HARİTA MÜHENDİSLİĞİ DEĞERLENDİRMELERİ

Editör: Prof.Dr. Mustafa YILMAZ

yaz
yayınları

Harita Mühendisliği Değerlendirmeleri

Editör

Prof.Dr. Mustafa YILMAZ

yaz yayınları

2025

Harita Mühendisliği Değerlendirmeleri

Editör: Prof.Dr. Mustafa YILMAZ

© YAZ Yayınları

Bu kitabın her türlü yayın hakkı Yaz Yayınları'na aittir, tüm hakları saklıdır. Kitabın tamamı ya da bir kısmı 5846 sayılı Kanun'un hükümlerine göre, kitabı yayınlayan firmanın önceden izni alınmaksızın elektronik, mekanik, fotokopi ya da herhangi bir kayıt sistemiyle çoğaltılamaz, yayınlanamaz, depolanamaz.

E_ISBN 978-625-5838-85-8

Ekim 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

info@yazyayinlari.com

İÇİNDEKİLER

Education–Sector Interaction in the Map Sector: A Quantitative Study on Student and Graduate Expectations.....1
Fuat BAŞÇİFTÇİ, Sevgi BÖGE

Chlorophyll-A Normalised Fluorescence Line Height and Particulate Organic Carbon Analysis in Marmara Sea Coasts: Environmental Monitoring and Assessment.....21
Nehir UYAR

Heuristically Optimized Speech Recognition Models Enhanced by Synthetic Data Augmentation for GIS Applications 42
Ahmet Emin KARKINLI, Erkan BEŞDOK

"Bu kitapta yer alan bölümlerde kullanılan kaynakların, görüşlerin, bulguların, sonuçların, tablo, şekil, resim ve her türlü içeriğin sorumluluğu yazar veya yazarlarına ait olup ulusal ve uluslararası telif haklarına konu olabilecek mali ve hukuki sorumluluk da yazarlara aittir."

EDUCATION–SECTOR INTERACTION IN THE MAP SECTOR: A QUANTITATIVE STUDY ON STUDENT AND GRADUATE EXPECTATIONS

Fuat BAŞÇİFTÇİ[1]

Sevgi BÖGE[2]

1. INTRODUCTION

The surveying sector has undergone a rapid transformation process in recent years due to technological developments. Innovative technologies such as UAVs (Unmanned Aerial Vehicles), GNSS (Global Navigation Satellite Systems), LIDAR (Light Detection and Ranging), GIS (Geographic Information Systems), and artificial intelligence applications are reshaping the way business is conducted in the sector (Erkek and Cankurt, 2019; Breunig et al., 2020). This transformation directly affects the qualifications and training processes of technical personnel who will work in the sector. In this context, the perceptions, expectations, and level of adaptation to technology of students enrolled in the Map and Cadastre Program at vocational schools are of critical importance in terms of sectoral development (Başçiftçi and Böge, 2021).

Vocational schools are educational institutions that focus on practical training and aim to train intermediate-level personnel for the sector (Kaysi and Gürol, 2017; SUBÜ, 2023). The attitudes of

[1] Assoc. Prof. Dr., Karamanoglu Mehmetbey University, Vocational School of Technical Sciences, Map and Cadastre Programme, fuatbasciftci@kmu.edu.tr, ORCID: 0000-0002-5791-0676.

[2] Lecturer, Selcuk University, Kadinhani Faik İcil Vocational School, Map and Cadastre Programme, sboge@selcuk.edu.tr, ORCID: 0000-0001-6159-9721.

students enrolled in these programs toward the sector play a decisive role in shaping both education policies and partnerships with the sector (Yıldırım and Şimşek, 2018; Vuoriainen et al., 2025). Students' internship experiences, practical course content, and interactions with industry representatives are among the factors that directly influence their professional development. Therefore, systematic analysis of students' sectoral expectations is necessary for both academic and sectoral planning (Böge and Başçiftçi, 2025).

Students' perceptions of the sector are not limited to technical qualifications. Broader issues such as working conditions, employment opportunities, career development, social rights, and gender equality also play a role in shaping these perceptions (Atlı and Tulunay Ateş, 2023). In particular, female students' assessments of equal opportunities in the sector provide important data for questioning the effectiveness of gender policies in the sector. In this context, one of the main objectives of the study is to examine the impact of demographic variables (gender, age, class level, graduation status) on sectoral perceptions.

Perceptions of technological developments are another important factor that determines students' professional vision and potential contribution to the sector. Being open to new technologies, adapting to digital transformation in the sector, and demanding access to continuous education opportunities are factors that directly affect students' professional motivation and integration into the sector. Therefore, a detailed analysis of students' attitudes toward technology will be guiding in terms of bringing sectoral innovations together with education.

This study aims to reveal the sectoral expectations of students in the Map and Cadastre Program, their perceptions of technological developments, and their relationships with demographic variables. Data obtained from 361 participants were

evaluated using descriptive statistics, reliability analysis, thematic classifications, and demographic breakdowns. The research offers both educational institutions and industry representatives the opportunity to better understand the student profile and shape their strategic planning accordingly. Additionally, this study comprehensively evaluates the perceptions of students studying in the field of Map and Cadastre regarding the sector, providing a thorough assessment in terms of educational policies, sector strategies, and technological adaptation processes. The findings establish a data-driven foundation for sectoral development and educational reforms while also making students' voices visible in the academic sphere.

2. METHOD

This study was conducted based on quantitative research methods within the scope of a descriptive and correlational survey model. The descriptive approach aims to reveal students' sectoral expectations and perceptions of technology in their current state, while the correlational approach aims to analyze the meaningful relationships between these variables and demographic differences (Karasar, 2018).

2.1. Target Population and Study Sample

The population of the study consists of a total of **361 participants** who are either second-year (2^{nd} year) students or graduates of the Map and Cadastre Program at various vocational schools affiliated with universities in Türkiye. The demographic distribution of the participants is as follows:

- **Gender:** 44.32% female, 55.68% male
- **Age group:** 35.18% are between 18–21 years old, 31.30% are between 22–25 years old, 24.93% between 26–30 years old, and 8.59% are aged 30 and above

- **Educational status:** 40.17% are 2nd year students, while 59.83% are graduates

Reason for preference: The most common reason was "Because it offers good job opportunities" (34.35%)

2.2. Data Collection Tool

A questionnaire was used as the data collection tool in this study. The questionnaire consists of two main sections. The first section includes five questions aimed at identifying the demographic distribution of the participants. The second section includes an 18-item Likert-type questionnaire developed by the researchers and validated through expert opinions to ensure content validity. The items were structured according to a 5-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree) to assess the participants' opinions. The 5-point Likert scale items were rated as follows: Strongly Agree (4.21–5.00), Agree (3.41–4.20), Neutral (2.61–3.40), Disagree (1.81–2.60), and Strongly Disagree (1.00–1.80). The second section of the questionnaire focuses on five sub-themes:

1. Expectations Regarding Education
2. Technological Expectations
3. Working Conditions and Career Expectations
4. Sector Expectations
5. Future Expectations

2.3. Data Analysis

The survey data were analyzed using the SPSS statistical software. A significance level of 0.05 was adopted as the criterion for statistical tests. To assess the reliability of the data in the second part of the questionnaire, Cronbach's Alpha coefficient was calculated. The reliability coefficient obtained was 0.868, indicating a high level of internal consistency. Based on this

value, the scale can be considered highly reliable. An independent samples t-test was conducted to determine whether there were significant differences between 2nd year students and graduates in terms of their expectations regarding the map sector. To analyze participants' responses, frequency distributions, standard deviations, and arithmetic means were calculated and presented in tabular form. The overall mean score for participants' opinions and expectations in the second part of the questionnaire was $\overline{X}$ = 3.829. This result indicates that both 2nd year and graduate students expressed a high level of expectations and positive views toward the map sector.

3. FINDINGS

This section presents the statistical analyses, thematic patterns, and demographic breakdowns derived from the survey data collected from 2nd year and graduate students in the Map and Cadastre Program. The analyses are supported by descriptive statistics and reliability measurements to ensure the robustness of the findings.

3.1. Descriptive Statistics

To reveal the general tendencies of the participants, descriptive statistics—namely the mean ($\overline{X}$) and standard deviation (std)—for each item in the second part of the questionnaire are presented in Table 1.

Table 1. Descriptive Statistics of Participants' Responses to Survey Items (N = 361)

Scale Items	$\overline{X}$	Results	std
6. I believe that educational support from the sector in professional areas would have a positive impact.	4.17	"Agree"	0.94
7. The knowledge gained during the internship is effective in developing professional skills.	4.18		1.03

8. In terms of sectoral alignment, I believe that practical training is more advantageous.	4.56	"Strongly agree"	0.85
9. Sector representatives should engage in greater collaboration with vocational schools.	4.46		0.84
10. The sector needs to adopt emerging technologies—such as UAVs, GNSS, LiDAR, GIS technologies, and artificial intelligence applications—more extensively	3.35	"Neutral"	0.97
11. Continuous training opportunities should be established to keep pace with technological innovations.	4.46	"Strongly agree"	0.84
12. I believe that the working conditions in the sector (such as occupational safety, career development, social benefits, salary, and work-life balance) are satisfactory	2.89	"Neutral"	1.23
13. I believe that the title of "Map and Cadastre Technician", which I have earned (or will earn), has positive aspects within the sector	3.81	"Agree"	1.00
14. I believe that the sector should support its personnel in order to participate in international projects.	4.40	"Strongly agree"	0.78
15. I believe that employment opportunities after graduation are at a sufficient level	2.97	"Neutral"	1.27
16. I believe that the workload in the field is distributed in a balanced manner.	3.06		1.10
17. The sector provides equal opportunities for the development of both male and female personnel	2.51	"Disagree"	1.30
18. I believe that career day events should be organized where experienced professionals in the sector meet with students	4.18	"Agree"	0.89
19. Recruitment processes in the sector are conducted in a transparent manner.	3.26	"Neutral"	1.28
20. I do not appreciate being asked to work overtime when there is a shortage of staff in the institution or company I work for	3.65	"Agree"	1.14

21. Public institutions should create more employment opportunities for map and cadastre technicians	4.46	"Strongly agree"	0.90
22. Networking within the sector should be strengthened	4.31		0.84
23. Criticism of my education, contributions to the company, and dedication has a negative effect on me.	3.24	"Neutral"	1.28
Overall Average	**3.83**	**"Agree"**	**1.03**

An examination of Table 1 reveals the perspectives of the 361 students who participated in the survey regarding the map and cadastre sector, and provides insight into their sectoral expectations. The overall mean score was $\overline{X} = 3.83$, corresponding to the 'Agree' level. This finding indicates that participants generally exhibit positive attitudes toward the sector.

Among the items with the highest levels of agreement, the following statements stand out: "In terms of sectoral adaptation, I believe that practical training is more advantageous" ($\overline{X} = 4.56$, std = 0.85); "Continuous training opportunities should be established to keep pace with technological innovations" ($\overline{X} = 4.46$, std = 0.84); and "Sector representatives should engage in greater collaboration with vocational schools" ($\overline{X} = 4.46$, std = 0.84). These findings suggest that students expect greater support in adapting to the sector and accessing current technologies.

Nevertheless, it is noteworthy that students expressed ambivalent or negative opinions regarding "working conditions in the sector" ($\overline{X} = 2.89$, std = 1.23), "employment opportunities" ($\overline{X} = 2.97$, std = 1.27), "transparency in recruitment processes" ($\overline{X} = 3.26$, std = 1.28), and "gender equality" ($\overline{X} = 2.51$, std = 1.30). In particular, the low mean score for gender equality indicates a perceived lack of equal opportunities between male and female professionals in the sector.

The mean score for the statement "the sector should adopt new technologies more widely" ($\overline{X} = 3.35$, std = 0.97) corresponds

to a “neutral" response level, which may indicate that students are not sufficiently aware of existing technological practices in the field. In this regard, it is essential to introduce technological advancements in the sector more effectively to students and to ensure their integration into academic curricula.

Additionally, participants expressed a high level of agreement with the statements “Public institutions should create more employment opportunities for map and cadastre technicians” ($\overline{X}$ = 4.46, std = 0.90) and “The sector should support its personnel to participate in international projects” ($\overline{X}$ = 4.40, std = 0.78). These findings highlight students’ expectations regarding public sector employment and career opportunities at the international level. The relatively high standard deviation values observed in certain items (e.g. item 15: std = 1.27; item 17: std = 1.30) indicate notable divergences in participants’ responses. Such variation may suggest that the student population is heterogeneous in terms of experience, knowledge level, or personal expectations.

These findings clearly reveal the need for stronger collaboration between the sector and educational institutions, the importance of technological and practice-oriented training, demands for improved employment opportunities, and persistent issues related to gender balance within the sector. In this context, enhancing cooperation between the sector and academic institutions, facilitating students’ access to technological advancements, and developing inclusive policies are of critical importance for sectoral development.

The results of the independent samples t-test—conducted to examine whether there are differences in the views and expectations of second-year and graduate students in the Map and Cadastre Program regarding the map sector—are presented in

Table 2. The table includes mean scores ($\overline{X}$), standard deviations (std), t-test values (t), and significance levels (p).

Table 2. The views and expectations of 2nd year and graduate students regarding the map sector

Variables	Groups	N	$\overline{X}$	std	t test		
					t	df	p
6. I believe that the sector's support for professional education would make a positive contribution.	2nd year	145	4.01	.92	-2.812	359	.005
	Graduate	216	4.29	.93			
7. The knowledge acquired during the internship period is effective in the development of professional skills.	2nd year	145	4.10	1.04	-1.118	359	.264
	Graduate	216	4.23	1.02			
8. In terms of adaptation to the sector, I believe that practice-oriented education is more advantageous.	2nd year	145	4.29	1.01	-5.125	359	.000
	Graduate	216	4.74	.66			
9.Sector representatives should engage in greater collaboration with vocational schools.	2nd year	145	4.19	1.01	-5.082	359	.000
	Graduate	216	4.63	.66			
10. The sector needs to adopt emerging technologies (such as UAVs, GNSS, LiDAR, GIS technologies, and artificial intelligence applications) more extensively.	2nd year	145	4.01	1.15	-5.657	359	.000
	Graduate	216	4.58	.75			
11.Continuous training opportunities should be established to keep pace with technological innovations.	2nd year	145	4.17	1.02	-5.489	359	.000
	Graduate	216	4.65	.62			
12. I believe that the working conditions in the sector (such as occupational safety, career development,	2nd year	145	3.25	1.08	4.679	359	.000
	Graduate	216	2.65	1.27			

social benefits, salary, and work-life balance) are satisfactory.							
13. I believe that the title I have earned "Map and Cadastre Technician" has positive aspects within the sector.	2nd year	145	3.92	.89	1.608	359	.109
	Graduate	216	3.75	1.06			
14. I believe that the sector should support its personnel to participate in international projects.	2nd year	145	4.15	.88	-5.251	359	.000
	Graduate	216	4.57	.65			
15. I believe that employment opportunities after graduation are at a sufficient level.	2nd year	145	3.25	1.09	3.475	359	.001
	Graduate	216	2.78	1.34			
16. I believe that the workload in the field is distributed in a balanced manner.	2nd year	145	3.24	1.02	2.582	359	.010
	Graduate	216	2.94	1.13			
17. The sector provides equal opportunities for the development of both male and female personnel.	2nd year	145	2.92	1.31	5.072	359	.000
	Graduate	216	2.23	1.22			
18. I believe that experienced professionals in the sector should organize career day meetings with students.	2nd year	145	4.06	.94	-2.112	359	.035
	Graduate	216	4.26	.85			
19. Recruitment processes in the sector are conducted in a transparent manner.	2nd year	145	3.33	1.14	.859	359	.391
	Graduate	216	3.21	1.37			
20. I do not appreciate being asked to work overtime when there is a shortage of staff in my institution/company.	2nd year	145	3.68	1.09	.300	359	.764
	Graduate	216	3.64	1.18			
21. Public institutions should create more employment opportunities for map and cadastre technicians	2nd year	145	4.22	.94	-4.196	359	.000
	Graduate	216	4.62	.83			

22. Networking within the sector should be strengthened.	2nd year	145	4.08	.93	-4.307	359	.000
	Graduate	216	4.46	.74			
23. Criticism of my education, contributions to the company, and dedication affects me negatively.	2nd year	145	3.32	1.28	.994	359	.321
	Graduate	216	3.18	1.28			

Analysis of Table 2 reveals that 2nd year students tend to exhibit a positive and idealistic outlook toward the map sector and its future prospects.

- Compared to graduates, they reported higher levels of satisfaction regarding working conditions, workload balance, employment opportunities, and gender equality in the sector.
- They strongly support the view that internship duration contributes to the development of professional skills and believe that the technician title has positive aspects within the sector.
- They gave high scores on issues such as educational support from the sector, use of new technologies, continuous learning opportunities, and cooperation between the sector and vocational schools. However, compared to graduates, their level of expectation remained lower.

This table shows that 2nd year students, having not yet experienced the sector intensively, do not fully perceive the challenges of professional life and therefore maintain higher expectations.

Graduate students have a more critical and realistic perspective shaped by their sectoral experiences.

- Compared to 2nd year students, graduates reported higher expectations regarding professional training support from

the sector, the importance of practical education, wider use of new technologies (e.g. UAVs, GNSS, Lidar, GIS, artificial intelligence), creation of continuous learning opportunities, support for participation in international projects, and strengthening of intra-sectoral communication networks.

- In contrast, they expressed lower levels of satisfaction regarding working conditions, employment opportunities, workload balance, and equal opportunity.
- Regarding the contribution of internship duration, the positive aspects of the technician title, transparency in recruitment processes, and attitudes toward overtime work, their views are like those of 2nd year students

This situation indicates that graduates have a clearer awareness of the challenges, shortcomings, and areas for improvement they encounter in professional life.

3.2. Reliability Analysis

To evaluate the internal consistency of the second part of the survey instrument employed in this study, Cronbach's Alpha was computed. This section comprises 18 items aimed at capturing the sector-related perceptions and expectations of both 2nd year students and graduates enrolled in map and cadastre programs. The overall Cronbach's Alpha coefficient obtained was 0.868, indicating a high level of reliability for the questionnaire. In the literature, Cronbach's Alpha values of 0.70 and above are generally considered acceptable, while values exceeding 0.80 are interpreted as reflecting high internal consistency (Tabachnick and Fidell, 2013). The reliability values for the five subthemes addressed in the second section of the questionnaire:

- Educational Expectations: $\alpha = 0.826$
- Technological Expectations: $\alpha = 0.850$
- Working Conditions and Career Expectations: $\alpha = 0.717$
- Sectoral Expectations: $\alpha = 0.512$
- Future Expectations: $\alpha = 0.584$

When the reliability values of the thematic subdimensions of the questionnaire are examined:

- The $\alpha = 0.826$ value for **the Educational Expectations** subdimension indicates that this theme has high reliability.
- **The Technological Expectations** theme also has a highly reliable structure with an a value of 0.850.
- **The Work Conditions and Career Expectations** subtheme yielded a value of 0.717, which is considered acceptable in terms of reliability.

In contrast, the Cronbach's Alpha values for **the Sectoral Expectations** ($\alpha = 0.512$) and **Future Expectations** ($\alpha = 0.584$) subthemes fall within the range generally considered low.

4. FINDINGS

This study has revealed the sectoral expectations of students and graduates of the Map and Cadastre Program, their attitudes toward technology, and the relationships with demographic variables. The findings generally indicate that participants hold a positive attitude toward the sector, yet highlight significant areas for improvement in working conditions, employment opportunities, gender equality, and technological awareness.

The overall mean score obtained in the study corresponds to the “Agree” level, reflecting students’ high motivation toward their profession and the sector. In particular, the high agreement rates regarding the importance of practical training, strengthening sector–vocational school collaboration, and expanding lifelong learning opportunities support the need for sector–education integration (Jackson, 2016; Tereshchenko et al., 2024). This indicates that students consider practical training and professional development opportunities critical in the process of adapting to the sector.

However, the neutral or negative evaluations regarding working conditions, employment opportunities, and the transparency of recruitment processes suggest that the challenges faced by graduates in professional life are not adequately anticipated by students. Indeed, the t-test results indicate that graduates report lower levels of satisfaction in these areas, suggesting that sectoral experience leads to more realistic expectations. Similarly, the more negative perceptions expressed by graduates regarding gender equality reveal that existing issues related to equal opportunities for men and women in the sector are more clearly observed in practice. This result indicates the need to develop policies aimed at increasing the representation of women in the sector and ensuring equal access to opportunities (Powell et al., 2009; Parmaxi et al., 2024).

The findings related to technological innovation indicate that participants value the use of technology but perceive the current level as insufficient. In particular, the high level of agreement regarding the need for broader adoption of emerging technologies (e.g. UAVs, GNSS, LIDAR, GIS, artificial intelligence) suggests that the sector should offer more opportunities to students within the digital transformation process. In this context, it is recommended that curriculum

content be aligned with current technologies and supported through sectoral collaborations (Schwab, 2024; Zou et al., 2025).

This study reveals the need to strengthen education–sector collaboration in the Map and Cadastre field, enhance technological infrastructure and practical training opportunities, improve working conditions, and ensure gender equality. The findings offer both educational institutions and sector representatives an opportunity to engage in strategic planning that considers the needs and expectations of students and graduates.

5. CONCLUSION AND RECOMMENDATIONS

This study has comprehensively revealed the sectoral expectations of students and graduates of the Map and Cadastre Program, their attitudes toward technological developments, and the relationships with demographic variables. The findings generally indicate that participants have a positive outlook toward the sector; however, there is a need for improvement particularly in areas such as working conditions, employment opportunities, gender equality, and technological awareness. According to the study results, participants strongly emphasized that practical training plays a critical role in professional adaptation and that collaboration between the sector and vocational schools should be strengthened. Graduates showed higher levels of agreement regarding the need for wider adoption of new technologies (e.g. UAVs, GNSS, LIDAR, GIS, artificial intelligence) and support for participation in international projects. Conversely, the more negative perceptions of graduates concerning working conditions, employment opportunities, and gender equality suggest that the challenges encountered in professional life are more clearly observed within this group.

The reliability analyses of the survey indicated that the subdimensions “Expectations Education” and “Technological

Expectations" demonstrated high internal consistency. However, low reliability values were obtained for the subdimensions "Sectoral Expectations" and "Future Expectations."

Based on the findings of the research, the following recommendations may be proposed:

1. Strengthening Education–Sector Collaboration: Regular partnerships should be established between vocational schools and sector representatives through internship programs, practical training projects, and career events. Such initiatives can better equip students to align with sectoral expectations and enhance their professional readiness.
2. Enhancing Technology Integration: Curriculum content should be aligned with the sector's digital transformation process, and technologies such as UAVs, GNSS, LIDAR, GIS, and artificial intelligence should be actively incorporated into classroom and laboratory settings.
3. Improving Working Conditions: Occupational safety, social benefits, work–life balance, and career development opportunities should be strengthened within the sector. These improvements may encourage long-term employment of graduates and enhance workforce retention.
4. Promoting Gender Equality: Gender equality policies should be established within the sector to ensure equal opportunities for both male and female employees, and these policies should be monitored regularly.
5. Expanding Employment Opportunities: Public and private sector institutions should increase employment opportunities for map and cadastre technicians, and

ensure transparency and objective criteria in recruitment processes for graduates.

In conclusion, this study presents significant findings that serve as a bridge between the map and cadastre sector and vocational education institutions. The results offer guidance for enhancing the qualifications of human resources in the sector, accelerating technological adaptation, and implementing equitable employment policies.

REFERENCES

Atlı, K. M., Tulunay Ateş, Ö. (2023). Examination of Social Gender Equality Studies. *Marmara University Journal of Women and Gender Studies*, 7(2), 21–36. https://doi.org/10.29228/mukatcad.33 (in Turkish)

Başçiftçi, F., ve Böge, S. (2021). A Research on the Expectations of the Map Sector from Map and Cadastre Technician. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences,* 8(14), 175-189. (in Turkish)

Böge S., Başçiftçi, F., (2025). İşletmede Mesleki Eğitim (3+1) Modelinin Meslek Yüksekokullarının Harita ve Kadastro Programı Öğrencilerinin Mesleki Gelişimlerine Etkisi. *Mühendislikte Trend Akademik Çalışmalar,* ss. 6-15, ISBN: 978-625-5954-92-3

Breunig, M., Bradley, P. E., Jahn, M., Kuper, P., Mazroob, N., Rösch, N., Al-Doori, M., Stefanakis, E., Jadidi, M. (2020). Geospatial data management research: Progress and future directions. *ISPRS International Journal of Geo-Information*, 9(2), 95. https://doi.org/10.3390/ijgi9020095

Erkek, B., Cankurt, İ. (2019). Harita ve Kadastro Sektörünün Geleceği ve Geliştirilecek Politikalar. *17. Türkiye Harita Bilimsel ve Teknik Kurultayı*, Ankara. https://www.tkgm.gov.tr/sites/default/files/2020-10/1_8841_full.pdf

Jackson, D. (2016). Re-conceptualising graduate employability: The importance of pre-professional identity. *Higher Education Research & Development*, 35(5), 925-939. https://doi.org/10.1080/07294360.2016.1139551

Karasar, N. (2018). Bilimsel Araştırma yöntemi. (33. Bsm.) Ankara: Nobel Akademik Yayıncılık

Kaysi, F., Gürol, A. (2017). Assessment of Workplace Practical Training Processes of Vocational School Students. *The Journal of Kesit Academy*, 8, 266–280. (in Turkish)

Parmaxi, A., Christou, E., Valdés, J. F., Hevia, D. M. P., Perifanou, M., Econimides A. A., Maaj, J., Manchenko, M. (2024). Gender equality in science, technology, engineering and mathematics: Industrial vis-a-vis academic perspective. *Discover Education*, 3(3). https://doi.org/10.1007/s44217-023-00082-7

Powell, A., Bagilhole, B., Dainty, A. (2009). How women engineers do and undo gender: Consequences for gender equality. *Gender, Work & Organization*, 16(4), 411–428. https://doi.org/10.1111/j.1468-0432.2008.00406.x

Sakarya Uygulamalı Bilimler Üniversitesi (SUBÜ). (2023). 3+1 Uygulamalı Eğitim Modeli. https://meyok.subu.edu.tr/tr/31-uygulamali-egitim-modeli

Schwab, K. (2024). The Fourth Industrial Revolution: what it means, how to respond1. *In Handbook of research on strategic leadership in the Fourth Industrial Revolution* (pp. 29-34). Edward Elgar Publishing.

Tabachnick, B. G., Fidell, L. S. (2013). Using Multivariate Statistics (C. Campanella, Ed.; 6. bs). Pearson Education

Tereshchenko, E., Salmela, E., Melkko, E., Phang, S. K., Happonen, A. (2024). Emerging best strategies and capabilities for university–industry cooperation. *Journal of Innovation and Entrepreneurship*, 13(28). https://doi.org/10.1186/s13731-024-00386-4

Vuoriainen, A., Rikala, P., Heilala, V., Lehesvuori, S., Oz, S., Kettunen, L., & Hämäläinen, R. (2025). The six C's of successful higher education-industry collaboration in

engineering education: a systematic literature review. *European Journal of Engineering Education*, 50(1), 26-50. https://doi.org/10.1080/03043797.2024.2432440

Yıldırım, A., Şimşek, H. (2018). Sosyal bilimlerde nitel araştırma yöntemleri. Seçkin Yayıncılık.

Zou, Y., Kuek, F., Feng, W., Cheng, X. (2025). Digital learning in the 21st century: Trends, challenges, and innovations. *Frontiers in Education*, 10. https://doi.org/10.3389/feduc.2025.1562391

CHLOROPHYLL-A NORMALISED FLUORESCENCE LINE HEIGHT AND PARTICULATE ORGANIC CARBON ANALYSIS IN MARMARA SEA COASTS: ENVIRONMENTAL MONITORING AND ASSESSMENT

Nehir UYAR[1]

1. INTRODUCTION

Marine sediments cover most of the Earth's surface, creating one of our planet's largest and most crucial living environments. This vast, often-overlooked habitat is home to an enormous and incredibly varied population of microscopic life, primarily Bacteria and Archaea. These simple, single-celled organisms are far from insignificant; they are essential for keeping the global ecological system in balance.

The sheer quantity of this microbial life is astonishing. Early work by researchers like Whitman et al. (1998) suggested that prokaryotes living in these deep marine sediments could make up anywhere from half to more than the previously estimated total prokaryotic biomass on Earth. More conservative, yet still impressive, findings from subsequent studies (e.g., Parkes et al., 2000) estimate that these sediment-dwellers account for roughly 10% to 33% of all living biomass on the planet.

[1] Lecturer, Zonguldak Bülent Ecevit University, Zonguldak Vocational School, Architecture and Urban Planning, nehiruyar@beun.edu.tr, ORCID: 0000-0003-3358-3145.

This massive microbial presence isn't just a numerical oddity; it has deep and lasting ecological consequences. The Bacteria and Archaea in marine sediments are indispensable drivers of the global biogeochemical cycles, which govern the flow of fundamental elements like carbon, nitrogen, and sulfur. Through their various metabolic processes, these microorganisms are the hidden engine that transforms and moves nutrients and energy across the entire biosphere, ultimately influencing both ocean and land ecosystems.

Due to their significant importance, substantial research effort is presently devoted to determining the secrets of these sedimentary microbial communities (D'Hondt et al., 2004). Current research is directed at determining their distributions, diversities, and metabolic functions. The research attempts to unravel the intricate two-way street: how the physical and chemical properties of the sediments influence the microbial population activity, and, conversely, how the processes of the microbes alter the composition of their immediate chemical surroundings.

Such studies are essential for recreating historical environmental circumstances, forecasting future ecological shifts, and understanding the functions that these microbes play in Earth's current environmental processes. Scientists are learning more about how life survives and adapts to some of the harshest and least energy-rich environmental situations on Earth by better comprehending the complexities of the microbial-geochemical connection.

Marine sediments cover almost all of the Earth’s surface, forming an important and very large habitat for an extraordinarily diverse community of microorganisms largely composed of Bacteria and Archaea. These sedimentary prokaryotes may be more important than previously thought; earlier estimates

suggested that they make up at least 50% of the total prokaryotic biomass of the earth, while later research indicates that they may make up as much as one-tenth to one-third of the total living biomass of the earth. This vast substratum of microorganisms is of great importance for the stability of the earth, for it forms the hidden engine by means of which are carried on the fundamental biogeochemical cycles, such as those of carbon, nitrogen, and sulphur, which regulate the transformation and movement in the biosphere of energy and foodstuffs. But notwithstanding the importance of the subject, our knowledge of the class of microorganisms in question is very little, so far as regards their special identity and their distribution, and scarcely any given of the sedimentary habitats, especially in the ecologically precarious regions such as that of the Sea of Marmara. This semi-enclosed body of water is of importance as forming a link in the chain of the connection of the Black Sea with the Gulf of the Aegean, affecting the shipping interests of over 20 million people in its watershed, but is at the present in a more or less rapidly deteriorating condition from the influence and effects of increasing pollution. Such factors as rapid urbanization, industrial waste, domestic sewage effluent not subjected to purification effective, run-off from agricultural land, oil carriers and aerial contaminants in addition to those brought in appreciable quantities from the Black Sea, combine to produce in part serious effects by gradually effecting the estuarine and gulfs, and raising very serious question from the environmental and health point of view as affecting the marine biota as a whole, that it is imperative that some knowledge be obtained and measures of at least partial amelioration of the pollution, which is now in a serious effectual condition, before it is too late.

Coastal pollution has become a significant public health concern in recent years, with waters being heavily contaminated by chemical pollutants like PAHs, petroleum, heavy metals, and

detergents (Verep et al., 2007, 2012; Uncumuşaoğlu et al., 2012). Among these, anionic detergents pose a serious threat to marine life; they form a surface layer that blocks the transfer of oxygen from the air into the water, leading to oxygen deprivation. Just a small concentration of these surfactants (1 to 20 parts per million) can be lethal to fish. Simultaneously, scientists are focusing on ocean color as a key indicator of marine ecosystem health and climate change impacts (Wernand et al. 2011). Because the color of the sea reflects the physical and biological properties of the water, changes in that color are closely monitored using various advanced satellite methods. These techniques rely on measuring features like the backscattering coefficient, spectral curvature, normalized reflectance, chlorophyll a levels, and normalized fluorescence line height (e.g., Cannizzaro et al., 2009; Hu et al., 2005). Research on the biological pump of the ocean, which uses the export flux of Particulate Organic Carbon (POC) as a stand-in for surface primary production, is frequently associated with this monitoring (Lins et al., 2014). Approximately 50 petagrams of carbon are produced by phytoplankton worldwide each year, with 10–20% of that POC being exported from the upper ocean to the deep mesopelagic layer (Bisson et al., 2020).

The current approaches for determining POC flux include remote sensing techniques (Allison et al., 2010), in situ optical measurements (Kiko et al., 2017), sediment trapping methods (Buesseler et al., 2007a), and radioisotope methods (Zhou et al., 2020). Sea Surface Temperature (SST) is one of the essential parameters that should be used when judging climate models (Hurrell & Trenberth, 1999). Besides, SST simulation itself is quite tricky because it needs to consider numerous physical phenomena involving cloud-related processes, atmospheric and oceanic heat transfer, atmospheric water vapour, sea ice cover, etc (Randall et al., 2007).

The increased use of satellite images, which can hardly be neglected in measuring the degree of pollution of the Marmara Sea, has become very important since it is the means of studying with efficiency the changes occurred on the surface of the sea and to make studies of pollution. In particular it is by means of remote sensing that the quality of the water, the changes in colour and the traces of pollution present in relation to its study are ascertained. The pollution of the Marmara Sea can be detected in satellite images as violent changes of colour or specific images such as abrupt colour changes, or as specific images may be observed which are indicative of the mixture of industrial wastes, fertilizers and generally chemicals connected with agricultural pursuits or general pollution introduced into the sea. The tracing of the said pollution gives information, it is evident, concerning the sources of pollution, the dissipation of the pollution and the general condition of the pollution. The usefulness of this method of examination of pollution, both as regards the polluted state of the Marmara Sea, and generally to aid in the formation of schemes for environmental management, is now so obvious, and indeed has become an absolute necessity. It follows, therefore, that, being in possession of a more detailed analysis of the pollution in the water of the Sea of Marmara, that the preventive and remedial steps for its better state of health may be worked upon more effectively.

2. WORKING AREA

The Sea of Marmara, one of the important water sources in Turkey, experiences the effects of many pollution causes arising from industry, urban places and rivers that flow into it. In recent years, the pollution that is detected has reached serious dimensions in the Sea of Marmara. Every types of industry waste, agricultural plant food products, house waste and other pollution

sources are thrown into the sea and the water quality is damaged seriously. Moreover, heavy marine traffic, ports, ship transportation also the causes of deterioration of the general quality of the sea water. This increasing pollutions cause a lot of pressure on the sea population and marine life. Such a situation is a danger for the ecological balance and for the health of the organisms living in the Sea of Marmara. So, in order to protect Sea of Marmara, environmental protection laws and methods which are the innocents by their side must be given immediately.

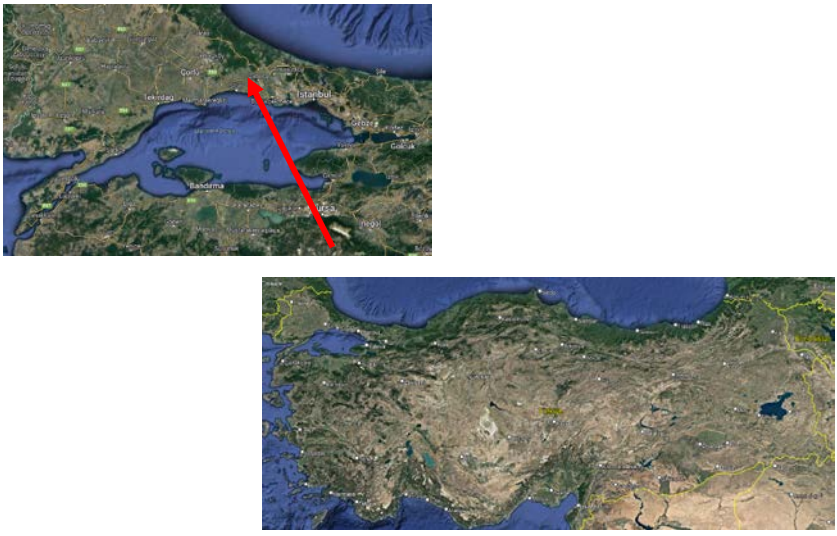


Figure 1. Google Earth image of Marmara Sea

3. METHODS AND RESULTS

The study's approach uses Google Earth Engine and MODIS-Aqua satellite data to analyze Particulate Organic Carbon and Chlorophyll-a Normalized Fluorescence Line Height throughout the Sea of Marmara's shores. Five-year data sets from 2016 to 2021 were the study's primary emphasis, and unique methods for satellite images analysis were created. Water quality indicators were determined using MODIS-Aqua satellite data, and the evolution of Particulate Organic Carbon and Chlorophyll-

a Normalized Fluorescence Line Height over time was investigated. Important conclusions regarding the Marmara Sea's environmental health were obtained by this analysis procedure, which also provided a detailed understanding of the region's water quality impacts. This approach offered a useful instrument for remote coastal monitoring and evaluation.

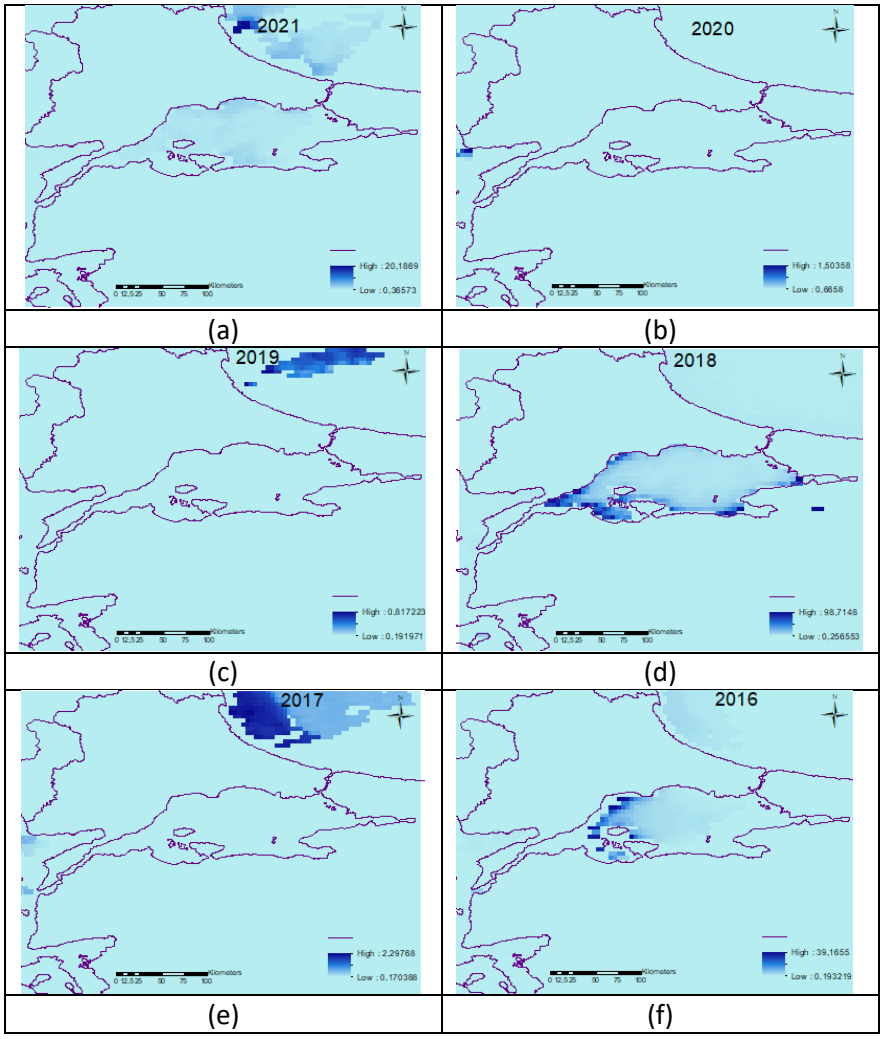
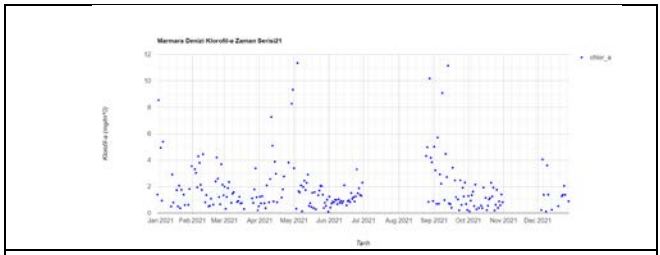
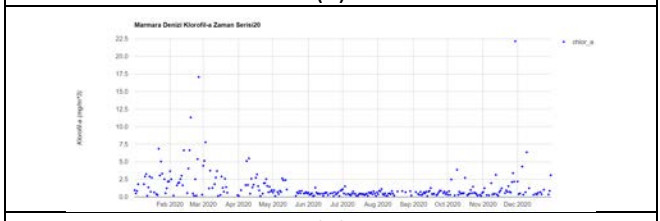


Figure 2. Chlorophyll concentration thematic maps dated (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016

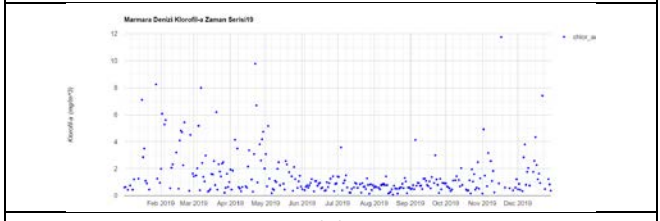
Figure 2 displays thematic maps of chlorophyll concentration from 2016 to 2021. The years 2018 and 2019 saw the highest and lowest readings, respectively.



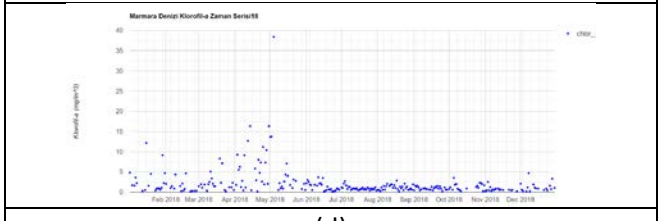
(a)



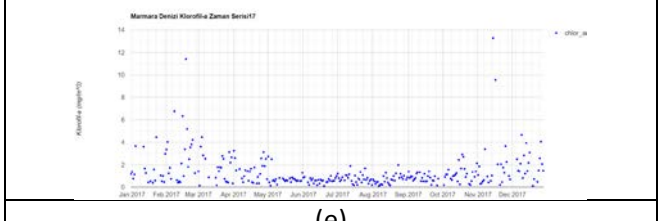
(b)



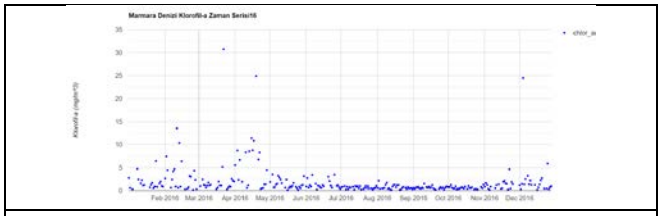
(c)



(d)



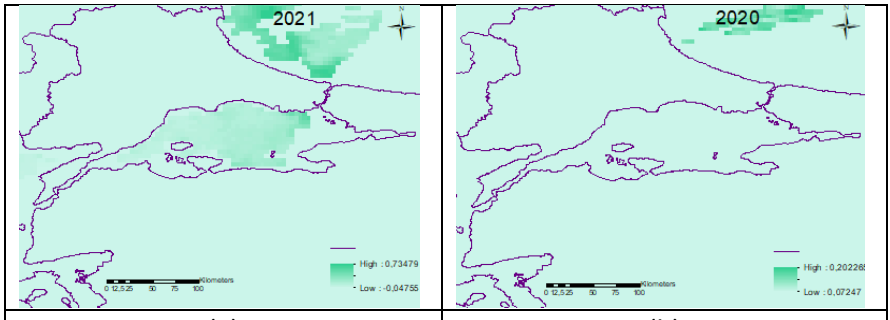
(e)



(f)

Figure 3. Chlorophyll concentration graphs for (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016

The monthly changes in chlorophyll content from 2016 to 2021 are shown in Figure 3. There are discernible seasonal and interannual variations over the course of these six years, with some months continuously showing higher or lower chlorophyll levels according on the year. May and September of 2021 had the highest chlorophyll concentrations, which could be related to seasonal phytoplankton blooms that are usually brought on by ideal light and temperature conditions. Conversely, February and December had the lowest levels, which may have been brought on by the colder water temperatures and less sunshine throughout the winter, which can restrict the growth of phytoplankton. Peak chlorophyll levels were recorded in March and December of 2020, suggesting that both early spring and late in the year may have seen productive circumstances.



(a) (b)

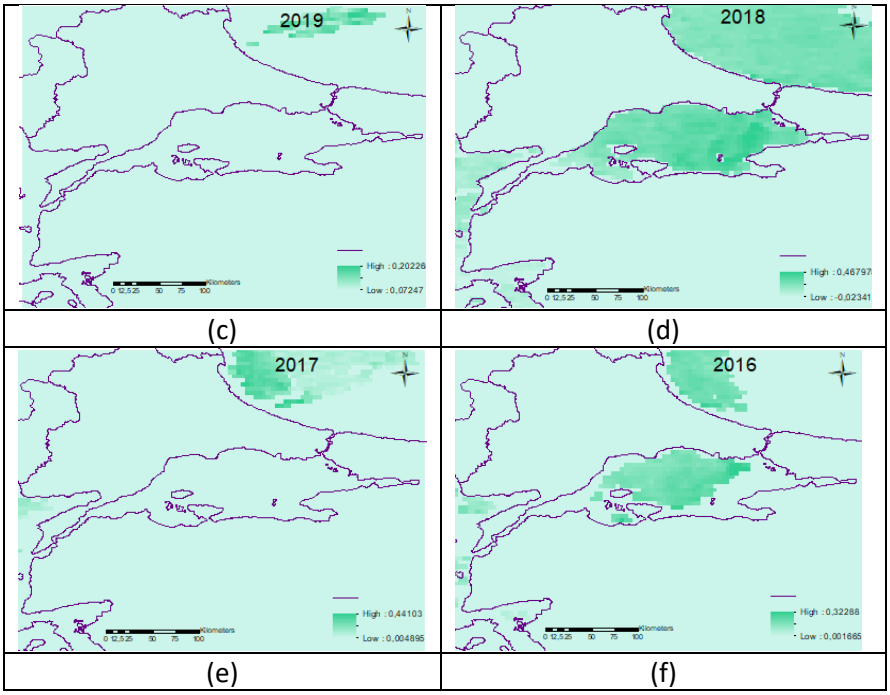
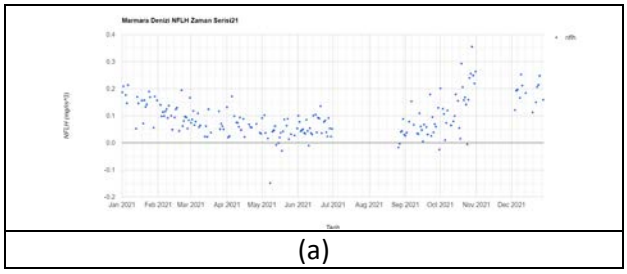


Figure 4. NFHL thematic maps of (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016

Figure 4 shows the thematic maps of NFHL concentration for 2016-2021. The highest values were observed in 2018 and the lowest values were observed in 2017.



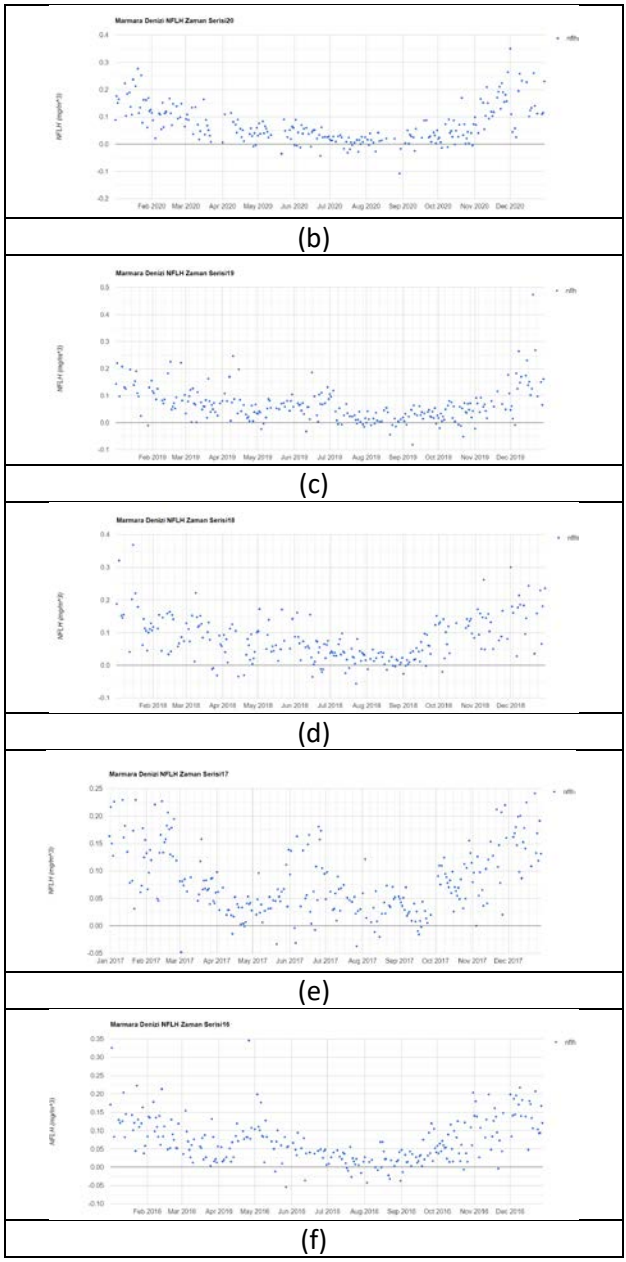


Figure 5. NFHL graphs for (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016

NFHL charts for 2016-2021 are given in Figure 5. In 2021, the highest value was observed in October and the lowest

value in May. In 2020 and 2019, the highest value was observed in December and the lowest value in August. In 2018, the highest value was seen in January and the lowest value was seen in August. In 2017, the highest value was observed in December and January, and the lowest value was observed in May. In 2016, the highest value was observed in January and May, and the lowest value was observed in August.

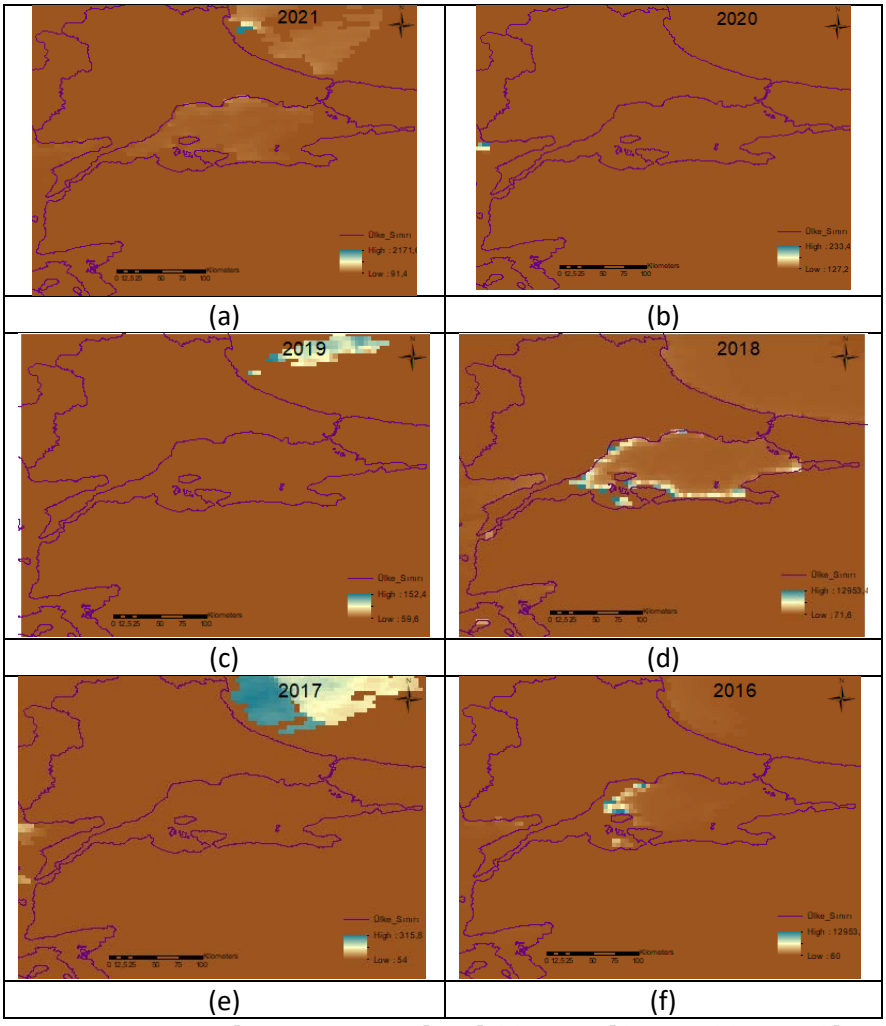
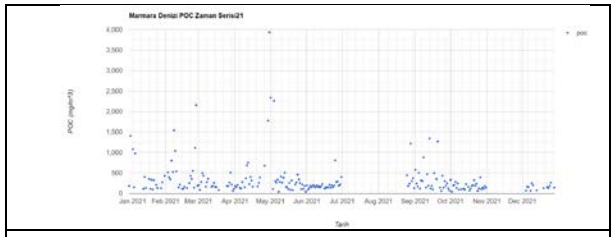
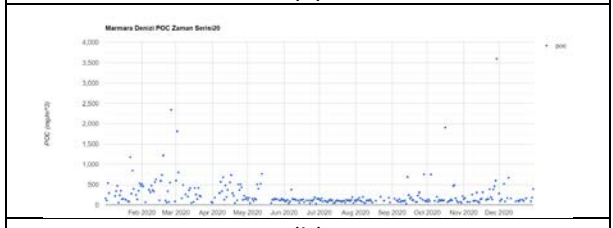


Figure 6. POC thematic maps dated (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016

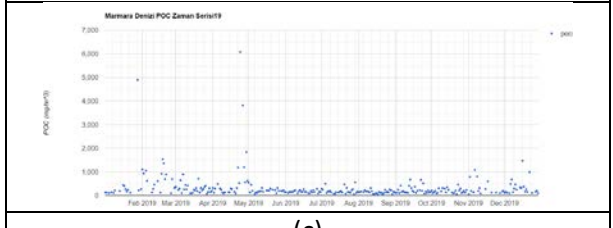
POC concentration thematic maps for 2016-2021 are given in Figure 6. The highest values were observed in 2018 and the lowest values were observed in 2017.



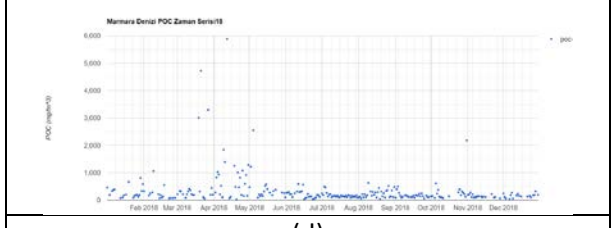
(a)



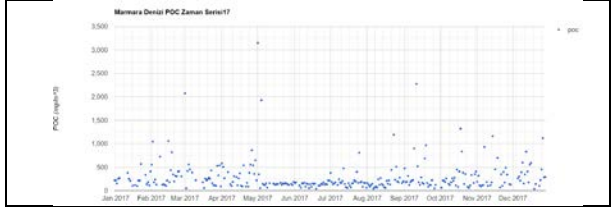
(b)



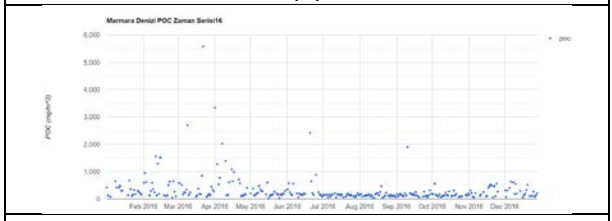
(c)



(d)

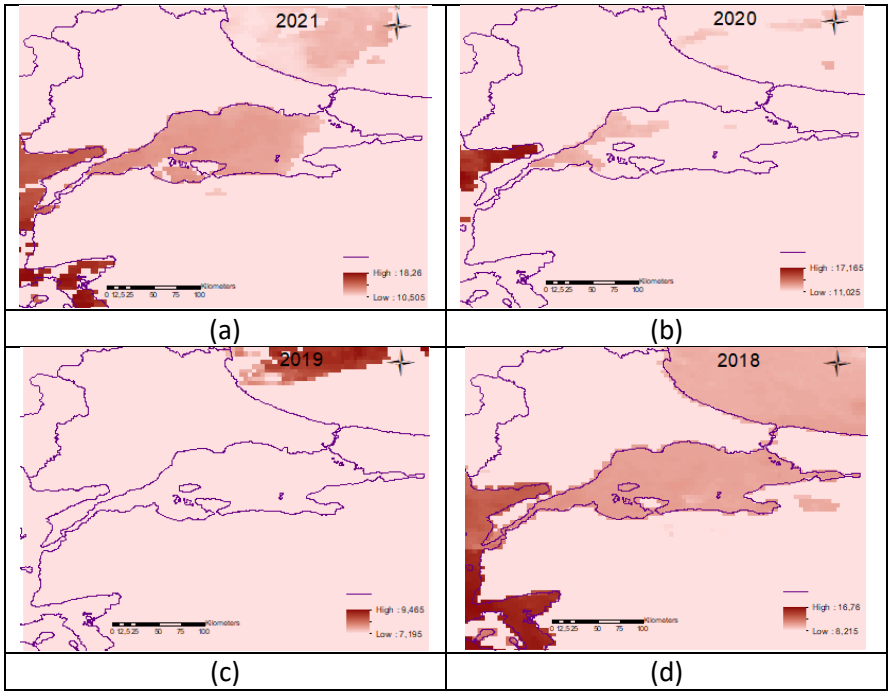


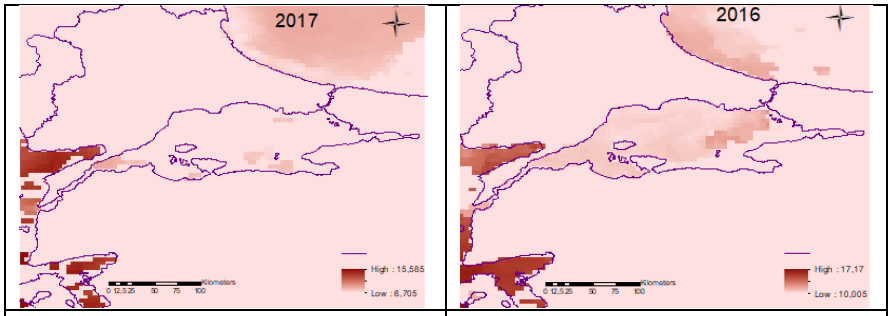
(e)



(f)

Figure 7. (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016 POC analysis graphs





(e) (f)

SST thematic maps for the years (a) 2021, (b) 2020, (c) 2019, (d) 2018, (e) 2017 and (f) 2016 are shown in Figure 8.

Figure 7 displays POC charts for the years 2016 through 2021. In 2021, January had the lowest value and May had the highest. In 2020, the peak value was in December, and the lowest was in August. In 2019, the highest value was in May, and the lowest was in June. In 2018, the highest values were in April and May, and the lowest was in December. In 2017, the highest value was in May, and the lowest was in June. In 2016, the highest value was in April, and the lowest was in August.

Figure 8 shows the thematic maps of SST concentration for 2016-2021. The highest values were observed in 2021 and the lowest values were observed in 2017.

(a)

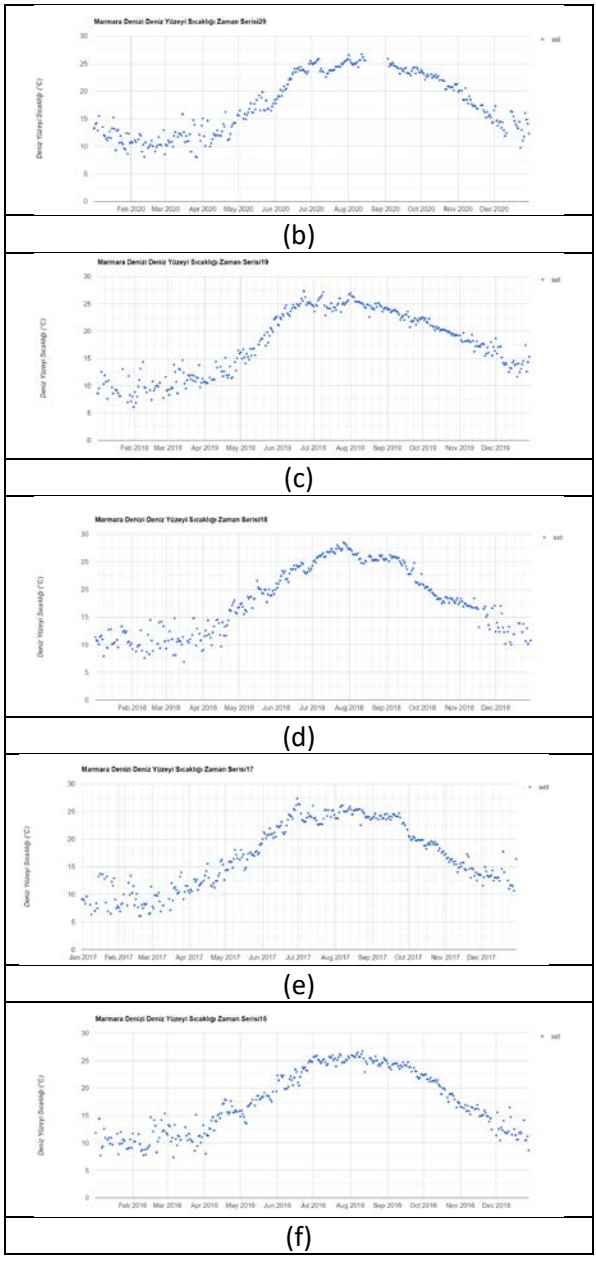


Figure 9. SST graphs for (a) 2021 (b) 2020 (c) 2019 (d) 2018 (e) 2017 (f) 2016

SST charts for 2016-2021 are shown in Figure 9. In 2021, the highest value was reached in July and September, while the

lowest value was recorded in December. In 2020, the top value occurred in August and September, and the bottom value was noted in March. In 2019, the peak value appeared in August and September, and the lowest value was in March. In 2018, the highest value was noted in August and September, with the lowest in January. In 2017, the highest value took place in July, and the lowest value was in March. In 2016, the peak value was seen in August and September, while January and February had the lowest values.

4. CONCLUSION

The analyses and investigations show that MODIS-Aqua satellite data is an effective tool for determining the Chlorophyll-a Normalised Fluorescence Line Height and Particulate Organic Carbon levels in the coastal areas of the Marmara Sea. This study evaluated the change in seawater quality over time by analyzing data sets obtained over a five-year period. It revealed the effects of environmental factors. These analyses are an important step in understanding the health of aquatic ecosystems in the region and developing environmental management strategies. Monitoring the water quality of the Marmara Sea with remote sensing techniques could play a critical role in identifying and addressing future environmental problems. The data obtained in this study can serve as the foundation for further research on monitoring coastal ecosystems and environmental health.

Chlorophyll a values were generally higher in December and March. NFHL values were typically high in December and January but low in May. POC values were generally high in May and low in June and August.

REFERENCES

Allison D.B., Stramski D., Mitchell B.G. (2010). Seasonal and interannual variability of particulate organic carbon within the Southern Ocean from satellite ocean color observations J. Geophys. Res., 115 p. C06002

Bisson K., Siegel D.A., DeVries T. (2020). Diagnosing mechanisms of ocean carbon export in a satellite-based food web model Front. Mar. Sci., 7 p. 505

Buesseler K.O., Antia A.N., Chen M., Fowler S.W., Gardner W.D., Gustafsson O., Harada K., Michaels A.F., Rutgers van der Loeff M., Sarin M., Steinberg D.K., (2007). Trull an assessment of the use of sediment traps for estimating upper ocean particle fluxes J. Mar. Res., 65 pp. 345-416

Cannizzaro, J. P., Hu, C., English, D. C., Carder, K. L., Heil, C. A., & Müller-Karger, F. E. (2009). Detection of Karenia brevis blooms on the west Florida shelf using in situ backscattering and fluorescence data. Harmful Algae, 8(6), 898–909.

Carvalho, G. A., Minnett, P. J., Banzon, V. F., Baringer, W., & Heil, C. A. (2011). Long-term evaluation of three satellite ocean color algorithms for identifying harmful algal blooms (Karenia brevis) along the west coast of Florida: A matchup assessment. Remote Sensing of Environment, 115(1), 1–18.

D'Hondt S., Jørgensen B.B., Miller D.J., Batzke A., Blake R., Cragg B.A, et al. (2004). Distribution of microbial activities in deep subseafloor sediments Science, 306 (2004), pp. 2216-2221

Hu, C., Muller-Karger, F. E., Taylor, C., Carder, K. L., Kelble, C., Johns, E., & Heil, C. A. (2005). Red tide detection and tracing using MODIS fluorescence data: A regional

example in SW Florida coastal waters. Remote Sensing of Environment, 97(3), 311–321.

Hurrell, J. W., & Trenberth, K. E. (1999). Global sea surface temperature analyses: Multiple problems and their implications for climate analysis, modeling, and reanalysis. Bulletin of the American Meteorological Society, 80(12), 2661–2678.

Kiko R., Biastoch A., Brandt P., Cravatte S., Hauss H., Hummels R., Kriest I., Marin F., McDonnell A.M.P., Oschlies A., Picheral M., Schwarzkopf F.U., Thurnherr A.M., Stemmann L. (2017). Biological and physical influences on marine snowfall at the equator Nat. Geosci., 10 pp. 852-858

Lins L., Guilini K., Veit-Köhler G., Hauquier F., Alves R.M.S., Esteves A.M., Vanreusel A. (2014). The link between meiofauna and surface productivity in the Southern Ocean Deep Sea Res Part II Top. Stud. Oceanogr., 108 pp. 60-68

Orhon D. (1995). Evaluation of the impact from the Black Sea on the pollution of the Marmara Sea Water Sci. Technol., 32 (7) (1995), pp. 191-198

Parkes R., Cragg B., Wellsbury P. (2000). Recent studies on bacterial populations and processes in subseafloor sediments, a review Hydrogeol. J., 8 (2000), pp. 11-28

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., et al. (2007). Climate models and their evaluation. In S. Solomon, et al. (Eds.), Climate change 2007: The physical science basis. Contribution of working Group I to the fourth assessment report of the intergovernmental panel on climate change (pp. 589–662). Cambridge University Press.

Spehar, R.L., Holcombe, G.W., Carlson, R.W., Drummod, R.A., Yount, J.D., Pickering, Q.H. (1979) Effect of pollution on fresh water fish. J Water Pollution Control Federation 51: 1616-1684.

Stumpf, R. P., Culver, M. E., Tester, P. A., Tomlinson, M., Kirkpatrick, G. J., Pederson, B. A., et al. (2003). Monitoring Karenia brevis blooms in the Gulf of Mexico using satellite ocean color imagery and other data. Harmful Algae, 2(2), 147–160.

Tomlinson, M. C., Wynne, T. T., & Stumpf, R. P. (2009). An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, Karenia brevis. Remote Sensing of Environment, 113(3), 598–609.

Uncumuşaoğlu A. A., Gürkan Ş., Özkan E. Y., Büyükışık H. B. (2012). Tirebolu kıyılarından (Giresun, Doğu Karadeniz) yakalanan Denizatı (Hippocampus hippocampus) karaciğer ve kas dokusunda biriken ağır metaller üzerine bir ön araştırma Fresenius Çevresi. Boğa. , 21 (11b) (2012) , s. 3418 – 3420

Unlu S., Topcuoglu S., Alpar B., Kirbasoglu C., Yilmaz Y. Z. (2006). Hydrological processes and sediment pollution in the semi-enclosed bays of the Marmara Sea, Turkey Geophys. Res. Abstract, 8 (2006), p. 00432

Verep B., Akın Ş., Mutlu C., Apaydın G., Ertuğral B., Çevik U. (2007). Karadeniz kıyısındaki deniz ürünleri kafeslerinde yetiştirilen gökkuşağı alabalığında (Onchorhynchus mykiss) iz elementlerin değerlendirilmesi Fresenius Çevresi. Boğa. , 16 (9a) (2007) , s. 1005 - 1011

Verep B., Mutlu C., Apaydın G., Çevik U. (2012). İyidere Çayı (Rize-Türkiye) tatlı su balık türleri, su ve sedimentlerinde

iz element analizleri Pak. J. Biol. Bilim. , 15 (14) (2012) , s. 658 – 665

Weith G. D., Konasewich D. E., (1975). Structure-activity Correlations in Studies of Toxicity and. Bioconcentration with Aquatic Organisms, Proceedings of Symposium organised in Ontario, 11–13 Mart 1975 (1975) (347 s.)

Wernand MR, Gieskes WWC (2011). 1600'den (Hudson) 1930'a (Raman) Ocean optics from 1600(Hudson) to 1930 (Raman); Shift in interpretation of natural wa-ter colouring, Union des oc´eanographes de France, 87 s. ISBN: 978-2-9510625-3-5.

Whitman W. B., Coleman D.C., Wiebe W.J. (1998). Prokaryotes, the unseen majority Proc. Natl. Acad. Sci. USA, 95, pp. 6578-6583

Zhou K., Dai M., Maiti K., Chen W., Chen J., Hong Q., Ma Y., Xiu P., Wang L., Xie Y. (2020). Impact of physical and biogeochemical forcing on particle export in the South China Sea Prog. Oceanogr., 187 Article 102403

HEURISTICALLY OPTIMIZED SPEECH RECOGNITION MODELS ENHANCED BY SYNTHETIC DATA AUGMENTATION FOR GIS APPLICATIONS[1]

Ahmet Emin KARKINLI[2]

Erkan BEŞDOK[3]

1. INTRODUCTION

Geographic Information Systems (GIS) serve as a fundamental tool in numerous fields, including urban planning, transportation, disaster management, environmental sciences, and agriculture, for the collection, analysis, visualization, and management of spatial data (Burrough, McDonnell, & Lloyd, 2015). By resolving complex spatial relationships, GIS supports decision-making processes; however, data collection and entry in field operations are often labor-intensive and inefficient due to operators' hands and eyes being continuously occupied (Goodchild, 2009). Speech recognition technology offers the potential to address this issue by enabling hands-free data entry and querying (Gaikwad, Gawali, & Yannawar, 2010). As a natural method of human-computer interaction, speech recognition can streamline data entry in GIS applications involving technical terminology, thereby enhancing the

---

1 Karkınlı, A.E., (2017), Coğrafi Bilgi Sistemlerinde Geovideo/ Audio Kullanımı, Produced from Erciyes University, Institute of Science, PhD Thesis.

2 Dr. Öğr. Üyesi, Niğde Ömer Halisdemir Üniv., Mühendislik Fakültesi, Harita Mühendisliği Bölümü akarkinli@ohu.edu.tr, ORCID: 0000-0001-7216-6251.

3 Prof. Dr., Erciyes Üniv., Mühendislik Fakültesi, Harita Mühendisliği Bölümü ebesdok@erciyes.edu.tr, ORCID: 0000-0001-9309-375X.

efficiency of field operations (Çivicioğlu, Karkınlı, & Beşdok, 2012). Isolated word recognition systems, which provide high accuracy with limited vocabularies, are particularly suitable for data entry and voice control in specialized domains (Arslan & Barışçı, 2020).

Speech recognition has been extensively studied in the international literature for both Turkish and other languages (Kurimo et al., 2006). Turkish, with its agglutinative structure and free word order, generates large vocabularies and complex morphological forms, posing significant challenges for automatic speech recognition (ASR) systems (Arslan & Barışçı, 2020; Kurimo et al., 2006). To overcome these challenges, isolated word recognition systems aim to achieve high accuracy by utilizing limited, context-specific vocabularies (L. R. Rabiner, 1989).

A speech recognition system first applies a feature extraction stage to derive a meaningful and compact representation from raw audio signals (Gaikwad et al., 2010). Commonly, methods such as Mel-Frequency Cepstral Coefficients (MFCC) are used to transform the frequency and temporal characteristics of the audio into vectors (Davis & Mermelstein, 1980). These feature vectors are then passed to a classification stage, where algorithms such as Hidden Markov Models (HMM), Dynamic Time Warping (DTW), and Gaussian Mixture Models (GMM) determine which word was spoken (Gaikwad et al., 2010).

Isolated word recognition has been a long-standing area of research, with classical methods like HMM, DTW, and GMM being successfully employed (L. Rabiner & Juang, 1993; Reynolds, 1995; Sakoe & Chiba, 1978). HMM is a powerful tool for modeling temporal dependencies in speech signals, and (L. R. Rabiner, 1989) established its mathematical foundation, making

it a widely adopted standard. DTW is an effective method for aligning speech signals with templates at low computational cost, as (Sakoe & Chiba, 1978) demonstrated in its practical application. GMM is used to model the statistical distributions of speech features, and (Reynolds, 1995) highlighted its effectiveness in speaker identification and word recognition.

However, these classical approaches face two significant challenges. First, the performance of models like GMM, which often rely on the Expectation-Maximization (EM) algorithm, can be limited by EM's tendency to converge to local optima (Dempster, Laird, & Rubin, 1977). Second, building robust models requires large, multi-repetition speech datasets, which are often impractical and time-consuming to collect in real-world scenarios.

To address these limitations, this study introduces a two-stage investigation. First, we conduct a comparative analysis of the classical DTW, HMM, and GMM methods using real speech data to establish a performance baseline. Second, we propose a novel framework to overcome data scarcity and model optimization issues. This framework introduces a Harmonic-Percussive Source Separation (HPSS)-based synthetic data generation approach to create realistic, repeated samples from a single recording. Subsequently, we utilize this augmented dataset to train hybrid GMM models, optimizing their parameters not only with EM but also with advanced heuristic algorithms, including Differential Evolution (DE), Self-adaptive DE (SaDE), and the Backtracking Search Algorithm (BSA), to achieve global optimization and enhance recognition accuracy.

This work is structured as follows: Section 2 reviews related work. Section 3 details the proposed methodology, including the system architecture, dataset design, and classification methods. Section 4 presents the experiments and

discussion, where the performances of the classical methods on real data and the hybrid models on synthetic data are evaluated and compared. Finally, Section 5 concludes the study with key findings and suggestions for future research.

2. RELATED WORK

This section reviews prior studies on automatic speech recognition (ASR), focusing on isolated word recognition techniques for Turkish and the application of heuristic optimization methods to enhance model performance (Gaikwad et al., 2010). Field operations for data collection are often inefficient due to operators' hands and eyes being continuously occupied (Goodchild, 2009). ASR aims to overcome these challenges by enabling hands-free data entry (Gaikwad et al., 2010). For instance, (Çivicioğlu et al., 2012) integrated ASR algorithms into Geographic Information Systems (GIS) and proposed a Differential Search Algorithm (DSA)-based approach to facilitate data entry in field applications. Other studies have also demonstrated the potential of ASR for technical terminology-based data entry in GIS (Arslan & Barışçı, 2020). However, the literature on ASR applications in GIS remains limited. (Goodchild, 2009) discussed the potential of voice-controlled systems for spatial data collection but did not focus on practical implementations. Similarly, (Mahmoudi, Camboim, & Brovelli, 2023) noted that voice-based interfaces could enhance user experience in GIS. In the Turkish context, ASR systems relying on technical dictionaries, such as transportation terminology, emerge as innovative solutions for field applications.

Isolated word recognition has been a cornerstone of Turkish ASR research, addressing the language's agglutinative morphology through constrained vocabularies. (Tüzün et al.,

1995) developed a speaker-independent system with a 12-word vocabulary, leveraging Hidden Markov Models (HMM) and Linear Predictive Coding (LPC) to achieve reliable results. (Arısoy & Arslan, 2004) proposed a radiology dictation system, achieving 87% accuracy with HMM-based models. (Asliyan, Gunel, & Yakhno, 2008) introduced a syllable-based Dynamic Time Warping (DTW) system for a 200-word vocabulary, reporting nearly 90% accuracy. (Salor, Pellom, Ciloglu, & Demirekler, 2007) created a Turkish speech corpus and adapted the SONIC recognizer, achieving competitive results. (Keser & Edizkan, 2009) employed a phoneme-based approach using the Common Vector Approach (CVA), demonstrating robust performance. (B. Tombaloğlu & Erdem, 2016) combined Mel-Frequency Cepstral Coefficients (MFCC) with Support Vector Machines (SVM), matching HMM performance. These studies highlight the success of classical methods in Turkish ASR.

Isolated word recognition has been extensively explored beyond Turkish, offering a comparative perspective. In English, (Wilpon, Rabiner, Lee, & Goldman, 1990) achieved over 95% accuracy using HMM for keyword recognition. (Juang & Rabiner, 1991) utilized HMM with cepstral features, reporting high accuracy for command-and-control applications. In Chinese, (Chen & Wang, 1995) employed neural networks, achieving 86.72% accuracy in isolated monosyllabic tone recognition tasks. In Hindi, (Bansal, Dev, & Jain, 2008) developed a vector quantization (VQ) and HMM-based system, attaining 92% accuracy, while (Kumar & Aggarwal, 2011) later leveraged the Hidden Markov Model Toolkit (HTK) for Hindi, reporting 94% accuracy. In Arabic, (Alotaibi, 2012) implemented an HMM-based system, achieving 98% accuracy. (Chauhan & Tanawala, 2015) compared MFCC and LPC for Gujarati, reporting 88% accuracy with MFCC-based HMM systems. These international

studies demonstrate the adaptability of classical methods across diverse linguistic structures.

Recent advancements in Turkish ASR have incorporated deep learning models. (Burak Tombaloğlu & Erdem, 2021) utilized Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, reducing error rates compared to HMM-GMM systems. (Polat & Oyucu, 2020) developed a comprehensive Turkish speech corpus to improve recognition accuracy. (Kheddar, Hemis, & Himeur, 2024) noted that deep learning has propelled advancements in automatic speech recognition, narrowing the human-machine interaction gap, but challenges like data scarcity persist, while isolated word recognition is supported by specific datasets without being highlighted as underexplored. (Palaz, Kanak, Bicil, & Doğan, 2005) introduced the TREN platform, adapting HMM for Turkish morphology. These developments highlight deep learning potential, though the complexity and data requirements of such models often make classical methods more practical for isolated word recognition.

Heuristic optimization techniques have emerged as a promising avenue to improve ASR model performance. The Expectation-Maximization (EM) algorithm, commonly used in GMM, can converge to local optima, limiting accuracy (Dempster et al., 1977). (Lin & Shuxun, 2005) optimized GMM parameters using a genetic algorithm and fuzzy approach, enhancing speaker recognition performance. (Najkar, Razzazi, & Sameti, 2010) applied Particle Swarm Optimization (PSO) to HMM parameters, achieving high accuracy in speech recognition tasks. (Benkhellat & Belmehdi, 2012) enhanced DTW alignments with GA. (Nirmal, Jayaswal, & Kachare, 2024) proposed a hybrid Bald Eagle-Crow Search Algorithm for GMM optimization in speaker verification frameworks. (Emdadi, Ahmadi Moughari, Yassaee Meybodi, & Eslahchi, 2019) explored an Ant Colony

Optimization-inspired approach for HMM parameter estimation. The Backtracking Search Algorithm (BSA), introduced by (Civicioglu, 2013), offers superior global search capabilities and has been applied to various numerical optimization problems, including potential use in GMM parameters. This review highlights the need for innovative methods to enhance speaker-dependent isolated word recognition systems for technical vocabularies and aims to address this gap with the proposed BSA-enhanced GMM approach.

3. METHODOLOGY

This section provides a detailed explanation of the data preparation processes and implementation methods used for the development of the speaker-dependent isolated word recognition system. The creation of the codebook, preprocessing of speech data, feature extraction, classification techniques, and heuristic optimization algorithms are the fundamental components that ensure the system's high accuracy. First, the feature extraction method using MFCC will be described; then, classical approaches based on DTW, HMM, and GMM will be discussed. Finally, the implementation details of the DE, SADE, and BSA techniques used to optimize GMM parameters will be presented.

3.1. Codebook, Speech Data Preparation, and Preprocessing

The processes of codebook creation, speech data collection, and preprocessing are meticulously designed to ensure that the system accurately recognizes words specific to transportation terminology. The codebook consists of 400 words selected from the “Highway Terms” section of the "Transportation and Communication Terms Dictionary” published by the Republic of Turkey Ministry of Transport, Maritime Affairs, and Communications. The words were chosen

with careful consideration of the agglutinative structure of Turkish and the phonetic diversity specific to technical terminology.

Speech data were recorded by a single speaker in a controlled environment to obtain a dataset suitable for the requirements of a speaker-dependent system. During the recording process, each term was sampled at 11,025 Hz in 16-bit PCM format and collected with 20 repetitions, consistent with the speaker-dependent scenario. To remove non-speech silences in the raw signal, the Short-Time Energy method was employed (Nandhini & Shenbagavalli, 2014). After calculating the energy for each frame using Equation (1), frames below a fixed threshold were marked as silence and removed.

$$E_n = \sum_{m=-\infty}^{\infty} (s[m]w[n-m])^2 = \sum_{m=-\infty}^{\infty} s^2[m]w^2[n-m] \quad (1)$$

Here, E_n represents the short-time energy of the n-th frame; m denotes the discrete time axis sample index; $s[m]$ is the amplitude of the raw speech signal at time m; $w[n-m]$ is the shifted version of the window function corresponding to the n-th frame over which the energy is calculated; and n indicates the center sample index of the frame being analyzed.

In general, high-frequency components in speech signals have lower amplitude values. If these high-frequency components are used as data in a speech recognition application without enhancement, the features they carry cannot be fully utilized. To make the features of high-frequency components in a speech signal more prominent, a pre-emphasis operation with the transfer function specified in Equation (2) is applied. The application of the transfer function given in Equation (2) in the time domain can be obtained as shown in Equation (3).

$$H(z) = 1 - \alpha z^{-1} \qquad 0 \le \alpha \le 1 \tag{2}$$

$$\tilde{s}[k] = s[k] - \alpha s[k-1] \tag{3}$$

Here, $s[k]$ represents the k-th sample amplitude of the raw speech signal to be processed; $\tilde{s}[k]$ is the same sample of the output signal after passing through the pre-emphasis filter; α is the pre-emphasis coefficient, typically chosen in the range of 0.9 to 1, which determines the degree of enhancement of high-frequency components; and z^{-1} represents a one-sample delay operator. When expressed in the time domain as in Equation (3), each new signal sample is obtained by subtracting the product of the previous sample and the coefficient α, thereby emphasizing high-frequency components. This process makes formant structures and phoneme transitions more distinct, thus enhancing the effectiveness of feature extraction (Saxena, Farooqui, & Ali, 2022).

3.2. Feature Extraction and Mel-Frequency Cepstral Coefficients

Feature extraction is a fundamental stage of the speech recognition system, transforming raw speech signals into a compact and meaningful representation for classification. In this study, Mel-Frequency Cepstral Coefficients (MFCC) are used to effectively capture the acoustic characteristics of Turkish speech. MFCC, as a method that mimics the frequency perception of the human auditory system, has become a standard feature extraction technique in the speech recognition literature (Davis & Mermelstein, 1980; Tiwari, 2010). In particular, the agglutinative structure and phonetic diversity of Turkish, such as vowel length variations and consonant clusters, make MFCC highly suitable for Turkish speech recognition systems (Arslan & Barışçı, 2020).

MFCC extraction involves a multi-step process that analyzes the speech signal in both the time and frequency domains. Initially, the speech signal undergoes a framing process. The signal is divided into short time frames of 20-30 milliseconds, with each frame created with a 50% overlap to ensure temporal continuity. This frame length is chosen to be short enough to capture the stationary properties of the speech signal while being long enough to provide meaningful acoustic information. For the local characteristics of Turkish speech, this duration is sufficient to analyze rapid articulatory changes, such as consonant-vowel transitions. To minimize spectral leakage, a Hamming window is applied to each frame. The Hamming window is defined by Equation (4):

$$W(n) = 0.53836 - 0.46164 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \quad (4)$$

Here, $W(n)$ represents the value of the window function at the n-th sample; n is the sample index within the frame ($0 \le n \le N-1$); and N is the total number of samples in the frame.

The framed and windowed signal is then subjected to the Fast Fourier Transform (FFT) to convert it into the frequency domain. FFT computes the frequency spectrum of each frame, extracting the frequency components of the signal. Then, to mimic the non-linear frequency perception of the human auditory system, a Mel-filter bank is applied. The Mel scale, which converts frequency from Hz to the Mel unit, is defined in Equation (5):

$$m = 2595 log_{10}(1 + \frac{f}{700}) \quad (5)$$

Here, m represents the Mel frequency, and f is the frequency in Hz. The Mel scale exhibits a linear distribution up to 1-2 kHz and a logarithmic distribution at higher frequencies.

For each frame, the energy values obtained from P triangular Mel filters are calculated using Equation (6):

$$E_p(n) = \sum_{k=0}^{N-1} \left| X_n(k) \right|^2 H_p(k) \quad (6)$$

Here, $E_p(n)$ is the energy output of the p-th filter for the n-th frame; $\left| X_n(k) \right|$ is the FFT amplitude spectrum of the n-th frame; $H_p(k)$ is the triangular weight function of the p-th Mel filter; k is the frequency index; N is the number of FFT points; and p is the filter index $(1 < p < P)$.

Finally, the logarithm of the filter bank energies is taken, and the Discrete Cosine Transform (DCT) is applied to obtain the MFCCs. The MFCCs are computed using Equation (7):

$$c_i(n) = \sum_{p=1}^{P} \log(E_p(n)) \cos\left(\frac{\pi i (p - 0.5)}{P}\right) \quad (7)$$

Here, $c_i(n)$ represents the i-th MFCC coefficient for the n-th frame; $E_p(n)$ is the energy output of the p-th filter; p is the filter index; P is the total number of filters; and i is the cepstral coefficient index. The features are stored in vector format for use in the classification stage (Memon, Lech, & He, 2009).

3.3. Feature Classification

Feature classification is one of the fundamental components of speech recognition systems and involves the process of labeling speech signals using previously extracted MFCC features. This problem is addressed within a supervised classification framework, where models learned from training data are used to recognize unknown samples in test data. In this section, three classical methods commonly employed in speech

recognition systems, DTW, HMM, and GMM, will be introduced in detail, and their basic working principles will be explained.

3.3.1. Dynamic time warping (DTW)

Dynamic Time Warping (DTW) is an alignment technique used to measure similarities between time series in speech recognition systems, and it is particularly effective in isolated word recognition scenarios (Sakoe & Chiba, 1978). DTW enables flexible alignment of two speech signals along the time axis, accommodating variations in the speaker's pronunciation speed or differences in word lengths. The core principle of DTW involves using dynamic programming to find the lowest cumulative distance between two time series (Mizutani & Dreyfus, 2021). In this process, the MFCC feature vectors of each test word are compared with the feature vectors of reference words in the codebook. The alignment begins with the construction of a distance matrix; this matrix contains the local distance between each frame and every other frame. The local distance is typically computed as the Euclidean distance shown in Equation (8):

$$d(i,j) = \sqrt{\sum_{k=1}^{d} \left(x_{i,k} - y_{j,k} \right)^2} \tag{8}$$

Here, $d(i,j)$ represents the local distance between the i-th test frame and the j-th reference frame; $x_{i,k}$, the k-th dimension of the test vector; $y_{j,k}$, the k-th dimension of the reference vector; and d, the dimension of the feature vector.

These local distances are combined using dynamic programming to construct the cumulative distance matrix. The cumulative distance, which is the essence of DTW, is calculated with a dynamic programming algorithm. In each cell, the smallest

distance from the previous alignment points is added to the local distance. This is expressed by Equation (9):

$$D(i,j) = d(i,j) + min\left[D(i-1,j), D(i,j-1), D(i-1,j-1)\right] \quad (9)$$

Here, $D(i,j)$ represents the cumulative distance between the i-th test frame and the j-th reference frame; $d(i,j)$ the local distance; and the min operator selects the smallest distance from the previous alignment points (diagonal, horizontal, or vertical). The alignment is completed by finding the shortest path from the starting point $(0,0)$ to the ending point (I,J), where I is the number of frames in the test signal and J is the number of frames in the reference signal.

One of the advantages of DTW is that it does not require model training, making it suitable for rapid prototype development; however, its computational cost increases with large datasets. This method offers a simple yet effective approach, especially for small to medium-sized codebooks, in both training and testing phases.

3.3.2. Hidden markov model (HMM)

Hidden Markov Models (HMMs) are powerful statistical models used in speech recognition systems to model time-varying statistical characteristics, making them particularly suitable for isolated word recognition applications. HMM is a supervised classification method that effectively models the transitions between observed states and the hidden states derived from observations. This method operates based on a Markov process, which represents the temporal evolution of speech signals within the system (L. R. Rabiner, 1989). HMM works on three fundamental problems: calculating the likelihood of an observation sequence, finding the most likely state sequence, and optimizing model parameters.

The core structure of HMM consists of a series of hidden states, transition probabilities between states, observation probabilities, and initial probabilities. The model components are defined as follows: a set of states $S = \{S_1, S_2, \ldots, S_N\}$, a transition probability matrix $A = \{a_{ij}\}$, observation probability distributions $B = \{b_i(o_t)\}$, and initial state probabilities $\pi = \{\pi_i\}$. Here, a_{ij} represents the transition probability from the i-th state to the j-th state; $b_i(o_t)$, the probability of generating observation o_t in the i-th state; and π_i, the initial probability of the i-th state. For continuous features, observation probabilities $b_i(o_t)$ are generally modeled using Gaussian distributions, as expressed in Equation (10):

$$b_i(o_t) = \sum_{m=1}^{M} c_{im} \mathsf{N}\left(o_t; \mu_{im}, \Sigma_{im}\right) \tag{10}$$

Here, c_{im} represents the weight of the m-th Gaussian component in the i-th state; $\mathsf{N}\left(o_t; \mu_{im}, \Sigma_{im}\right)$, the Gaussian probability density function; μ_{im}, the mean vector; Σ_{im}, the covariance matrix and M, the number of components.

HMM is used in speech recognition based on three fundamental problems. The first problem involves calculating the likelihood of the observation sequence $O = (O_1, O_2, \ldots, O_T)$ given the model $\lambda = \{A, B, \pi\}$ denoted as $P(O \mid \lambda)$. This is efficiently solved using the Forward algorithm, where the forward probabilities $\alpha_t(i)$ are computed as in Equation (11):

$$\alpha_t(i) = \left[\sum_{j=1}^{N} \alpha_{t-1}(j) a_{ji} \right] b_i(o_t) \tag{11}$$

Here, $\alpha_t(i)$ represents the probability of being in the i-th state at time t and producing the observations $O_1, O_2, \ldots, O_t$. Initially, $\alpha_1(i) = \pi_i b_i(O_1)$ is defined, and the total probability at the final step is obtained as $P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$.

The second problem involves determining the most likely state sequence $Q = (q_1, q_2, \ldots, q_T)$ given the observation sequence and the model $\lambda = \{A, B, \pi\}$. This is solved using the Viterbi algorithm (Rabiner, 1989). The Viterbi algorithm employs dynamic programming to find the state sequence with the highest probability, as expressed in Equation (12):

$$\delta_t(i) = \max_{1 \le j \le N} \left[\delta_{t-1}(j) a_{ji} \right] b_i(o_t) \qquad (12)$$

Here, $\delta_t(i)$ represents the maximum probability of producing the observations $O_1, O_2, \ldots, O_t$ while being in the i-th state at time t; a_{ji}, the transition probability from state j to state i; and $b_i(o_t)$, the probability of producing observation o_t in the i-th state. Initially, $\delta_1(i) = \pi_i \, b_i(O_1)$ is defined, and the most likely state sequence is determined through backtracking.

The third problem involves optimizing the model parameters $\lambda = \{A, B, \pi\}$ for the given observation sequence $O_1, O_2, \ldots, O_t$ in the best possible way. This is achieved using the Baum-Welch algorithm. The Baum-Welch algorithm iteratively updates the parameters to maximize the likelihood of the observations. In this process, both the Forward and Backward algorithms are utilized. The backward probabilities $\beta_t(i)$ are defined in Equation (13):

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij}\, b_j\left(o_{t+1}\right) \beta_{t+1}(j) \tag{13}$$

Here, $\beta_t(i)$ represents the probability of producing the observations $O_{t+1}, O_{t+2}, \ldots, O_T$ while being in the i-th state at time t. The Baum-Welch algorithm uses the values of $\alpha_t(i)$ and $\beta_t(i)$ to re-estimate the transition and observation probabilities, repeating this process until convergence.

The advantage of HMM lies in its ability to statistically model the dynamic structure of speech signals over time. However, its performance depends on the appropriate selection of observation probability distributions. Additionally, the computational cost of the process increases with large datasets, making the efficient implementation of the Baum-Welch algorithm significant. HMM has long been accepted as a standard method in speech recognition literature, successfully applied in both isolated word recognition and continuous speech recognition.

3.3.3. Gaussian Mixture Models (GMMs)

Gaussian Mixture Models (GMMs) are statistical methods used in speech recognition systems to model the probability density functions (PDFs) of feature vectors, making them a widely employed approach in supervised classification (Reynolds, 1995). GMM represents a dataset as a weighted combination of multiple Gaussian components rather than a single Gaussian distribution, enabling flexible modeling of complex data distributions. This method is particularly suitable for continuous features, such as MFCCs, as it effectively captures the clustering structure of data points. GMM assumes a latent variable indicating which component each data point belongs to and performs classification by estimating the probability of these latent variables.

The fundamental structure of GMM is defined as a linear combination of K Gaussian components. The probability density function of the model is expressed as follows in Equation (14):

$$p(x\ |\lambda) = \sum_{k=1}^{K} w_k\ \mathrm{N}\left(x\ |\mu_k, \Sigma_k\right) \tag{14}$$

Here, $p(x\ |\lambda)$ denotes the probability density of observation x; w_k, the weight of the k-th component (with $\sum_{k=1}^{K} w_k$ and $w_k \geq 0$); $\mathrm{N}\left(x\ |\mu_k, \Sigma_k\right)$, the probability density function of the k-th Gaussian component; μ_k, the mean vector; Σ_k, the covariance matrix; and K, the number of components. Each Gaussian component is modeled as an independent normal distribution, as given by Equation (15):

$$\mathrm{N}\left(x\ |\mu_k, \Sigma_k\right) = \frac{1}{(2\pi)^{\frac{d}{2}}\ |\Sigma_k|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)\right) \tag{15}$$

Here, d represents the dimension of the feature vector; $|\Sigma_k|$, the determinant of the covariance matrix; and $(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)$ the Mahalanobis distance. This structure serves as a powerful tool for modeling multimodal distributions within a dataset.

The parameters of GMM, $\lambda = \{w_k, \mu_k, \Sigma_k\}$, are optimized using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977). The EM algorithm consists of two main phases: the Expectation (E) phase and the Maximization (M) phase. In the E phase, the expected values of the latent variables (posterior probabilities) are calculated for each data point, as expressed in Equation (16):

$$\gamma_{ik} = \frac{w_k \, \mathrm{N}\left(x_i \mid \mu_k, \Sigma_k\right)}{\sum_{j=1}^{K} w_j \, \mathrm{N}\left(x_i \mid \mu_j, \Sigma_j\right)} \tag{16}$$

Here, γ_{ik} represents the posterior probability that the i-th data point belongs to the k-th component, and x_i represents the i-th data point.

In the M phase, these posterior probabilities are used to update the weights, mean vectors, and covariance matrices. The weights are updated as shown in Equation (17):

$$w_k^{\text{new}} = \frac{1}{N} \sum_{i=1}^{N} \gamma_{ik} \tag{17}$$

The mean vectors are updated according to Equation (18):

$$\mu_k^{\text{new}} = \frac{\sum_{i=1}^{N} \gamma_{ik} \, x_i}{\sum_{i=1}^{N} \gamma_{ik}} \tag{18}$$

The covariance matrices are updated as given in Equation (19):

$$\Sigma_k^{\text{new}} = \frac{\sum_{i=1}^{N} \gamma_{ik} (x_i - \mu_k^{\text{new}})(x_i - \mu_k^{\text{new}})^T}{\sum_{i=1}^{N} \gamma_{ik}} \tag{19}$$

This iterative process continues until the log-likelihood function no longer shows further improvement, thereby maximizing the model's fit to the dataset. In the GMM classification phase, the probability density value of a test data point is calculated by summing the contributions of all components. This value is used to determine which class the data

point belongs to according to the model. Separate GMMs can be trained for different clusters within the dataset, and the test data is assigned to the model that yields the highest probability. During this process, the selection of hyperparameters, such as the number of components and whether the covariance matrices are used in full or diagonal form, directly affects the model's performance. One of the advantages of GMM is its ability to flexibly model multimodal distributions, which makes it effective in capturing various feature variations. However, the model's success depends on the appropriate selection of the number of components and initial parameters. There is a risk of the EM algorithm getting trapped in local maxima. GMM is a widely used method in the speech recognition literature and is recognized as an effective tool, particularly for feature-based classification problems.

3.4. Heuristic Optimization Methods

Heuristic optimization algorithms are computational methods developed to solve complex and typically multidimensional optimization problems, drawing inspiration from nature or relying on problem-specific strategies. Adopting a population-based approach, these methods aim to systematically explore the solution space to converge toward a global optimum. In this section, the fundamental working principles of the Differential Evolution (DE), Self-Adaptive DE (SaDE), and Backtracking Search Algorithm (BSA) algorithms, which are used to find a solution closer to the global optimum for GMM parameters, are explained.

3.4.1. Differential evolution (DE)

Differential Evolution (DE), proposed by Storn and Price in 1997, is a meta-heuristic approach that effectively solves complex and multi-modal problems using population-based search techniques, particularly for continuous optimization problems. This method stands out for its advantages, such as fast

convergence, minimal control parameters, and the absence of derivative information requirements (Storn & Price, 1997).

The basic structure of the DE algorithm starts with a randomly initialized population. The population consists of N_p individuals, each representing an m dimensional solution vector: $y_{k,T} = \left[y_{s1,T}, y_{s2,T}, \ldots, y_{km,T} \right]$, where $k = 1, 2, \ldots, N_p$ and T represents the current iteration number. The algorithm consists of three main stages: mutation, crossover, and selection. First, for each individual in the population, a mutant vector is created. The most common mutation strategy, known as "DE/rand/1," is defined by Equation (20):

$$w_{k,T+1} = y_{s1,T} + \beta\left(y_{s2,T} - y_{s3,T} \right) \tag{20}$$

Here, $w_{k,T+1}$ is the mutant vector for the k-th individual; $y_{s1,T}, y_{s2,T}, y_{s3,T}$ are three vectors randomly selected from the population and distinct from each other; β, the scaling factor, typically takes a value in the range [0, 2], controlling the magnitude of the mutation step and influencing the algorithm's exploration capability.

After mutation, the crossover stage is performed. In this stage, an intermediate vector $z_{k,T+1} = \left[z_{k1,T+1}, z_{k2,T+1}, \ldots, z_{km,T+1} \right]$ is created between the mutant vector $w_{k,T+1}$ and the target vector $y_{k,T}$. The crossover operation, typically using a uniform (binomial) crossover with a crossover rate α (generally in the range $[0,1]$), is performed as shown in Equation (21):

$$z_{kj,T+1} = \begin{cases} w_{kj,T+1}, & \text{if rand}_j \le \alpha \quad \text{or} \quad j = j_{\text{select}}, \\ y_{kj,T}, & \text{otherwise,} \end{cases} \tag{21}$$

Here, rand_j, a random number in the range $[0,1]$, is generated, and j_{select} is a randomly chosen index to ensure that at least one dimension of the mutant vector is taken. The parameter α regulates the algorithm's exploration and exploitation balance; higher α values promote more exploration.

Finally, in the selection stage, a comparison is made between the intermediate vector $z_{kj,T+1}$ and the target vector $y_{k,T}$, and depending on their fitness, the better individual is selected for the next iteration. If the problem is a minimization problem, Equation (22) is used for selection:

$$y_{k,T+1} = \begin{cases} z_{k,T+1}, & \text{if } h\left(z_{k,T+1}\right) \le h\left(y_{k,T}\right), \\ y_{k,T}, & \text{otherwise.} \end{cases} \tag{22}$$

Here, $h(\cdot)$ represents the target function and yields a lower suitability value, indicating a better solution. However, during this process, all individuals in the population repeat the steps, continuing until a specified condition, such as the maximum number of iterations or the convergence of the target function, is met. Among the advantages of DE, the simplicity of parameter tuning stands out; it only requires a few control parameters, namely β, α, and N_p.

3.4.2. Self-adaptive differential evolution (SaDE)

Self-Adaptive Differential Evolution (SaDE) is an advanced version of the DE algorithm that automatically adjusts control parameters and mutation strategies during the optimization process, making it a more efficient meta-heuristic method. Proposed by Qin and Suganthan in 2005, SaDE aims to eliminate the need for manual parameter tuning, which is often necessary in standard DE for different problem types. SaDE provides better adaptation to various problem characteristics

during the evolution of the population by dynamically selecting both the scaling factor β β and the crossover rate α α, as well as the mutation strategy, based on global and local information obtained throughout the iterations (Qin & Suganthan, 2005).

SaDE, like DE, uses the same population structure: a population of N_p individuals, where each individual is represented as an m-dimensional solution vector: $y_{k,T} = \left[y_{s1,T}, y_{s2,T}, \ldots, y_{km,T} \right]$, where $k = 1, 2, \ldots, N_p$, and T represents the current iteration number. One of the key differences in SaDE is that it adaptively chooses between four different mutation strategies during iterations: DE/rand/1, DE/current-to-best/2, DE/rand/2, and DE/current-to-rand/1. These strategies vary in terms of the diversity they introduce to the population and the speed of convergence, allowing the algorithm to progress according to the specific conditions of the problem.

The first strategy provided is DE/rand/1, which is a basic strategy in DE. In this strategy, the difference between three randomly selected individuals from the population is used (Equation (20)). The second strategy, DE/current-to-best/2, directs the current individual toward the best individual while using two difference vectors, as shown in Equation (23):

$$w_{k,T+1} = y_{k,T} + \beta\left(y_{\text{best},T} - y_{k,T}\right) + \beta\left(y_{s1,T} - y_{s2,T}\right) + \beta\left(y_{s3,T} - y_{s4,T}\right) \quad (23)$$

Here, $y_{\text{best},T}$ represents the best individual in the population. The third strategy, DE/rand/2, involves creating a mutant vector using two different random vectors, as shown in Equation (24):

$$w_{k,T+1} = y_{s1,T} + \beta\left(y_{s2,T} - y_{s3,T}\right) + \beta\left(y_{s4,T} - y_{s5,T}\right) \quad (24)$$

The fourth strategy, DE/current-to-rand/1, performs a direct orientation from the current individual to a randomly selected individual and applies crossover, creating a mutant vector as shown in Equation (25):

$$w_{k,T+1} = y_{k,T} + \beta\left(y_{s1,T} - y_{k,T}\right) + \beta\left(y_{s2,T} - y_{s3,T}\right) \qquad (25)$$

These strategies demonstrate SaDE's ability to adapt to different problem types and optimization stages. In SaDE, the β and α parameters are adjusted separately for each individual; the β values are sampled from a normal distribution.

In the SaDE algorithm, the β and α parameters are dynamically adjusted based on the previous iteration. The β coefficients are randomly selected from a normal distribution $\beta \sim N(\mu\beta, 0.3)$; here, μ_β is calculated as the average of the β values that were effective in previous iterations and is updated at the end of each iteration. In a similar approach, the α parameters are derived from a distribution $\alpha \sim N(\mu_\alpha, 0.1)$, and μ_α is continuously optimized based on past success rates. This flexible adaptation mechanism allows the algorithm to adapt more effectively to various problem conditions and changing optimization needs.

3.4.3.Backtracking search optimization algorithm (BSA)

Backtracking Search Optimization Algorithm (BSA), proposed by Civcioğlu in 2013, is a population-based, evolutionary algorithm designed to solve numerical optimization problems. BSA operates with a single control parameter and has low sensitivity to this parameter, making it easily adaptable to different problem types. Additionally, the algorithm incorporates a memory mechanism and determines the search direction by randomly selecting individuals from past generations, thus

integrating previous experiences into the optimization process (Civicioglu, 2013).

N_p represents the population size, and the population consists of N_p individuals. Each individual, $y_{k,T} = \left[y_{s1,T}, y_{s2,T}, \ldots, y_{km,T} \right]$, is expressed as an m-dimensional solution vector. Here, $k = 1, 2, \ldots, N_p$ denotes the individual index, T indicates the current iteration number, and m represents the problem's dimension (number of variables). The population is randomly distributed within the lower bounds (low_j) and upper bounds (up_j) for each dimension, and this process is defined by Equation (26):

$$y_{kj,T} \sim \mathrm{U}\left(\mathrm{low}_j, \mathrm{up}_j\right) \tag{26}$$

Here, U denotes a uniform distribution; low_j and up_j represent the lower and upper bounds, respectively, for the j-th dimension. Additionally, BSA creates a historical population ($oldP$), which serves as the algorithm's memory and is randomly distributed, a process defined by Equation (27):

$$\mathrm{oldP}_{kj,T} \sim \mathrm{U}\left(\mathrm{low}_j, \mathrm{up}_j\right) \tag{27}$$

The algorithmic structure of BSA consists of five main stages: initialization, selection-I, mutation, crossover, and selection-II. In the first stage, the population ($y_{k,T}$) and the historical population ($oldP$) are randomly generated. In the selection-I stage, the historical population is updated, and the search direction is determined. During this stage, a random probability check decides whether the historical population will be updated. For this purpose, two random numbers are selected, and the following rule is applied according to Equation (28):

$$\text{if } a < b \text{ then oldP} := y_{k,T} \tag{28}$$

Here, a and b are random probability values selected from $U(0,1)$ uniform distribution, and $:=$ indicates the update operation. Subsequently, the individuals of the historical population are randomly reordered, a process defined by Equation (29):

$$\text{oldP} := \text{permuting}\left(\text{oldP}\right) \tag{29}$$

Here, *permuting* is a function that randomly reorders the individuals. In the mutation stage, BSA generates an intermediate trial population ($w_{k,T+1}$), which represents the candidate solutions for the next iteration. The mutation is performed using the difference between the historical population ($oldP$) and the current population ($y_{k,T}$), and this process is defined by Equation (30):

$$w_{k,T+1} = y_{k,T} + \beta\left(\text{oldP}_{k,T} - y_{k,T}\right) \tag{30}$$

Here, β is a scale factor that controls the amplitude of the search direction matrix ($\text{oldP}_{k,T} - y_{k,T}$) and, on average, is a random value from a uniform distribution with a variance of 1, calculated as $\beta = 3 \cdot randn.\ randn \sim N\left(0,1\right)$. This approach allows BSA to utilize both global and local search mechanisms in a controlled manner. The trial population $w_{k,T+1}$ is kept within the bounds defined by the control mechanism low_j and up_j for each dimension.

In the crossover stage, BSA creates a binary *map* matrix; this matrix has dimensions $N_p \times m$ and each element takes a value of either 0 or 1. The *map* determines which dimensions of

the current population ($y_{k,T}$) will be assigned to the trial population. The *map* matrix is generated using one of two different strategies, selected randomly or for each individual based on a specified probability. The crossover strategy of BSA is a complex structure, but it increases the diversity of the population.

In the selection-II stage, the fitness values $h(w_{k,T+1})$ of the trial population are compared with the fitness values $h(y_{k,T})$ of the current population, and the better individual is transferred to the next generation. Here, $h(\cdot)$ represents the objective function, and typically, it produces a value to be minimized. This comparison and selection process is expressed with Equation (31):

$$y_{k,T+1} = \begin{cases} w_{k,T+1}, & \text{if } h(w_{k,T+1}) \le h(y_{k,T}), \\ y_{k,T}, & \text{otherwise.} \end{cases} \tag{31}$$

This process continues until the maximum number of iterations or a stopping criterion is reached.

4. EXPERIMENTS AND DISCUSSION

This section presents the experimental results in two main stages, aligning with the core objectives of this study. The first stage provides a comparative analysis of classical methods (DTW, HMM, and GMM) using a real speech dataset to establish a performance benchmark. The second stage addresses the practical challenges of data collection by evaluating a novel framework centered on synthetic data generation. This part details the performance of GMM trained on data augmented via Harmonic-Percussive Source Separation (HPSS), first with the standard Expectation-Maximization (EM) algorithm, and

subsequently with advanced heuristic optimization approaches (DE, SaDE, and BSA).

4.1. Experimental Setup

In this study, a speaker-dependent isolated speech recognition engine was designed to feed the GIS database. The experiments were conducted using a 400-word Turkish codebook selected from the Highways Terminology section of the Transportation and Communication Terms Dictionary prepared by the Republic of Turkey Ministry of Transport, Maritime Affairs, and Communications. The codebook consists of professional terms related to transportation terminology and is presented in full in Table 1. All words in the codebook were sampled 20 times at 11,025 Hz with a fixed time interval of 1.5 seconds.

The speech recognition process was carried out in three main stages: preprocessing, feature extraction, and classification. In the preprocessing stage, the Short-time Energy method was used to separate voiced and unvoiced regions to reduce noise in the audio data and eliminate non-speech signals. Additionally, pre-emphasis was applied to enhance high-frequency components. As defined in Equation (2) and Equation (3), α value of $\alpha = 0.95$ was used. In the feature extraction stage, the MFCC (Mel-Frequency Cepstral Coefficients) method was preferred. The MFCC method processes audio signals by dividing them into 25 ms frames, applying a Hamming window function, and obtaining features using a Mel-frequency filter bank that mimics the human ear's frequency perception. Thirteen MFCC coefficients were calculated, with the first coefficient excluded, resulting in 12 feature values being used. The frame-rate for MFCC was set to 128, and the filter bank consisted of 32 filters.

The evaluation metrics used were recognition accuracy (number of correctly recognized words / total number of words)

and error rate. Recognition accuracy was calculated for each test based on the 400-word codebook, and the performance of the methods was compared according to these metrics.

Table 1. Codebook

1 ADA	81 CURUTME	161 HARKET	241 MALIYET	321 SEHIM
2 ACIK	82 DAĞ	162 HAREKETLI	242 MALZEME	322 SERVIS
3 AFUYMAN	83 DAIRESEL	163 HARITA	243 MANİVELA	323 SEVIYE
4 AGREGA	84 DAMPERLI	164 HAVA	244 MANSAP	324 SEVK
5 AGIR TASIT	85 DEBİ	165 HEMZEMIN	245 MARS	325 SEYYAR
6 AGIRLIK	86 DEBRİJAJ	166 HENDEK	246 MEKİK	326 SIKIŞLIK
7 AKARYAKIT	87 DEBUŞ	167 HEYELAN	247 MEMBA	327 SINIR
8 AKILLI	88 DEFORMASYON	168 HIZ	248 MENFEZ	328 SINIRI
9 ALAN	89 DELİCİ	169 HIZLAN	249 MERMER	329 SIVILAŞMA
10 ALISTIRMA	90 DEMIRYOLU	170 IDARE	250 MESAFE	330 SIVILDIR
11 ALIYMAN	91 DEPO	171 IFRAZ	251 MELEKI	331 SİNYALİZASYON
12 ALTGEÇIT	92 DEPREM	172 IHAL	252 MESNET	332 SIYAH
13 ALTITEMEL	93 DERİVASYON	173 IKTISAP	253 METRE	333 SOL
14 AMPATMAN	94 DERZ	174 ILERI	254 MICIR	334 SON
15 ANAKIRIŞ	95 DEVAMİ	175 IMAR	255 MIKTAR	335 SONDAJ
16 APLIKASYON	96 DEVER	176 İNŞAA	256 MINIBUS	336 SOMEL
17 ARABA	97 DİKAT	177 İNŞAAT	257 MOLOZ	337 SPIRAL
18 ARAÇ	98 DİLATASYON	178 ISKELE	258 MOTOR	338 STABILITE
19 ARAZÖZ	99 DİNGİL	179 İSTASYON	259 MOTORIN	339 STABILIZE
20 ARİYAT MALZEME	100 DİREKSİYON	180 İSTİKŞAF	260 MOZAİK	340 STANDART
21 ASALT	101 DISTRIBÜTOR	181 İŞ	261 MUAYENE	341 SU
22 ASMA	102 DOLUDUMA	182 ISARET	262 MÜRC	342 SÜRTÜNME
23 ASTAR	103 DOLGU	183 ISARETLEME	263 MÜCADELE	343 ŞANTİYE
24 ASINDIRICI	104 DÖKÜM	184 IŞÇI	264 MÜCAVIR	344 ŞARTNAME
25 ASINMA	105 DOZER	185 IŞGAL	265 MÜHENDIS	345 ŞASİ
26 ASINMIŞ	106 DONATI	186 IŞLETME	266 MÜLKIYET	346 ŞEF
27 AYIRICI SERİDİ	107 DONUŞ	187 İZ	267 MÜTEAHHIT	347 ŞERİT
28 AYIRMA	108 DOSEME	188 IZIN	268 NAKLIYE	348 SEV
29 AZ	109 DRENAJ	189 JEODEZİ	269 ORTA	349 ŞUBE
30 AZAMİ	110 DURAK	190 KALEME	270 ORTAK	350 TABAKA
31 BAĞLANTI	111 DURLAK	191 KALEM	271 OTORAY	351 TAHSIS
32 BAKIM	112 DURKLAMA	192 KALIP	272 OTOMASYON	352 TAKOGRAF
33 BANKET	113 DUVAR	193 KALİTE	273 ORTA	353 TALI
34 BARBAKAN	114 DÜSEY	194 KAMA	274 OTALAMA	354 TALVEG
35 BASIT	115 DÜZELTME	195 KAMYON	275 OTO	355 TANDEM
36 BAŞLANGIÇ	116 DÜZEY	196 KAMYONET	276 OTOBÜS	356 TANKER
37 BELGESİ	117 EGIM	197 KANTAR	277 OTOMOBİL	357 TAŞ
38 BENZİN	118 EGTIM	198 KAPALI	278 OTOYOL	358 TASI
39 BETON	119 EHLİYET	199 KAPLAMA	279 OYUK	359 TASIT
40 BEYAN	120 EKIP	200 KAR	280 ÖLÇME	360 TEHLIKE
41 BEZEME	121 EKSKA	201 KARAYOLU	281 PAN	361 TEHLIKELI
42 BIÇAK	122 EKSPERTIZ	202 KASA	282 ÖNLEME	362 TEKER
43 BİLET	123 ELEK	203 KATKI	283 PAPTA	363 TEKERLEK
44 BİLGİ	124 EMNİYET	204 KAVSAK	284 PALET	364 TEKNIK
45 BINDER	125 EMNİYET	205 KAYGAN	285 PALYE	365 TEKYON
46 BİNEK	126 EMNİYETSİZ	206 KAZA	286 PARAFIN	366 TEMEL
47 BİRİM	127 EMULSİYON	207 KAZI	287 PARALEL	367 TERSIYOL
48 BOKLET	128 ENKESIT	208 KAZIC	288 PARK	368 TIR
49 BİTİŞ	129 ENSİZ	209 KAZIK	289 PARLAMA	369 TIRMANMA
50 BİTÜM	130 ERİMEK	210 KAZIMA	290 PARSEL	370 TON
51 BOMBE	131 ENİYİYE	211 KAZIYICI	291 PASAJ	371 TRAFİK
52 BORDÜR	132 ETUD	212 KEMER	292 PATLAYICI	372 TRANSIT
53 BORU	133 FAUNA	213 KENAR	293 PERE	373 TRANVAY
54 BOSAJ	134 FERAĞ	214 KENT	294 PERİYODİK	374 TREN
55 BOS	135 FEZEYAN	215 KESIKLI	295 PGZAL	375 TUNEL
56 BOSALTMA	136 GABARI	216 KESIM	296 PİK	376 ULAŞIM
57 BOYA	137 GAGA	217 KESKI	297 PLAN	377 UYARI
58 BOYKESIT	138 GARA	218 KESON	298 PLANÇETE	378 VEKTÖR
59 BOYUNA	139 GARAJ	219 KILOGRAM	299 PLAKA	379 VİBRASYON
60 BOZUK	140 GEÇİ	220 KİLOMETRE	300 PLATFORM	380 VİDA
61 BOLUNMUŞ	141 GEÇİCİ	221 KIRMIZI	301 PLENT	381 YAĞIŞ
62 BRUKNER	142 GEÇİŞ	222 KIS	302 POLIMER	382 YAĞMUR
63 BUZ	143 GEÇİŞ	223 KİLİTLİ	303 PUSKURTME	383 YANDAN
64 BÜZ	144 GEÇİT	224 KİRİŞ	304 RADYE	384 YARALI
65 CAM	145 GERGİ	225 KÖ	305 RAKIM	385 YARMA
66 CAMIL	146 GIDON	226 KOMPLEKS	306 RAKORTMAN	386 YASAK
67 ÇAMUR	147 GİRİŞ	227 KONSOL	307 RAMPA	387 YASTIK
68 ÇARPMA	148 GÖRÜŞ	228 KONTROL	308 RAPOR	388 YAVAŞLA
69 CED	149 HIZSINIRI	229 KORIDOR	309 RASTER	389 YAYA
70 ÇEKİCİ	150 GRADASYON	230 KORKULUK	310 REFUJ	390 YAZ
71 ÇELİK	151 GREYDER	231 KORNİŞ	311 REGLAJ	391 YERİ
72 ÇERÇEVE	152 GROBETON	232 KOROZYON	312 RIJT	392 YERLEŞİM
73 ÇEVRE	153 GUSE	233 KOT	313 RODMIKS	393 YETERLİLİK
74 ÇİĞ	154 GÜCLÜ	234 KÖPRÜ	314 ROGAR	394 YOK
75 ÇİZGI	155 GÜNDÜZ	235 KOSE	315 ROMORK	395 YOL
76 ÇİZGİSEL	156 GÜVENLI	236 KUM	316 ROPER	396 YÖNETİCİ
77 ÇOK	157 GÜVENSIZ	237 KURP	317 RUZGAR	397 YÜK
78 ÇÖKME	158 HACIM	238 KUSUR	318 SADEME	398 YÜZEY
79 ÇÖRTEN	159 HAFIF	239 MAKİNA	319 SAĞ	399 ZAYIF
80 ÇUKUR	160 HARCKET	240 MALC	320 SARKMA	400 ZORUNLU

4.2. Speech Recognition Results

The performance of Dynamic Time Warping (DTW), Hidden Markov Models (HMM), and Gaussian Mixture Models (GMM) was evaluated under a unified experimental setup to ensure a fair comparison. For all three approaches, the dataset consisted of a 400-word Turkish codebook, utilizing 15 samples per word for training (a total of 6000 samples) and 5 samples per word for testing (a total of 2000 samples). The workflow shared across the methods included a preprocessing stage, where Short-time Energy was used for silence removal and a pre-emphasis filter was applied to enhance high-frequency components. Subsequently, in the feature extraction stage, 12 MFCC coefficients were derived from 25 ms frames processed with a 32-filter Mel-frequency bank. The integrated workflow, highlighting both the common and method-specific stages, is detailed in the pseudocode presented in Figure 1.

Algorithm 1 Isolated Word Recognition Algorithms (HMM, GMM, DTW)

Input:
  $TrainSet = \{t_1, \ldots, t_n\}$, $TrainLabels = \{\ell_1, \ldots, \ell_n\}$, $TestSet = \{s_1, \ldots, s_m\}$
Output:
  Labels ▷ Predicted label for each test utterance
1. Preprocessing
CleanTrain ← ∅; *CleanTest* ← ∅
for each x in $TrainSet \cup TestSet$ **do**
  $x_1 \leftarrow \text{RemoveSilence}(x)$; $x_2 \leftarrow \text{PreEmphasisFilter}(x_1)$
  if $x \in TrainSet$ **then** $CleanTrain \leftarrow CleanTrain \cup \{x_2\}$
  else $CleanTest \leftarrow CleanTest \cup \{x_2\}$
  end if
end for
2. Feature Extraction (MFCC)
for each t in *CleanTrain* **do** $MFCC_Train[t] \leftarrow \text{ComputeMFCC}(t)$
end for
for each s in *CleanTest* **do** $MFCC_Test[s] \leftarrow \text{ComputeMFCC}(s)$
end for
3. Training and Recognition

3a. HMM-Based Approach

// Model Training (Baum–Welch)
Models ← ∅
for each unique ℓ in *TrainLabels* **do**
  $Data_\ell \leftarrow \{MFCC_Train[t] : \text{label}(t) = \ell\}$
  $\text{HMM}_\ell \leftarrow \text{InitializeHMM}(\text{numStates=N})$
  $\text{HMM}_\ell \leftarrow \text{BaumWelchTrain}(\text{HMM}_\ell, Data_\ell)$
  $Models[\ell] \leftarrow \text{HMM}_\ell$
end for
// Recognition (Viterbi)
for each s in *CleanTest* **do**
  $bestScore \leftarrow -\infty$; $bestLabel \leftarrow$ None
  for each (ℓ, HMM_ℓ) in *Models* **do**
    $score \leftarrow \text{ViterbiLogLik}(\text{HMM}_\ell, MFCC_Test[s])$
    if $score > bestScore$ **then**
      $bestScore \leftarrow score$; $bestLabel \leftarrow \ell$
    end if
  end for
  $Labels[s] \leftarrow bestLabel$
end for

3b. GMM-Based Approach

// Model Training (EM)
Models ← ∅
for each unique ℓ in *TrainLabels* **do**
  $Data_\ell \leftarrow \{MFCC_Train[t] : \text{label}(t) = \ell\}$
  $\text{GMM}_\ell \leftarrow \text{InitializeGMM}(\text{numComponents=M})$
  $\text{GMM}_\ell \leftarrow \text{EM_TrainGMM}(\text{GMM}_\ell, Data_\ell)$
  $Models[\ell] \leftarrow \text{GMM}_\ell$
end for
// Classification
for each s in *CleanTest* **do**
  $bestScore \leftarrow -\infty$; $bestLabel \leftarrow$ None
  for each (ℓ, GMM_ℓ) in *Models* **do**
    $score \leftarrow \text{LogLikelihood}(\text{GMM}_\ell, MFCC_Test[s])$
    if $score > bestScore$ **then**
      $bestScore \leftarrow score$; $bestLabel \leftarrow \ell$
    end if
  end for
  $Labels[s] \leftarrow bestLabel$
end for

3c. DTW-Based Approach

// Recognition
for each s in *CleanTest* **do**
  $bestScore \leftarrow +\infty$; $bestLabel \leftarrow$ null
  for each t in *CleanTrain* **do**
    $d \leftarrow \text{DTW_Dist}(MFCC_Test[s], MFCC_Test[s])$
    if $d < bestScore$ **then**
      $bestScore \leftarrow d$; $bestLabel \leftarrow \text{LabelOf}(t)$
    end if
  end for
  $Labels[s] \leftarrow bestLabel$
end for

return *Labels*

Figure 1. Pseudocode for the isolated word recognition workflow

The primary performance metrics, including recognition accuracy and computation time, are summarized in Table 2. The results indicate a significant performance variance among the models. HMM achieved the highest accuracy at 98.70%, correctly identifying 1974 out of 2000 test samples. This superior performance can be attributed to HMM's inherent strength in modeling the temporal sequences present in speech signals. DTW, a template-matching approach, provided a robust baseline performance with 89.05% accuracy, successfully recognizing 1781 samples. GMM, trained using the Expectation-Maximization algorithm with 12 Gaussian components over 400 epochs, yielded the lowest accuracy of 82.10%. This is likely because the standard GMM acts as a "bag-of-frames" model, disregarding the crucial temporal order of acoustic features, which is critical for speech recognition.

In terms of computational cost, the recognition times on the MATLAB platform were comparable, though influenced by different factors. DTW's time scaled with the number of training templates (9.123–17.657 s), while HMM's test time depended on the number of models (10.829–29.088 s), and GMM's was influenced by the model complexity and training epochs (11.945–18.309 s).

Table 2. Comparative performance metrics of DTW, HMM, and GMM

Method	Success Rate (%)	Correctly Recognized	Recognition Time (s, 95% CI)
HMM	98.70	1974 / 2000	10.829 – 29.088
DTW	89.05	1781 / 2000	9.123 – 17.657
GMM	82.10	1642 / 2000	11.945 – 18.309

A visual analysis of the results further supports these findings. Figure 2 presents the comparative recognition success heatmaps, where the scarcity of failed trials (white lines) for HMM is visually evident compared to the higher error density in

DTW and particularly GMM. Furthermore, Figure 3 illustrates the word-level accuracy distributions. The chart confirms HMM's consistency, with nearly all 400 words achieving high recognition rates (0.8-1.0 range), whereas the accuracy for words in DTW and GMM is more broadly distributed, indicating lower reliability for certain words.

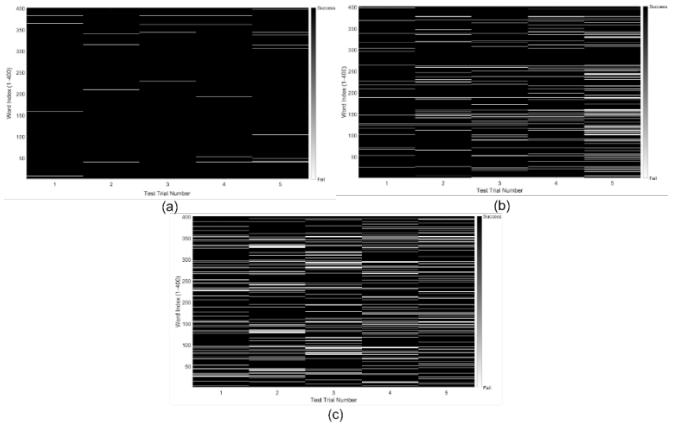


Figure 2. Comparative recognition success heatmaps for (a) HMM, (b) DTW, and (c) GMM. Black indicates success and white indicates failure for each of the five test trials per word.

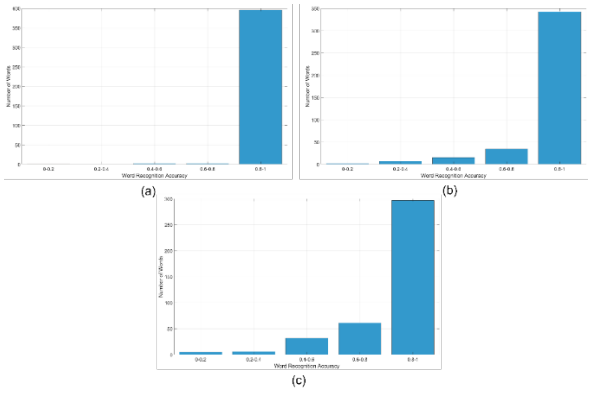


Figure 3. Comparative distribution of word-level accuracy for (a) HMM, (b) DTW, and (c) GMM. The height of the bar shows the number of words falling into each accuracy bracket.

4.3. Synthetic Data Generation and Hybrid GMM-Based Speech Recognition Results

In large-vocabulary, speaker-dependent speech recognition systems, it is practically challenging to compile comprehensive datasets that include multiple repetitions for each word. The real-world data collection process is both time-consuming and often fails to create a statistically adequate observation pool, due to difficulties in capturing pronunciation variations arising from different psychological or physiological conditions and in predicting the level and characteristics of ambient noise. To overcome the challenges posed by multi-repetition data collection, this study presents an innovative approach that generates synthetic data using only a single real audio recording for each word and trains hybrid models optimized with heuristic algorithms on this data. This method offers a practical solution by reducing the collection time for a 400-word dataset from approximately 6.2 hours to 30 minutes.

At the core of this approach is the Harmonic-Percussive Source Separation (HPSS) technique, which separates an audio signal into its harmonic (tonal) and percussive (transient) components. The HPSS method was preferred because it produces fewer artifacts and preserves speech naturalness compared to other time-scale modification techniques. After separation, natural pronunciation differences are simulated by scaling the signal in the time domain while preserving its frequency structure. Experimental observations have shown that scaling factors randomly selected from the U(0.90, 1.10) interval produce the most natural results. With this method, 19 synthetic repetitions were generated from the single real recording of each word, thus expanding the training dataset to a size of 400x20. Additionally, to simulate the recording environment, random white noise of 30-40 dB was added to the signals. The variation in frame lengths created by the generated synthetic data for the

word "ADA" is shown in Figure 4. The test dataset, in turn, was compiled to consist of 5 real observations for each word (400x5).

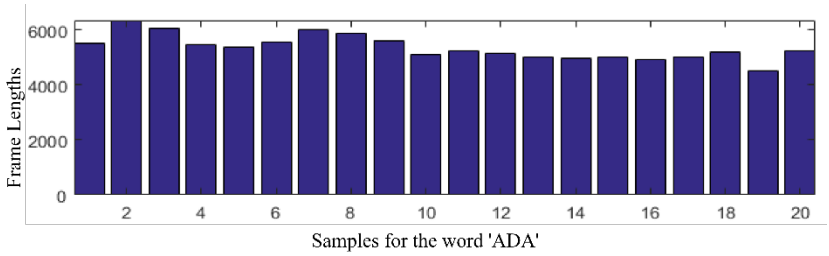


Figure 4. Frame lengths of 20 synthetic samples generated from a single real sample for the 'ADA' audio signal.

The performance of this synthetically enriched dataset was evaluated using Gaussian Mixture Models (GMM) due to their lower computational load compared to HMM and higher performance than DTW. First, the GMM parameters were trained with the standard Expectation-Maximization (EM) algorithm. The results presented in Table 3 clearly demonstrate the success of synthetic data enrichment. Specifically, when the Reference Size (the number of synthetic repetitions used) exceeds 12 and the Number of Mixed-Gaussians exceeds 8, the recognition accuracy surpasses the 98% level, reaching a rate considered quite high according to the literature. This demonstrates that synthetic data generated from a single sample significantly increases the model's generalization capability. The performance variation of the EM-based GMM under different parameters is shown in Figure 5, while the detailed statistical results are presented in Table 4.

Table 3. Speech recognition performance values (x100%) with EM-based GMM.

	Reference Size (Number of Repetitions)									
Number of Mixed-Gaussians	2	4	6	8	10	12	14	16	18	20
2	0.7945	0.8570	0.8545	0.8570	0.8590	0.8550	0.8500	0.8560	0.8560	0.8555
4	0.8560	0.8505	0.8560	0.8585	0.8550	0.9445	0.9465	0.9420	0.9450	0.9460
6	0.8595	0.8545	0.8565	0.8520	0.8515	0.9695	0.9705	0.9695	0.9675	0.9710
8	0.8550	0.8560	0.8575	0.8520	0.8570	0.9810	0.9825	0.9820	0.9805	0.9805
10	0.8535	0.8530	0.8590	0.8545	0.8520	0.9860	0.9845	0.9850	0.9850	0.9845

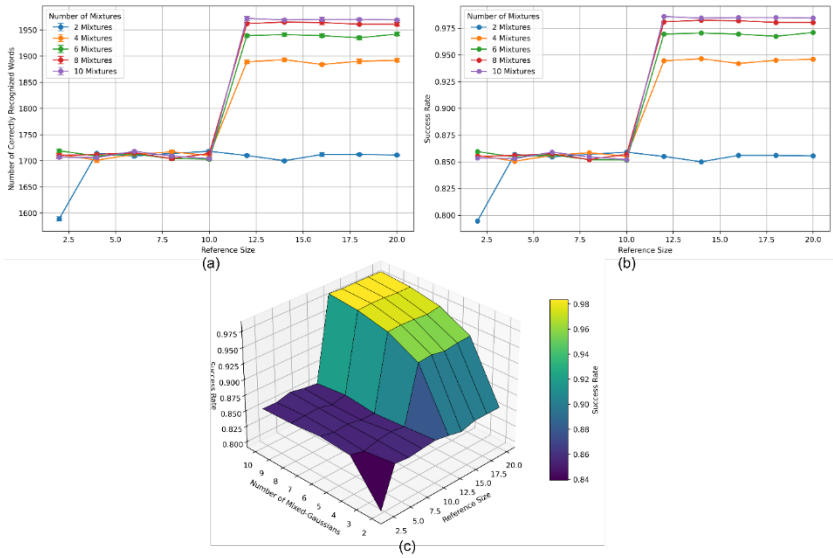


Figure 5. Speech recognition performance with EM-based GMM; (a) and (b) The effect of reference size on speech recognition, (c) The effect of reference size and the number of mixed-gaussians used on speech recognition.

Table 4. Detailed statistical data on the speech recognition performance with EM-based GMM.

Number of Mixed-Gaussians	Statistics (for 95% CI)	Reference Size (Number of Repetitions)									
		2	4	6	8	10	12	14	16	18	20
2	mean	1589.000	1714.000	1709.000	1714.000	1718.000	1710.000	1700.000	1712.000	1712.000	1711.000
	std	2.043	0.542	1.767	1.998	1.497	0.950	0.896	2.021	0.838	0.677
	lower	1584.996	1712.939	1705.536	1710.085	1715.066	1708.139	1698.244	1708.039	1710.357	1709.674
	upper	1593.004	1715.061	1712.464	1717.915	1720.934	1711.861	1701.756	1715.961	1713.643	1712.326
4	mean	1712.000	1701.000	1712.000	1717.000	1710.000	1889.000	1893.000	1884.000	1890.000	1892.000
	std	1.871	2.407	0.508	1.524	2.125	1.725	1.944	1.084	2.336	1.929
	lower	1708.333	1696.283	1711.005	1714.012	1705.835	1885.619	1889.191	1881.876	1885.422	1888.219
	upper	1715.667	1705.717	1712.995	1719.988	1714.165	1892.381	1896.809	1886.124	1894.578	1895.781
6	mean	1719.000	1709.000	1713.000	1704.000	1703.000	1939.000	1941.000	1939.000	1935.000	1942.000
	std	1.585	0.784	1.247	1.848	1.384	1.368	1.736	1.526	1.801	1.702
	lower	1715.893	1707.463	1710.557	1700.377	1700.288	1936.319	1937.598	1936.009	1931.471	1938.664
	upper	1722.107	1710.537	1715.443	1707.623	1705.712	1941.681	1944.402	1941.991	1938.529	1945.336
8	mean	1710.000	1712.000	1715.000	1704.000	1714.000	1962.000	1965.000	1964.000	1961.000	1961.000
	std	2.110	1.543	2.317	1.138	0.681	1.101	0.728	2.157	0.594	1.753
	lower	1705.864	1708.975	1710.458	1701.769	1712.665	1959.841	1963.573	1959.772	1959.836	1957.565
	upper	1714.136	1715.025	1719.542	1706.231	1715.335	1964.159	1966.427	1968.228	1962.164	1964.435
10	mean	1707.000	1706.000	1718.000	1709.000	1704.000	1972.000	1969.000	1970.000	1970.000	1969.000
	std	1.595	2.139	0.898	2.214	1.203	2.009	1.092	2.268	1.151	0.830
	lower	1703.874	1701.808	1716.240	1704.661	1701.642	1968.062	1966.860	1965.555	1967.744	1967.373
	upper	1710.126	1710.192	1719.760	1713.339	1706.358	1975.938	1971.140	1974.445	1972.256	1970.627

To overcome the tendency of the EM algorithm to converge to local optima and to further improve the success rate, the GMM parameters were also optimized in a hybrid structure

using three different heuristic optimization algorithms. These hybrid models accelerated the convergence process by using the solution vectors obtained from EM as the initial population. The results obtained with BSA are presented in Table 5 and Table 6. BSA-GMM provided 52 additional correct recognitions (a 2.60% improvement) compared to the EM-based model.

Table 5. Speech recognition performance values (100x %) with BSA-based GMM.

	Reference Size (Number of Repetitions)									
Mix Gaussian	2	4	6	8	10	12	14	16	18	20
2	0.7950	0.8580	0.8540	0.8580	0.8580	0.8555	0.8495	0.8550	0.8565	0.8560
4	0.8570	0.8510	0.8555	0.8580	0.8555	0.9455	0.9460	0.9410	0.9455	0.9470
6	0.8605	0.8550	0.8555	0.8525	0.8525	0.9705	0.9715	0.9700	0.9670	0.9700
8	0.8560	0.8565	0.8585	0.8525	0.8575	0.9815	0.9815	0.9830	0.9810	0.9800
10	0.8540	0.8540	0.8595	0.8555	0.8530	0.9865	0.9840	0.9860	0.9860	0.9835

Table 6. Statistical data on the speech recognition performance with BSA-based GMM; a 95% confidence interval, alpha=0.05, and a normal distribution model were used.

Number of Mixed-Gaussians	Statistics (for 95% CI)	Reference Size (Number of Repetitions)									
		2	4	6	8	10	12	14	16	18	20
2	mean	1590.000	1716.000	1708.000	1716.000	1716.000	1711.000	1699.000	1710.000	1713.000	1712.000
	std	2.057	2.059	1.721	1.118	1.895	2.219	1.751	2.465	2.453	0.833
	lower	1585.969	1711.964	1704.628	1713.809	1712.285	1706.650	1695.569	1705.169	1708.192	1710.367
	upper	1594.031	1720.036	1711.372	1718.191	1719.715	1715.350	1702.431	1714.831	1717.808	1713.633
4	mean	1714.000	1702.000	1711.000	1716.000	1711.000	1891.000	1892.000	1882.000	1891.000	1894.000
	std	0.546	0.821	2.347	2.407	0.922	1.221	1.599	1.044	1.421	1.892
	lower	1712.929	1700.390	1706.400	1711.282	1709.193	1888.607	1888.867	1879.954	1888.214	1890.291
	upper	1715.071	1703.610	1715.600	1720.718	1712.807	1893.393	1895.133	1884.046	1893.786	1897.709
6	mean	1721.000	1710.000	1711.000	1705.000	1705.000	1941.000	1943.000	1940.000	1934.000	1940.000
	std	1.501	1.932	1.552	0.503	1.289	1.484	1.306	1.209	1.501	1.390
	lower	1718.059	1706.213	1707.958	1704.015	1702.473	1938.091	1940.441	1937.631	1931.058	1937.275
	upper	1723.941	1713.787	1714.042	1705.985	1707.527	1943.909	1945.559	1942.369	1936.942	1942.725
8	mean	1712.000	1713.000	1717.000	1705.000	1715.000	1963.000	1963.000	1966.000	1962.000	1960.000
	std	0.681	1.047	2.387	0.553	0.580	1.066	1.665	2.482	2.485	2.486
	lower	1710.666	1710.948	1712.322	1703.916	1713.863	1960.910	1959.737	1961.136	1957.129	1955.127
	upper	1713.334	1715.052	1721.678	1706.084	1716.137	1965.090	1966.263	1970.864	1966.871	1964.873
10	mean	1708.000	1708.000	1719.000	1711.000	1706.000	1973.000	1968.000	1972.000	1972.000	1967.000
	std	0.720	1.829	1.548	0.846	2.386	0.984	2.498	1.665	0.867	1.274
	lower	1706.589	1704.415	1715.966	1709.341	1701.324	1971.072	1963.104	1968.736	1970.302	1964.504
	upper	1709.411	1711.585	1722.034	1712.659	1710.676	1974.928	1972.896	1975.264	1973.698	1969.496

Similarly, the results of the model trained using the DE algorithm are shown in Table 7 and Table 8. DE-GMM exhibited the highest performance among the heuristic methods, with 55 additional correct recognitions (2.75% improvement).

Table 7. Speech recognition performance values (100x %) with DE/Rand/1/Bin based GMM.

	Reference Size (Number of Repetitions)									
MixGaussian	2	4	6	8	10	12	14	16	18	20
2	0.7950	0.8575	0.8535	0.8575	0.8580	0.8560	0.8510	0.8565	0.8565	0.8545
4	0.8565	0.8515	0.8565	0.8595	0.8545	0.9440	0.9470	0.9425	0.9460	0.9465
6	0.8600	0.8535	0.8560	0.8530	0.8525	0.9705	0.9695	0.9705	0.9685	0.9720
8	0.8560	0.8570	0.8580	0.8530	0.8560	0.9815	0.9830	0.9825	0.9810	0.9810
10	0.8545	0.8525	0.8595	0.8555	0.8515	0.9870	0.9850	0.9860	0.9840	0.9840

Table 8. Statistical data on the speech recognition performance with DE/Rand/1/Bin based GMM; a 95% confidence interval, alpha=0.05, and a normal distribution model were used.

Number of Mixed-Gaussians	Statistics (for 95% CI)	Reference Size (Number of Repetitions)									
		2	4	6	8	10	12	14	16	18	20
2	mean	1590.000	1715.000	1707.000	1715.000	1716.000	1712.000	1702.000	1713.000	1713.000	1709.000
	std	1.257	1.339	1.274	0.935	1.106	1.939	1.465	0.767	2.432	2.283
	lower	1587.536	1712.375	1704.502	1713.168	1713.832	1708.200	1699.129	1711.497	1708.234	1704.526
	upper	1592.464	1717.625	1709.498	1716.832	1718.168	1715.800	1704.871	1714.503	1717.766	1713.474
4	mean	1713.000	1703.000	1713.000	1719.000	1709.000	1888.000	1894.000	1885.000	1892.000	1893.000
	std	1.150	2.207	2.274	1.740	0.643	1.462	0.636	1.018	1.239	1.873
	lower	1710.747	1698.674	1708.542	1715.590	1707.740	1885.135	1892.754	1883.004	1889.571	1889.328
	upper	1715.253	1707.326	1717.458	1722.410	1710.260	1890.865	1895.246	1886.996	1894.429	1896.672
6	mean	1720.000	1707.000	1712.000	1706.000	1705.000	1941.000	1939.000	1941.000	1937.000	1944.000
	std	1.536	0.943	1.807	0.855	2.016	0.978	2.029	1.427	0.857	2.455
	lower	1716.989	1705.152	1708.458	1704.324	1701.049	1939.084	1935.022	1938.203	1935.321	1939.189
	upper	1723.011	1708.848	1715.542	1707.676	1708.951	1942.916	1942.978	1943.797	1938.679	1948.811
8	mean	1712.000	1714.000	1716.000	1706.000	1712.000	1963.000	1966.000	1965.000	1962.000	1962.000
	std	0.972	0.912	2.473	2.086	2.466	0.730	0.752	1.965	0.605	1.303
	lower	1710.095	1712.213	1711.152	1701.912	1707.166	1961.569	1964.526	1961.149	1960.813	1959.446
	upper	1713.905	1715.787	1720.848	1710.088	1716.834	1964.431	1967.474	1968.851	1963.187	1964.554
10	mean	1709.000	1705.000	1719.000	1711.000	1703.000	1974.000	1970.000	1972.000	1968.000	1968.000
	std	0.815	2.458	2.286	2.308	1.443	1.669	2.204	1.838	1.913	0.918
	lower	1707.402	1700.183	1714.519	1706.476	1700.171	1970.730	1965.681	1968.398	1964.250	1966.200
	upper	1710.598	1709.817	1723.481	1715.524	1705.829	1977.270	1974.319	1975.602	1971.750	1969.800

The results of the tests conducted with the SaDE algorithm are presented in Table 9 and Table 10. SaDE-GMM once again demonstrated a superior performance to the baseline EM model, with 31 additional correct recognitions (1.55% improvement).

Table 9. Speech recognition performance values (100x %) with SaDE-based GMM.

	Reference Size (Number of Repetitions)									
MixGaussian	2	4	6	8	10	12	14	16	18	20
2	0.7935	0.8575	0.8550	0.8575	0.8585	0.8560	0.8510	0.8555	0.8570	0.8550
4	0.8565	0.8515	0.8570	0.8580	0.8560	0.9450	0.9475	0.9425	0.9460	0.9465
6	0.8605	0.8555	0.8575	0.8525	0.8520	0.9705	0.9700	0.9705	0.9685	0.9700
8	0.8560	0.8570	0.8565	0.8525	0.8575	0.9820	0.9835	0.9810	0.9810	0.9810
10	0.8540	0.8525	0.8595	0.8540	0.8530	0.9870	0.9840	0.9855	0.9855	0.9850

Table 10. Statistical data on the speech recognition performance with SaDE-based GMM; a 95% confidence interval, alpha=0.05, and a normal distribution model were used.

Number of Mixed-Gaussians	Statistics (for 95% CI)	Reference Size (Number of Repetitions)									
		2	4	6	8	10	12	14	16	18	20
2	mean	1587.000	1715.000	1710.000	1715.000	1717.000	1712.000	1702.000	1711.000	1714.000	1710.000
	std	0.602	2.059	1.670	0.709	2.424	2.281	1.854	1.243	1.170	1.796
	lower	1585.821	1710.964	1706.727	1713.610	1712.249	1707.529	1698.366	1708.564	1711.706	1706.480
	upper	1588.179	1719.036	1713.273	1716.390	1721.751	1716.471	1705.634	1713.436	1716.294	1713.520
4	mean	1713.000	1703.000	1714.000	1716.000	1712.000	1890.000	1895.000	1885.000	1892.000	1893.000
	std	1.105	1.771	1.897	2.439	2.287	2.325	2.361	2.369	0.851	1.475
	lower	1710.834	1699.530	1710.282	1711.220	1707.517	1885.443	1890.372	1880.357	1890.333	1890.108
	upper	1715.166	1706.470	1717.718	1720.780	1716.483	1894.557	1899.628	1889.643	1893.667	1895.892
6	mean	1721.000	1711.000	1715.000	1705.000	1704.000	1941.000	1940.000	1941.000	1937.000	1940.000
	std	1.472	0.714	1.151	1.613	1.497	2.136	1.608	2.375	0.788	2.393
	lower	1718.114	1709.601	1712.744	1701.838	1701.066	1936.814	1936.848	1936.345	1935.455	1935.309
	upper	1723.886	1712.399	1717.256	1708.162	1706.934	1945.186	1943.152	1945.655	1938.545	1944.691
8	mean	1712.000	1714.000	1713.000	1705.000	1715.000	1964.000	1967.000	1962.000	1962.000	1962.000
	std	1.131	1.395	2.277	1.410	2.491	1.678	0.990	1.979	2.150	1.598
	lower	1709.783	1711.266	1708.537	1702.235	1710.117	1960.711	1965.060	1958.122	1957.786	1958.868
	upper	1714.217	1716.734	1717.463	1707.765	1719.883	1967.289	1968.940	1965.878	1966.214	1965.132
10	mean	1708.000	1705.000	1719.000	1708.000	1706.000	1974.000	1968.000	1971.000	1971.000	1970.000
	std	0.586	2.291	1.670	0.948	1.412	1.209	1.915	1.020	1.661	1.970
	lower	1706.852	1700.509	1715.727	1706.142	1703.233	1971.629	1964.246	1969.001	1967.745	1966.139
	upper	1709.148	1709.491	1722.273	1709.858	1708.767	1976.371	1971.754	1972.999	1974.255	1973.861

The primary reason for this performance increase brought by heuristic optimization is that evolutionary algorithms explore the parameter space more effectively, yielding higher log-likelihood values and thereby defining more discriminative cluster boundaries. A comparative visualization of the success rates of the methods at their peak performance (for 10-Mixed Gaussians) is provided in Figure 6. This graph clearly demonstrates the marginal yet consistent improvement that the heuristic algorithms provide over the standard EM approach.

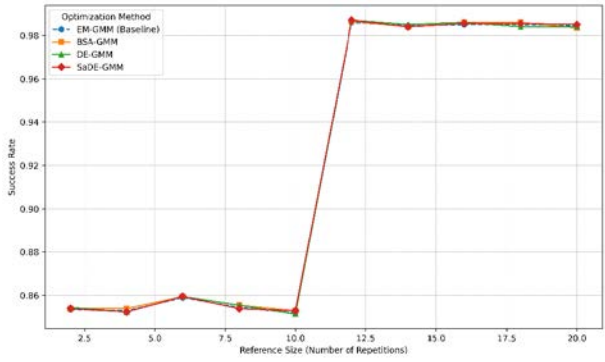


Figure 6. Comparative graph of the peak recognition performances of GMM models optimized with EM and heuristic algorithms.

This study offers a practical and effective solution to the challenge of data collection for large-vocabulary, speaker-dependent isolated word recognition problems. The generation of synthetic data via HPSS from a single real sample, combined with the optimization of GMM parameters using heuristic algorithms, has yielded a system that is both more efficient and higher-performing than classical methods in the literature that rely on multi-repetition real observation data. Although the training times for evolutionary algorithms are inherently longer than for EM-based solutions, their convergence behavior has produced more stable results. The obtained results demonstrate that high-accuracy speech recognition systems can be developed even under limited data conditions, revealing that the combination of HPSS-based synthetic data generation and heuristic optimization presents a powerful alternative for practical speech technologies.

5. CONCLUSION

This study presented a comprehensive, two-stage investigation into isolated word recognition for a large-vocabulary, speaker-dependent Turkish dataset designed for Geographic Information Systems (GIS) applications. The research aimed to both establish a performance benchmark for classical methods and propose a practical solution to the significant challenge of large-scale data collection.

In the first stage, a comparative analysis of Dynamic Time Warping (DTW), Gaussian Mixture Models (GMM), and Hidden Markov Models (HMM) was conducted on a real speech dataset. The findings clearly demonstrated the superiority of HMM, which achieved a recognition accuracy of 98.70%, outperforming both DTW (89.05%) and GMM (82.10%). This result confirms HMM's robustness in modeling the temporal characteristics of

speech for high-accuracy recognition tasks when sufficient real data is available.

The second and primary stage of this work addressed the data scarcity problem. We introduced an innovative framework centered on synthetic data generation using Harmonic-Percussive Source Separation (HPSS), which successfully augmented a dataset from a single real utterance per word. The effectiveness of this approach was validated as a GMM trained with the standard EM algorithm on this synthetic data achieved an accuracy exceeding 98%. Furthermore, to mitigate the risk of EM converging to local optima, GMM parameters were optimized using hybrid heuristic algorithms (DE, BSA, and SaDE). These hybrid models demonstrated a consistent, albeit marginal, improvement over the baseline, with DE-GMM yielding the best performance enhancement of up to 2.75%.

The key contribution of this research is the demonstration of a practical and efficient framework that enables the development of high-performance speech recognition systems without the need for extensive and time-consuming real-world data collection. The combination of HPSS-based data augmentation and heuristic-enhanced GMM provides a powerful alternative for developing robust systems in specialized, low-resource domains.

Despite these promising results, the study has limitations, including its focus on a speaker-dependent and isolated-word context. Future research could extend this framework to speaker-independent scenarios and investigate its applicability to continuous speech recognition. Additionally, applying the proposed synthetic data generation technique to train state-of-the-art deep learning models could be a promising avenue for further performance gains. This work validates that a data-centric and

optimized modeling approach can effectively overcome practical barriers in deploying specialized speech recognition technologies.

Acknowledgments

This study was supported by the TÜBİTAK project no. 110Y309, titled "Georeferanslanmış Konuşma Tanıma Tabanlı GIS Veri Toplama Aracı" (2013).

REFERENCES

Alotaibi, Y. A. (2012). Comparing ANN to HMM in implementing limited Arabic vocabulary ASR systems. *International Journal of Speech Technology, 15*(1), 25-32. doi:10.1007/s10772-011-9107-3

Arısoy, E., & Arslan, L. M. (2004). Turkish radiology dictation system. *Training, 91469*, 1562.

Arslan, R. S., & Barışçı, N. (2020). A detailed survey of Turkish automatic speech recognition. *Turkish journal of electrical engineering and computer sciences, 28*(6), 3253-3269.

Asliyan, R., Gunel, K., & Yakhno, T. (2008). *Syllable based turkish speech recognition using dynamic time warping and multilayer perceptron.* Paper presented at the 2008 IEEE 16th Signal Processing, Communication and Applications Conference.

Bansal, P., Dev, A., & Jain, S. B. (2008). Optimum HMM combined with vector Quantization for Hindi Speech word Recognition. *IETE Journal of Research, 54*(4), 239-243. doi:10.4103/0377-2063.44216

Benkhellat, Z., & Belmehdi, A. (2012). *Genetic algorithms in speech recognition systems.* Paper presented at the Proceedings of the International Conference on Industrial Engineering and Operations Management.

Burrough, P. A., McDonnell, R. A., & Lloyd, C. D. (2015). *Principles of Geographical Information Systems*: OUP Oxford.

Chauhan, H., & Tanawala, B. (2015). Comparative study of MFCC and LPC algorithms for Gujrati isolated word recognition. *International Journal of Innovative Research*

in Computer and Communication Engineering, 3(2), 822-826.

Chen, S.-H., & Wang, Y.-R. (1995). Tone recognition of continuous Mandarin speech based on neural networks. *IEEE Trans. Speech Audio Process., 3*, 146-150.

Civicioglu, P. (2013). Backtracking Search Optimization Algorithm for numerical optimization problems. *Applied Mathematics and Computation, 219*(15), 8121-8144. doi:https://doi.org/10.1016/j.amc.2013.02.017

Çivicioğlu, P., Karkınlı, A. E., & Beşdok, E. (2012). *Konuşma tanıma algoritmalarının CBS'de kullanımı ve DSA tabanlı bir konuşma tanıma uygulaması*. Paper presented at the IV. Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu, Zonguldak.

Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 28*(4), 357-366. doi:10.1109/TASSP.1980.1163420

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological), 39*(1), 1-22.

Emdadi, A., Ahmadi Moughari, F., Yassaee Meybodi, F., & Eslahchi, C. (2019). A novel algorithm for parameter estimation of Hidden Markov Model inspired by Ant Colony Optimization. *Heliyon, 5*(3), e01299. doi:https://doi.org/10.1016/j.heliyon.2019.e01299

Gaikwad, S. K., Gawali, B. W., & Yannawar, P. (2010). A review on speech recognition technique. *International Journal of Computer Applications, 10*(3), 16-24.

Goodchild, M. F. (2009). Geographic information systems and science: today and tomorrow. *Annals of GIS, 15*(1), 3-9. doi:10.1080/19475680903250715

Juang, B. H., & Rabiner, L. R. (1991). Hidden Markov models for speech recognition. *Technometrics, 33*(3), 251-272.

Keser, S., & Edizkan, R. (2009, 9-11 April 2009). *Phonem-based isolated Turkish word recognition with subspace classifier*. Paper presented at the 2009 IEEE 17th Signal Processing and Communications Applications Conference.

Kheddar, H., Hemis, M., & Himeur, Y. (2024). Automatic speech recognition using advanced deep learning approaches: A survey. *Information Fusion, 109*, 102422. doi:https://doi.org/10.1016/j.inffus.2024.102422

Kumar, K., & Aggarwal, R. (2011). Hindi speech recognition system using HTK. *International Journal of Computing and Business Research, 2*(2), 2229-6166.

Kurimo, M., Puurula, A., Arisoy, E., Siivola, V., Hirsimäki, T., Pylkkönen, J., . . . Saraclar, M. (2006). *Unlimited vocabulary speech recognition for agglutinative languages.* Paper presented at the Proceedings of the human language technology conference of the NAACL, main conference.

Lin, L., & Shuxun, W. (2005, 30 Oct.-1 Nov. 2005). *Genetic algorithms and fuzzy approach to Gaussian mixture model for speaker recognition.* Paper presented at the 2005 International Conference on Natural Language Processing and Knowledge Engineering.

Mahmoudi, H., Camboim, S., & Brovelli, M. A. (2023). Development of a Voice Virtual Assistant for the Geospatial Data Visualization Application on the Web.

ISPRS International Journal of Geo-Information, 12(11). doi:10.3390/ijgi12110441

Memon, S., Lech, M., & He, L. (2009, 17-18 Feb. 2009). *Using information theoretic vector quantization for inverted MFCC based speaker verification.* Paper presented at the 2009 2nd International Conference on Computer, Control and Communication.

Mizutani, E., & Dreyfus, S. (2021). On using dynamic programming for time warping in pattern recognition. *Information Sciences, 580*, 684-704. doi:https://doi.org/10.1016/j.ins.2021.08.075

Najkar, N., Razzazi, F., & Sameti, H. (2010). A novel approach to HMM-based speech recognition systems using particle swarm optimization. *Mathematical and Computer Modelling, 52*(11), 1910-1920. doi:https://doi.org/10.1016/j.mcm.2010.03.041

Nandhini, S., & Shenbagavalli, A. (2014). Voiced/unvoiced detection using short term processing. *International Journal of Computer Applications, 975*, 8887.

Nirmal, A., Jayaswal, D., & Kachare, P. H. (2024). A Hybrid Bald Eagle-Crow Search Algorithm for Gaussian mixture model optimisation in the speaker verification framework. *Decision Analytics Journal, 10*, 100385. doi:https://doi.org/10.1016/j.dajour.2023.100385

Palaz, H., Kanak, A., Bicil, Y., & Doğan, M. U. u. (2005, 2005//). *A DCOM-Based Turkish Speech Recognition System: TREN – Turkish Recognition ENgine.* Paper presented at the Computer and Information Sciences - ISCIS 2005, Berlin, Heidelberg.

Polat, H., & Oyucu, S. (2020). Building a speech and text corpus of Turkish: large corpus collection with initial speech recognition results. *Symmetry, 12*(2), 290.

Qin, A. K., & Suganthan, P. N. (2005, 2-5 Sept. 2005). *Self-adaptive differential evolution algorithm for numerical optimization.* Paper presented at the 2005 IEEE Congress on Evolutionary Computation.

Rabiner, L., & Juang, B.-H. (1993). *Fundamentals of speech recognition*: Prentice-Hall, Inc.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE, 77*(2), 257-286. doi:10.1109/5.18626

Reynolds, D. A. (1995). Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication, 17*(1), 91-108. doi:https://doi.org/10.1016/0167-6393(95)00009-D

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 26*(1), 43-49. doi:10.1109/TASSP.1978.1163055

Salor, Ö., Pellom, B. L., Ciloglu, T., & Demirekler, M. (2007). Turkish speech corpora and recognition tools developed by porting SONIC: Towards multilingual speech recognition. *Computer Speech & Language, 21*(4), 580-593.

Saxena, G. D., Farooqui, N. A., & Ali, S. (2022). Extricate features utilizing mel frequency cepstral coefficient in automatic speech recognition system. *Int. J. Eng. Manuf, 12*(6), 14-21.

Storn, R., & Price, K. (1997). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization, 11*(4), 341-359. doi:10.1023/A:1008202821328

Tiwari, V. (2010). MFCC and its applications in speaker recognition. *International journal on emerging technologies, 1,* 19.

Tombaloğlu, B., & Erdem, H. (2016, 16-19 May 2016). *Development of a MFCC-SVM Based Turkish Speech Recognition system.* Paper presented at the 2016 24th Signal Processing and Communication Application Conference (SIU).

Tombaloğlu, B., & Erdem, H. (2021). Turkish speech recognition techniques and applications of recurrent units (LSTM and GRU). *Gazi University Journal of Science, 34*(4), 1035-1049.

Tüzün, Ö. B., Erzin, E., Demirekler, M., Memişoğlu, A. T., Uğur, M. S., & Çetin, A. E. (1995, 1995//). *A Speaker Independent Isolated Word Recognition System for Turkish.* Paper presented at the Speech Recognition and Coding, Berlin, Heidelberg.

Wilpon, J. G., Rabiner, L. R., Lee, C.-H., & Goldman, E. (1990). Automatic recognition of keywords in unconstrained speech using hidden Markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 38*(11), 1870-1878.

HARİTA MÜHENDİSLİĞİ DEĞERLENDİRMELERİ

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