Gazi AKADEMİK YAYINCILIK

AIS ARTIFICIAL INTELLIGENCE STUDIES

ARTIFICIAL INTELLIGENCE STUDIES

Comparative Assessment of Machine Learning Models in EEG-Based Sentence Classification

Emrullah Şahin*[a], Durmuş Özdemir[b]

ABSTRACT

In this study conducted a detailed analysis of widely used machine learning algorithms for sentence classification on TSEEG (Turkish EEG) dataset, examining both classification accuracy and computational efficiency. Given the complex nature of EEG signals, the study also investigates the necessity and impact of preprocessing techniques—such as normalization, filtering, and feature extraction—on classification performance. Among the models evaluated, the Support Vector Machine (SVM) achieved the highest test accuracy (99.17%) and F1 score (99.12%), demonstrating strong reliability and processing speed. To further validate the reliability and robustness of the SVM model, cross-validation was used, confirming its stability across different data subsets. This study not only emphasizes the critical role of preprocessing in enhancing EEG data analysis but also provides practical benchmarks on accuracy and processing time, offering valuable guidance for model selection in similar research. By identifying effective algorithmic choices, this work supports future research in making informed decisions regarding preprocessing and model selection across diverse EEG datasets and application domains.

[a,*] Kutahya Dumlupınar University, Engineering Faculty, Dept. of Software Engineering 43000 - Kütahya, Türkiye ORCID: 0000-0002-3390-6285

[b] Kutahya Dumlupınar University, Engineering Faculty, Dept. of Computer Engineering 43000 - Kütahya, Türkiye ORCID: 0000-0002-9543-4076

[*] Corresponding author.
e-mail: emrullah.sahin@dpu.edu.tr

Keywords: EEG sentence classification, Machine learning, Model comparison

Anahtar Kelimeler: EEG cümle sınıflandırma, Makine öğrenimi, Model karşılaştırma

EEG Tabanlı Cümle Sınıflandırmada Makine Öğrenimi Modellerinin Karşılaştırmalı Değerlendirilmesi

ÖZ

Bu çalışma, TSEEG (Türkçe EEG) veri kümesi üzerinde cümle sınıflandırması için yaygın olarak kullanılan makine öğrenimi algoritmalarının sınıflandırma doğruluğu ve hesaplama verimliliği açısından kapsamlı bir analizini sunmaktadır. EEG sinyallerinin karmaşık yapısı göz önüne alındığında, normalizasyon, filtreleme ve öznitelik çıkarımı gibi ön işleme tekniklerinin sınıflandırma performansı üzerindeki etkisi ayrıntılı bir şekilde incelenmiştir. Çalışmada değerlendirilen modeller arasında Destek Vektör Makineleri (SVM), %99,17 test doğruluğu ve %99,12 F1 skoru ile en iyi performansı sergilemiştir. Ayrıca, çapraz doğrulama yöntemi kullanılarak SVM modelinin farklı veri alt kümeleri üzerindeki istikrarı ve güvenilirliği teyit edilmiştir. Bu çalışma, yalnızca EEG veri analizinde ön işlemenin kritik rolünü vurgulamakla kalmayıp, aynı zamanda doğruluk ve işlem süresi açısından pratik kıyaslama değerleri sunarak benzer araştırmalarda model seçimi için değerli bir rehberlik sağlamaktadır. Etkili algoritmik seçimlerin belirlenmesiyle bu çalışma, farklı EEG veri kümeleri ve uygulama alanlarında ön işleme ve model seçimine yönelik bilinçli kararların alınmasına destek olmaktadır.

Submitted: 05.12.2024
Revised: 14.12.2024
Accepted: 18.12.2024

doi: 10.30855/AIS.2024.07.02.06

To cite this article: E. Şahin and D. Özdemir, "Comparative Assessment of Machine Learning Models in EEG-Based Sentence Classification." *Artificial Intelligence Studies,* vol.6, no.1, pp. 97-114, 2024. 10.30855/AIS.2024.07.02.06

1. Introduction (Giriş)

In recent years, rapid advancements in artificial intelligence (AI) and machine learning (ML) have catalyzed substantial progress across diverse application domains [1], [2], including image processing [3], [4], [5], natural language processing [6], [7], and predictive modeling [8], [9], [10]. Enhanced capabilities of these methods to analyze and interpret complex data structures have facilitated the development of innovative solutions in sectors such as healthcare [11], [12], [13], automotive [14], and security [15]. The ability of deep learning [16], [17], [18] and ML algorithms [19], [20], [21] to extract meaningful insights from intricate data has fueled technological advancements, rendering human-machine interaction increasingly natural and efficient. Notably, Brain-computer interfaces (BCIs) have emerged as a transformative technology, with the potential to directly translate neural activity from the human brain into computer systems. This capability offers promising applications in sectors like healthcare, communication, and education, enabling unprecedented forms of interaction and control [22].

BCI systems aim to analyze individuals' mental activities and convey them to the external environment via computer systems, thereby enabling human thoughts to interact with digital systems. This capability presents significant opportunities, particularly in communication and control applications. Among various BCI systems, electroencephalography (EEG) technology is especially prominent due to its non-invasiveness, portability, and relatively low cost. The societal implications of BCI and EEG-based technologies are broad and transformative. For instance, advancements in BCI systems can greatly enhance the quality of life for individuals with disabilities, while also offering innovative solutions across sectors that rely on human-machine interaction. In healthcare, these systems can facilitate early diagnosis and continuous monitoring of neurological conditions. In education, they hold potential for personalizing learning processes and optimizing efficiency in instructional methods. Such applications not only enhance individual quality of life but also amplify the positive societal impacts of technology by broadening access and increasing functionality across multiple domains [23], [24].

Currently, a major focus of BCI (Brain-Computer Interface) research is the translation of electrical activity from the brain into various commands or text. EEG (electroencephalography) technology is particularly prominent in this field, offering both accessibility and a non-invasive means for such transformations. The non-invasive nature of EEG enhances user comfort and expands accessibility to a broader user base [25]. Moreover, the communication possibilities facilitated by EEG-based translation are especially critical for individuals with limited mobility or speech capabilities, as this technology enables users to interact with digital systems solely through mental activities, thereby promoting greater independence. In this context, the transformation of EEG signals into commands or text is seen as a forward-looking solution with substantial potential in communication, education, and accessibility, promising significant advancements in these areas [26].

Providing accurate and reliable analysis in the conversion of EEG signals into various commands or text remains one of the most complex challenges in the field. A key limitation is the scarcity of high-quality, comprehensive EEG and other biomedical datasets, which constrains progress in this domain [27]. Additionally, ethical and privacy concerns hinder data sharing and standardization, complicating efforts to derive reliable insights from EEG signals. The sensitivity of EEG data to environmental noise and biological artifacts necessitates rigorous preprocessing steps to ensure the accuracy and reliability of analysis. In this regard, preprocessing techniques—such as noise reduction, feature extraction, and signal modeling—are essential to transforming raw EEG signals into meaningful information, thereby optimizing the performance of classification algorithms.

Enhancing these processes requires the optimization of preprocessing methods, such as artifact and noise reduction, and the development of novel techniques that can further improve data quality and classification accuracy. Such advancements will contribute to making EEG-based BCI applications more reliable and effective. Achieving more precise analyses of EEG signals will not only enhance the accuracy of BCI applications but also bolster the reliability and accessibility of BCI systems through improved EEG analysis [28]. Consequently, each innovation in processing EEG signals and transforming them into meaningful information is critical to expanding the future application potential of BCI systems.

In this study examined the effectiveness of commonly used machine learning algorithms for classifying Turkish EEG data, focusing on both accuracy and computational efficiency. At the same time, it explored the critical role of preprocessing techniques in enhancing classification performance, providing valuable insights for future EEG-based research and applications.

The organizational structure of this study is as follows: Section 1.1 reviews the related literature. Section 2, Materials and methods, describes the data collection process, preprocessing steps, and classification methods employed in the study. Section 3 presents the results, including classification accuracy and performance analysis. Finally, Section 4 discusses the overall conclusions, addresses the study's limitations, and provides recommendations for future research.

1.1. Related works (İlgili çalışmalar)

This literature review discusses various approaches aimed at enhancing the classification performance of EEG signals. First, studies focused on noise reduction in EEG signals and methods for improving classification accuracy are reviewed. Next, advanced machine learning and deep learning approaches that facilitate the recognition of complex patterns within EEG data are examined. Lastly, research analyzing EEG signals with a focus on Turkish language-specific characteristics is explored. This review aims to provide a comprehensive overview of the primary approaches in the literature within the field of EEG-based classification.

Noise Reduction and Classification Enhancement: Reducing artifacts and noise in EEG signals is essential for achieving reliable classification outcomes. Deng et al. demonstrated the effectiveness of noise reduction techniques by achieving 67% accuracy in classifying the /ba/ and /ku/ syllables in EEG signals using the SOBI algorithm combined with spectral analysis methods [29]. Brigham and Kumar [30] introduced a robust approach to signal cleaning for phonetic decomposition in EEG signals through the application of independent component analysis (ICA) and k-nearest neighbor (KNN) algorithms.

Chi et al. achieved over 70% accuracy in classifying five phonemes using Naive Bayes and LDA algorithms, demonstrating that classification performance improves when EEG signals are sourced from regions outside the frontal and occipital areas [31]. In multi-word classification tasks with EEG, Garcia et al. achieved over 40% accuracy using methods such as Naive Bayes, SVM, and Random Forests, underscoring the need for more advanced methods in complex classification tasks [32]. In the healthcare context, Sarmiento et al. developed a BCI protocol based on mental imagery of open-mid and closed vowels for patients with amputations or paralysis. By calculating the power spectral density (PSD) in EEG signals from the tongue region (21 electrodes), they achieved 84-94% accuracy with SVM [33].

In recent studies, Haider et al. reported accuracy rates of 96.1%, 97.1%, and 94.8% using SVM, LDA, and KNN algorithms, respectively, on a 10-class task, highlighting the importance of signal processing techniques in multiclass classification [34]. Park et al. achieved 80.41% accuracy using CNN for vowel classification with EEG signals, reflecting both the challenges and advances in this field [35]. Collectively, these studies indicate that advanced signal processing and machine learning techniques are crucial for effectively classifying linguistic structures with EEG signals.

Advanced Machine Learning and Deep Learning Approaches: The application of deep learning algorithms for classifying complex structures within EEG signals is gaining traction. Panachakel et al. demonstrated the efficacy of deep learning in this domain, achieving 57% accuracy in phoneme and word classification from EEG signals using a combination of deep neural networks (DNN) and discrete wavelet transform (DWT) [36]. Similarly, Moctezuma et al. achieved 95% ±4 accuracy in classifying imagined commands in Spanish, such as "up," "down," "right," "left," and "select," by employing random forests (RF), common average referencing (CAR), and discrete wavelet transform (DWT) methods [37]. This study highlights the effectiveness of hybrid models in enhancing EEG classification performance.

In more advanced classification studies, Kumar et al. achieved 96.09% accuracy in classifying 10 distinct characters using a combination of CNN and LSTM [38], while Ullah et al. reported an accuracy rate of 95.2% using a deep convolutional neural network (DCNN) for the recognition of 26 different characters [39]. Yang et al. attained 96.41% accuracy, 84.26% precision, and 96.52% recall in a 5-class word classification task using the EOWGMO-MDADenseNet-AM model, further underscoring the utility of complex deep learning models in EEG-based text generation [40]. Additionally, Ali et al. combined CNN, LSTM, and XGBoost algorithms to achieve 96.89% accuracy in multi-class classification tasks, demonstrating the high performance of deep learning approaches in classifying linguistic and textual structures from EEG signals [41].

Turkish Language-Based EEG Studies: Phonetic and morphological characteristics of a language significantly influence the transcription of EEG signals. The structural features of different languages may lead to variations in brain signal patterns; thus, there is a need for language-specific EEG studies [42]. Research on agglutinative languages, such as Turkish, suggests that EEG-based language processing systems should be adapted to accommodate the unique linguistic structure of Turkish. However, a review of the literature reveals that research specific to Turkish is limited, indicating a need for more comprehensive studies.

In this context, Eroğlu et al. investigated Turkish reading skills by recording the brain signals of 17 participants with the EMOTIV EPOC+ EEG device before, during, and after a computer-assisted training process. In this study, letters of the Turkish alphabet were rotated 180 degrees and presented in a distorted form to participants, and their reading performance with these altered texts was analyzed. Theta band activity, particularly recorded in Broca's area (electrodes F7 and FC5), was found to correlate with changes in reading speed and error rates [43]. As one of the initial studies associating Turkish language learning and reading skills with EEG data, this research provides a valuable contribution to the literature.

Kutlu Onay [44] developed a brain-computer interface (BCI) system for recognizing Turkish letters using EEG signals recorded during motor and non-motor imagery tasks. In this study, EEG signals from 8 participants were collected with the Neuroelectrics Enobio-8 device. Participants were presented with images resembling Turkish letters and images of objects beginning with those letters. The EEG signals gathered during motor imagery were translated into text by associating letters with the participants' thoughts. This study highlights the challenges posed by inter-individual variability in EEG signals, which complicates the development of generalized systems; however, it demonstrates that personalized systems can achieve successful outcomes.

Expanding beyond Turkish letter recognition, Barua et al. developed a lightweight and highly accurate model for Turkish sentence classification using EEG signals. In their study, EEG data from 20 participants were recorded with a 14-channel EMOTIV EPOC+ device as 20 standard Turkish sentences were presented to participants in both listening and viewing modes. The extracted features were processed using innovative techniques such as Dynamic Dimensional Binary Pattern (DSBP) and Multilevel Discrete Wavelet Transform (MDWT) and then classified using k-nearest neighbor (k-NN) and support vector machines (SVM). The study achieved accuracy rates of 98.81% in viewing mode and 98.19% in listening mode, marking a significant advancement in EEG-based language processing and the development of Turkish EEG datasets [45].

Focusing on vowel recognition alongside Turkish sentence classification, Haltaş and Erguzen developed an EEG-based BCI system for recognizing Turkish vowels. In this study, two distinct BCI systems were tested to identify the common Turkish vowels "A," "E," and "I" using EEG signals. The system employing discrete wavelet transform (DWT) and support vector machines (SVM) achieved the highest accuracy, reaching 80.2%. These findings indicate the potential for EEG-based systems to enhance expressive capabilities for individuals with speech disabilities [46].

In language processing, Demir [47] designed a brain keyboard interface using EEG signals to address communication challenges faced by individuals with speech impairments. In this study, letter predictions were made by analyzing power spectral density (PSD) in the 8-30 Hz range, and four different imagery paradigms were tested. Achieving 98.66% accuracy with the long short-term memory (LSTM) algorithm, the study highlights the strong potential of brain-keyboard interfaces in advancing human-machine interaction.
Although machine learning and deep learning algorithms are effective for classifying linguistic structures using EEG signals, research specifically targeting the Turkish language remains limited. This

study aims to identify the most effective machine learning algorithms for EEG-based Turkish sentence classification and to propose robust approaches for this application. Additionally, the contribution of various preprocessing steps to classification performance is evaluated, providing insights that can serve as references for related processes in the field.

2. Material and Methods (Materyal ve Yöntem)

In this study employed a range of machine learning algorithms to address the challenges of multiclass classification in EEG signal analysis. Given the high dimensionality and susceptibility of EEG data to noise, these algorithms are selected for their ability to enhance classification performance. Each algorithm leverages unique methodological principles designed to maximize accuracy and robustness in handling EEG data. The following subsections detail the algorithms applied, data collection procedures, preprocessing steps, and classification methodologies used in this research. A graphical summary of the study is presented in Figure 1.

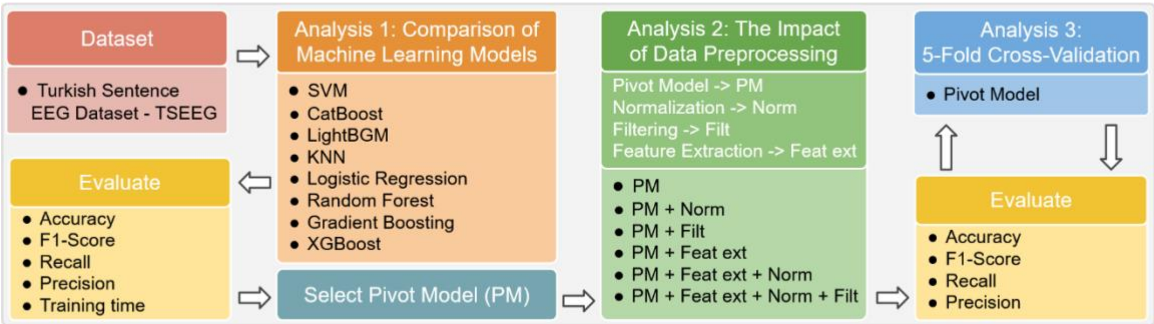

Figure 1. Graphical summary of the study's methodology and workflow
(Çalışmanın metodolojisi ve iş akışının grafiksel özeti)

2.1. Tree-based methods (Ağaç-tabanlı yöntemler)

Tree-based ensemble methods, such as Random Forest, Gradient Boosting, and advanced algorithms like CatBoost, LightGBM, and XGBoost, are frequently used for processing complex, high-dimensional data, including EEG signals. These methods utilize tree-based structures to enhance classification accuracy and robustness. Each algorithm incorporates specific optimizations tailored to improve performance on challenging datasets, making them well-suited for EEG data analysis.

2.1.1. Random forest (Random forest)

Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their predictions to create a more robust and accurate model [48]. Each tree operates independently, and the final classification outcome is determined through a majority voting process based on the predictions of all trees, as expressed mathematically in Equation 1.

$$\hat{y} = \text{mode}\{T_b(x): b = 1,2, \dots, B\} \tag{1}$$

where:

- $\hat{y}$ is the model's final prediction,
- $T_b(x)$ denotes the prediction from the b-th tree,
- B represents the total number of trees.

Random Forest is effective in reducing variance, thus minimizing the risk of overfitting, which is particularly beneficial in stabilizing classification performance with complex data, such as EEG signals.

2.1.2. Gradient boosting and its enhanced versions (Gradient boosting ve geliştirilmiş versiyonları)

Gradient Boosting is a sequential ensemble method that incrementally builds new trees, with each tree designed to correct the errors of its predecessors. This approach involves iteratively updating the model parameters in the direction of the negative gradient of the loss function, as mathematically expressed in Equation 2.

$$F_m(x) = F_{m-1}(x) + \eta \cdot g_m(x) \quad (2)$$

In the context of gradient boosting, the model at the m-th iteration is represented by $F_m(x)$, the learning rate is denoted by η, and $g_m(x)$ is the negative gradient of the loss function. Several enhanced versions of gradient boosting, including CatBoost, LightGBM, and XGBoost, introduce modifications to this structure for different types of data and applications.

CatBoost: CatBoost employs a distinctive target encoding methodology tailored to the specific characteristics of categorical variables, which are a pervasive feature of numerous datasets [49]. Instead of one-hot encoding, CatBoost encodes categorical variables based on target values to capture the relationships within categories directly. This encoding can be mathematically expressed as Equation 3.

$$\text{Encoded Value} = \frac{\sum_{i=1}^{n} y_i}{n + \lambda} \quad (3)$$

Here, the encoding scheme adjusts category-specific values by balancing with a regularization term, λ, to mitigate overfitting. This approach enables CatBoost to efficiently handle high-cardinality categorical data, enhancing both accuracy and training speed.

LightGBM: LightGBM (Light Gradient Boosting Machine) is designed for both high efficiency and scalability, which is critical when dealing with large datasets [50]. LightGBM achieves this through two primary innovations:

- ✓ Gradient-based One-Side Sampling (GOSS): In contrast to the conventional approach of utilizing the entirety of the training data, GOSS employs a selective strategy whereby data points with pronounced gradients are retained. This approach is typically more effective in optimizing the model. By focusing on a subset of the most informative samples, LightGBM reduces the computational complexity of the model while maintaining its accuracy.
- ✓ Exclusive Feature Bundling (EFB): EFB groups features that are mutually exclusive, that is, features that do not co-occur in the data, into single features. This process of bundling reduces the effective dimensionality of the dataset, thereby facilitating the construction of trees at a faster rate while maintaining the quality of the model.

Together, GOSS and EFB make LightGBM a highly efficient and scalable solution, particularly advantageous for high-dimensional data like EEG signals. The techniques facilitate the expeditious training of LightGBM while concurrently ensuring the maintenance of competitive accuracy, even when confronted with voluminous datasets.

XGBoost: XGBoost is designed to optimize the gradient boosting process in terms of both speed and accuracy. The second-order Taylor expansion is introduced to approximate the loss function, thereby facilitating more accurate updates in each iteration [51]. This process is mathematically expressed in Equation 4.

$$L(\theta) \approx \sum_{i=1}^{n} [g_i \cdot h(x_i) + \frac{1}{2} h_i \cdot h(x_i)^2] + \Omega(h) \quad (4)$$

where g_i and h_i are the first and second derivatives of the loss function, respectively. This precise gradient update, in conjunction with regularization terms (L1 and L2), enables XGBoost to regulate model complexity, avert overfitting, and enhance generalization, which is especially advantageous for EEG data with intricate structures.

2.2. Support vector machine (Destek vektör makinesi)

The objective of Support Vector Machines (SVM) is to determine a hyperplane that maximizes the margin of separation between classes [52]. For multiclass classification, SVM constructs independent hyperplanes for each class using either a one-versus-rest or a one-versus-one strategy. The mathematical formulation of the hyperplane optimization problem is expressed in Equation 5.

$$\left(min_{w,b}\frac{1}{2}\|\boldsymbol{w}\|^2\right) \text{ subject to } (\,y_i(\boldsymbol{w}\cdot x_i + b) \geq 1, i = 1, \dots, n) \tag{5}$$

where:

- w is the normal vector of the hyperplane,
- b is the intercept,
- y_i denotes the class label.

SVM is advantageous for EEG data due to its capacity to establish distinct separations between classes, thereby enhancing classification accuracy.

2.3. K-nearest neighbors (K-en yakın komşu)

The KNN algorithm determines the class of a given sample based on the k most similar samples in the feature space [53]. Similarity is determined by calculating the Euclidean distance between data points. The mathematical formulation of the Euclidean distance is expressed in Equation 6.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6}$$

where $d(x,y)$ represents the distance between points x and y. KNN is effective for EEG data classification by utilizing local relationships, enabling accurate identification of complex patterns within high-dimensional datasets.

2.4. Logistic regression (Lojistik Regresyon)

Logistic regression predicts class probabilities by applying the generalized sigmoid function, known as the SoftMax function, for multiclass classification problems [54]. The SoftMax function is mathematically expressed in Equation 7.

$$P(y = c \mid x) = \frac{e^{\beta_c \cdot x}}{\sum_{j=1}^{C} e^{\beta_j \cdot x}} \tag{7}$$

where:

- $P(y = c \mid x)$ denotes the probability that observation x belongs to class c,
- β_c is the coefficient vector for class c,
- C is the total number of classes.

Logistic regression provides a robust approach for calculating probabilities in EEG classification, effectively mapping class likelihoods across the high-dimensional data structure.

2.5. Dataset (Veri kümesi)

In this study, we utilized the Turkish Sentence-EEG (TSEEG) dataset, originally collected by Barua et al. and publicly available on the Kaggle platform. This dataset consists of EEG recordings corresponding to 1,600 samples, each associated with a unique Turkish sentence presented in two modes: demonstration (visual) and listening (auditory). EEG signals for each sample were recorded from 14 channels placed across 16 scalp locations using the EMOTIV EPOC+ mobile system. Each recording lasted 15 seconds at a sampling rate of 128 Hz, resulting in 1,920 data points (15 seconds × 128 Hz) per channel. Consequently, each sample is represented as a data array of dimensions 1920×14, and the entire dataset forms a structure of 1600×14×1920. This dataset supports both channel-wise analysis and 20-class classification, with each sentence corresponding to a unique class. For further technical details and background, readers are encouraged to consult the original publication by Barua et al [45].

2.5.1. Data preprocessing (Veri önişleme)

Preprocessing is a crucial step in EEG signal analysis, essential for enhancing data quality by reducing noise, isolating relevant information, and standardizing features to enable robust classification. The preprocessing pipeline in this study includes bandpass filtering, feature extraction through frequency analysis, and normalization.

Bandpass Filtering**:** The initial step involves applying a bandpass filter to remove frequencies outside the relevant EEG range, thereby improving the signal-to-noise ratio. A 5th-order Butterworth [55] bandpass filter is used to retain signals within the 1–40 Hz range, effectively capturing essential EEG rhythms and filtering out irrelevant frequencies. The filter's transfer function $H(s)$ is expressed in Equation 8.

$$H(s) = \frac{\omega_c^n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + \omega_c^n} \tag{8}$$

where:

- ω_c represents the cutoff frequencies in radians per second,
- n is the filter order (set to 5 in this study),
- s is the complex frequency variable.

This filtering stage preserves the primary EEG bands (delta, theta, alpha, beta, gamma), enabling a clearer signal for feature extraction.

Feature Extraction using Frequency Analysis**:** Following filtering, features are extracted in the frequency domain using Fast Fourier Transform (FFT). FFT decomposes the time-domain EEG signal into its frequency components, allowing power calculation within specific EEG frequency bands [56]. Each band is associated with different cognitive and physiological states, as outlined below:

- Delta (1–4 Hz): Linked to deep sleep and unconscious processes.
- Theta (4–8 Hz): Associated with drowsiness, meditation, and certain types of learning.
- Alpha (8–13 Hz): Related to relaxation and reduced mental workload.
- Beta (13–30 Hz): Associated with active thinking, concentration, and anxiety.
- Gamma (30–40 Hz): Linked to high-level cognitive functions and focused attention.

For each frequency band, power is calculated by summing the squared magnitudes of FFT values within that band's frequency range. This calculation is mathematically expressed in Equation 9.

$$P_{\text{band}} = \sum_{f \in F_{\text{band}}} |X(f)|^2 \tag{9}$$

where:

- P_{band} is the power of the specified frequency band,
- F_{band} represents the frequency range for the band (e.g., delta, theta),
- $X(f)$ is the amplitude of the EEG signal at frequency f.

For each channel, five features corresponding to the delta, theta, alpha, beta, and gamma bands are extracted, resulting in a total of 5×14=70 features per sample. Thus, the original data dimension of 1600×14×1920 (samples, channels, time points) is reduced to 1600×70, where each sample is represented by a 70-dimensional feature vector.

Normalization: The final step in preprocessing is normalization, which ensures consistent scaling across all features. In this study, each feature undergoes a logarithmic transformation followed by z-score normalization [57] to manage variability and adjust for skewness. The logarithmic transformation is applied first, as expressed in Equation 10.

$$\text{feature}_{\text{log}} = \log(1 + \text{feature}) \tag{10}$$

This transformation compresses the range of feature values, reducing the influence of large values and achieving a more symmetrical distribution. Subsequently, z-score normalization is applied to the transformed features, as expressed in Equation 11.

$$\text{feature}_{\text{normalized}} = \frac{\text{feature}_{\text{log}} - \mu}{\sigma} \tag{11}$$

where:

- μ is the mean of the logarithmically transformed feature,
- σ is the standard deviation of the transformed feature.

This combined normalization approach provides a uniform feature scale across all samples, improving model robustness and performance during classification.

2.6. Performance metrics (Performans metrikleri)

To comprehensively evaluate the performance of the EEG classification model, four key metrics were utilized: Precision, Recall, F1 Score, and Accuracy. Each metric provides unique insights into the model's effectiveness, capturing various aspects of classification performance [58]. Their formal mathematical definitions are provided below.

2.6.1. Precision (Kesinlik)

Precision (P) evaluates the model's capability to accurately classify positive instances among all the instances it has labeled as positive. It is defined mathematically as Equation 12.

$$P = \frac{T_p}{T_p + F_p} \tag{12}$$

where T_p represents true positives (correctly identified positive instances) and F_p denotes false positives (instances incorrectly classified as positive). High precision indicates the model's specificity, minimizing false positives.

2.6.2. Recall (Duyarlılık)

Recall (R), also known as sensitivity or true positive rate, quantifies the model's ability to capture all actual positive instances. It is formally defined as Equation 13.

$$R = \frac{T_p}{T_p + F_n} \tag{13}$$

Here, F_n represents false negatives, or actual positives that were missed by the model. High recall implies that the model effectively identifies relevant positive instances, minimizing false negatives.

2.6.3. F1 score (F1 skoru)

The F1 Score (F_1) combines precision and recall by calculating their harmonic mean, balancing the trade-off between these two metrics. It is defined mathematically as Equation 14.

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{14}$$

The F1 Score is particularly useful in cases with class imbalance, where both false positives and false negatives impact performance significantly.

2.6.4. Accuracy (Doğruluk)

Accuracy (A) provides an overall measure of the model's effectiveness by calculating the proportion of correctly classified instances among all instances. It is formally expressed as Equation 15.

$$A = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{15}$$

where T_n represents true negatives (correctly identified negative instances). While accuracy is a straightforward measure of overall correctness, it can be less informative for imbalanced datasets, making precision, recall, and F1 Score more critical in those contexts.

These metrics offer a comprehensive evaluation of the model's classification performance, capturing both overall and class-specific insights essential for assessing EEG data accurately.

3. Results (Sonuçlar)

This section outlines the experimental results, organized to offer a detailed understanding of the influence of preprocessing steps, the role of cross-validation in model evaluation, and the comparative performance of the machine learning algorithms employed in this research.

3.1. Implementation details (Uygulama detayları)

All experiments were conducted on a workstation equipped with an NVIDIA 3060 Ti GPU, an Intel i7-11700F processor, and 24GB DDR4 RAM, using Python as the programming environment. This setup provided a consistent hardware and software environment, ensuring that each algorithm was evaluated under equivalent conditions for fair comparison. The classification algorithms were initialized with the following baseline parameters:

- SVM: Linear kernel.
- KNN: 5 neighbors.
- Logistic Regression: Maximum iterations set to 1000.
- Random Forest, Gradient Boosting, XGBoost, LightGBM: 100 estimators.
- CatBoost: 100 iterations.

These parameters were selected based on standard settings in the literature, providing each model with an optimal starting configuration. Further tuning and cross-validation were applied as necessary, but these baseline settings allowed for uniform initialization across models to ensure consistent testing conditions.

3.2. Impact of preprocessing steps (Önişleme adımlarının etkileri)

To gain a comprehensive understanding of the impact and necessity of preprocessing on EEG signals, various tests were conducted using the SVM algorithm as a baseline model. The experiments were based on the TSEEG dataset, which was split into training and test sets at an 85:15 ratio. The results are presented in Table 1. reflect the performance on the test set, assessing the impact of each preprocessing step (normalization, filtering, and feature extraction) on classification outcomes.

Table 1. The impact of different preprocessing steps on classification performance
(Farklı ön işleme adımlarının sınıflandırma performansına etkisi)

Model	Precision	Recall	F1 Score	Accuracy
SVM	0.6430	0.6471	0.6224	0.6375
SVM + Normalization	0.6851	0.6893	0.6683	0.6708
SVM + Filtering	0.1121	0.1029	0.0846	0.0875
SVM + Feature Extraction	0.9379	0.938	0.9331	0.9417
SVM + Feature Extraction + Normalization	0.9638	0.9655	0.9621	0.9667
SVM + Filtering + Feature Extraction + Normalization	0.9919	0.9913	0.9912	0.9917

The results illustrate a clear progression in model performance as different preprocessing steps are incorporated. The baseline SVM model, without any preprocessing, achieves a moderate classification accuracy of 63.75%, reflecting the raw data's limitations in distinguishing EEG patterns effectively. Adding normalization enhances model performance across all metrics, with accuracy reaching 67.08%, as standardizing feature scales reduces variability and helps the model focus on relevant patterns. In contrast, filtering alone results in a significant drop in performance, suggesting that while it removes noise, using filtering without other steps may obscure critical signal features necessary for effective classification.

Feature extraction, by itself, dramatically improves performance, reaching 94.17% accuracy. This suggests that isolating specific EEG frequency bands provides the model with concentrated, informative features, enabling better class separation. When feature extraction is combined with normalization, performance further increases to 96.67% accuracy, benefiting from both enriched signal information and scale consistency across features.

Finally, incorporating the full preprocessing pipeline—including filtering, feature extraction, and normalization—yields the highest performance, with accuracy peaking at 99.17%. This result suggests that the complete preprocessing approach optimizes signal clarity and scale, resulting in the highest precision, recall, F1 Score, and accuracy. These findings underscore the value of comprehensive preprocessing for EEG classification, with each step contributing uniquely to model robustness and accuracy.

3.3. Cross-validation results (Çapraz doğrulama sonuçları)

To ensure robust evaluation and mitigate the potential influence of data splitting on model performance, 5-fold Stratified Cross-Validation was conducted using the SVM model with a linear kernel. This cross-validation approach maintains the proportion of classes in each fold, providing a balanced representation of the data and enhancing the reliability of the results. Table 2. summarizes the performance metrics (Precision, Recall, F1 Score, and Accuracy) across each fold, along with the average results.

Table 2. Performance of SVM with 5-fold stratified cross-validation
(5 katlı stratified çapraz doğrulama ile performans değerlendirmesi)

Fold index	Precision	Recall	F1 Score	Accuracy
1	0.9777	0.9752	0.9756	0.9750
2	0.9971	0.9969	0.9969	0.9969
3	0.9851	0.9844	0.9844	0.9844
4	0.9910	0.9906	0.9906	0.9906
5	0.9971	0.9969	0.9969	0.9969
Average	0.9896	0.9888	0.9889	0.9888

The results demonstrate consistently high performance across all folds, indicating the model's stability and robustness. Each fold achieves near-identical metrics, with accuracy values ranging from 97.50% to 99.69%, and precision, recall, and F1 Score values also maintaining high levels. The average metrics (Precision: 0.9896, Recall: 0.9888, F1 Score: 0.9889, Accuracy: 0.9888) underscore the reliability of the SVM model when using StratifiedKFold validation, suggesting strong generalizability across different subsets of the data.

This cross-validation approach confirms that the model is well-suited for the classification task, showing minimal variance between folds and maintaining a high level of accuracy and consistency in performance across each metric.

3.4. Comparison of machine learning models (Makine öğrenmesi modellerinin karşılaştırması)

The performance of multiple machine learning algorithms on the TSEEG dataset was evaluated using key metrics—Precision, Recall, F1 Score, and Accuracy—to assess each model's ability to generalize to unseen EEG data. The TSEEG dataset was partitioned into training and test sets with an 85:15 ratio, consistent with the methodology outlined in previous sections. The performance results for both training and test sets are summarized in Table 3. Additionally, the implications of each metric are analyzed in detail, with references to graphical representations provided to enhance clarity and facilitate interpretation.

Table 3. Performance comparison of machine learning models on the TSEEG dataset
(Makine öğrenimi modellerinin TSEEG veri kümesi üzerindeki performans karşılaştırması)

Dataset	Models	Precision	Recall	F1 Score	Accuracy
Train	CatBoost	1.0	1.0	1.0	1.0
	LightGBM	1.0	1.0	1.0	1.0
	SVM	1.0	1.0	1.0	1.0
	KNN	0.9773	0.9759	0.9762	0.9765
	Logistic Regression	1.0	1.0	1.0	1.0
	Random Forest	1.0	1.0	1.0	1.0
	Gradient Boosting	1.0	1.0	1.0	1.0
	XGBoost	1.0	1.0	1.0	1.0
Test	CatBoost	0.9595	0.9672	0.9625	0.9667
	LightGBM	0.9694	0.9771	0.9704	0.9792
	SVM	**0.9919**	**0.9913**	**0.9912**	**0.9917**
	KNN	0.9647	0.9709	0.9665	0.9667
	Logistic Regression	0.9845	0.9807	0.9818	0.9833
	Random Forest	0.9704	0.9802	0.972	0.9792
	Gradient Boosting	0.9537	0.9565	0.9537	0.9542
	XGBoost	0.9698	0.9746	0.9711	0.975

The results in Table 3 indicate that SVM and Logistic Regression emerge as the leading models, demonstrating superior test accuracy, recall, and F1 scores. SVM achieves the highest classification performance on the test set, with an accuracy of 99.17%, closely followed by Logistic Regression at 98.33%. Both models maintain near-perfect scores across all metrics on the training set, suggesting strong generalizability and high classification accuracy. Although most models (except KNN) achieve perfect performance on the training set, they show slight drops on the test set, indicating they may not generalize perfectly to unseen data, though they still perform well. The KNN model, meanwhile, demonstrates lower training performance compared to other models, reflecting limitations in both training and generalization capacity.

3.4.1. Detailed performance analysis by metrics (Metriğe göre detaylı performans analizi)

Accuracy: SVM achieves the highest accuracy on the test set (99.17%), indicating its strong capability to correctly classify EEG signals across classes (see Figure 2 for detailed training and test accuracy comparisons). Logistic Regression also performs exceptionally well, achieving 98.33% accuracy. LightGBM and XGBoost follow closely with 97.92% and 97.5%, respectively, indicating robust classification abilities, though slightly less precise than SVM. CatBoost and Random Forest perform well, achieving accuracy levels above 96%, while Gradient Boosting and KNN, at 95.42% and 96.67%, respectively, show slightly lower test accuracy, potentially due to sensitivity to the variability in EEG data. Figure 2 highlights these accuracy differences, illustrating the training-test gap for each model, with KNN showing the greatest discrepancy.

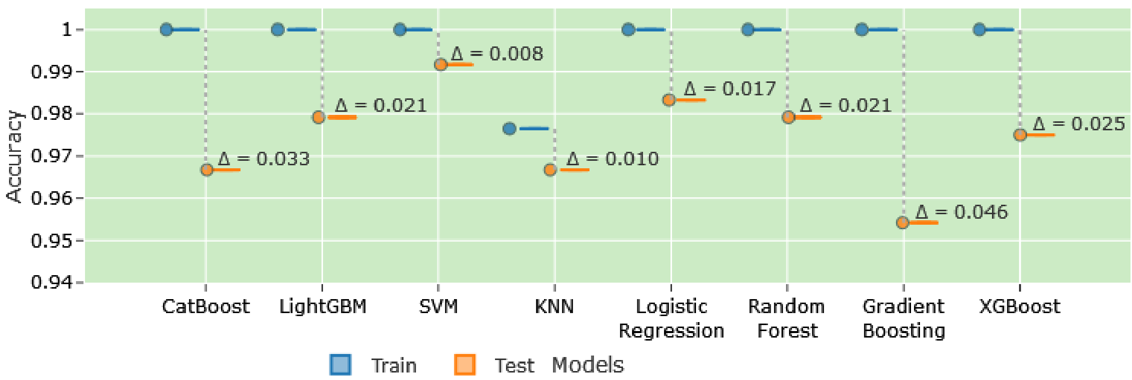

Figure 2. Comparison of accuracy in training and test operations of models
(Modellerin eğitim ve test işlemlerindeki doğruluk karşılaştırması)

F1 Score: The F1 Score, which balances precision and recall, further highlights SVM's strong performance with a score of 0.9912, indicating optimal performance in balancing precision and recall (see Figure 3). Logistic Regression and LightGBM achieve F1 scores of 0.9818 and 0.9704, respectively, showing well-rounded performance. CatBoost, despite achieving perfect scores on the training set, has a test F1 Score of 0.9625, suggesting a slight decline in balance between precision and recall on unseen data. Gradient Boosting, with the lowest F1 Score at 0.9537, may be less effective in cases where a balanced classification metric is critical. The training-test F1 Score differences in Figure 3. clearly illustrate how each model's performance varies on unseen data.

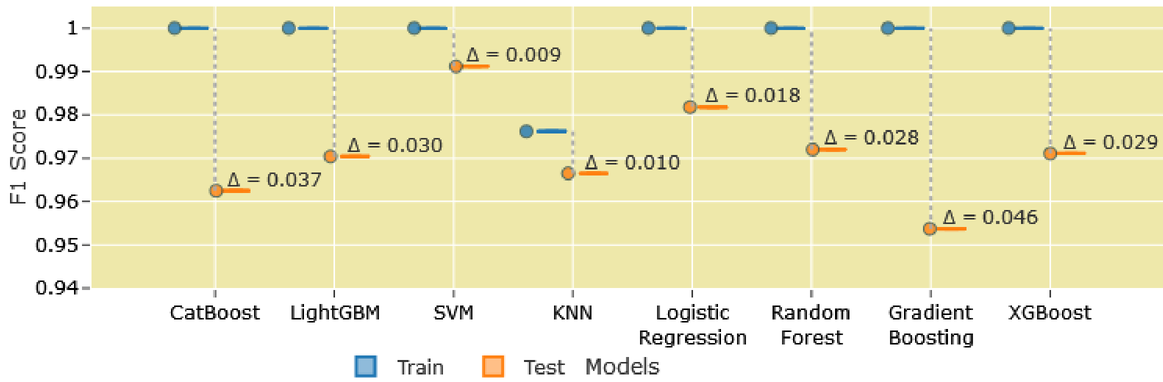

Figure 3. Comparison of f1-score in training and test operations of models
(Modellerin eğitim ve test işlemlerindeki f1 skor karşılaştırması)

Precision: Precision reflects each model's capacity to avoid false positives. SVM demonstrates the highest test precision at 0.9919, indicative of its reliability in correctly identifying positive instances (see Figure 4 for training and test precision comparisons). Logistic Regression and LightGBM also achieve high precision values (0.9845 and 0.9694), underscoring their effectiveness in minimizing false classifications. Conversely, Gradient Boosting exhibits the lowest test precision (0.9537), suggesting a slightly higher susceptibility to false positives compared to other models. The precision differences between training and test sets are clearly visualized in Figure 4, showing the generalization capability of each model on unseen data.

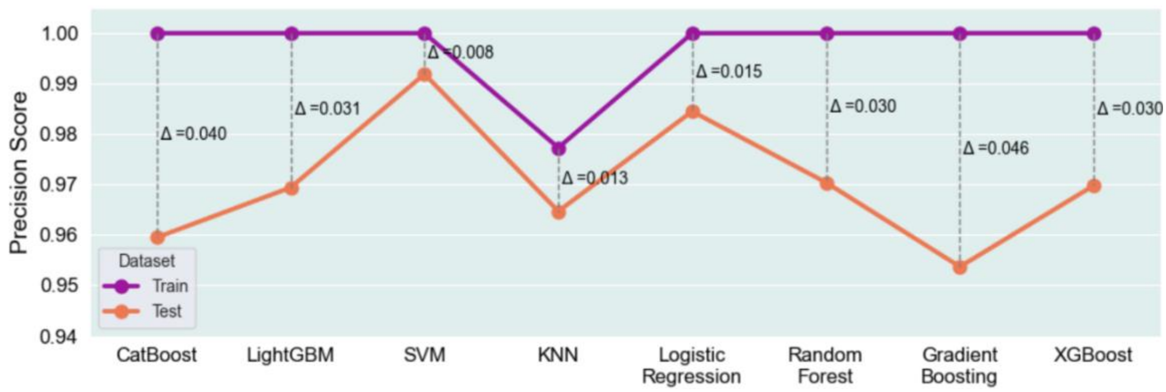

Figure 4. Comparison of precision in training and test operations of models
(Modellerin eğitim ve test işlemlerindeki kesinlik karşılaştırması)

Recall: Recall, which measures the ability to capture true positives, is high across most models, with SVM leading at 0.9913, followed closely by LightGBM (0.9771) and Random Forest (0.9802). Figure 5 demonstrates these recall values, where SVM and LightGBM exhibit minimal discrepancy between training and test recall, suggesting robust true positive capture rates. Lower recall in KNN and Gradient Boosting (0.9709 and 0.9565, respectively) indicates these models may not capture all relevant instances as effectively as SVM or LightGBM. The training-test gap is particularly visible in Figure 5, where the KNN model shows the greatest recall discrepancy.

Figure 5. Comparison of recall in training and test operations of models
(Modellerin eğitim ve test işlemlerindeki duyarlılık karşılaştırması)

3.4.2. Training time comparison (Eğitim süresi karşılaştırması)

The training times for each model, displayed on a logarithmic scale in Figure 6, reveal significant computational differences among algorithms. KNN and SVM are the fastest models, requiring only 0.1 seconds and 0.2 seconds respectively, closely followed by Logistic Regression, which also trains within 0.2 seconds. XGBoost, while still relatively efficient, requires 0.7 seconds, indicating slightly greater computational demand. Random Forest and LightGBM take 1.6 seconds and 1.8 seconds respectively, reflecting their more complex ensemble structures. CatBoost requires a more substantial 12.6 seconds, while Gradient Boosting has the longest training time, taking 131.6 seconds, largely due to its iterative boosting process. These training times illustrate that, while ensemble models generally offer robust performance, they can incur significant computational costs, particularly for real-time applications.

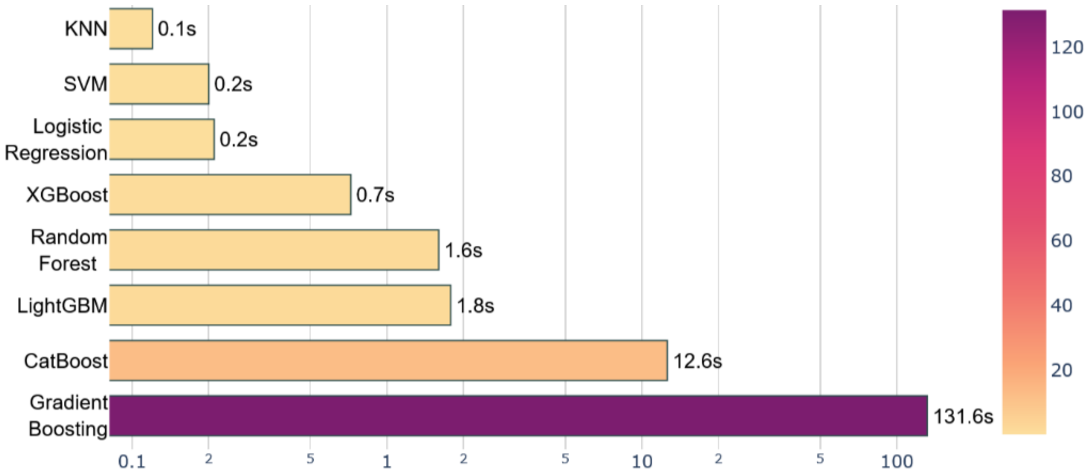

Figure 6. Training time comparison of models, displayed in log scale
(Modellerin eğitim süresi karşılaştırması, logaritmik ölçekte gösterilmiştir)

3.4.3. Overall evaluation and summary (Genel değerlendirme ve özet)

From the overall performance analysis, SVM emerges as the most effective model for EEG classification on the TSEEG dataset. It achieves the highest accuracy (99.17%) and F1 Score on the test set, with minimal training-test discrepancies across all metrics, demonstrating both precision and reliability in handling EEG data. Logistic Regression also performs remarkably well, showing high generalization capability with minimal training time. LightGBM provides a balanced alternative, maintaining robust classification performance, though with slightly higher computational demand. In contrast, while CatBoost and Random Forest exhibit high classification accuracy, their longer training times may limit scalability in time-sensitive contexts. KNN, despite its simplicity and fast training time, shows lower overall performance and higher discrepancies between training and test metrics, indicating a limited ability to generalize effectively. Finally, Gradient Boosting, although powerful, has the highest training time, which may reduce its practicality in scenarios where efficiency is critical. SVM and Logistic Regression stand out as top choices for EEG classification, balancing high performance with computational efficiency. LightGBM also serves as a strong candidate when ensemble methods are preferred. This analysis underscores the importance of selecting models that not only excel in accuracy but also align with the computational constraints of real-world applications.

4. Conclusion (Sonuç)

This study provides a comprehensive evaluation of frequently used machine learning algorithms for EEG-based classification on the TSEEG dataset, focusing on both classification accuracy and computational efficiency. Among the tested models, the Support Vector Machine (SVM) achieved the best performance, with a test accuracy of 99.17% and an F1 score of 99.12%. It demonstrated a balanced trade-off between accuracy and time efficiency, outperforming other models such as Logistic Regression, LightGBM, and CatBoost across all key metrics. The robustness and reliability of the SVM model were further validated through 5-fold cross-validation, consistently delivering stable results across different data splits. Additionally, this study analyzed the impact of various preprocessing steps on classification performance, with a particular emphasis on the SVM model. A structured preprocessing pipeline—including normalization, filtering, and feature extraction—was shown to play a pivotal role in enhancing classification accuracy. These steps reduced noise and variability in the EEG signals, optimizing the SVM's ability to process and classify data effectively. The study emphasizes that preprocessing is an indispensable component for reliable EEG-based classification. While achieving high accuracy remains paramount, the time efficiency demonstrated by SVM further underscores its suitability for real-world applications, especially as EEG datasets continue to grow in size and complexity. By balancing precision, robustness, and computational speed, SVM emerged as a strong candidate for future EEG-based systems.

While this study focuses on the TSEEG dataset, which provides valuable insights into EEG-based language classification, it is recognized that the dataset may not fully encompass the variability of EEG signals across different languages, cultural contexts, or user demographics. This highlights an exciting opportunity for future work to expand the scope by incorporating more diverse datasets, which would enhance the generalizability and applicability of these findings. Although advanced deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), were not included in this study due to the moderate size of the dataset, their potential to capture complex temporal and spatial patterns in EEG data remains promising. As larger and more complex datasets become available, exploring these models with appropriate regularization techniques could unlock deeper insights and advance the capabilities of EEG-based classification systems.

Future research should focus on developing diverse EEG datasets that encompass multiple languages, including Turkish and other underrepresented languages, to facilitate comprehensive cross-linguistic and cross-cultural EEG studies. Such datasets would enable rigorous testing of model generalizability and adaptation across various linguistic and demographic groups, enhancing the applicability of EEG classification systems. Moreover, future studies could explore integrating advanced deep learning models, such as CNNs and RNNs, particularly when larger and more complex datasets become available. These models would allow deeper exploration of the temporal and spatial structures of EEG signals, unlocking new possibilities for EEG-based applications.

In summary, this study provides a foundational framework for EEG classification, highlighting the critical role of preprocessing, model accuracy, and computational efficiency. The findings serve as a baseline for future research and point towards the development of universal, scalable, and culturally adaptive EEG-based classification systems. By addressing the limitations of existing datasets and incorporating advanced methodologies, future work can advance the field towards language-inclusive, real-time EEG solutions that align with diverse real-world demands.

Conflict of Interest Statement (Çıkar Çatışması Beyanı)

No conflicts of interest have been declared by the authors.

References (Kaynaklar)

[1] F. Aydemir and S. Arslan, “A system design with deep learning and IoT to ensure education continuity for post-COVID,” *IEEE Transactions on Consumer Electronics*, vol. 69, no. 2, pp. 217–

225, 2023.
[2] E. Şahin and M. F. Talu, "Denim Kumaşından Otomatik Yüksek Çözünürlüklü Bıyık Desen Sentezi," *Computer Science*, vol. IDAP-2022, pp. 86–100, 2022.
[3] S. A. Güven and M. F. Talu, "Brain MRI high resolution image creation and segmentation with the new GAN method," *Biomed Signal Process Control*, vol. 80, p. 104246, 2023.
[4] G. Arslan, F. Aydemir, and S. Arslan, "Enhanced license plate recognition using deep learning and block-based approach," *Journal of Scientific Reports-A*, no. 058, pp. 57–82, 2023.
[5] V. Kaya, "Classification of waste materials with a smart garbage system for sustainable development: a novel model," *Front Environ Sci*, vol. 11, p. 1228732, 2023.
[6] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimed Tools Appl*, vol. 82, no. 3, pp. 3713–3744, 2023.
[7] Y. Yao, J. Duan, K. Xu, Y. Cai, Z. Sun, and Y. Zhang, "A survey on large language model (llm) security and privacy: The good, the bad, and the ugly," *High-Confidence Computing*, p. 100211, 2024.
[8] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, "An introductory review of deep learning for prediction models with big data," *Front Artif Intell*, vol. 3, p. 4, 2020.
[9] D. Özdemir, S. Dörterler, and D. Aydın, "A new modified artificial bee colony algorithm for energy demand forecasting problem," *Neural Comput Appl*, vol. 34, no. 20, pp. 17455–17471, 2022.
[10] N. Yagmur, İ. Dag, and H. Temurtas, "Classification of anemia using Harris hawks optimization method and multivariate adaptive regression spline," *Neural Comput Appl*, vol. 36, no. 11, pp. 5653–5672, 2024.
[11] P. Rajpurkar, E. Chen, O. Banerjee, and E. J. Topol, "AI in health and medicine," *Nat Med*, vol. 28, no. 1, pp. 31–38, 2022.
[12] E. Şahin, D. Özdemir, and H. Temurtaş, "Multi-objective optimization of ViT architecture for efficient brain tumor classification," *Biomed Signal Process Control*, vol. 91, p. 105938, 2024.
[13] S. Altun and M. F. Talu, "A new approach for Pap-Smear image generation with generative adversarial networks," *Journal of the Faculty of Engineering and Architecture of Gazi University*, vol. 37, no. 3, pp. 1401–1410, 2022.
[14] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: A survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 315–329, 2020.
[15] G. Bendiab, A. Hameurlaine, G. Germanos, N. Kolokotronis, and S. Shiaeles, "Autonomous vehicles security: Challenges and solutions using blockchain and artificial intelligence," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 3614–3637, 2023.
[16] V. Kaya, İ. Akgül, and Ö. Z. Tanır, "IsVoNet8: a proposed deep learning model for classification of some fish species," *J Agric Sci (Belihuloya)*, vol. 29, no. 1, pp. 298–307, 2023.
[17] Ç. Erçelik and K. Hanbay, "Beyin Tümörü Sınıflandırmada Histogram Eşitleme Yönteminin Bazı Derin Öğrenme Modellerine Etkileri," *Computer Science*, vol. 8, no. 2, pp. 83–92, 2023.
[18] E. ŞAHiN, N. N. Arslan, and D. Özdemir, "Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning," *Neural Comput Appl*, pp. 1–107, 2024.
[19] F. Aydemir and S. Arslan, "COVID-19 PANDEMİ SÜRECİNDE ÇOCUKLARIN EL YIKAMA ALIŞKANLIĞININ NESNELERİN İNTERNETİ TABANLI SİSTEM İLE İZLENMESİ," *Mühendislik Bilimleri ve Araştırmaları Dergisi*, vol. 3, no. 2, pp. 161–168, 2021.
[20] V. Kaya, S. Tuncer, and A. Baran, "Detection and classification of different weapon types using deep learning," *Applied Sciences*, vol. 11, no. 16, p. 7535, 2021.
[21] S. Dörterler, S. Arslan, and D. Özdemir, "Unlocking the potential: A review of artificial intelligence applications in wind energy," *Expert Syst*, p. e13716, 2024.
[22] U. Chaudhary, N. Birbaumer, and A. Ramos-Murguialday, "Brain–computer interfaces for communication and rehabilitation," *Nat Rev Neurol*, vol. 12, no. 9, pp. 513–525, 2016.
[23] S. N. Abdulkader, A. Atia, and M.-S. M. Mostafa, "Brain computer interfacing: Applications and challenges," *Egyptian Informatics Journal*, vol. 16, no. 2, pp. 213–230, 2015.
[24] X. Gu *et al.*, "EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications," *IEEE/ACM Trans Comput Biol Bioinform*, vol. 18, no. 5, pp. 1645–1666, 2021.
[25] K. Värbu, N. Muhammad, and Y. Muhammad, "Past, present, and future of EEG-based BCI applications," *Sensors*, vol. 22, no. 9, p. 3331, 2022.
[26] U. Shah, M. Alzubaidi, F. Mohsen, A. Abd-Alrazaq, T. Alam, and M. Househ, "The role of artificial

intelligence in decoding speech from EEG signals: a scoping review," *Sensors*, vol. 22, no. 18, p. 6975, 2022.
[27] H. Banville, O. Chehab, A. Hyvärinen, D.-A. Engemann, and A. Gramfort, "Uncovering the structure of clinical EEG signals with self-supervised learning," *J Neural Eng*, vol. 18, no. 4, p. 046020, 2021.
[28] V. Roy, P. K. Shukla, A. K. Gupta, V. Goel, P. K. Shukla, and S. Shukla, "Taxonomy on EEG artifacts removal methods, issues, and healthcare applications," *Journal of Organizational and End User Computing (JOEUC)*, vol. 33, no. 1, pp. 19–46, 2021.
[29] S. Deng, R. Srinivasan, T. Lappas, and M. D'Zmura, "EEG classification of imagined syllable rhythm using Hilbert spectrum methods," *J Neural Eng*, vol. 7, no. 4, p. 046006, 2010.
[30] K. Brigham and B. V. K. V. Kumar, "Imagined speech classification with EEG signals for silent communication: a preliminary investigation into synthetic telepathy," in *2010 4th International Conference on Bioinformatics and Biomedical Engineering*, IEEE, 2010, pp. 1–4.
[31] X. Chi, J. B. Hagedorn, D. Schoonover, and M. D'Zmura, "EEG-based discrimination of imagined speech phonemes," *Int J Bioelectromagn*, vol. 13, no. 4, pp. 201–206, 2011.
[32] A. A. T. García, C. A. R. García, and L. V. Pineda, "Toward a silent speech interface based on unspoken speech," in *International Conference on Bio-inspired Systems and Signal Processing*, SciTePress, 2012, pp. 370–373.
[33] L. C. Sarmiento, P. Lorenzana, C. J. Cortes, W. J. Arcos, J. A. Bacca, and A. Tovar, "Brain computer interface (BCI) with EEG signals for automatic vowel recognition based on articulation mode," in *5th ISSNIP-IEEE Biosignals and Biorobotics Conference (2014): Biosignals and Robotics for Better and Safer Living (BRC)*, IEEE, 2014, pp. 1–4.
[34] A. A. Haider, N. Goga, A. Vasilateanu, A. M. Muslim, K. A. Ali, and M.-V. Dragoi, "English and Romanian Brain-to-Text Brain-Computer Interface Word Prediction System," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022.
[35] H. Park and B. Lee, "Multiclass classification of imagined speech EEG using noise-assisted multivariate empirical mode decomposition and multireceptive field convolutional neural network," *Front Hum Neurosci*, vol. 17, p. 1186594, 2023.
[36] J. T. Panachakel, A. G. Ramakrishnan, and T. V Ananthapadmanabha, "Decoding imagined speech using wavelet features and deep neural networks," in *2019 IEEE 16th India Council International Conference (INDICON)*, IEEE, 2019, pp. 1–4.
[37] L. A. Moctezuma, A. A. Torres-García, L. Villaseñor-Pineda, and M. Carrillo, "Subjects identification using EEG-recorded imagined speech," *Expert Syst Appl*, vol. 118, pp. 201–208, 2019.
[38] P. Kumar and E. Scheme, "A deep spatio-temporal model for EEG-based imagined speech recognition," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2021, pp. 995–999.
[39] S. Ullah and Z. Halim, "Imagined character recognition through EEG signals using deep convolutional neural network," *Med Biol Eng Comput*, vol. 59, no. 5, pp. 1167–1183, 2021.
[40] J. Yang *et al.*, "Thoughts of brain EEG signal-to-text conversion using weighted feature fusion-based multiscale dilated adaptive DenseNet with attention mechanism," *Biomed Signal Process Control*, vol. 86, p. 105120, 2023.
[41] S. Ali, W. Mumtaz, and A. Maqsood, "EEG based thought-to-text translation via deep learning," in *2023 7th International Multi-Topic ICT Conference (IMTIC)*, IEEE, 2023, pp. 1–8.
[42] F. Bai, A. S. Meyer, and A. E. Martin, "Neural dynamics differentially encode phrases and sentences during spoken language comprehension," *PLoS Biol*, vol. 20, no. 7, p. e3001713, 2022.
[43] G. Eroğlu, M. Çetin, and S. Balcisoy, "Electroencephalographic identifiers of reading abilities in turkish language," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2018, pp. 1–4.
[44] F. Kutlu Onay, "Translation of motor and non-motor imaginary activity EEG signals to text," PhD. , Karadeniz Technical University, Trabzon, 2020.
[45] P. D. Barua *et al.*, "Automated EEG sentence classification using novel dynamic-sized binary pattern and multilevel discrete wavelet transform techniques with TSEEG database," *Biomed Signal Process Control*, vol. 79, p. 104055, 2023.
[46] K. Haltas and A. Erguzen, "Enhancing Speech Impairment Support: Designing an EEG-Based BCI System for Turkish Vowel Recognition.," *Traitement du Signal*, vol. 41, no. 3, 2024.
[47] M. Demir, "Brain keyboard interface design with deep learning method using EEG signals," MSc, Dumlupinar University, Kütahya, 2024.
[48] L. Breiman, "Random forests," *Mach Learn*, vol. 45, pp. 5–32, 2001.

[49] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Adv Neural Inf Process Syst*, vol. 31, 2018.

[50] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Adv Neural Inf Process Syst*, vol. 30, 2017.

[51] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[52] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[53] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, Springer, 2003, pp. 986–996.

[54] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[55] S. Butterworth, "On the theory of filter amplifiers," *Wireless Engineer*, vol. 7, no. 6, pp. 536–541, 1930.

[56] J. W. Cooley and J. W. Tukey, "An algorithm for the machine calculation of complex Fourier series," *Math Comput*, vol. 19, no. 90, pp. 297–301, 1965.

[57] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *J Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[58] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.