Karadeniz Fen Bilimleri Dergisi, 15(2), 633-647, 2025. DOI: 10.31466/kfbd.1538305

KFBD

Karadeniz Fen Bilimleri Dergisi
The Black Sea Journal of Sciences
ISSN (Online): 2564-7377 https://dergipark.org.tr/tr/pub/kfbd

Araştırma Makalesi / Research Article

Predicting Stroke Risk with Machine Learning and Hyperparameter Optimization

Burak ÖZKANAT[1], Evin ŞAHİN SADIK[2*]

Abstract

Stroke is a serious medical condition that causes the death of brain cells due to insufficient blood flow due to blockage or rupture in the blood vessels leading to the brain. Stroke is the most common cause of death and disability in adults after heart attack and cancer, causing individuals to not only die but also live with permanent disabilities. In this study, 12 features and 7 different machine learning methods belonging to 5100 individuals in an open-source dataset were used to predict stroke risk. Hyperparameter optimization was applied to increase the performance of machine learning methods and the best parameters were selected. When the results were examined, the random forest algorithm was able to detect the risk of stroke with an accuracy of 96.98%, which is higher than other studies in literature. This study discusses the effective use of machine learning algorithms to predict stroke risk and efforts to improve model performance. The results obtained may help in more accurate determination of stroke risk and taking preventive measures.

Keywords: Classification, Hyperparameter optimization, Stroke, Machine learning.

Makine Öğrenimi ve Hiperparametre Optimizasyonuyla İnme Riskinin Tahmin Edilmesi

Öz

İnme, beyne giden damarlarda meydana gelen tıkanma veya yırtılma sonucu yetersiz kan akışıyla beyin hücrelerinin ölümüne neden olan ciddi bir tıbbi durumdur. İnme, bireylerin hayatını kaybetmesinin yanı sıra kalıcı sakatlıklarla yaşamlarını sürdürmelerine de yol açabilen erişkinlerde kalp krizi ve kanserden sonra en yaygın ölüm ve sakatlık sebebidir. Bu çalışma kapsamında, inme riskini tahmin etmek için açık kaynak bir veri setindeki 5100 bireye ait 12 öznitelik ve 7 farklı makine öğrenmesi yöntemi kullanılmıştır. Makine öğrenmesi yöntemlerinin performansını arttırmak için hiperparametre optimizasyonu uygulanmış ve en iyi parametreler seçilmiştir. Sonuçlar incelendiğinde Rastgele orman algoritması ile literatürdeki diğer çalışmalara göre daha yüksek olan %96,98 oranında bir doğruluk ile inme riski tespiti yapılabilmiştir. Bu çalışmada, inme riskini tahmin etmek için makine öğrenimi algoritmalarının etkin kullanımı ve model performansını iyileştirmeye yönelik çabalar tartışılmaktadır. Elde edilen sonuçlar inme riskinin daha doğru belirlenmesine ve önleyici tedbirlerin alınmasına yardımcı olabilir.

Anahtar Kelimeler: Sınıflandırma, Hiperparametre optimizasyonu, İnme, Makine öğrenimi.

[1]Kütahya Dumlupınar University, Electrical and Electronics Enginnering Department, Kütahya, Turkey, burak.ozkanat0@ogr.dpu.edu.tr
[2]Kütahya Dumlupınar University, Electrical and Electronics Enginnering Department, Kütahya, Turkey, evin.sahin@dpu.edu.tr

[*]Sorumlu Yazar/Corresponding Author

Geliş/Received: 25.08.2024 **Kabul/Accepted:** 25.04.2025 **Yayın/Published:** 15.06.2025

Özkanat, B., Şahin Sadık, E. (2025). Predicting Stroke Risk with Machine Learning and Hyperparameter Optimization. *Karadeniz Fen Bilimleri Dergisi*, 15(2), 633-647.

1. Introduction

Stroke is a life-threatening condition that can cause loss or deterioration of brain functions as a result of sudden interruption of blood circulation to the brain or leakage of blood to an area of the brain (Katan & Luft, 2018). Stroke risk factors include hypertension, high cholesterol, diabetes, obesity, smoking, sedentary lifestyle and family history of stroke (Xia et al., 2019). Stroke remains one of the leading causes of death and disability worldwide, and the risk of stroke is predicted to continue to increase over the next decade and beyond. Approximately 17 million people in the world and 140 thousand people in our country have a stroke every year. According to the current Turkish Statistical Institute (TUIK) report, 33.4% of the 565,594 people who died in 2021 died due to circulatory system diseases, and 18.9% of this rate consisted of deaths due to cerebrovascular diseases (TUİK, 2021). The report published by the European Stroke Organization (ESO) and the European Stroke Alliance for Europe (SAFE) draws attention to the 780 thousand new stroke cases in Europe annually. This number is expected to be around 4 million 630 thousand in 2035 (SAFE, 2019). The rate of recovery and permanent damage after a stroke may vary depending on factors such as the individual's stroke type, severity, effectiveness of early interventions and suitability of the rehabilitation program. After a stroke, some patients may experience permanent disorders such as speech disorders, swallowing disorders, memory and cognitive activity disorders, depression, anxiety, paralysis or movement disorders (Delpont et al., 2018). In order to minimize post-stroke damage, it is necessary to pay attention to factors such as healthy nutrition, keeping diabetes and cholesterol under control, managing stress, keeping blood pressure under control, and not smoking and drinking alcohol (Marsh & Keyrouz, 2010; Pandian et al., 2018).

Nowadays, artificial intelligence techniques have begun to be used in addition to traditional methods in detecting strokes and monitoring the parameters that cause strokes. These methods can help predict the risk of stroke more effectively and enable preventive measures to be taken. Because machine learning algorithms are effective at analyzing large data sets, they can help determine individuals' risk of stroke by combining specific risk factors and clinical data to predict stroke risk, such as hypertension, high cholesterol, diabetes, and personal health data.

In this article, various machine learning algorithms used to predict stroke risk, as well as hyperparameter optimization and studies to improve model performance, are also discussed. These studies may allow a more precise determination of stroke risk and early preventive measures. In the remainder of the study, studies in literature on this subject have been examined, and in the next section, dataset, signal preprocessing, data balancing, data division, classification algorithms, and evaluation metrics are given under the title of material method. The results of the classification

algorithms are shown in the results section. The results were evaluated and discussed in the conclusion and evaluation section.

2. Related Works

In recent years, machine learning (ML) methods have been widely used in disease prediction. To date, various algorithms including Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbors (KNN), Naive Bayes (NB) and Artificial Neural Networks (ANN) have been applied to publicly available stroke datasets for early diagnosis in stroke prediction.

For example, Tazin et al. (Tazin et al., 2021) used RF, DT, Voting Classifier (VC), and LR on a dataset of 5,110 individuals, reporting the highest accuracy of 96% with RF. Sailasya & Kumari (Sailasya & Kumari, 2021) tested six different algorithms (LR, DT, RF, KNN, SVM, NB) on the same dataset and achieved 82% accuracy with NB. Similarly, Imran et al. (Imran et al., 2022) employed DT, RF, and AdaBoost and reached 95% accuracy with both DT and RF

Emon et al. (Emon et al., 2020) implemented models like SGD, GBC, XGB, and Weighted Voting (WV), achieving 97% accuracy with the WV method. Singh et al. (Singh et al., 2022) also reported 95% accuracy using XGB. Additionally, Kansadub et al. (Kansadub et al., 2015) applied DT, NB, and ANN to over 68,000 records, obtaining 75% with DT and 74% with ANN. In their work on predicting modified Rankin Scale (mRS) scores, Monteiro et al. (Monteiro et al., 2018) compared LR, DT, SVM, RF, and XGB algorithms, reporting 93% accuracy with RF. Likewise, Nwosu et al. (Nwosu et al., 2019) evaluated DT, RF, and ANN on a large dataset of 29,072 individuals, reporting 74–75% accuracy with RF and ANN.

In addition to traditional models, some researchers have explored more advanced techniques such as Penalized Logistic Regression, Stochastic Gradient Boosting (SGB), and Decentralized SGD (DSGD), reaching accuracies up to 97% (Arslan et al., 2016; Penafiel et al., 2020). These studies demonstrate that both traditional and ensemble-based models can be effective in stroke classification when properly tuned and trained on representative data.

Building upon the findings of previous studies, this research not only evaluates seven commonly used ML algorithms LR, KNN, NB, SVM, DT, RF and ANN on the same open-source dataset, but also integrates advanced techniques such as hyperparameter optimization and class balancing. Unlike many earlier works that applied default model settings, this study employs GridSearchCV and Keras-Tuner to fine-tune parameters and utilizes SMOTE to mitigate class imbalance. As a result, the RF model achieves 96.98% accuracy, surpassing or matching the highest performance reported in related studies.

These findings reinforce the importance of data preparation and model tuning and demonstrate that even commonly used algorithms can outperform more complex approaches when properly optimized.

3. Materials and Methods

In this section, the study's basic building blocks and methodological approaches are discussed. The general scheme of the study is given in Figure 1. Under this section, details of the data set used are given first. The data preprocessing section explains the processes carried out to prepare the data for analysis, cleaning, standardization and attribute processes. The methods used to eliminate class imbalances in balancing the data are mentioned. In the classification algorithms section, algorithms used for stroke prediction and their features are discussed. The data evaluation section explains cross-validation, hyperparameter optimization, evaluation metrics and steps for comparing algorithms.

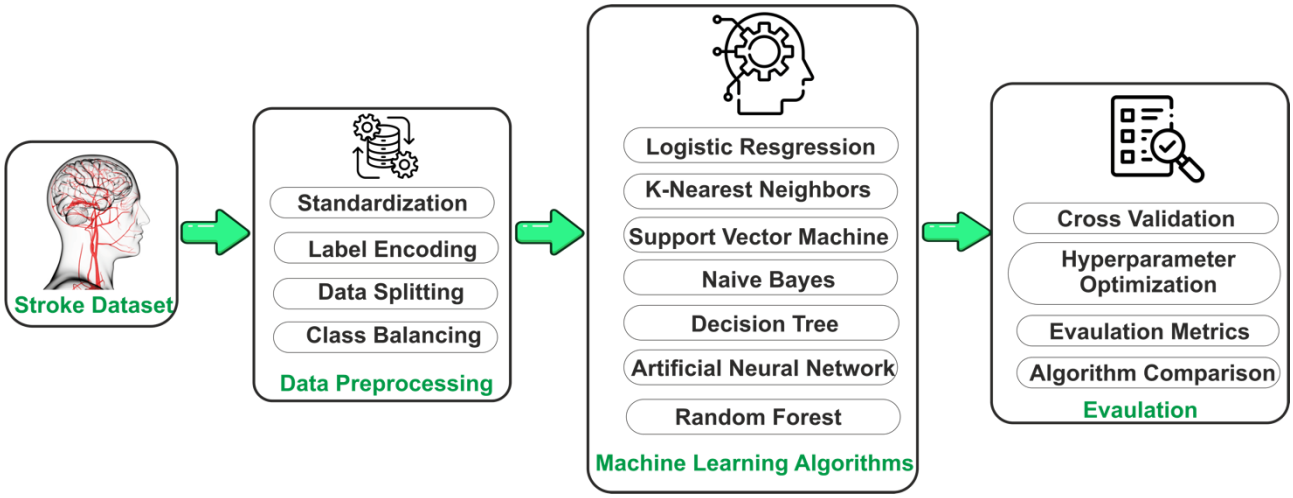

Figure 1. Overview of the stroke prediction workflow using machine learning algorithms.

3.1. Dataset

The dataset used in the study is an open-source stroke dataset obtained from Kaggle (Federico Soriano Palacios, n.d.). There are 12 features of 5100 people in the data set. ID, age, hypertension, heart disease, and stroke status are expressed as integers among these attributes. The other five features, gender, marriage status, type of work, housing type and smoking status, are categorical variables. Finally, the average glucose level and body mass index are decimal numbers.

3.2. Data Preprocessing

The data was subjected to some processes in the signal preprocessing step to make it suitable for machine learning algorithms. Five of the ten attributes in the data set (age, hypertension status, heart disease status, average glucose level and body mass index) contain numerical values. The other five features (gender, marriage status, type of work, housing type and smoking status) consist of categorical variables. For machine learning algorithms to process, categorical variables were converted to numerical values by label encoding. The missing data of 201 people in the body mass index attribute in the data set was filled with the median value. Additionally, one person whose gender was unclear was removed from the dataset.

3.3. Addressing Class Imbalance with SMOTE

The data used in machine learning algorithms training creates the problem of overfitting when the data of one class is more than the other classes when teaching two or multiple classes. In other words, while the model learns the data of one class very well, it ignores the other class or classes. This situation is also present in our data set. In our data set, the number of patients who had a stroke is 249, while the number of individuals who did not have a stroke is 4860. In order to prevent overfitting, synthetic data was generated using the SMOTE method (Chawla et al., 2002). The data of people who had a stroke were synthetically generated to create additional samples, and the dataset was balanced. The original version of the data is shown in Figure 2 a) and the version after the SMOTE process is shown in Figure 2 b).

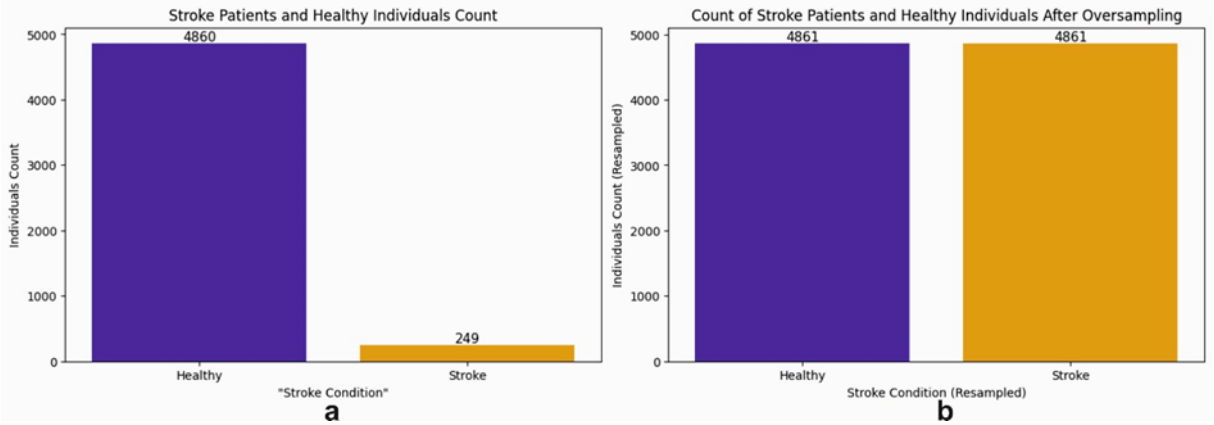

Figure 2. a) Original Data Set b) SMOTE applied dataset.

3.4. Data Splitting

Once you have completed data preprocessing and balanced the unbalanced dataset using oversampling, the next step is to build the model. The data was split into two sets for training the model: 67% for training data and 33% for test data. After the split, various classification algorithms were applied to train the model. The classification algorithms used for this purpose include LR, KNN, SVM, NB, DT, RF, and ANN.

3.5. Classification Algorithms

Within the scope of the study, the stroke dataset was evaluated using seven different machine learning methods, including LR, KNN, SVM, NB, DT, ANN and RF. The working logic of these algorithms is given below, one by one.

Logistic Regression: A statistical model is used to classify two-class (binary) data. It is an algorithm that gives the results as probabilities and classifies them according to a specific threshold value (Hosmer Jr et al., 2013).

K-Nearest Neighbour : Determines the class of an example based on the classes of its k closest neighbours. It selects the most common class of the k nearest neighbours based on the distance criterion (usually Euclidean distance) (Zhang & Zhou, 2007).

Support Vector Machine: Separates classes from each other by creating a hyperplane in the feature space. It aims to provide the best separation by maximizing the margin (distance) between two classes (Joachims, 1998).

Naïve Bayes: It is a probabilistic classification algorithm that assumes that each feature is independent of each other when calculating class probabilities (Rish, 2001).

Decision Tree: A decision tree is a model that is classified by a series of decision rules. It reaches the result by dividing the data set into branches according to certain features' values. It is an algorithm that resembles human thought and decision-making (Quinlan, 1987).

Artificial Neural Network: It contains many processing units (neurons) arranged in layers inspired by the functionality of neurons in the human brain. Can model complex patterns and data relationships (Agatonovic-Kustrin & Beresford, 2000).

Random Forest: It is an ensemble learning model with many decision trees. Each tree is trained with its own samples, feature subsets, and votes for classification (Pal, 2005).

3.6. Hyperparameter Optimization

Hyperparameter optimization is the selection of the best parameters that will increase performance in machine learning models. There are different hyperparameter optimization methods. In this study, the GridSearchCV method (Shamrat et al., 2020) in the Scikit-learn library was applied to machine learning methods. Using this method, all possible hyperparameter combinations were investigated and the parameters that provided the best model performance were determined. For the artificial neural network, optimum hyperparameters were determined using Keras-tuner (O'Malley et al., 2019). This process helped us fine-tune the model and achieve the best possible performance by adjusting hyperparameters such as the number of layers, neurons, and learning rate.

3.7. Classification Metrics

Within the scope of the study, the data were trained with 6 different machine learning algorithms and artificial neural networks and these trained models were compared using seven accuracy metrics. These metrics are: confusion matrix, accuracy, precision, recall, F1 score, ROC curve and Area Under the Curve (AUC) value. The equations for accuracy, precision, recall and F1 score metrics are given in Equation 1, Equation 2, Equations 3 and 4, respectively. The TP expression given in the equations represents true positive data, TN expression represents true negative data, FP expression represents false positive data and FN expression represents false negative data.

$$Accuracy = \frac{TP + TN}{TN + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Re}call = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision+Recall}}{\text{Precision} \times \text{Recall}} \tag{4}$$

4. Findings and Discussion

In this section, the performance of all machine learning models applied to the stroke dataset is evaluated based on multiple classification metrics, including accuracy, precision, recall, F1-score, and AUC. The analysis is supported with visual representations such as confusion matrices and ROC curves, followed by a comparative discussion of model results and their alignment with previous studies. Table 1 summarizes the performance metrics and optimal hyperparameters for each algorithm. As observed, the RF algorithm achieved the highest accuracy 96.98% and AUC 0.993, followed closely by KNN and DT. On the other hand, NB and LR showed relatively lower performance, particularly in accuracy and AUC, yet they remain valuable due to their simplicity and interpretability. These results form the basis for the detailed model-by-model analysis, confusion matrix interpretation, and ROC curve discussion presented in the subsequent sections.

Table 1. Performance evaluation metrics of all algorithms.

Algorithms	***Hyperparameter***	***Accuracy***	***Precision***	***Recall***	***F1-Score***	***AUC***
NB	model: BernoulliNB	0.7828	0.7519	0.7985	0.7745	0.866
LR	C: 0.001, penalty: l2, solver: lbfgs	0.7946	0.7626	0.8120	0.7865	0.873
ANN	layers: [1024, 512, 1536], learning_rate: 0.001, epochs: 10	0.9156	0.9234	0.9080	0.9156	0.976
SVM	C: 100, kernel: rbf	0.9299	0.9372	0.9227	0.9299	0.981
DT	criterion: gini, max_depth: None, min_samples_leaf: 2, min_samples_split: 5	0.9386	0.9416	0.9351	0.9383	0.951
KNN	metric: Manhattan, n_neighbors: 1	0.9430	0.9246	0.9590	0.9415	0.942
RF	**criterion: gini, max_depth: None, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 200**	**0.9698**	**0.9837**	**0.9835**	**0.9700**	**0.993**

The performance differences between models can be partially attributed to the hyperparameter optimization process applied using GridSearchCV and Keras-Tuner. For example, the Random Forest model achieved its best results with 200 estimators, ‘gini’ criterion, and default maximum depth, which allowed the model to grow deeper trees and capture complex patterns. Similarly, the ANN model performed well with a deep architecture consisting of three layers ([1024, 512, 1536] units) and a learning rate of 0.001, which balanced convergence speed and accuracy. On the other hand, SVM’s strong performance (AUC = 0.981) can be linked to its high ‘C’ value (100) and RBF kernel, which are known to enhance class separation in non-linear spaces. These results show that proper hyperparameter tuning significantly enhances model accuracy and generalization capability.

When the training and test data of the study are evaluated and the confusion matrices given in Figure 3 are examined, it is possible to see that the matrices are listed from least successful to most successful. NB algorithm results in Figure 3 a) LR in Figure 3 b) SVM in Figure 3 c) DT in Figure 3

d) KNN in Figure 3 e) RF in Figure 3 f) Confusion matrices of the algorithms are given. These matrices clearly present the success level of each classification algorithm and show in detail the number of correct and incorrect classifications of each algorithm. Confusion matrices provide a valuable guide to understanding which classes algorithms predict better or worse. In this way, it can be understood which algorithm requires more improvement in which classes.

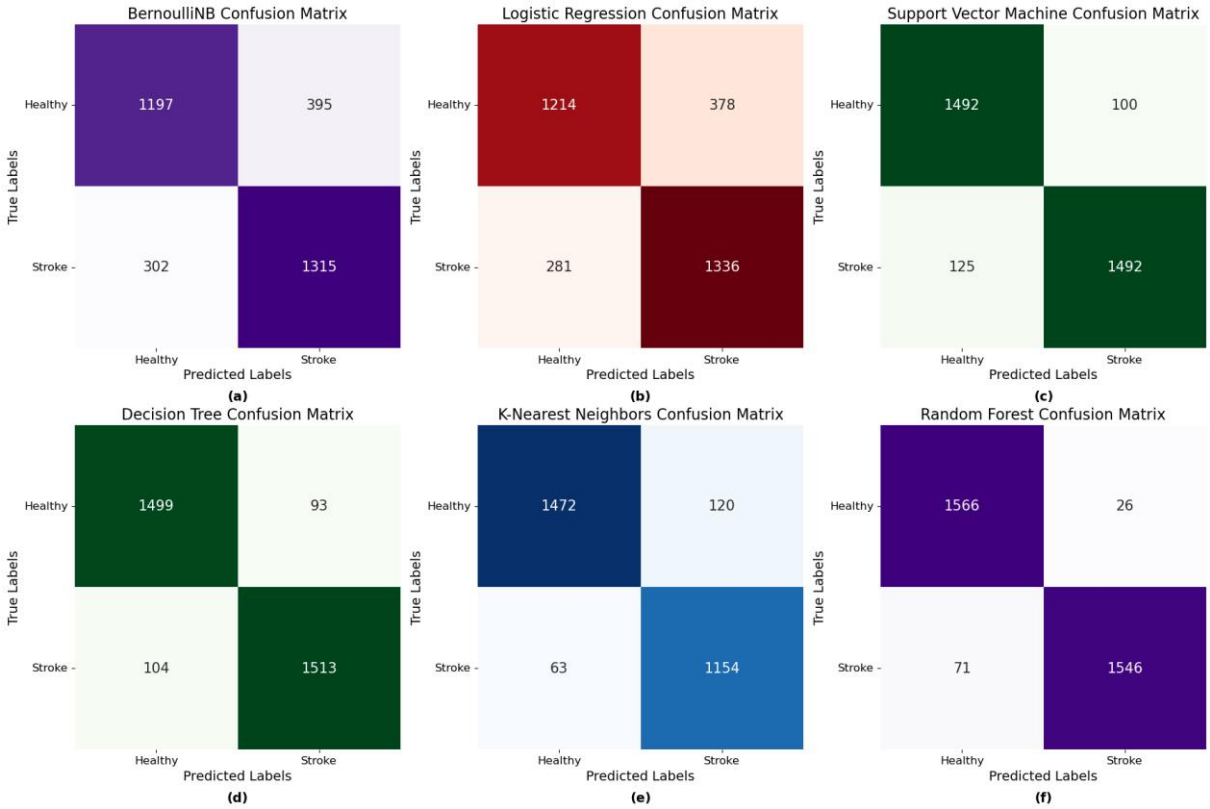

Figure 3. Confusion matrices for each machine learning a) NB b) LR c) SVM d) DT e) KNN f) RF.

The confusion matrix of the data evaluated in the artificial neural network algorithm is given in Figure 4. This matrix visualizes the classification performance of the ANN algorithm in detail. Each cell contains actual and predicted class labels, indicating the number of correct and incorrect algorithm classifications.

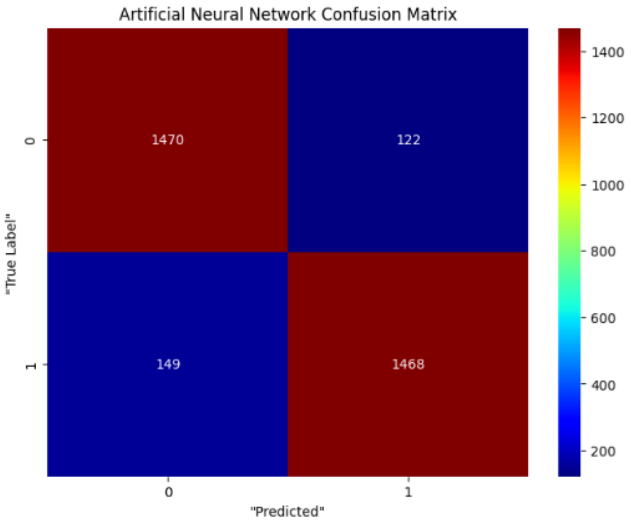

Figure 4. Confusion Matrix of ANN.

These quantitative results are further supported by confusion matrices and ROC curves, which visually illustrate the classification capabilities of each model. The ROC curve showing the classification results is given in Figure 5. The RF algorithm has the highest accuracy and AUC values and stands out in classification performance. This can be achieved thanks to the RF algorithm's ability to capture complex structures and interactions in the data set. Precision and recall values are also quite high, indicating that the algorithm's ability to identify and predict the positive class accurately is strong. F1-Score is the harmonic mean of precision and recall values, and a high F1-Score indicates that the model has a balanced performance in both metrics.

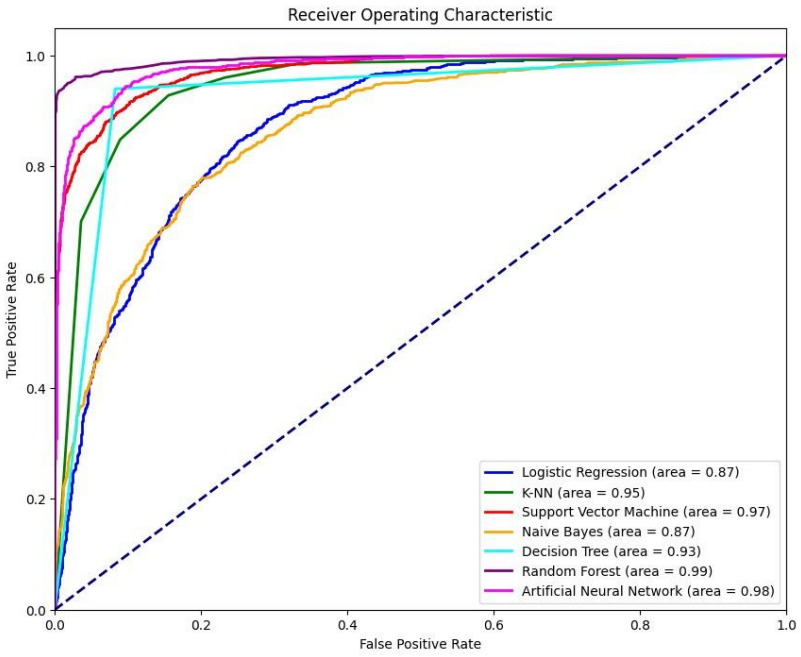

Figure 5. ROC curves of all machine learning algorithms used in the study.

When other algorithms are examined, it is seen that ANN and KNN algorithms exhibit high performance, especially the recall value of the KNN algorithm exhibits high performance. This shows that KNN effectively recognizes examples that need to be classified as positive without missing them. The decision tree and SVM algorithm also attract attention with its high AUC value and show that the classification ability of the model is strong. Although simpler algorithms such as NB and LR appear to perform relatively poorly in accuracy and other metrics, these methods can be advantageous in terms of speed and interpretability.

To evaluate the effectiveness of the proposed study, its results were compared with those reported in the literature using similar or identical stroke datasets. Table 2 presents a summary of machine learning algorithms used in previous studies and their corresponding highest accuracy rates. These studies vary in terms of dataset composition and classification methods, including the Kaggle stroke dataset also used in this research. By analyzing Table 2, the relative success of different approaches can be observed, and it becomes evident that the proposed model outperforms others in terms of classification accuracy.

Table 2. Comparison of stroke classification results.

Reference	Method	Dataset	Highest Accuracy (%)
Sergio Penafiel et al., (Penafiel et al., 2020)	DSGD, RF, MLP, SVM, K-NN, NB	Original Dataset non-stroke: 22140, stroke: 5736	DSGD:85, RF:84, MLP:82, SVM:82, K-NN:79, NB:61
Tahia Tazin et al., (Tazin et al., 2021)	RF, DT, VC, LR	Stroke Prediction Dataset (Kaggle) stroke: 249 non-stroke: 4861	RF:96, DT:94, VC:91, LR:79
Gangavarapu Sailasya et al., (Sailasya & Kumari, 2021)	LR, DT, RF, K-NN, SVM, NB	Stroke Prediction Dataset (Kaggle) stroke: 249 non-stroke: 4861	LR:78 DT:66, RF:73, K-NN:80, SVM:80, NB:82
Bahtiar Imran et al., (Imran et al., 2022)	DT, RF, AdaBoost	Stroke Prediction Dataset (Kaggle) stroke: 249 non-stroke: 4861	DT:95, RF: 95, AdaBoost: 91.7
Utkrisht Singh et al., (Singh et al., 2022)	XGB, RF, SVM, AdaBoost, DT	Stroke Prediction Dataset (Kaggle) stroke: 249 non-stroke: 4861	XGB:95, RF:92, SVM:91, AdaBoost:89, DT:87
Ting Zuo et al., (Zuo et al., 2024)	Dueling DQN	Stroke screening data from a hospital in China	Dueling DQN: 89
Muhammad Raihan Firmansyah et al., (Firmansyah & Astuti, 2024)	KNN	Stroke Prediction Dataset (Kaggle)	KNN: 95

Alomoush et al., (Alomoush et al., 2024)	Modified Mountain Gazelle Optimizer (mMGO) + kNN	Stroke Prediction Dataset (Kaggle)	mMGO + kNN : 95
This study	**NB, LR, ANN, KNN, SVM, DT, RF**	Stroke Prediction Dataset (Kaggle) stroke: 249 non-stroke: 4861	**NB:78, LR:80, ANN:92, KNN:94, SVM:92, DT:93, RF:96.8**

As seen in Table 2, the RF model in this study achieved an accuracy of 96.98%, which surpasses the results of comparable studies such as Tazin et al. (2021), who reported 96%, and Singh et al. (2022), who achieved 95% with XGBoost. While the differences may appear small numerically, they are significant in medical classification tasks where even marginal improvements can enhance early detection and prevention.

This performance improvement is largely related to the hyperparameter optimization strategies implemented in this study. In particular, using GridSearchCV for tree-based models and Keras-Tuner for ANN enabled more effective learning by adapting the models to the dataset features. Using SMOTE helped address the class imbalance issue in the stroke dataset and improved the generalization ability of the models.

The ANN model in this study outperformed similar models in previous studies such as Kansadub et al. (Kansadub et al., 2015), showing an accuracy of 91.56% compared to their accuracy of 74%. This improvement is likely due to the optimized learning parameters. Overall, the findings show that the performance of traditional algorithms can be significantly improved through model design, parameter tuning, and data preprocessing strategies.

5. Conclusions and Recommendations

This work explored that the stroke risk prediction can be significantly improved through the application of ML algorithms combined with hyperparameter optimization. Among the models evaluated, the RF algorithm outperformed others with the highest accuracy (96.98%) and AUC (0.993), indicating its strong capability to model complex relationships within the data. The success of RF can be largely attributed to hyperparameter fine-tuning, which enhanced its ability to generalize and reduced the risk of overfitting.

Hyperparameter optimization played a crucial role across all models tested. Classical algorithms like LR and NB, although initially less accurate, showed competitive performance after proper tuning, suggesting that model selection alone is not sufficient how the model is trained matters

equally. GridSearchCV and Keras-Tuner allowed the identification of optimal configurations for each algorithm, resulting in noticeable improvements in classification metrics, particularly for RF, KNN, and ANN.

One of the key strengths of this study is the integration of preprocessing strategies such as class balancing using SMOTE, which contributed to more stable and fair evaluations. Furthermore, unlike many previous studies that relied on default hyperparameters, this research emphasized systematic tuning, which appears to be a decisive factor in achieving state-of-the-art results.

Despite promising outcomes, this study has certain limitations. The dataset used is imbalanced and relatively small in stroke-positive samples, even after SMOTE adjustment. In real-world clinical environments, datasets may contain noise, missing values, or more complex patterns, which could affect model performance. Therefore, it is recommended that future studies validate these results using larger, more diverse datasets possibly from hospital records or multi-institutional databases.

Additionally, while the RF model performed best, it is computationally more expensive than simpler models. In time-critical medical applications, lighter models like LR or NB with acceptable accuracy and faster inference times might be more appropriate. This highlights the importance of aligning model selection with specific application requirements.

In future studies, researchers can focus on ensemble or hybrid models that combine the strengths of multiple classifiers to improve model performance. They can also use more advanced hyperparameter optimization techniques such as Bayesian optimization or genetic algorithms to further improve model accuracy. The use of deep learning architectures such as Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) networks can also be useful, especially when working with temporal or image-based stroke indicators.

In conclusion, this study confirms that both traditional and advanced ML models can effectively predict stroke risk when appropriately tuned and supported by data preprocessing strategies. The findings offer a valuable foundation for developing real-time, ML-based stroke screening systems in healthcare settings.

Authors' Contributions

All authors contributed equally to the study.

Statement of Conflicts of Interest

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The author declares that this study complies with Research and Publication Ethics.

References

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717–727.

Alomoush, W., Houssein, E. H., Alrosan, A., Abd-Alrazaq, A., Alweshah, M., & Alshinwan, M. (2024). Joint opposite selection enhanced Mountain Gazelle Optimizer for brain stroke classification. *Evolutionary Intelligence*, *17*(4), 2865–2883.

Arslan, A. K., Colak, C., & Sarihan, M. E. (2016). Different medical data mining approaches based prediction of ischemic stroke. *Computer Methods and Programs in Biomedicine*, *130*, 87–92.

Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Delpont, B., Blanc, C., Osseby, G. V, Hervieu-Bègue, M., Giroud, M., & Béjot, Y. (2018). Pain after stroke: a review. *Revue Neurologique*, *174*(10), 671–674.

Emon, M. U., Keya, M. S., Meghla, T. I., Rahman, M. M., Al Mamun, M. S., & Kaiser, M. S. (2020). Performance analysis of machine learning approaches in stroke prediction. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 1464–1469.

Federico Soriano Palacios. (n.d.). *Stroke Prediction Dataset*. https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset?resource=download

Firmansyah, M. R., & Astuti, Y. P. (2024). Stroke classification comparison with KNN through standardization and normalization techniques. *Advance Sustainable Science, Engineering and Technology*, *6*(1), 2401012.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Imran, B., Wahyudi, E., Subki, A., Salman, S., & Yani, A. (2022). Classification of stroke patients using data mining with adaboost, decision tree and random forest models. *ILKOM Jurnal Ilmiah*, *14*(3), 218–228.

Joachims, T. (1998). *Making large-scale SVM learning practical*. Technical report.

Kansadub, T., Thammaboosadee, S., Kiattisin, S., & Jalayondeja, C. (2015). Stroke risk prediction model based on demographic data. *2015 8th Biomedical Engineering International Conference (BMEiCON)*, 1–3.

Katan, M., & Luft, A. (2018). Global burden of stroke. *Seminars in Neurology*, *38*(02), 208–211.

Marsh, J. D., & Keyrouz, S. G. (2010). Stroke prevention and treatment. *Journal of the American College of Cardiology*, *56*(9), 683–691.

Monteiro, M., Fonseca, A. C., Freitas, A. T., e Melo, T. P., Francisco, A. P., Ferro, J. M., & Oliveira, A. L. (2018). Using machine learning to improve the prediction of functional outcome in ischemic stroke patients. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(6), 1953–1959.

Nwosu, C. S., Dev, S., Bhardwaj, P., Veeravalli, B., & John, D. (2019). Predicting stroke from electronic health records. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5704–5707.

O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., & Invernizzi, L. (2019). Keras tuner. *Retrieved May*, *21*, 2020.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222.

Pandian, J. D., Gall, S. L., Kate, M. P., Silva, G. S., Akinyemi, R. O., Ovbiagele, B. I., Lavados, P. M., Gandhi, D. B. C., & Thrift, A. G. (2018). Prevention of stroke: a global perspective. *The Lancet*, *392*(10154), 1269–1278.

Penafiel, S., Baloian, N., Sanson, H., & Pino, J. A. (2020). Predicting stroke risk with an interpretable classifier. *IEEE Access*, *9*, 1154–1166.

Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, *27*(3), 221–

234.
Rish, I. (2001). An empirical study of the naive Bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, *3*(22), 41–46.
SAFE. (2019). *Avrupa İnme Faaliyet Planı Raporu*. https://actionplan.eso-stroke.org/wp-content/uploads/2021/03/SAP-Turkish-s.pdf
Sailasya, G., & Kumari, G. L. A. (2021). Analyzing the performance of stroke prediction using ML classification algorithms. *International Journal of Advanced Computer Science and Applications*, *12*(6).
Shamrat, F. M. J. M., Ghosh, P., Sadek, M. H., Kazi, M. A., & Shultana, S. (2020). Implementation of machine learning algorithms to detect the prognosis rate of kidney disease. *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 1–7.
Singh, U., Jena, A. K., & Haque, M. T. (2022). An Ensemble Learning Approach and Analysis for Stroke Prediction Dataset. *2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, 1–8.
Tazin, T., Alam, M. N., Dola, N. N., Bari, M. S., Bourouis, S., & Monirujjaman Khan, M. (2021). Stroke disease detection and prediction using robust learning approaches. *Journal of Healthcare Engineering*, *2021*.
TUİK. (2021). *Ölüm ve Ölüm Nedeni İstatistikleri*.
Xia, X., Yue, W., Chao, B., Li, M., Cao, L., Wang, L., Shen, Y., & Li, X. (2019). Prevalence and risk factors of stroke in the elderly in Northern China: data from the National Stroke Screening Survey. *Journal of Neurology*, *266*, 1449–1458.
Zhang, M.-L., & Zhou, Z.-H. (2007). ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recognition*, *40*(7), 2038–2048.
Zuo, T., Li, F., Zhang, X., Hu, F., Huang, L., & Jia, W. (2024). Stroke classification based on deep reinforcement learning over stroke screening imbalanced data. *Computers and Electrical Engineering*, *114*, 109069.