak aşırı öğrenme riskini azaltmakta ve genel öğrenme performansını artırmaktadır. Model bu parametrelerle cuda kullanmadan CPU üzerinde eğitimi 9 saat 25 dakika, cuda kullanılarak GPU üzerinde 2 dakika 34 saniyede eğitilmiştir. Model eğitilirken

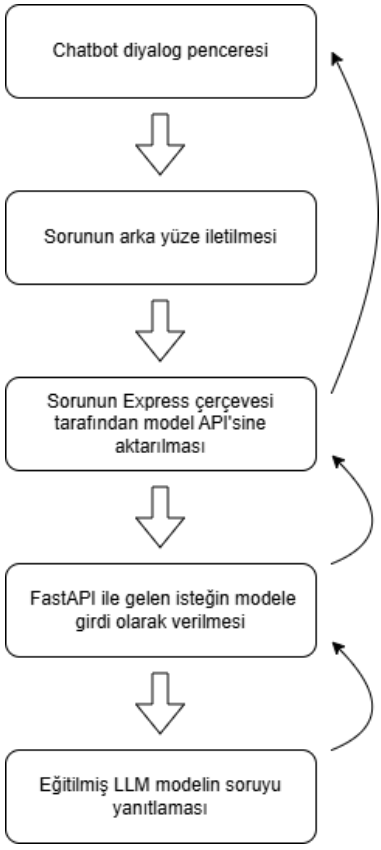
colab sanal ortamı kullanılmıştır. Ekran kartında Tesla T4 grafik kartı tercih edilmiştir.

5. GELİŞTİRİLEN WEB UYGULAMASI

Bu bölümde web uygulamasının oluşturulmasında kullanılan teknolojilerin ne amaçla kullanıldığı ve web uygulamasının mimarisi detaylandırılmıştır. Kullanıcıdan alınan doğal dil girdisinin, geliştirilen web tabanlı sohbet botu aracılığıyla işlenme süreci Şekil 5'te akış diyagramı şeklinde sunulmuştur. Bu diyagram, kullanıcı arayüzünden başlayarak modelin cevap üretimine kadar olan tüm süreci adım adım göstermektedir.

Web uygulaması istemci sunucu mimarisi olarak düşünülmüştür. Bu sayede arka yüz ve ön yüz taraflarını birbirinden ayırarak bağımsız halde çalışması hedeflenmiştir. Sunucu tarafında Bölüm 4' te bahsedildiği üzere Nodejs arayüzü olan Express kullanılmıştır. Express ile bir RESTful API tasarlanıp bu API sayesinde arka taraf ve ön taraf arasında iletişim sağlanmıştır. Express kullanılırken Sequelize açık kaynak kütüphanesi Nesne İlişkisel Eşleme (Object- Relational Mapping - ORM) yöntemi kullanılarak veri tabanı tabloları Javascript sınıfları ile otomatik olarak oluşturulur ayrıca bir SQL kodu yazılmadan tablolar oluşturulup ve SQL sorguları kullanılmıştır.

Ön yüz kısmı gerçeklenirken React ortamında React kütüphaneler topluluğu olan Material UI kütüphanesi kullanılarak tasarım gerçeklenmiştir. Ön yüzde temel olarak kullanıcıların chatbot ile diyalog kurabildiği bir pop-up sekmesi, kuaförlerin görüntülendiği, kullanıcının kuaförlerden rezervasyon yapabileceği sayfalar bulunmaktadır.

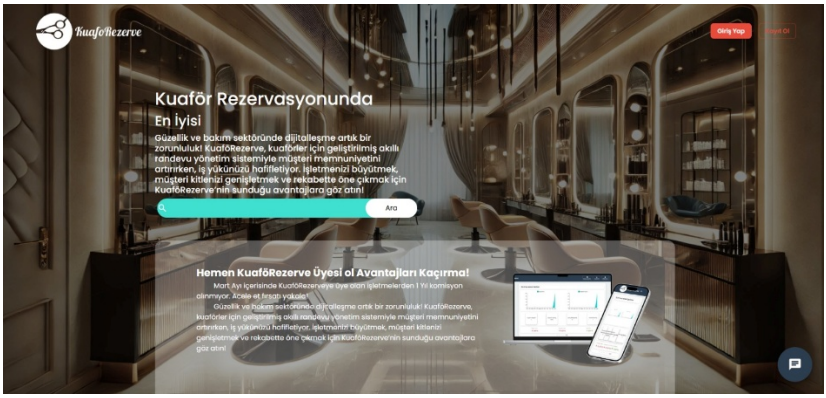


Şekil 5. Akış Diyagramı

Eğitilen model ve ön yüz tarafında iletişim aşamalı olarak gerçekleşmektedir. Şekil 5'te gösterildiği gibi kullanıcılar soracakları soruyu ön yüzde tasarlanan pop-up içerisine yazar ve Express sunucusuna istekte bulunmaktadır. Express sunucusu bu isteğin içindeki soruyu istek yardımı ile FastAPI ile oluşturulmuş modele veri girişini sağlayan RESTful API'ye tekrar istek göndererek ön yüz ile model arasındaki iletişimi sağlamaktadır. Model soruyu değerlendirdikten sonra oluşturduğu metni cevaplar şeklinde sırayla FastAPI, Express ve ön yüz tarafına iletilmektedir. Ön yüze gelen cevap içerisindeki metin kullanıcıya pop-up içerisinde bir diyalog olarak görünmektedir.

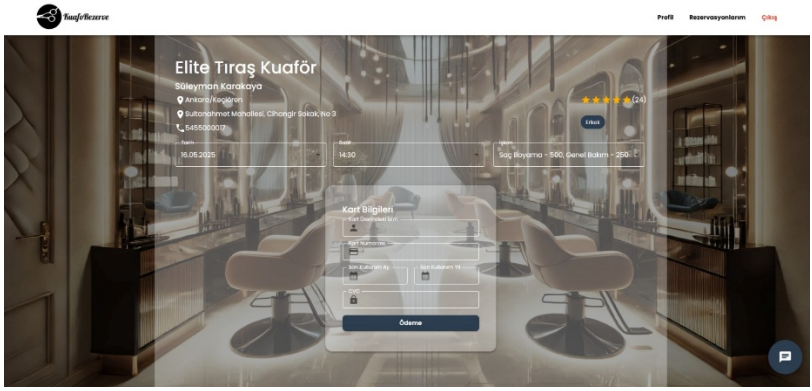
6. SONUÇLAR VE TARTIŞMA

Bu çalışmada geliştirilen chatbotun kapsamlı bir değerlendirmesi yapılmıştır. Gerçekleştirilen modelleme süreci, LLM tabanlı chatbot yapısının başarıyla entegre edildiği bir web uygulamasıyla tamamlanmıştır. Uygulamanın genel kullanıcı arayüzüne ek olarak kullanıcı sorularına yanıt üretebilen chatbot ekranı da tasarlanmıştır. Aşağıda, geliştirilen web platformunun genel yapısı ve özel olarak chatbot bileşenine ait ekran görüntüleri sunularak sistemin bütünsel görünümü ortaya konulmuştur. Şekil 6'da sunulan ana ekran, geliştirilen KuafoRezerve uygulamasının başlangıç arayüzünü temsil etmektedir ve görselde göründüğü üzere sayfanın sağ alt bölümünde chatbot ile iletişime geçilebilmektedir. Şekil 7'de yer alan ekran ile kullanıcılar tercih ettikleri kuaförü seçebilmektedir. Şekil 8'de gösterilen arayüz ise kullanıcıların rezervasyon işlemlerini tamamlamalarına olanak sağlamaktadır.



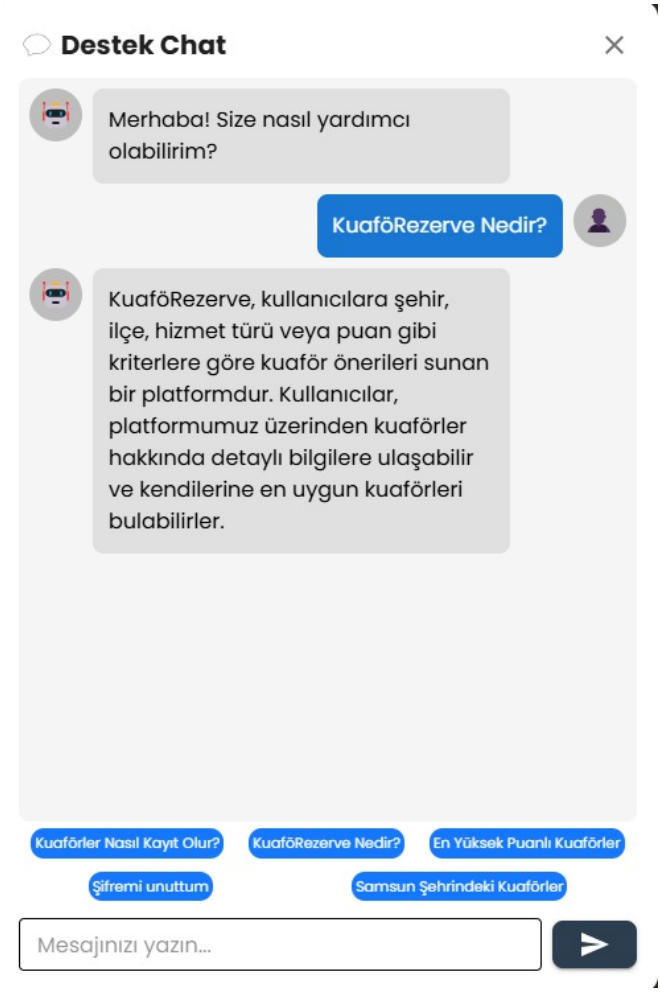
Şekil 6. Uygulama Ana Ekranı

Şekil 7. Kuaför Bulma Ekranı

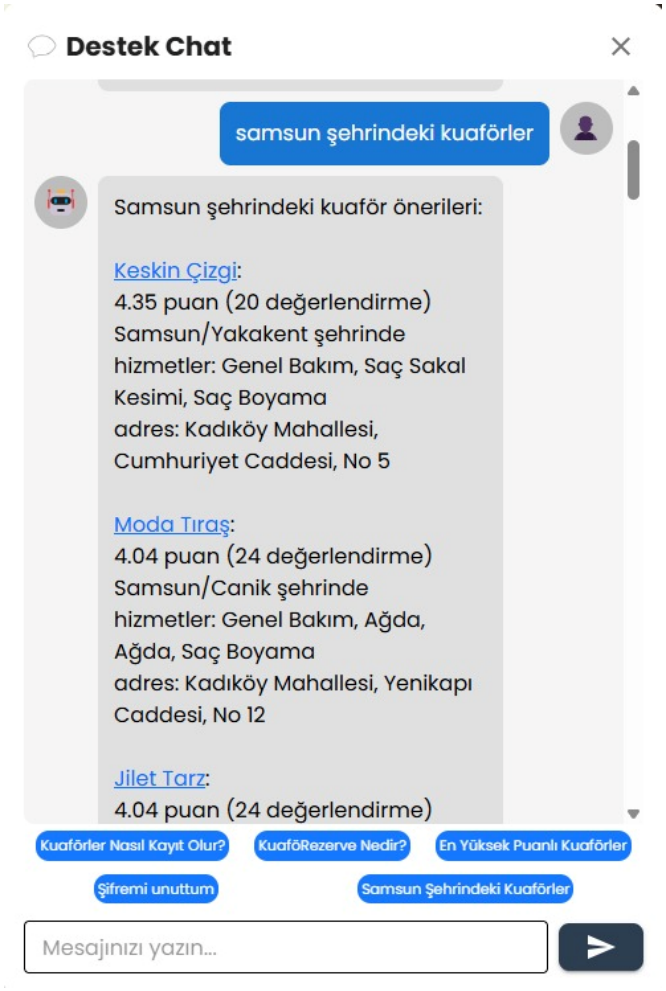


Şekil 8. Rezervasyon Tamamlama Ekranı

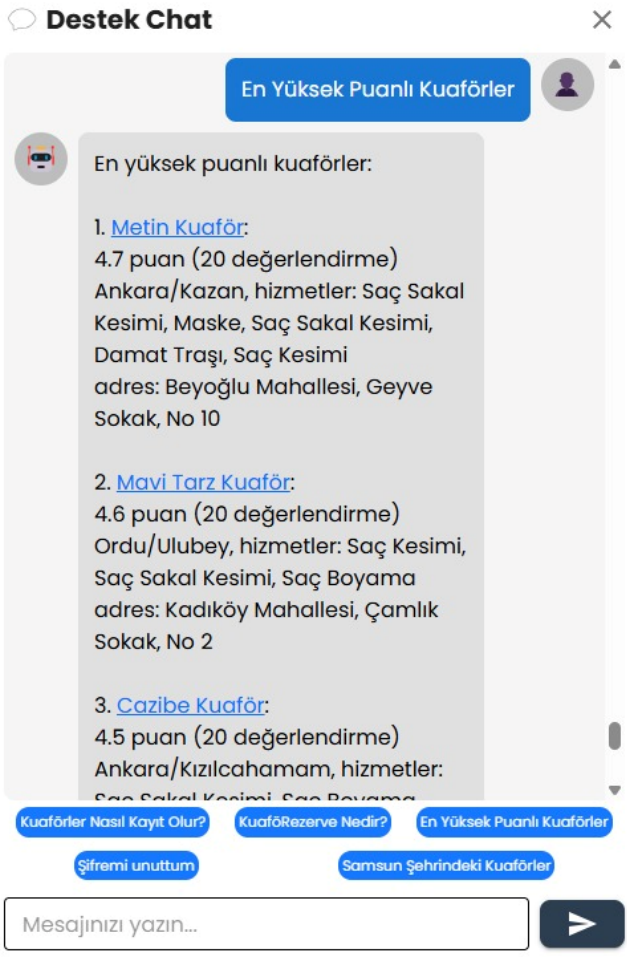
Şekil 9, 10 ve 11'de sırasıyla geliştirilen chatbot arayüzü sunulmaktadır. Bu arayüz, kullanıcıların platform hakkında bilgi almasına, şehirdeki kuaför önerilerini görüntülemesine ve en yüksek puanlı kuaförlere erişmesini sağlamaktadır. Chatbot, kullanıcı sorularına yanıt vererek gerekli sayfaya yönlendirme sağlamaktadır.



Şekil 9. Örnek Sohbet-1



Şekil 10. Örnek Sohbet-2



Şekil 11. Örnek Sohbet-3

7. GENEL DEĞERLENDİRMELER VE GELECEK ÇALIŞMALAR

Kuaförlere yapılan rezervasyonlarda zorluklar gözlenmiş olup bu sorunları gidermek amacı ile kullanıcıların chatbot desteği alarak rezervasyon yapmasını sağlayan bir uygulama geliştirilmesi amaçlanmıştır. Sohbet botunun kullanıcıların yönelttiği soruları doğru biçimde işleyip en uygun yanıtları

oluşturması temel hedef olarak belirlenmiştir. Gerçekleştirilen testler doğrultusunda chatbot çoğu senaryoda kullanıcı taleplerini doğru anlamlandırmış ve başarılı cevaplar sunduğu gözlemlenmiştir. Bu bağlamda, belirlenen hedeflerin büyük ölçüde başarıyla gerçekleştirildiği söylenebilir. Bu çalışmada, Türkçe dilinde bir sohbet botu geliştirmek amacıyla BERT tabanlı bir model kullanılmıştır. Ancak, dilenirse GPT, RoBERTa veya T5 gibi farklı büyük dil modelleri de entegre edilerek sistemin yanıt çeşitliliği artırılabilir. Kuaförlere özgü bir rezervasyon veri kümesinin literatürde veya açık kaynaklarda bulunmaması nedeniyle bu alandaki eksiklikten kaynaklı zorluklar yaşanmış ve bunun sonucunda özgün bir veri kümesi tarafımızdan oluşturulmuştur. Bu katkılar, doğal dil işleme alanında Türkçe dilinde kaynak niteliğinde kullanılabilecek durumdadır. Geliştirilen KuafoRezerve sistemi şu an için yalnızca web tabanlı bir platformda çalışmakta olup gelecekte diğer dijital platformlara entegrasyonu planlanmaktadır. Ayrıca sistem, yalnızca kuaför hizmetleriyle sınırlı kalmayıp güzellik salonları, spa merkezleri, sağlık ve bakım hizmetleri, halı sahalar gibi farklı sektörlere uyarlanabilecek esnek bir yapıya sahiptir.

KAYNAKÇA

Acharya, S., Sornalakshmi, K., Paul, B., & Singh, A. (2022, October). Question answering system using NLP and BERT. *In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 925-929). IEEE.

Akarsu, K., & Er, O. (2023). Artificial intelligence based Chatbot in E-health system. *Artificial Intelligence Theory and Applications*, 3(2), 113-122.

Albayrak, N., Özdemir, A., & Zeydan, E. (2018, May). An overview of artificial intelligence based chatbots and an example chatbot application. *In 2018 26th signal processing and communications applications conference (SIU)* (pp. 1-4). IEEE.

Arısoy, A. (2024). Natural Language Processıng Algorıthms And Performance Comparıson. *Yalvaç Akademi Dergisi*, 9(2), 106-121.

Banu, S., & Patil, S. D. (2020, October). An intelligent web app chatbot. *In 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 309-315). IEEE.

Bratić, D., Šapina, M., Jurečić, D., & Žiljak Gršić, J. (2024). Centralized database access: transformer framework and llm/chatbot integration-based hybrid model. *Applied System Innovation*, 7(1), 17.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. *In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies*, (pp. 4171-4186).

Gemirter, C. B., & Goularas, D. (2021). A Turkish question answering system based on deep learning neural networks. *Journal of Intelligent Systems: Theory and Applications*, 4(2), 65-75.

Gökkaya, E. S., & İşcan, Ö. F. (2025). Yapay Zeka Uygulamalarının Çalışanlar Üzerindeki Etkisinin Teknoloji Kabul Modeli İle Ölçümlenmesi: Chatbot Örneği. *Journal of Economics Business and International Relations-JEBI*, 4(1), 1-18.

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.

Oliveira, P. F., & Matos, P. (2023). Introducing a chatbot to the web portal of a higher education institution to enhance student interaction. *Engineering Proceedings*, 56(1), 128.

İşeri, İ., Aydın, Ö., & Tutuk, K. (2021). Müşteri hizmetleri yönetiminde yapay zekâ temelli chatbot geliştirilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (29), 358-365.

Işık, A. H., & Yağcı, A. (2020). Sequence to Sequence LSTM Modeli ile Telegram Bot Uygulaması. *Gazi Mühendislik Bilimleri Dergisi*, 6(1), 32-39.

Kanodia, N., Ahmed, K., & Miao, Y. (2021, November). Question answering model based conversational chatbot using bert model and google dialogflow. *In 2021 31st International Telecommunication Networks and Applications Conference (ITNAC)* (pp. 19-22). IEEE.

Kesarwani, S., & Juneja, S. (2023, April). Student chatbot system: A review on educational chatbot. *In 2023 7th international conference on trends in electronics and informatics (ICOEI)* (pp. 1578-1583). IEEE.

Kesgin, H. T., Öztunç, O., & Diri, B. (2023). Ensemble Learning Approach to Chatbot Design Based on Paraphrase Detection. *Kocaeli Journal of Science and Engineering*, 6(2), 129-137.

Khin, N. N., & Soe, K. M. (2020, November). Question answering based university chatbot using sequence to sequence model. *In 2020 23rd conference of the oriental COCOSDA international committee for the co-ordination and standardisation of speech databases and assessment techniques (O-COCOSDA)* (pp. 55-59). IEEE.

Li, B., Jiang, N., Sham, J., Shi, H., & Fazal, H. (2019, September). Real-world conversational AI for hotel bookings. *In 2019 Second International Conference on Artificial Intelligence for Industries (AI4I)* (pp. 58-62). IEEE.

Nirala, K. K., Singh, N. K., & Purani, V. S. (2022). A survey on providing customer and public administration based services using AI: chatbot. *Multimedia Tools and Applications*, 81(16), 22215-22246.

Nsaif, W. S., Salih, H. M., Saleh, H. H., & Al-Nuaimi, B. T. (2024). Chatbot development: Framework, platform, and assessment metrics. *The Eurasia Proceedings of Science Technology Engineering and Mathematics*, 27, 50-62.

Tanwar, P., Bansal, K., Sharma, A., & Dagar, N. (2023, November). AI based chatbot for healthcare using machine learning. *In 2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* (pp. 751-755). IEEE.

Toprak, G., & Rasheed, J. (2022). Machine learning based natural language processing for Turkish venue recommendation

chatbot application. *Avrupa Bilim ve Teknoloji Dergisi*, (38), 501-506.

AKILLI TARIM UYGULAMALARI KAPSAMINDA MULTİSPEKTRAL KAMERALAR İLE DOMATES HASTALIK ANALİZİ

Mahmut DURGUN[1]

1. GİRİŞ

Küresel nüfus artışı ve iklim değişikliği baskıları altında, tarımsal üretimde sürdürülebilirlik ve verimlilik arayışı, geleneksel yöntemlerden teknoloji tabanlı yaklaşımlara doğru radikal bir dönüşümü zorunlu kılmaktadır. Bu dönüşümün merkezinde, sıklıkla birbirinin yerine kullanılsa da nüanslara sahip olan "hassas tarım" ve "akıllı tarım" kavramları yer almaktadır. Hassas tarım (Precision Agriculture), bilgi teknolojilerini kullanarak tarladaki değişkenlikleri ölçen, analiz eden ve gübre, su gibi girdilerin mekânsal ve zamansal olarak optimize edilmesini sağlayan bir yönetim stratejisidir(Getahun, Kefale, & Gelaye, 2024). Akıllı tarım (Smart Farming) ise bu yaklaşımı bir adım öteye taşıyarak; Nesnelerin İnterneti (IoT), yapay zeka (AI), robotik ve büyük veri analitiği gibi teknolojilerin entegrasyonuyla, sadece tarlayı değil, tüm çiftlik operasyonlarını kapsayan, veriye dayalı, bütüncül ve otonom karar alma süreçlerini ifade etmektedir(Sharma & Shivandu, 2024). Bu teknolojik paradigma, özellikle domates (Solanum lycopersicum L.) gibi ekonomik değeri yüksek ancak hastalık ve

[1] Doç. Dr., Tokat Gaziosmanpaşa Üniversitesi, Turhal Uygulamalı Bilimler Fakültesi, Elektronik Ticaret ve Yönetimi Bölümü, ORCID: 0000-0002-5010-687X.

zararlılara karşı hassas ürünlerin yönetiminde kritik bir rol oynamaktadır.

Domates, dünya genelinde taze tüketim ve endüstriyel işleme açısından stratejik bir üründür; ancak üretim süreci, biyotik stres faktörlerinin sürekli tehdidi altındadır. Fungal (örn. Erken Yanıklık, Alternaria solani), bakteriyel ve viral (örn. Domates Lekeli Solgunluk Virüsü - TSWV) patojenler, bitki fizyolojisini bozarak hem verim miktarında hem de ürün kalitesinde dramatik düşüşlere neden olabilmektedir (Nazarov, Baleev, Ivanova, Sokolova, & Karakozova, 2020). Örneğin, Samsun ilinde yürütülen bir saha çalışmasında, TSWV enfeksiyonunun domates veriminde %42,1 oranında azalmaya ve pazarlanabilir ürün değerinde %95,5'lik bir kayba yol açtığı, bunun da yaklaşık 0,9 milyon dolarlık ekonomik zarara tekabül ettiği rapor edilmiştir (Sevik & Arli-Sokmen, 2012). Geleneksel hastalık teşhis yöntemleri olan manuel gözlem ve laboratuvar testleri, genellikle zaman alıcıdır, uzmanlık gerektirir ve insan hatasına açıktır; bu durum, hastalıkların erken evrede tespit edilerek yayılmasının önlenmesini zorlaştırmaktadır(Jafar, Bibi, Naqvi, Sadeghi-Niaraki, & Jeong, 2024).

Bu zorlukların aşılmasında, multispektral görüntüleme teknolojileri devrim niteliğinde bir çözüm sunmaktadır. İnsan gözünün algılayabildiği görünür ışık (RGB) spektrumunun ötesine geçen bu kameralar; Yeşil, Kırmızı, Kırmızı-Kenar (Red-Edge) ve Yakın Kızılötesi (NIR) gibi spesifik dalga boylarında veri toplayarak bitki sağlığının hassas bir şekilde izlenmesine olanak tanır(Abdulridha, Ampatzidis, Kakarla, & Roberts, 2020). Bitki dokusundaki klorofil içeriği, hücre yapısı ve su stresi gibi fizyolojik değişimler, yaprakların spektral yansıma özelliklerini değiştirmektedir. Multispektral kameralar, Normalize Edilmiş Fark Vejetasyon İndeksi (NDVI) ve Normalize Edilmiş Kırmızı Kenar İndeksi (NDRE) gibi vejetasyon indekslerini kullanarak, hastalık belirtileri henüz insan gözüyle fark edilemezken

(asemptomatik evre) stresi tespit edebilmekte ve müdahale imkanı sağlamaktadır(J. Lu, Ehsani, Shi, de Castro, & Wang, 2018; Verma, Chug, & Singh, 2018).

Bu kitap bölümünün temel amacı, akıllı tarım uygulamaları çerçevesinde multispektral görüntüleme teknolojilerinin domates hastalıklarının tespiti ve analizindeki rolünü kapsamlı bir şekilde incelemektir. Bölüm kapsamında; hassas tarım teknolojilerinin teorik temelleri, domates bitkisinde yaygın görülen hastalıkların spektral imzaları, multispektral veri toplama süreçleri ve elde edilen verilerin işlenerek anlamlı tarımsal içgörülere dönüştürülmesi süreçleri ele alınacaktır. Ayrıca, bu teknolojilerin hastalık yönetiminde sağladığı ekonomik ve çevresel avantajlar, literatürdeki güncel vaka analizleri ve deneysel bulgular ışığında tartışılacaktır.

2. MULTİSPEKTRAL KAMERA SİSTEMLERİ VE SPEKTRAL BANTLAR

2.1. Multispektral kamera teknolojisinin çalışma prensibi

Multispektral görüntüleme (MSI), elektromanyetik spektrumun görünür ve görünür olmayan bölgelerindeki verilerin, belirli ve ayrık dalga boyu aralıklarında (bantlarda) toplanması prensibine dayanır. İnsan gözünün algıladığı görünür ışık spektrumunun ötesine geçen bu teknoloji, genellikle 3 ila 10 arasında değişen sayıda spektral banttan veri yakalar(J. M. Amigo, 2019). Hiperspektral görüntülemeden (HSI) temel farkı, yüzlerce dar ve sürekli bant yerine, daha az sayıda ancak tarımsal analiz için stratejik öneme sahip daha geniş bantlara odaklanmasıdır. Bu özellik, multispektral sistemlerin daha az veri üretmesini sağlayarak işleme hızını artırır ve maliyet etkin bir çözüm sunar(B. Lu, Dao, Liu, He, & Shang, 2020).

Bu kameraların çalışma mimarisinde, gelen ışığı sensöre ulaşmadan önce belirli dalga boylarına ayıran optik filtreler (örneğin, filtre tekerlekleri veya desenli filtreler) veya ışığa duyarlı özel sensör dizileri kullanılır(J. M. Amigo & Grassi, 2019). Elde edilen veriler, radyometrik kalibrasyon süreçlerinden geçirilerek dijital sayılardan (DN) mutlak yüzey yansıma değerlerine dönüştürülür ve bu sayede bitki sağlığının hassas bir şekilde analiz edilmesine olanak tanır(Pourazar, Samadzadegan, & Dadrass Javan, 2019).

2.2. Tarımda yaygın kullanılan spektral bantlar

Tarımsal analizlerde kullanılan multispektral kameralar, bitki fizyolojisindeki değişimleri tespit etmek amacıyla elektromanyetik spektrumun spesifik bölgelerine odaklanır.

2.2.1. Görünür ışık (RGB) bantları Bu bölge 400 nm ile 700 nm arasındaki dalga boylarını kapsar ve Mavi (450-520 nm), Yeşil (520-600 nm) ve Kırmızı (600-700 nm) bantlarından oluşur. Bitkilerdeki klorofil pigmenti, fotosentez süreci için mavi ve kırmızı ışığı güçlü bir şekilde emerken, yeşil ışığı yansıtır; bu durum sağlıklı bitkilerin insan gözüne yeşil görünmesinin nedenidir(Rehman vd., 2024). RGB bantları, bitki morfolojisi ve pigmentasyon değişimlerinin görsel analizi için temel teşkil eder.

2.2.2. Kırmızı kenar (red edge) bandı Kırmızı kenar bandı (700-730 nm), kırmızı ışık emilim bölgesi ile yakın kızılötesi yansıma bölgesi arasındaki geçiş aralığında yer alır. Bu bant, bitki stresine ve klorofil içeriğindeki değişimlere karşı son derece hassastır(Filella & Penuelas, 1994). Klorofil miktarındaki en ufak bir azalma, bu bölgedeki spektral eğride kaymalara neden olduğu için, hastalıkların ve besin eksikliklerinin erken dönem tespitinde kritik bir rol oynar(Rehman vd., 2024).

2.2.3. Yakın kızılötesi (NIR) bandı NIR bandı (700-1300 nm), bitki sağlığı analizlerinde en sık kullanılan bölgelerden biridir. Sağlıklı bitki yapraklarının içindeki süngerimsi mezofil

tabakası, bu dalga boyundaki ışığı yüksek oranda yansıtır. Bu banttaki yansıma değerleri, bitki canlılığı (vigour) ve biyokütle yoğunluğu ile doğrudan ilişkilidir(De Silva & Brown, 2024; Sarvakar & Thakkar, 2024).

2.3. Spektral bantların bitki sağlığı ve stres durumlarıyla ilişkisi

Multispektral bantlardan elde edilen veriler, bitkilerin biyotik ve abiyotik stres faktörlerine verdiği tepkileri izlemek için kullanılır. Sağlıklı bir bitki, yüksek klorofil içeriği ve sağlam hücresel yapısı sayesinde görünür ışığı (özellikle kırmızıyı) emerken, NIR ışığını güçlü bir şekilde yansıtır(Berger vd., 2022). Ancak hastalık veya stres durumunda bu denge bozulur.

Örneğin, domates bitkisinde fungal enfeksiyonlar (örn. Erken Yanıklık) veya su stresi meydana geldiğinde, yaprak dokusundaki klorofil yıkıma uğrar. Bu durum, kırmızı banttaki emilimin azalmasına ve yansımanın artmasına neden olurken; hücresel yapının bozulması NIR yansımasının düşmesine yol açar (Muhammad vd., 2025). Bu spektral değişimler, Normalize Edilmiş Fark Vejetasyon İndeksi (NDVI) veya Normalize Edilmiş Kırmızı Kenar İndeksi (NDRE) gibi algoritmalar aracılığıyla sayısallaştırılarak, hastalık belirtileri henüz insan gözüyle görülemeyen "asemptomatik" evredeyken tespit edilebilir(Nguyen vd., 2021). Özellikle NDRE, yoğun bitki örtüsünde doygunluğa ulaşan NDVI'ya kıyasla, geç dönem hastalıkların ve azot eksikliğinin tespitinde daha hassas sonuçlar sunmaktadır(de Castro, Shi, Maja, & Peña, 2021).

3. DOMATES BİTKİSİNDE YAYGIN OLARAK GÖRÜLEN HASTALIKLAR

Domates (*Solanum lycopersicum*), küresel ölçekte ekonomik değeri yüksek bir ürün olmasına rağmen, üretim süreci

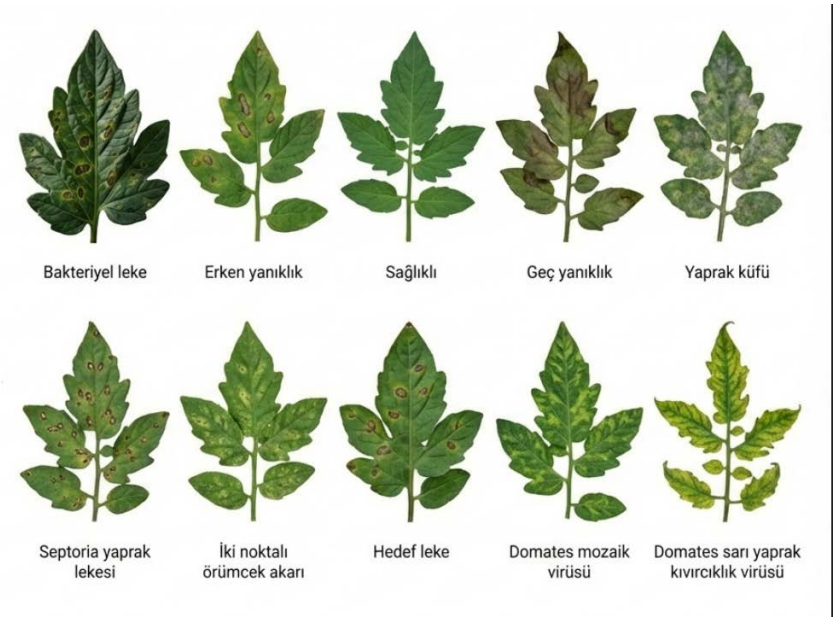
boyunca çeşitli biyotik stres faktörlerinin tehdidi altındadır. Hastalık etmenleri, bitkinin fizyolojik süreçlerini bozarak verim ve kalite kayıplarına yol açmaktadır. Bu hastalıkların sınıflandırılması ve bitki dokularında yarattığı görsel ve spektral değişimlerin anlaşılması, uzaktan algılama teknolojilerinin başarısı için temel teşkil etmektedir.

3.1. Fungal, bakteriyel ve viral domates hastalıklarının sınıflandırılması

Domates hastalıkları temel olarak fungal, bakteriyel ve viral patojenler olmak üzere üç ana kategoride incelenmektedir. Fungal hastalıklar, domates üretiminde en sık karşılaşılan sorunların başında gelmektedir. Özellikle *Alternaria solani* tarafından oluşturulan Erken Yanıklık (Early Blight) ve *Phytophthora infestans* kaynaklı Geç Yanıklık (Late Blight), yaprak, gövde ve meyvede ciddi hasarlara neden olmaktadır. Ayrıca, *Septoria lycopersici*'nin neden olduğu Septoria Yaprak Lekesi ve *Botrytis cinerea* kaynaklı Kurşuni Küf, bitkinin fotosentetik alanını daraltan diğer önemli fungal enfeksiyonlardır(Sanoubar & Barbanti, 2017). Bakteriyel hastalıklar arasında, *Xanthomonas* türlerinin neden olduğu Bakteriyel Leke (Bacterial Spot) ve *Pseudomonas syringae* pv. *tomato* kaynaklı Bakteriyel Benek (Bacterial Speck) öne çıkmaktadır. Bu hastalıklar genellikle tohum kaynaklı olup, nemli koşullarda hızla yayılmakta ve meyve kalitesini doğrudan etkileyen lezyonlar oluşturmaktadır(Catara & Bella, 2020). Viral hastalıklar ise vektör böcekler veya mekanik yollarla taşınan ve sistemik enfeksiyonlara yol açan patojenlerdir. Domates Lekeli Solgunluk Virüsü (Tomato Spotted Wilt Virus - TSWV), Domates Sarı Yaprak Kıvırcıklığı Virüsü (Tomato Yellow Leaf Curl Virus - TYLCV) ve Domates Mozaik Virüsü (ToMV) en yıkıcı viral etmenler arasındadır. Örneğin, TSWV enfeksiyonunun domates veriminde %40 ila %90 oranında kayba neden olabildiği rapor edilmiştir(Sevik & Arli-Sokmen, 2012).

3.2. Hastalıkların bitki yaprakları üzerindeki görsel ve fizyolojik etkileri

Hastalık etmenleri, bitki yapraklarında karakteristik görsel semptomlar ve fizyolojik bozulmalar meydana getirir. Fungal enfeksiyonlar genellikle nekrotik lezyonlar, kloroz (sararma) ve yaprak dökümü ile sonuçlanır (Şekil 1). Örneğin, Erken Yanıklıkta yapraklarda iç içe geçmiş halkalar (hedef tahtası görünümü) oluşurken, Geç Yanıklıkta suda haşlanmış gibi görünen gri-yeşil lekeler ve ardından doku ölümü gözlenir(Ally, Neetoo, Ranghoo-Sanmukhiya, & Coutinho, 2023). Bakteriyel enfeksiyonlarda ise genellikle yaprak kenarlarında yanıklıklar, sarı haleli küçük siyah lekeler ve vasküler dokularda bozulmalar meydana gelir. Viral hastalıklar ise yapraklarda mozaik desenleri, şekil bozuklukları (kıvrılma, küçülme), bodurluk ve bronzlaşma gibi sistemik belirtilerle kendini gösterir(Muhammad vd., 2025).



Şekil 1. Bitki yapraklarında karakteristik görsel semptomlar ve fizyolojik bozulmalar

Fizyolojik düzeyde bu patojenler; klorofil pigmentinin yıkımına, hücre duvarı bütünlüğünün bozulmasına, su içeriğinin azalmasına ve fotosentez kapasitesinin düşmesine neden olmaktadır(Chhabra, Kaur, Vij, & Gaur, 2020).

3.3. Hastalıkların spektral yansıma özelliklerini değiştirmesi

Bitki hastalıkları, yaprakların elektromanyetik spektrumu yansıtma biçimini (spektral imza) dramatik şekilde değiştirmektedir. Sağlıklı bitki örtüsü, fotosentez için görünür ışık (VIS) bölgesindeki kırmızı ve mavi ışığı emerken, yakın kızılötesi (NIR) ışığı güçlü bir şekilde yansıtır. Ancak hastalık durumunda bu denge bozulur. Klorofil kaybı (kloroz), görünür bölgedeki (özellikle kırmızı bantta, ~680 nm) ışık emilimini azaltarak yansımanın artmasına neden olur. Hücresel yapının bozulması ve su stresi ise NIR bölgesindeki (700-1300 nm) yansımanın azalmasına yol açar(Xu, Ying, Fu, & Zhu, 2007). Örneğin, *Xanthomonas* enfeksiyonu sonucu oluşan su emmiş (water-soaked) lezyonlar, özellikle 1400 nm civarındaki kısa dalga kızılötesi (SWIR) bantlarında belirgin spektral değişimler yaratır(Zhang, Vinatzer, & Li, 2024). Bu spektral farklılıklar, hastalıkların insan gözüyle görülmeden önce (asemptomatik evre) tespit edilmesine olanak tanır.

4. MULTİSPEKTRAL GÖRÜNTÜLER KULLANILARAK DOMATES BİTKİSİNDE HASTALIK ANALİZİ

Multispektral görüntüleme, bitki sağlığının izlenmesinde ve hastalıkların erken teşhisinde hassas tarımın en etkili araçlarından biridir. Bu süreç, verinin toplanmasından işlenmesine kadar titiz bir metodoloji gerektirir.

4.1. Multispektral veri toplama süreci

Veri toplama süreci, genellikle İnsansız Hava Araçları (İHA) veya yer tabanlı platformlar üzerine monte edilmiş multispektral kameralar aracılığıyla gerçekleştirilir. Bu kameralar; Yeşil, Kırmızı, Kırmızı Kenar (Red-Edge) ve Yakın Kızılötesi (NIR) gibi spesifik dalga boylarını kaydeden sensörlere sahiptir. Veri toplama sırasında güneş açısı, bulutluluk durumu ve uçuş yüksekliği gibi çevresel faktörler kritik öneme sahiptir. Tutarlı veriler elde etmek için genellikle öğle saatlerinde ve açık hava koşullarında çekim yapılması tercih edilir. Ayrıca, elde edilen görüntülerin doğru bir şekilde radyometrik kalibrasyonunun yapılabilmesi için, bilinen yansıma değerlerine sahip referans panellerinin (beyaz/gri referanslar) uçuş öncesinde veya sırasında görüntülenmesi gerekmektedir(Daniels vd., 2023).

4.2. Ön işleme ve görüntü işleme adımları

Ham multispektral görüntülerin anlamlı verilere dönüştürülmesi için bir dizi ön işleme adımı uygulanır. İlk olarak, sensör kaynaklı gürültülerin giderilmesi ve radyometrik düzeltmelerin yapılması gerekir. Bu aşamada, dijital sayılar (DN), kalibrasyon panelleri kullanılarak mutlak yansıma değerlerine dönüştürülür(Daniels vd., 2023). Ardından, çoklu kamera lenslerinden kaynaklanan hizalama sorunlarını gidermek için geometrik düzeltme ve görüntü kaydı (co-registration) işlemleri uygulanır. Görüntü işleme aşamasında ise, bitki olmayan arka planın (toprak, gölge vb.) görüntüden ayrıştırılması (segmentasyon) önemlidir. Bu işlem genellikle belirli eşik değerleri veya bitki indeksleri kullanılarak yapılır(Tan, Lu, & Jiang, 2021). Son olarak, bitki sağlığını temsil eden vejetasyon indeksleri (NDVI, GNDVI, NDRE vb.) hesaplanarak hastalık analizi için özellik haritaları oluşturulur.

4.3. Sağlıklı ve hastalıklı bitkilerin spektral olarak ayırt edilmesi

Sağlıklı ve hastalıklı bitkilerin ayrımı, spektral yansıma eğrilerindeki farklılıkların analizi ile sağlanır. Sağlıklı domates yaprakları, yüksek klorofil içeriği nedeniyle görünür ışığı emerken, NIR bölgesinde yüksek yansıma gösterir. Buna karşın hastalıklı dokularda, klorofil yıkımı nedeniyle kırmızı banttaki yansıma artarken, hücresel hasar nedeniyle NIR yansıması düşer. Bu zıtlık, Normalize Edilmiş Fark Vejetasyon İndeksi (NDVI) gibi indekslerde belirgin bir düşüş olarak kendini gösterir(Sarvakar & Thakkar, 2024). Ayrıca, termal görüntüleme ile entegre edilen multispektral analizler, hastalıklı bitkilerin stoma kapanması sonucu artan yaprak sıcaklığını tespit ederek (transpirasyonun azalması), görsel belirtiler oluşmadan önce biyotik stresi ortaya çıkarabilmektedir(Scutelnic, Muradore, & Daffara, 2024).

5. AKILLI TARIM UYGULAMALARINDA MULTİSPEKTRAL VERİLERİN ANALİZİNDE KULLANILAN YAPAY ZEKÂ YÖNTEMLERİ

Multispektral görüntüleme ile elde edilen büyük ve karmaşık veri setlerinin işlenmesinde geleneksel istatistiksel yöntemler yetersiz kalabilmektedir. Bu noktada yapay zekâ (YZ) algoritmaları, hastalıkların otomatik tespiti ve sınıflandırılmasında devreye girmektedir.

5.1. Makine öğrenmesi ve derin öğrenme yaklaşımları

Makine öğrenmesi (ML) ve derin öğrenme (DL), tarımsal veri analizinde kullanılan iki temel YZ yaklaşımıdır. Geleneksel ML yöntemleri (Destek Vektör Makineleri - SVM, Rastgele Orman - RF, K-En Yakın Komşu - KNN), genellikle uzmanlar tarafından belirlenen ve el ile çıkarılan özelliklere (renk, doku,

şekil) dayanır. Örneğin, Tan ve arkadaşları (2021), domates yaprak hastalıklarının sınıflandırılmasında SVM, KNN ve RF algoritmalarını karşılaştırmış ve el ile çıkarılan özelliklerle SVM'nin yüksek başarı oranlarına ulaştığını rapor etmiştir(Tan vd., 2021). Ancak bu yöntemler, özellik çıkarımı aşamasında insan müdahalesine ihtiyaç duyar ve büyük veri setlerinde genelleme yetenekleri sınırlı olabilir(Tan vd., 2021). Derin öğrenme (DL) ise, özellikle Evrişimli Sinir Ağları (CNN) aracılığıyla, ham veriden özellikleri otomatik olarak öğrenme yeteneğine sahiptir. Bu sayede, karmaşık ve düzensiz arka planlara sahip tarla görüntülerinde bile hastalıkları yüksek doğrulukla tespit edebilir(De Silva & Brown, 2024).

5.2. Domates hastalıklarının sınıflandırılmasında kullanılan algoritmalar

Domates hastalıklarının sınıflandırılmasında en sık kullanılan DL mimarileri arasında VGG16, ResNet50, InceptionV3 ve DenseNet121 bulunmaktadır. Yapılan çalışmalarda, ResNet50 ve DenseNet121 gibi derin mimarilerin, %90'ın üzerinde doğruluk oranlarıyla domates hastalıklarını sınıflandırabildiği gösterilmiştir(Gatla vd., 2024). Örneğin, "E-TomatoDet" gibi özelleştirilmiş modeller, global ve lokal özellikleri birleştirerek domates yaprak hastalıklarını %97.2 doğrulukla tespit edebilmektedir(Sun vd., 2025). Ayrıca, YOLO (You Only Look Once) gibi nesne algılama algoritmaları, hastalıkların sadece varlığını değil, yaprak üzerindeki konumunu da gerçek zamanlı olarak belirleyebilmekte ve bu da robotik ilaçlama sistemleri için zemin hazırlamaktadır(Khan, Shen, & Liu, 2025).

5.3. Multispektral veri ile yapay zekânın entegrasyonu

Yapay zekâ modellerinin multispektral verilerle entegrasyonu, sadece RGB görüntülerine kıyasla daha zengin bir veri havuzu sunar. RGB görüntüler sadece yüzeyel renk

değişimlerini yakalarken, multispektral veriler bitkinin içsel fizyolojik durumunu yansıtan NIR ve Red-Edge bantlarını da içerir. Bu veriler, CNN veya Vision Transformer (ViT) gibi modellere girdi olarak verildiğinde, modelin ayırt etme gücü artar. Örneğin, De Silva ve Brown (2024), multispektral görüntülerle eğitilen ViT modellerinin, domates hastalıklarını tespit etmede CNN modellerine kıyasla daha yüksek performans (%89.92 doğruluk) gösterdiğini belirtmiştir. Ayrıca, spektral indekslerin (NDVI, NDRE) yapay zekâ modellerine ek özellik olarak sunulması, özellikle erken dönem hastalık tespitinde model hassasiyetini artırmaktadır(Zhong vd., 2025).

6. MULTİSPEKTRAL KAMERALAR KULLANILARAK DOMATES HASTALIK ANALİZİNİN AVANTAJLARI, SINIRLILIKLARI VE GELECEKTEKİ ARAŞTIRMA ALANLARI

6.1. Yöntemin tarımsal uygulamalardaki avantajları

Multispektral görüntüleme destekli hastalık analizi, geleneksel gözlem yöntemlerine göre devrim niteliğinde avantajlar sunar. En önemli avantajı, hastalıkların "asemptomatik" dönemde, yani insan gözüyle görülür belirtiler oluşmadan önce tespit edilebilmesidir. Bu erken uyarı sistemi, üreticilere enfeksiyon yayılmadan müdahale etme şansı verir, böylece kimyasal ilaç kullanımı azalır ve verim kayıpları minimize edilir(Zhang vd., 2024). Ayrıca, İHA'lar ile geniş alanların hızlı ve yüksek çözünürlüklü taranabilmesi, iş gücü maliyetlerini düşürür ve tarla genelinde homojen bir izleme sağlar(Radočaj, Radočaj, Plaščak, & Jurišić, 2025).

6.2. Teknik ve ekonomik sınırlılıklar

Teknolojinin sunduğu fırsatlara rağmen bazı sınırlılıklar mevcuttur. Teknik açıdan, değişen ışık koşulları, rüzgar ve karmaşık arka planlar (toprak, yabancı otlar), görüntü kalitesini ve model doğruluğunu olumsuz etkileyebilir. Ayrıca, multispektral ve özellikle hiperspektral verilerin işlenmesi yüksek hesaplama gücü gerektirir(Scutelnic vd., 2024). Ekonomik açıdan ise, multispektral sensörlerin ve bu verileri işleyecek donanımların maliyeti, küçük ölçekli çiftçiler için hala yüksek bir bariyer teşkil etmektedir. Veri analizi için gereken teknik uzmanlık ihtiyacı da teknolojinin yaygınlaşmasını zorlaştıran bir diğer faktördür.

6.3. Gelecekte yapılabilecek araştırma ve geliştirme alanları

Gelecekteki araştırmalar, daha uygun maliyetli ve kullanıcı dostu sensörlerin geliştirilmesine odaklanmalıdır. "Uç Yapay Zeka" (Edge AI) teknolojileri, verilerin buluta yüklenmeden cihaz üzerinde (örneğin drone üzerinde) işlenmesini sağlayarak gerçek zamanlı karar vermeyi hızlandırabilir(Gatla vd., 2024). Ayrıca, farklı sensör tiplerinin (termal, floresan, LiDAR ve multispektral) birleştirildiği "sensör füzyonu" çalışmaları, hastalık teşhisinin güvenilirliğini artıracaktır(Łągiewska & Panek-Chwastyk, 2025). Laboratuvar ortamında eğitilen modellerin gerçek tarla koşullarına adaptasyonunu sağlayan transfer öğrenme tekniklerinin geliştirilmesi de önemli bir araştırma alanıdır.

7. SONUÇ

Multispektral görüntüleme teknolojileri, akıllı tarım uygulamaları çerçevesinde domates hastalıklarının yönetiminde kritik bir dönüşüm sağlamaktadır. İnsan gözünün algı sınırlarını

aşan bu teknoloji, bitkilerin spektral imzalarındaki mikroskobik değişimleri analiz ederek hastalıkların erken evrede teşhis edilmesine olanak tanımaktadır. Fungal, bakteriyel ve viral hastalıkların yarattığı fizyolojik stresin, NIR ve Red-Edge gibi bantlarda yarattığı yansıma değişimleri, üreticilere proaktif bir yönetim imkanı sunar.

Yapay zekâ ve derin öğrenme algoritmalarının bu sürece entegrasyonu, veri analizini otomatikleştirerek insan hatasını ortadan kaldırmakta ve hastalık tespit doğruluğunu %99 seviyelerine kadar çıkarmaktadır. Bu teknolojiler, sadece verim kayıplarını önlemekle kalmayıp, agrokimyasal kullanımını optimize ederek çevresel sürdürülebilirliğe ve gıda güvenliğine doğrudan katkı sağlamaktadır. Gelecekte, sensör maliyetlerinin düşmesi ve otonom sistemlerin (robotik ve İHA) yaygınlaşması ile birlikte, multispektral analiz tabanlı hastalık yönetiminin modern tarımın standart bir parçası haline gelmesi beklenmektedir.

KAYNAKÇA

Abdulridha, J., Ampatzidis, Y., Kakarla, S. C., & Roberts, P. (2020). Detection of target spot and bacterial spot diseases in tomato using UAV-based and benchtop-based hyperspectral imaging techniques. *Precision Agriculture*, *21*(5), 955–978. doi:10.1007/s11119-019-09703-4

Ally, N. M., Neetoo, H., Ranghoo-Sanmukhiya, V. M., & Coutinho, T. A. (2023). Greenhouse-grown tomatoes: microbial diseases and their control methods: a review.

Amigo, J. M. (2019). Chapter 1.1 - Hyperspectral and multispectral imaging: setting the scene. Içinde J. M. B. T.-D. H. in S. and T. Amigo (Ed.), *Hyperspectral Imaging* (C. 32, ss. 3–16). Elsevier. doi:https://doi.org/10.1016/B978-0-444-63977-6.00001-8

Amigo, J. M., & Grassi, S. (2019). Chapter 1.2 - Configuration of hyperspectral and multispectral imaging systems. Içinde J. M. B. T.-D. H. in S. and T. Amigo (Ed.), *Hyperspectral Imaging* (C. 32, ss. 17–34). Elsevier. doi:https://doi.org/10.1016/B978-0-444-63977-6.00002-X

Berger, K., Machwitz, M., Kycko, M., Kefauver, S. C., Van Wittenberghe, S., Gerhards, M., … Schlerf, M. (2022). Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review. *Remote Sensing of Environment*, *280*, 113198. doi:10.1016/j.rse.2022.113198

Catara, V., & Bella, P. (2020). Bacterial diseases. Içinde *Integrated pest and disease management in greenhouse crops* (ss. 33–54). Springer.

Chhabra, R., Kaur, S., Vij, L., & Gaur, K. (2020). Exploring

physiological and biochemical factors governing plant pathogen interaction: A review. *Int. J. Curr. Microbiol. App. Sci*, *9*(11), 1650–1666.

Daniels, L., Eeckhout, E., Wieme, J., Dejaegher, Y., Audenaert, K., & Maes, W. H. (2023). Identifying the optimal radiometric calibration method for UAV-based multispectral imaging. *Remote Sensing*, *15*(11), 2909.

de Castro, A. I., Shi, Y., Maja, J. M., & Peña, J. M. (2021). UAVs for Vegetation Monitoring: Overview and Recent Scientific Contributions. *Remote Sensing*, *13*(11), 2139. doi:10.3390/rs13112139

De Silva, M., & Brown, D. (2024). *Tomato Disease Detection Using Multispectral Imaging with Deep Learning Models*. Içinde *2024 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)* (ss. 1–9). IEEE. doi:10.1109/icABCD62167.2024.10645256

Filella, I., & Penuelas, J. (1994). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *International Journal of Remote Sensing*, *15*(7), 1459–1470. doi:10.1080/01431169408954177

Gatla, A., Reddy, S. R. V. P., Mandru, D., Thouti, S., Kavitha, J., Souissi, A. S. E., … Flah, A. (2024). Optimizing Edge AI for Tomato Leaf Disease Identification. *Engineering, Technology & Applied Science Research*, *14*(4), 16061–16068.

Getahun, S., Kefale, H., & Gelaye, Y. (2024). Application of Precision Agriculture Technologies for Sustainable Crop Production and Environmental Sustainability: A Systematic Review. *The Scientific World Journal*, *2024*(1). doi:10.1155/2024/2126734

Jafar, A., Bibi, N., Naqvi, R. A., Sadeghi-Niaraki, A., & Jeong, D. (2024). Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations. *Frontiers in Plant Science*, *15*. doi:10.3389/fpls.2024.1356260

Khan, Z., Shen, Y., & Liu, H. (2025). Objectdetection in agriculture: A comprehensive review of methods, applications, challenges, and future directions. *Agriculture*, *15*(13), 1351.

Łągiewska, M., & Panek-Chwastyk, E. (2025). Integrating remote sensing and autonomous robotics in precision agriculture: Current applications and workflow challenges. *Agronomy*, *15*(10), 2314.

Lu, B., Dao, P., Liu, J., He, Y., & Shang, J. (2020). Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sensing*, *12*(16), 2659. doi:10.3390/rs12162659

Lu, J., Ehsani, R., Shi, Y., de Castro, A. I., & Wang, S. (2018). Detection of multi-tomato leaf diseases (late blight, target and bacterial spots) in different stages by using a spectral-based sensor. *Scientific Reports*, *8*(1), 2793. doi:10.1038/s41598-018-21191-6

Muhammad, A., Khan, Z. U., Khan, J., Mashori, A. S., Ali, A., Jabeen, N., … Li, F. (2025). A comprehensive review of crop stress detection: destructive, non-destructive, and ML-based approaches. *Frontiers in Plant Science*, *16*. doi:10.3389/fpls.2025.1638675

Nazarov, P. A., Baleev, D. N., Ivanova, M. I., Sokolova, L. M., & Karakozova, M. V. (2020). Infectious plant diseases: etiology, current status, problems and prospects in plant protection. *Acta Naturae*, *12*(3), 46–59.

doi:10.32607/actanaturae.11026

Nguyen, C., Sagan, V., Maimaitiyiming, M., Maimaitijiang, M., Bhadra, S., & Kwasniewski, M. T. (2021). Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning. *Sensors*, *21*(3), 742. doi:10.3390/s21030742

Pourazar, H., Samadzadegan, F., & Dadrass Javan, F. (2019). Aerial multispectral imagery for plant disease detection: radiometric calibration necessity assessment. *European Journal of Remote Sensing*, *52*(sup3), 17–31. doi:10.1080/22797254.2019.1642143

Radočaj, D., Radočaj, P., Plaščak, I., & Jurišić, M. (2025). Evolution of Deep Learning Approaches in UAV-Based Crop Leaf Disease Detection: A Web of Science Review. *Applied Sciences*, *15*(19), 10778.

Rehman, M., Pan, J., Mubeen, S., Ma, W., Luo, D., Cao, S., … Chen, P. (2024). Morpho-physio-biochemical, molecular, and phytoremedial responses of plants to red, blue, and green light: a review. *Environmental Science and Pollution Research*, *31*(14), 20772–20791. doi:10.1007/s11356-024-32532-6

Sanoubar, R., & Barbanti, L. (2017). Fungal diseases on tomato plant under greenhouse condition. *European Journal of Biological Research*, *7*(4), 299–308.

Sarvakar, K., & Thakkar, M. (2024). Different Vegetation Indices Measurement Using Computer Vision BT - Applications of Computer Vision and Drone Technology in Agriculture 4.0. Içinde S. S. Chouhan, U. P. Singh, & S. Jain (Ed.) (ss. 133–163). Singapore: Springer Nature Singapore. doi:10.1007/978-981-99-8684-2_9

Scutelnic, D., Muradore, R., & Daffara, C. (2024). A

multispectral camera in the VIS–NIR equipped with thermal imaging and environmental sensors for non invasive analysis in precision agriculture. *HardwareX*, *20*, e00596.

Sevik, M. A., & Arli-Sokmen, M. (2012). Estimation of the effect of Tomato spotted wilt virus (TSWV) infection on some yield components of tomato. *Phytoparasitica*, *40*(1), 87–93. doi:10.1007/s12600-011-0192-2

Sharma, K., & Shivandu, S. K. (2024). Integrating artificial intelligence and Internet of Things (IoT) for enhanced crop monitoring and management in precision agriculture. *Sensors International*, *5*, 100292. doi:10.1016/j.sintl.2024.100292

Sun, H., Fu, R., Wang, X., Wu, Y., Al-Absi, M. A., Cheng, Z., … Sun, Y. (2025). Efficient deep learning-based tomato leaf disease detection through global and local feature fusion. *BMC Plant Biology*, *25*(1), 311.

Tan, L., Lu, J., & Jiang, H. (2021). Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. *AgriEngineering*, *3*(3), 542–558.

Verma, S., Chug, A., & Singh, A. P. (2018). *Prediction Models for Identification and Diagnosis of Tomato Plant Diseases*. Içinde *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (ss. 1557–1563). IEEE. doi:10.1109/ICACCI.2018.8554842

Xu, H. R., Ying, Y. B., Fu, X. P., & Zhu, S. P. (2007). Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. *Biosystems Engineering*, *96*(4), 447–454.

Zhang, X., Vinatzer, B. A., & Li, S. (2024). Hyperspectral

imaging analysis for early detection of tomato bacterial leaf spot disease. *Scientific Reports*, *14*(1), 27666. doi:10.1038/s41598-024-78650-6

Zhong, X., Li, H., Cai, Y., Deng, Y., Xu, H., Tian, J., … Wang, L. (2025). Early Detection of Tomato Gray Mold Based on Multispectral Imaging and Machine Learning. *Horticulturae*, *11*(9), 1073.

TÜRKİYE VE DÜNYADA

BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ

yaz
yayınları

YAZ Yayınları
M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar / AFYONKARAHİSAR
Tel : (0 531) 880 92 99
yazyayinlari@gmail.com • www.yazyayinlari.com