*Fırat Üni Deny. ve Say. Müh. Derg., 4(2), 446-459, 2025*

Fırat Üniversitesi Deneysel ve Hesaplamalı Mühendislik Dergisi

# Medikal Veri İşlemede Makine Öğrenme Yaklaşımları: Felç için Akıllı Teşhis Sistemi Önerisi

Azra Şilan PERİ[1], Nida KATI[2], Ferhat UÇAR[3*]

[1,3]Yazılım Mühendisliği, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye.
[2]Metalurji ve Malzeme Mühendisliği, Teknoloji Fakültesi, Fırat Üniversitesi, Elazığ, Türkiye.
[1]azraperi497@gmail.com, [2]nkati@firat.edu.tr, [3]fucar@firat.edu.tr

Geliş Tarihi: 7.05.2025
Kabul Tarihi: 12.06.2025

Düzeltme Tarihi: 10.06.2025

doi: https://doi.org/10.62520/fujece.1694558
Araştırma Makalesi

Alıntı: A. Ş. Peri, N. Katı, F. Uçar, "Medikal veri işlemede makine öğrenme yaklaşımları: felç için akıllı teşhis sistemi önerisi ", Fırat Üni. Deny. ve Hes. Müh. Derg., vol. 4, no 2, pp. 446-459, Haziran 2025.

**Öz**

Bu çalışma, felç teşhisi için makine öğrenmesi ve derin öğrenme tabanlı bir akıllı teşhis sistemi önermektedir. Sağlık sektöründe yapay zekânın (AI) kullanımı, büyük veri analitiği ve dijitalleşme ile birlikte artmaktadır. Felç, dünya genelinde yaygın bir nörolojik hastalık olup erken teşhisle ölüm ve sakatlık oranları önemli ölçüde azaltılabilir. Çalışmada, Kaggle platformundaki 4909 bireyi kapsayan "Felç Tahmin Veri Seti" kullanılmıştır. Bu veri seti, yaş, cinsiyet, hipertansiyon, kalp hastalığı, yaşam tarzı gibi 12 giriş özelliği ve felç durumunu gösteren bir çıkış özelliği içermektedir. Veri ön işleme adımları olarak eksik verilerin ortalama ile doldurulması, kategorik verilerin One-Hot Encoding ile sayısallaştırılması, Min-Max Ölçeklendirme ve SMOTE ile sınıf dengesizliği çözülmüştür. Çalışmada, 15 farklı makine öğrenmesi ve derin öğrenme algoritması (Random Forest, Voting Classifier, Histogram Gradient Boosting, SVM, MLP vb.) değerlendirilmiş; performansları doğruluk, hassasiyet, geri çağırma, F1-skoru ve ROC-AUC metrikleriyle ölçülmüştür. Voting Classifier, %98,5 doğruluk ve 0,99 AUC ile en yüksek performansı göstermiştir. Random Forest ve Histogram Gradient Boosting gibi ağaç tabanlı modeller de yüksek doğruluk oranlarıyla dikkat çekmiştir. Hiperparametre optimizasyonu için GridSearchCV ve RandomizedSearchCV kullanılmış, aşırı öğrenmeyi önlemek için erken durdurma, düzenlileştirme ve dropout teknikleri uygulanmıştır. Bulgular, topluluk öğrenme yöntemlerinin felç teşhisinde geleneksel yöntemlere üstünlük sağladığını göstermektedir. Çalışma, yapay zeka tabanlı klinik karar destek sistemlerinin sağlık sektörüne entegrasyonunun önemini vurgulamakta ve gelecekte daha büyük veri setleriyle model performansının artırılabileceğini önermektedir.

**Anahtar kelimeler:** Felç teşhisi, Makine öğrenmesi, Derin öğrenme, Topluluk öğrenmesi, Klinik karar destek sistemleri

[*] Yazışılan yazar

İntihal Kontrol: Yes – Turnitin
Şikayet: fujece@firat.edu.tr

Firat University Journal of Experimental and Computational Engineering

FUJECE

# Machine Learning Approaches in Medical Data Processing: A Proposal for an Intelligent Stroke Diagnosis System

Azra Şilan PERİ[1], Nida KATI[2], Ferhat UCAR[3*]

[1,3]Software Engineering, Faculty of Technology, Fırat University, Elazığ, Türkiye.
[2]Metallurgy and Materials Engineering, Faculty of Technology, Fırat University, Elazığ, Türkiye.
[1]azraperi497@gmail.com, [2]nkati@firat.edu.tr, [3]fucar@firat.edu.tr

Received: 7.05.2025
Accepted: 12.06.2025

Revision: 10.06.2025

doi: https://doi.org/10.62520/fujece.1694558
Research Article

Citation: A. Ş. Peri, N. Katı, F. Uçar, "Machine learning approaches in medical data processing: a proposal for an ıntelligent stroke diagnosis system", Firat Univ. Jour.of Exper. and Comp. Eng., vol. 4, no 2, pp. 446-459, June 2025.

## Abstract

This study proposes an intelligent diagnostic system based on machine learning and deep learning for stroke detection. The use of artificial intelligence (AI) in healthcare is increasing alongside big data analytics and digitalization. Stroke, a prevalent neurological disease worldwide, can have its mortality and disability rates significantly reduced through early diagnosis. The study utilizes the "Stroke Prediction Dataset" from Kaggle, encompassing 4909 individuals. This dataset includes 12 input features such as age, gender, hypertension, heart disease, and lifestyle factors, along with one output feature indicating stroke status. Data preprocessing steps involved filling missing values with the mean, converting categorical data to numerical format using One-Hot Encoding, applying Min-Max Scaling, and addressing class imbalance with SMOTE. Fifteen different machine learning and deep learning algorithms (e.g., Random Forest, Voting Classifier, Histogram Gradient Boosting, SVM, MLP) were evaluated, with performance measured using accuracy, precision, recall, F1-score, and ROC-AUC metrics. The Voting Classifier achieved the highest performance with 98.5% accuracy and an AUC of 0.99. Tree-based models like Random Forest and Histogram Gradient Boosting also demonstrated high accuracy. Hyperparameter optimization was performed using GridSearchCV and RandomizedSearchCV, while early stopping, regularization, and dropout techniques were applied to prevent overfitting. The findings highlight the superiority of ensemble learning methods over traditional approaches in stroke diagnosis. The study underscores the importance of integrating AI-based clinical decision support systems into healthcare and suggests that model performance could be further enhanced with larger datasets in the future.

**Keywords:** Stroke diagnosis, Machine learning, Deep learning, Ensemble learning, Clinical decision support systems

*Corresponding author

Plagiarism Checks: Yes – Turnitin
Complaints: fujece@firat.edu.tr

## 1. Introduction

Artificial intelligence (AI) enables computer systems to learn from examples, allowing them to mimic the learning process, one of the basic functions of the human brain. Deep learning, as a sub-branch of machine learning, stands out with its ability to analyze large datasets and identify complex patterns. Especially in healthcare, AI-based systems are widely used to speed up diagnostic processes, support treatment decisions and improve patient care [1-2]. Healthcare is becoming increasingly digitalized and the need for big data analytics is growing as patient data moves to electronic media. The Health 4.0 concept promotes the use of artificial intelligence in disease diagnosis, supporting personalized treatment approaches [3]. Artificial intelligence provides successful results in many fields such as medical imaging, analysis of patient records and evaluation of genetic data, and increases the diagnostic accuracy of physicians with clinical decision support systems [4].

Stroke is one of the most common neurological diseases worldwide, affecting approximately 13.7 million people and causing 5.5 million deaths each year [5]. Ischemic strokes account for 87% of total strokes and this proportion has been increasing over the years [6]. Stroke is a sudden health problem that can seriously affect the quality of life of patients and cause permanent disability. However, with early diagnosis and rapid intervention, stroke-related deaths and disability rates can be prevented by approximately 80%. [7].

Recent studies show that machine learning and deep learning algorithms offer significant advantages in stroke diagnosis [8-9]. Compared to traditional diagnostic methods, artificial intelligence models can make disease predictions over large datasets and help diagnose diseases faster. In particular, combining methods such as deep neural networks (DNN), convolutional neural networks (CNN) and recurrent neural networks (RNN) with medical image analysis significantly improves diagnostic accuracy [1], [10]. The use of machine learning (ML) and deep learning (DL) based models for stroke diagnosis has become an important area of research in recent years [11]. Performance comparisons of various algorithms have been made and the most accurate methods have been identified. Caliskan and Ates used common machine learning algorithms such as Logistic Regression (LR), Decision Tree (DT), Support Vector Machines (SVM) and k-Nearest Neighbors (KNN) for stroke risk assessment. According to the results obtained, the DT model showed the highest success with 91% accuracy rate. Other models were evaluated in terms of classification performance with 89% accuracy for SVM, 81% for KNN and 75% for LR. [12]. In a similar study, Oğuz et al. performed a comparative analysis using 13 different machine learning models for early diagnosis and risk classification of stroke. As a result of the experiments, the Random Forest Classifier (Random Forest) was found to be the most successful model with a 99.425% accuracy rate [13].

In addition to machine learning-based approaches, DL and ML techniques also provide effective results in medical image analysis and classification processes. Alhatemi et al. conducted a study on the analysis of brain MRI images using DNN. Popular deep learning models such as DenseNet121, ResNet50, Xception, MobileNet, VGG16 and EfficientNetB2 were used in the study. The highest accuracy rate of 98.8% was obtained with the EfficientNetB2 model. The model’s sensitivity, precision and F1-score were also quite high, indicating that deep learning has great potential in stroke diagnosis [14]. Nancy et al. developed a stroke diagnosis model based on Deep Kernel Extreme Learning Machine (DKELM-AS) using electroencephalogram (EEG) data. The Fast Hartley Transform (FHT) was used to extract features from EEG signals and the DKELM-AS model achieved 95.2% accuracy. EEG-based diagnostic systems offer a faster and more cost-effective alternative to conventional imaging methods [15]. In another study focused on image analysis, UmaMaheswaran et al. developed a model for optimal feature selection in acute stroke diagnosis by analyzing computed tomography (CT) images. Feature extraction methods such as Local Binary Pattern (LBP), Gabor Filter and Discrete Wavelet Transform (DWT) were used. The XGBoost model achieved the best results with a 97% accuracy and a 0.015% false positive rate. The model showed a higher accuracy than traditional support vector machine (SVM), artificial neural network (ANN) and naive Bayes (NB) methods [16]. Ensemble machine learning techniques also offer significant advantages for stroke diagnosis. Srivinas et al. developed an ensemble learning-based diagnosis method by combining Random Forest, Extra Trees and Histogram Gradient Boosting (HGB) models. By combining the predictions from the individual classifiers, the model improved accuracy and enabled more reliable results in clinical practice. The study also suggested that the use of swarm intelligence-based optimization techniques in the future could further improve model performance [17]. These studies emphasize the importance of artificial intelligence (AI) and big data analytics in stroke diagnosis. Artificial intelligence models can support clinicians in the early diagnosis process by analyzing individual patient data in detail. With the Health 4.0 concept, personalized healthcare services are becoming widespread and the integration of machine learning-based models into clinical decision support systems is becoming increasingly important [4].

In this study, unlike the literature, the performance of ensemble learning techniques and traditional machine learning models in stroke diagnosis is investigated in depth. In this context, we systematically evaluate the potential of ensemble methods such as Random Forest, Extreme Randomized Trees and Histogram-based Gradient Boosting to offer superior performance compared to individual classifiers. Furthermore, the practicality and generalizability of these models in clinical applications are tested on real-world data to provide a new perspective to the literature.

The datasets used in this study consist of a comprehensive pool of data including patient histories, biometric measurements and lifestyle factors [18]. During the modeling process, detailed analyses were performed on these data to evaluate the performance of AI-based diagnostic systems. By emphasizing the critical importance of early diagnosis, the study aims to demonstrate the integration of AI-based clinical decision support systems into the healthcare sector and their transformative impact on clinical applications.

In the rest of the paper, the study consists of Material and Method, Results and Discussion, and Conclusion sections. In the Materials and Method section, the datasets, algorithms and modeling processes used are explained in detail. In the Results and Discussion section, the performance of the proposed models is evaluated and comparative analysis with the literature is presented. The conclusion summarizes the findings of the study, discusses its contributions to clinical practice and provides recommendations for future research.

## 2. Material and Method

The use of artificial intelligence and machine learning algorithms in stroke diagnosis increases the importance of big data analysis in the healthcare industry. Machine learning (ML) has become an effective tool in pattern recognition and data classification processes in disease diagnosis. Deep learning (DL) offers the ability to learn from more complex datasets using neural network structures. In this study, various ML and deep learning algorithms are used to determine the model with the highest accuracy rate in stroke diagnosis.

### Dataset Description

In this study, the "Stroke Prediction Dataset" [18] on Kaggle, which covers 4909 individuals, is used. This dataset consists of 13 attributes in total, 12 different input attributes for classification and prediction about the risk of having a stroke and a single output attribute that provides information about the stroke status. Each attribute in the dataset represents various information about the health status and lifestyle of individuals and reveals important factors affecting the risk of stroke. Table 1 provides detailed descriptions of each attribute.

**Table 1.** Dataset descriptions

| Attribute | Description |
|---|---|
| id | Identification Number |
| gender | Female/Male |
| age | age |
| hypertension | Available :1, Not Available :0 |
| heart_disease | Available :1, Not Available:0 |
| ever_married | Yes/No |
| work_type | Private Sector / State / Unemployed |
| residence_type | Rural/Urban |
| avg_glucose_level | Reference Value |
| bmi | Reference Value |
| smoking_status | Yes/No |
| stroke | Stroke:1, No stroke:0 |

Analyzing these variables together provides a better understanding of the relationships between individuals' health status and their risk of stroke. Stroke is a common health problem worldwide and such data plays an important role in formulating health policies and directing health services [4-5].

In the data analysis process, data cleaning, visualization and application of various machine learning algorithms were performed. In this process, the general structure of the dataset was examined by considering the missing data processing with appropriate methods. The distribution of the data was visualized using histograms and box plots for continuous variables and bar charts for categorical variables (Figure 1).

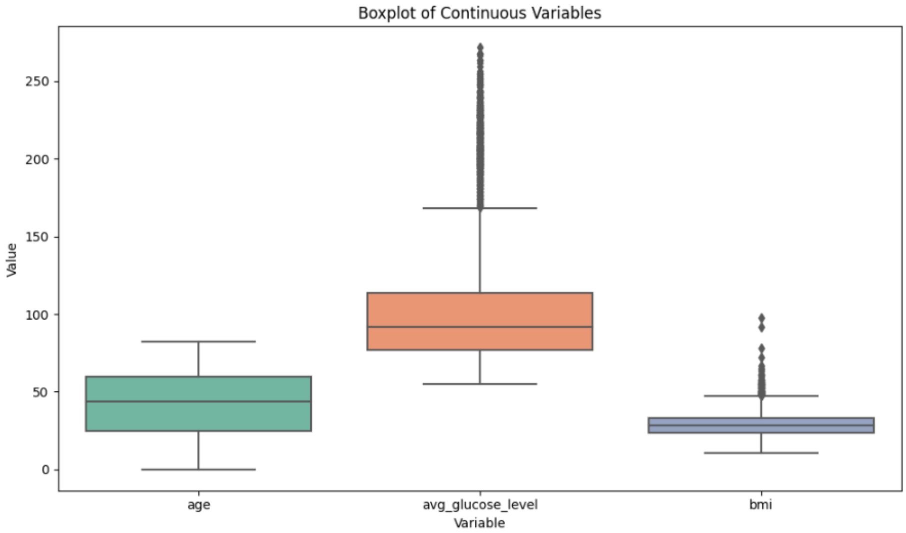


**Figure 1.** Boxplot presentation of age, bmi and avg_glucose_level features

In addition, correlation analyses were performed to evaluate the relationships between variables (Figure 2). The findings contributed to the development of medical decision support systems and enabled the analysis of determinants of health outcomes. These analyses provided important information for personalizing treatment plans, increasing the efficiency of healthcare services and improving clinical decision-making processes [3]. Data-driven approaches have the potential to help develop future health applications and optimize patient care processes.

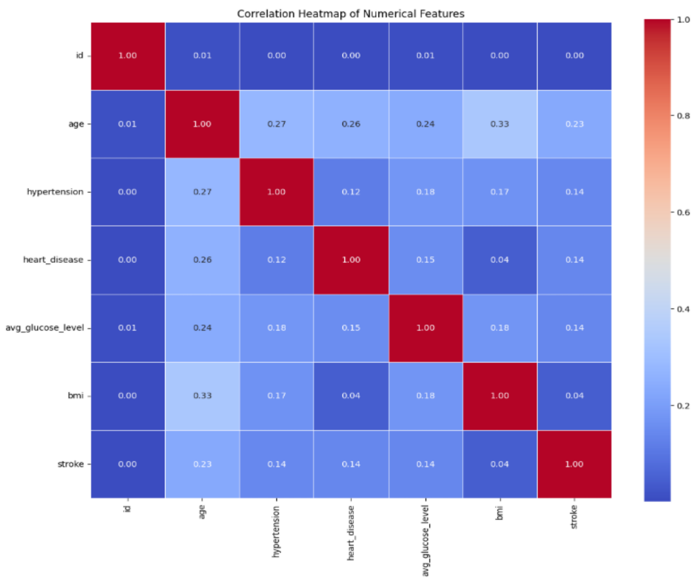


**Figure 2.** Heatmap of the Correlation Matrix for Numerical Variables

**Data Preprocessing Methods**

In the data analysis process, it is often not possible to use the raw data directly. Therefore, a series of pre-processing steps were applied to improve the quality of the data and the performance of the machine learning models.

During the initial data quality analysis, it was observed that the dataset contained missing values. Upon inspection, these were found exclusively in the "Body Mass Index" (BMI) variable, with a total of 201 records lacking this information. To maintain the integrity of the dataset and prevent a negative impact on model performance, these missing entries were imputed using the mean of the BMI variable. This mean imputation method was chosen as it is a standard technique that preserves the central tendency of the variable without significantly distorting its overall distribution. This process ensured that the statistical properties of the dataset were maintained, providing a more reliable foundation for the modeling phase.

The categorical variables in the dataset (e.g. gender, marriage status, smoking) should be converted into the numerical format required by the machine learning models. For this purpose, the One-Hot Encoding method was applied. This method digitized categorical data by creating new binary (0/1) variables for each categorical variable.

Numerical features in the dataset (age, BMI, glucose level) may negatively affect the performance of machine learning models due to their different scales. To solve this problem, these variables are scaled between 0 and 1 using the Min-Max Scaling method. This method performs normalization by considering the minimum and maximum values of each feature. Scaling is especially critical for distance-based algorithms (e.g. K-Nearest Neighbor) and gradient-based methods, as it ensures that models give equal weight to all features.

Since the proportion of individuals who suffered a stroke in the dataset is quite low, the problem of class imbalance arises. This can lead to models not learning enough about the minority class (stroke survivors). SMOTE (Synthetic Minority Over-sampling Technique) was used to solve this problem. SMOTE is a technique that balances the class distribution by adding synthetic examples to the minority class. This method generates new samples by interpolating between data points in the minority class, thus allowing the model to better learn the minority class [19]. These pre-processing steps made the dataset suitable for machine learning models and contributed to improving model performance.

**Modeling Process and Algorithms Used**

In this study, 15 different algorithms based on machine learning (ML) and deep learning (DL) were evaluated to optimize stroke diagnosis. The algorithms used in this study are grouped into three main categories according to their learning methods. These categories are 1) ensemble learning models, 2) support vector machines and neural networks, and 3) tree-based and statistical models. The aim of the study is to contribute to the analysis of health data by identifying the method with the highest accuracy as its overall schema visually described in Figure 3.

**Community Learning Models**

Ensemble learning models aim to achieve higher accuracy by combining the predictions of multiple algorithms. The following algorithms are used in this category:

- **Random Forest (RF):** It is a model created by combining multiple decision trees. Trees trained with random subsets produce the final prediction by voting [20]. Each tree is trained on a different random subset of the data (a technique known as bootstrapping), and each split in the tree considers only a random subset of features. This methodology promotes diversity among the trees and significantly reduces the model's risk of overfitting. It is known for its robustness against extreme learning in medical datasets.
- **Gradient Boosting (GB)**: Minimizes the error rate by training decision trees sequentially [21]. Delivers strong performance on complex datasets [22].

- **Histogram Gradient Boosting (HGB)**: It is an optimized version of Gradient Boosting and provides computational efficiency on large datasets [23].
- **AdaBoost (AB)**: Strengthens weak learners by weighting examples with high error rates [24]. Advantageous in medical diagnostics.
- **Voting Classifier (VC)**: Make decisions by combining the predictions of different models [25]. Accuracy is improved by majority or weighted voting [26].

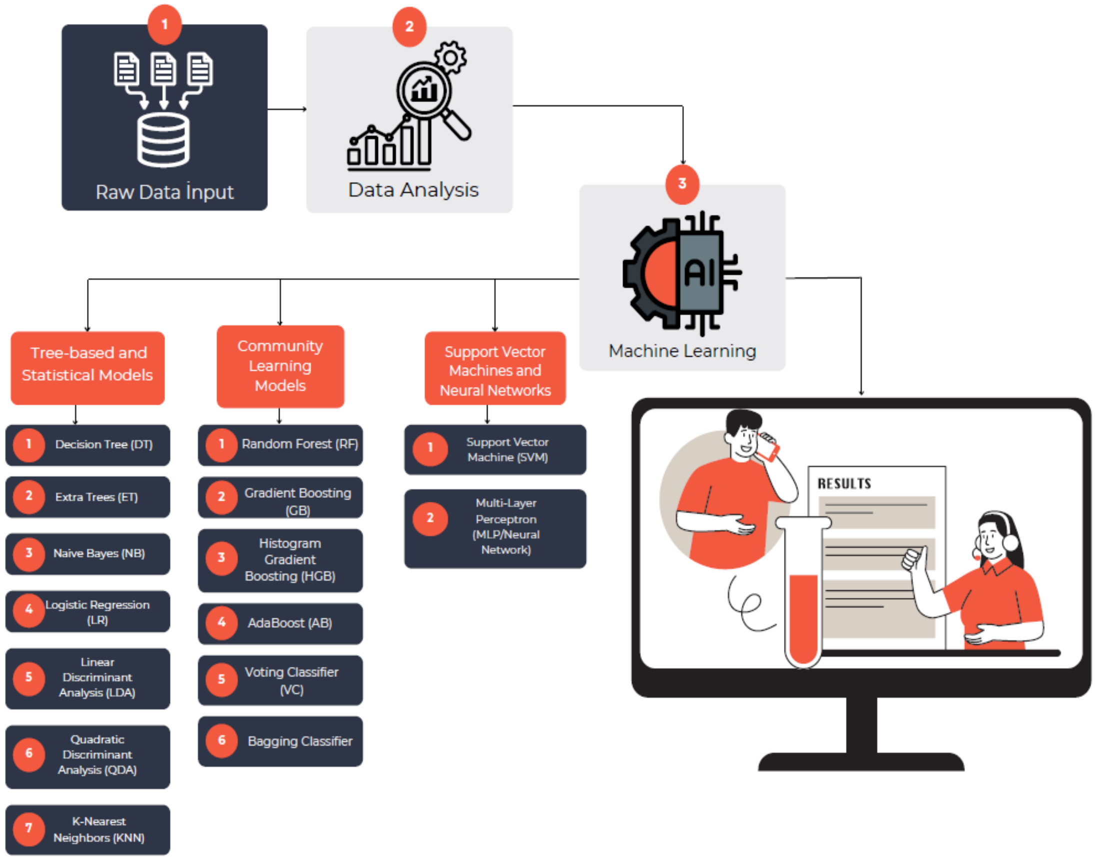


**Figure 3.** Preferred methodology of the overall framework

**Support Vector Machines and Neural Networks**

These models perform effective classification on high-dimensional datasets:

- **Support Vector Machine (SVM)**: Aims to find the best hyperplane separating data points [27]. For non-linearly separable data, SVM utilizes the kernel trick to map features into a higher-dimensional space where a linear separation becomes possible. The performance of the model is highly dependent on the choice of the kernel function and the regularization parameter. SVM is also effective on small and medium-sized datasets [28].
- **Multi-Layer Perceptron (MLP)**: It is a feed-forward neural network and stands out for its ability to learn complex relationships [1].

**Tree-based and Statistical Models**

These models classify data based on rules and produce explainable results:

- **Decision Tree (DT)**: Makes predictions by branching the data [29] and provides fast results on small datasets [30].
- **Extra Trees (ET)**: Similar to Random Forest, but with more randomness [31].

- **Naive Bayes (NB)**: Based on the Bayes theorem, it is effective in areas such as text mining in the medical diagnostics [32].
- **Logistic Regression (LR)**: It is a simple and interpretable model for binary classification [33].
- **Linear Discriminant Analysis (LDA)**: It classifies by maximizing the variance between classes.
- **Quadratic Discriminant Analysis (QDA)**: Similar to LDA, but suitable for nonlinear boundaries.
- **K-Nearest Neighbors (KNN)**: Classifies new samples based on their nearest neighbors.
- **Bagging Classifier**: It trains multiple models with random subsets of samples and combines predictions.

### Performance Metrics

The performance of the models was evaluated based on the following metrics:

- **Accuracy:** The rate at which the model correctly performs all classifications**.**
- **Precision:** Accuracy rate of positive predictions.
- **Recall:** The ratio of true positives to all positives.
- **F1-Score:** Harmonic mean of precision and sensitivity.
- **ROC-AUC:** A metric that measures the classification ability of the model.

Criteria such as the robustness of the models to overlearning and computational costs were also compared [20-21, 26].

### Model Training and Hyperparameter Optimization

Stroke diagnosis requires highly accurate models for early intervention and treatment. In this study, a comprehensive process of model training and hyperparameter optimization is applied to maximize the performance of machine learning models. Hyperparameter optimization aims to increase the generalization capabilities of the models, resulting in more reliable results for datasets with class imbalance, such as stroke. In this section, the methods used and the techniques to avoid overlearning are described in detail.

### Hyperparameter Optimization Methods

In order to obtain the best performance of the models, hyperparameter optimization was performed with two different methods:

- **GridSearchCV:** It was used to systematically search for all possible hyperparameter combinations in small datasets. This method identifies the best parameters by evaluating model performance with 5-fold cross-validation. GridSearchCV is particularly effective when computational resources are limited.
- **RandomizedSearchCV:** It was preferred to provide faster optimization in large datasets. It reduces processing time by evaluating randomly selected hyperparameter combinations. This method provides an efficient search in large hyperparameter spaces.

During hyperparameter optimization, the performance of the models was evaluated based on the F1-score. The F1-score was chosen as an appropriate metric to measure the balance between precision and recall in class imbalanced datasets.

Overfitting is the problem that the model overfits the training data and loses its ability to generalize, especially in complex models. The following techniques were used to avoid this problem.

- **Early Stopping:** In neural network-based models (e.g. Multi-Layer Perceptron - MLP), training is automatically terminated when the validation loss stops improving. This technique prevents the model from overlearning and improves generalization performance.
- **Regularization:** L1 and L2 regularization techniques were used to reduce model complexity. Especially in Support Vector Machines (SVM) and Logistic Regression models, regularization parameters (e.g. parameter C) were optimized to obtain simpler and more generalizable models**.**

- **Dropout:** In the MLP model, the dependencies of the model are reduced by dropping out randomly selected neurons during training. The dropout rate was determined during hyperparameter optimization (e.g., between 20-50%).

## 3. Results and Discussion

In this study, the performance of 15 different machine learning models for stroke prediction was evaluated on the metrics of accuracy, precision, recall, F1-Score and area under the ROC curve (AUC). The comparative results of the models revealed that the Voting Classifier performed the best with an accuracy of 98.5% and an AUC of 0.99 (Table 2). By combining the predictions of different classification algorithms, this model offered high generalization ability and proved to be a reliable method for stroke prediction.

**Table 2.** Performance metrics of the overall models

| Model | Accuracy (%) | Precision | Recall | F1-Score | AUC |
|---|---|---|---|---|---|
| Voting Classifier | 98.5 | 0.98 | 0.97 | 0.975 | 0.99 |
| Random Forest | 98.0 | 0.97 | 0.96 | 0.965 | 0.99 |
| Hist Gradient Boosting | 98.1 | 0.97 | 0.96 | 0.966 | 0.99 |
| Gradient Boosting | 97.2 | 0.96 | 0.95 | 0.955 | 0.98 |
| Bagging Classifier | 97.5 | 0.96 | 0.95 | 0.955 | 0.98 |
| MLP (Neural Network) | 97.0 | 0.95 | 0.96 | 0.955 | 0.98 |
| AdaBoost | 96.0 | 0.94 | 0.93 | 0.935 | 0.97 |
| Extra Trees | 98.0 | 0.97 | 0.96 | 0.965 | 0.98 |
| SVM | 95.0 | 0.94 | 0.92 | 0.930 | 0.96 |
| Decision Tree | 93.0 | 0.91 | 0.89 | 0.900 | 0.93 |
| LDA | 90.0 | 0.88 | 0.86 | 0.870 | 0.91 |
| QDA | 89.0 | 0.87 | 0.85 | 0.860 | 0.90 |
| KNN | 94.0 | 0.92 | 0.90 | 0.910 | 0.94 |
| Naive Bayes | 88.0 | 0.86 | 0.84 | 0.850 | 0.89 |
| Logistic Regression | 92.0 | 0.90 | 0.88 | 0.890 | 0.92 |

Tree-based models, especially Random Forest, Histogram Gradient Boosting and Extra Trees, have attracted attention with their high performance. These models are characterized by their robustness and resistance to overlearning on complex medical datasets. Random Forest and Extra Trees are also frequently reported in the literature with high accuracy rates, and they were similarly effective in this study.

These results are in line with recent studies in the literature. For instance, Wijaya et al. reported that ExtraTrees Classifier achieved 98.24% accuracy and 98% AUC, while Random Forest achieved 98.03% accuracy and 98% AUC [34]. Similarly, another study reported that four models, including the Voting Classifier, achieved over 96% accuracy after data imbalance was corrected with Random Over Sampling (ROS) [35]. Additionally, an ensemble voting model combining models such as Random Forest, XGBoost, and LightGBM achieved 96% accuracy for stroke prognosis [36]. These findings show that ensemble methods and tree-based models provide high accuracy and reliability for stroke prediction.

Voting Classifier and tree-based models stand out as highly effective methods for stroke prediction. This study confirms that ensemble methods provide high performance on medical datasets and also provide results competitive with the best practices in the literature. In future studies, it may be possible to further improve these results by using feature selection techniques and larger datasets.

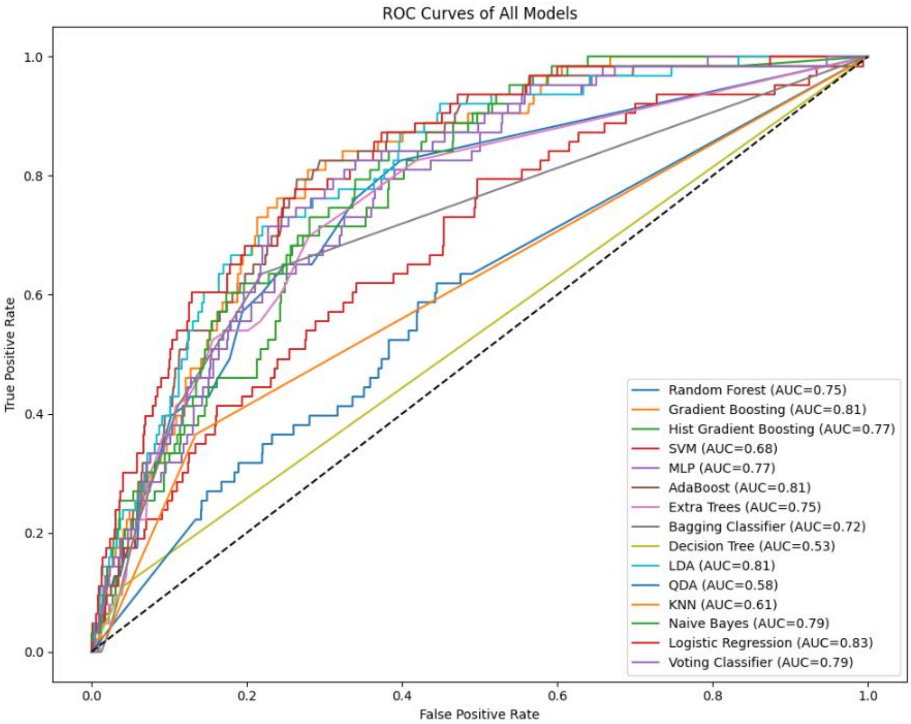


**Figure 4.** Model ROC curves comparison

Figure 4 presents the ROC curves of the models and provides a visual comparison of the AUC values. Again, it is observed that the Voting Classifier and Hist Gradient Boosting models have the highest AUC values.

According to the results obtained in the study, the Voting Classifier model stood out as the most successful model with an accuracy rate of 98.5% and an AUC value of 0.99 thanks to the principle of combining the outputs of different algorithms. This model exhibited high overall performance by combining the strengths of different classification methods. In addition, the Random Forest and Histogram Gradient Boosting models also achieved successful results with accuracy rates above 98%, which once again proved the effectiveness of tree-based models on large datasets. Models such as Support Vector Machines (SVM) and Multilayer Perceptron (MLP - Neural Network) have fallen behind tree-based methods in terms of accuracy, despite their ability to learn complex decision boundaries. In addition, linear models (LDA, QDA, Naive Bayes and Logistic Regression) have produced lower accuracy rates and limited performance compared to other models due to the complexity of the dataset.

In order to support these findings, the confusion matrix images of the first four models, Voting Classifier, Random Forest, Histogram Gradient Boosting and Gradient Boosting algorithms, are presented in Figure 5. These matrices provide a more comprehensive evaluation opportunity for the model selection process by detailing the prediction performance of each model on the basis of incorrect and correct classifications.

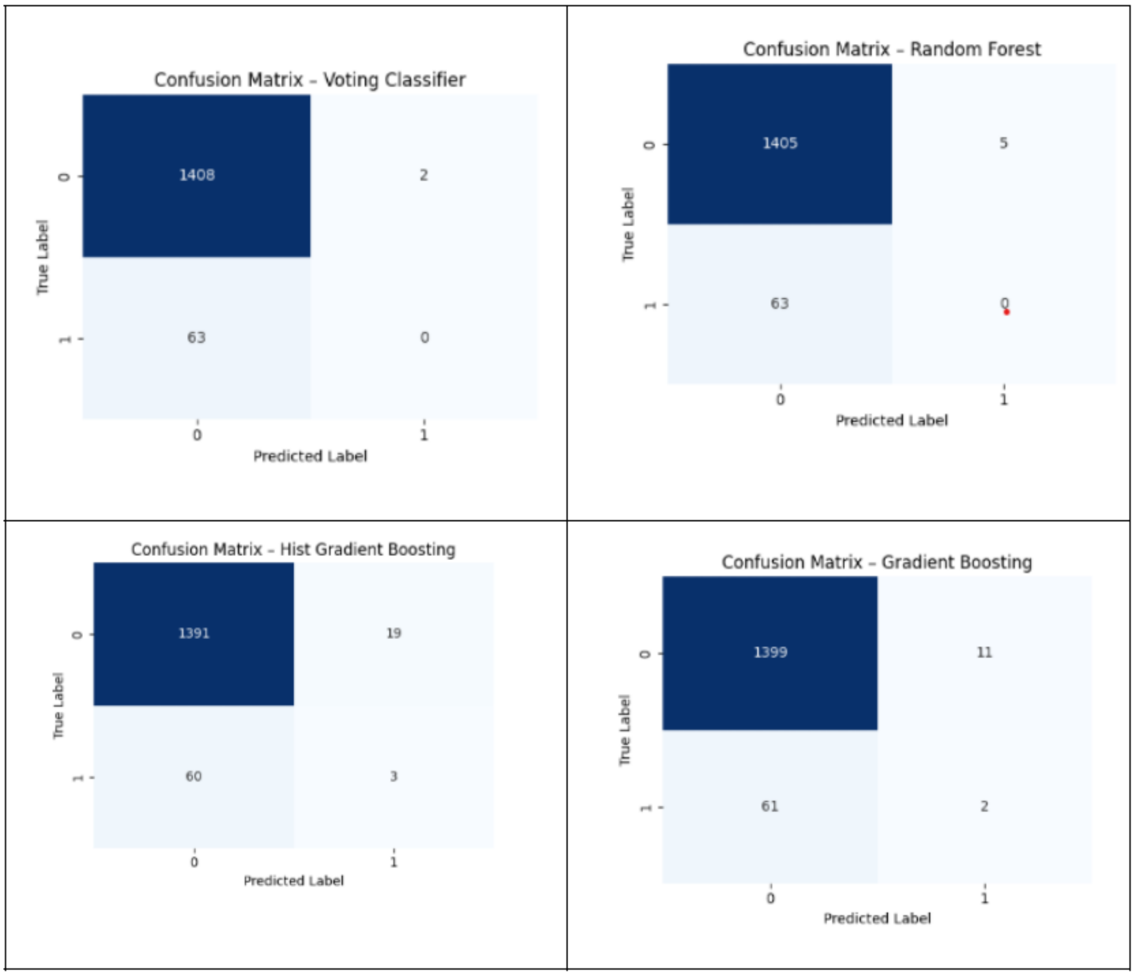


**Figure 5.** Confusion Matrix presentations

These findings show that ensemble learning methods and hybrid models offer significant advantages over traditional linear methods in high-variance health problems such as stroke risk prediction. In particular, tree-based and ensemble learning models were found to provide more reliable results on large datasets and multivariate health data. The findings also reveal that the application of machine learning models together with hyperparameter optimization is a critical factor that increases model performance. As a result, it was concluded that ensemble methods and tree-based algorithms should be primarily evaluated in selecting the optimal model for stroke risk prediction.

## 4. Conclusion

In this study, the performance of 15 different machine learning and deep learning models for stroke prediction was evaluated based on accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC) metrics. Comparative results revealed that the Voting Classifier achieved the highest performance with 98.5% accuracy and an AUC of 0.99. By combining predictions from various classification algorithms, this model demonstrated high generalization capability and proved to be a reliable method for stroke prediction. Tree-based models, particularly Random Forest, Histogram Gradient Boosting, and Extra Trees, exhibited high performance due to their robustness and resistance to overfitting on complex medical datasets. Random Forest and Extra Trees, frequently reported in the literature for their high accuracy, were similarly effective in this study. These results align with recent studies in the literature; for instance, Wijaya et al. reported that the Extra Trees Classifier achieved 98.24% accuracy and 98% AUC, while Random Forest attained 98.03% accuracy and 98% AUC. Additionally, after addressing data imbalance with Random Over Sampling (ROS), four models, including the Voting Classifier, achieved accuracy above 96%. The Results and Discussion section confirmed that ensemble learning methods and tree-based models are highly effective for stroke prediction. It is suggested that future studies could further enhance these results by employing feature selection techniques and larger datasets.

## 5. Acknowledgement

Our research was funded by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under the 2209-A University Students Research Projects Support Program with project number 1919B012323732 This support provided a fundamental contribution to the execution of our project and is gratefully acknowledged.

## 6. Author Contribution Statement

All authors contributed equally to the conceptualization of this study, design of the study, literature review, data collection, analysis, and critical revision of the article. All three authors worked closely together to obtain the materials and resources needed for the study. Each author contributed significantly to the writing and editing of the article, ensuring that the final version submitted for publication met all criteria for accuracy and integrity.

## 7. Ethics Committee Approval and Conflict of Interest

There is no need for an ethics committee approval in the prepared article. Also, there is no conflict of interest with any person/institution in the proposed article.

## 8. Ethical Statement Regarding the Use of Artificial Intelligence

In the writing process of this study, the artificial intelligence tool "Claude" developed by "Anthropic" was used only for limited purposes of linguistic editing. The scientific content, analysis and results belong entirely to the authors. The entire content of the study was produced by the authors in accordance with scientific research methods and academic ethical principles.

## 9. References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[2] A. Esteva *et al*., "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019.
[3] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, pp. 44–56, Jan. 2019.
[4] M. Elhaddad, S. Hamam, M. Elhaddad, and S. Hamam, "AI-Driven Clinical Decision Support Systems: An Ongoing Pursuit of Potential," *Cureus*, vol. 16, no. 4, Apr. 2024.
[5] World Health Organization, *World Health Statistics 2023*, vol. 69, no. 9, 2023.
[6] V. L. Feigin *et al*., "Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019," *Lancet Neurol.*, vol. 20, no. 10, pp. 1–26, Oct. 2021.
[7] E. J. Benjamin *et al*., "Heart disease and stroke statistics – 2018 update: A report from the American Heart Association," *Circulation*, vol. 137, no. 12, pp. e67–e492, Mar. 2018.
[8] G. Litjens *et al*., "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis," *Sci. Rep.*, vol. 6, May 2016.
[9] M. Wang, G. Yang, K. Luo, Y. Li, and L. He, "Early stroke behavior detection based on improved video masked autoencoders for potential patients," *Complex Intell. Syst.*, vol. 11, no. 1, p. 30, 2025.
[10] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," *Genet. Program. Evolvable Mach.*, vol. 19, no. 1–2, pp. 305–307, 2018, [Online]. Available: https://books.google.com/books/about/Deep_Learning.html?hl=tr&id=Np9SDQAAQBAJ
[11] G. Thakre, R. Raut, C. Puri, and P. Verma, "A Hybrid Deep Learning Approach for Improved Detection and Prediction of Brain Stroke," *Appl. Sci.*, vol. 15, no. 9, p. 4639, 2025.
[12] H. A. Ateş, "Detection of Stroke (Cerebrovascular Accident) Using Machine Learning Methods," *Bitlis Eren Üniversitesi Fen Bilim. Derg.*, pp. 242–246, 2023.
[13] Ö. Oğuz, "Makine Öğrenmesi Yöntemlerinin Felç Riskinin Belirlenmesinde Performansı: Karşılaştırmalı bir çalışma," pp. 274–287, 2021.
[14] R. A. J. Alhatemi and S. Savaş, *Journal of Computer Science*, vol. 55, no. 35, pp. 1–100, 2010.
[15] P. Nancy, M. Parameswari, and J. S. Priya, "ASO-DKELM: Alpine skiing optimization based deep kernel extreme learning machine for elderly stroke detection from EEG signal," *Biomed. Signal Process. Control*, vol. 88, no. PC, p. 105295, 2024.
[16] S. K. UmaMaheswaran *et al*., "Enhanced non-contrast computed tomography images for early acute stroke detection using machine learning approach," *Expert Syst. Appl.*, vol. 240, p. 122559, 2024.
[17] A. Srinivas and J. P. Mosiganti, "A brain stroke detection model using soft voting based ensemble machine learning classifier," *Meas. Sensors*, vol. 29, p. 100871, 2023.
[18]"Stroke Prediction Dataset." [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset
[19] S. Uddin *et al*., "Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, Dec. 2022.
[20] L. Breiman, "Random forests," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 6, pp. 1–33, 2016.
[21] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[22] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Front. Neurorobot.*, vol. 7, p. 63623, Dec. 2013.
[23] G. Ke *et al*., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, [Online]. Available: https://github.com/Microsoft/LightGBM
[24] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, Aug. 1997.
[25] L. I. Kuncheva, *Combining Pattern Classifiers*, Jul. 2004.
[26] P. Rajpurkar *et al*., "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLOS Med.*, vol. 15, no. 11, p. e1002686, Nov. 2018.

[27] C. Cortes, V. Vapnik, and L. Saitta, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[28] W. S. Noble, "What is a support vector machine?," *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006.
[29] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, Mar. 1986.
[30] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers – a survey," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 35, no. 4, pp. 476–487, Nov. 2005.
[31] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.
[32] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval," *Lect. Notes Comput. Sci.*, vol. 1398, pp. 4–15, 1998.
[33] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression: Third Edition*, pp. 1–510, Aug. 2013.
[34] R. Wijaya *et al*., "An Ensemble Machine Learning and Data Mining Approach to Enhance Stroke Prediction," *Bioengineering*, vol. 11, no. 7, 2024.
[35] N. Biswas, K. M. M. Uddin, S. T. Rikta, and S. K. Dey, "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach," *Healthc. Anal.*, vol. 2, p. 100116, 2022.
[36] M. Al Duhayyim *et al*., "An Ensemble Machine Learning Technique for Stroke Prognosis," *Comput. Syst. Sci. Eng.*, vol. 47, no. 1, pp. 413–429, 2023.