Annals of Healthcare Systems Engineering

www.ahse.reapress.com

Ann. Healthc. Syst. Eng. Vol. 3, No. 3 (2026) 204–215.

Paper Type: Original Article

A Machine Learning Models for Early Stroke Prediction

Sahel Shiravand[1], Abdolreza Fathi[1,*]

[1] Department of Computer Engineering, Razi University, Kermanshah, Iran; s.shiravand@stu.razi.ac.ir; a.fathi@razi.ac.ir.

Citation:

Received: 26 April 2025 *Revised: 19 July 2025* *Accepted: 10 December 2025*	Shiravand, S., & Fathi, A. (2026). A machine learning models for early stroke prediction. *Annals of Healthcare Systems Engineering*, *3*(3), 204-215.

Abstract

Stroke is a leading cause of mortality and long-term disability, and early prediction using machine learning models can support clinical decision-making. In this study, a hybrid, stable, and interpretable feature selection framework is employed together with the LightGBM algorithm for stroke prediction. First, SHAP values derived from an XGBoost model are used to quantify the importance of each feature. Then, the absolute correlation of each feature with the target variable stroke and its stability across cross-validation folds (i.e., the number of times it appears among the top 10 features) are considered as complementary criteria. These three measures are combined into a weighted composite score, and the top 10 features are selected for model training. The results show that, although some previous studies report higher overall accuracy, the proposed model achieves a Recall of 100% in identifying stroke patients, along with a Precision of 97.12% and an appropriate F1-score, providing a favorable balance between reducing missed cases and limiting false alarms.

Keywords: Stroke prediction, Machine learning, Feature selection, LightGBM, SHAP.

1 | Introduction

1.1 | Clinical and Societal Impact of Stroke

Stroke is a neurological condition caused by a disruption of blood supply to the brain, leading to rapid cell death and potential long-term disability. The Global Burden of Disease study reports over 7.25 million stroke-related deaths in 2021, with ischemic stroke contributing the largest share. Stroke affects 12 million people annually, and the lifetime risk is one in four individuals over the age of 25. This condition also incurs substantial economic costs, estimated at over US $890 billion, representing 0.66% of the global GDP [1], [2].

A brain clot or ruptured blood artery in the brain interrupts blood flow to a portion of the brain, resulting in a stroke, also known as a brain attack. Certain brain regions suffer harm or even die in both situations [2].

Corresponding Author: a.fathi@razi.ac.ir

https://doi.org/10.22105/ahse.v3i3.71

1.2 | Machine Learning, Data Challenges, and the Role of Feature Selection

Machine learning algorithms have shown promise in revolutionizing stroke prediction

by analyzing extensive datasets encompassing demographic information, medical histories, and physiological markers like age, blood pressure, and glucose levels [3], [4].

These models can capture complex, non-linear relationships that traditional statistical approaches may miss. However, stroke datasets are typically highly imbalanced, with far fewer stroke cases than non-stroke cases, making classical metrics such as overall accuracy insufficient to evaluate model performance [5], [6]. In addition, the presence of many potentially redundant or correlated variables introduces the risk of overfitting and reduces model interpretability. Under these conditions, robust feature selection becomes crucial: it can improve predictive performance, enhance generalization, and highlight the most influential factors contributing to stroke, which is important for clinical acceptance. At the same time, evaluation must emphasize sensitivity-oriented metrics, such as Recall, to ensure that true stroke cases are not overlooked [7], [8].

1.3 | Motivation, Research Gap, and Contribution of This Study

Despite the growing use of machine learning in stroke prediction, important gaps remain. Many existing studies rely on a single feature selection method or focus primarily on model accuracy without thoroughly analyzing feature importance or the stability of selected features across different validation folds. Moreover, the trade-off between Recall and Precision, which is crucial in high-risk medical applications, is often not explicitly addressed.

To address these limitations, this study proposes a hybrid, stable, and interpretable feature selection framework tailored to stroke prediction. First, an XGBoost model is employed to compute SHAP values and quantify the contribution of each feature. These values are combined with the absolute correlation of each feature with the target variable stroke and with a stability indicator that counts how frequently each feature appears among the top predictors across cross-validation folds. A weighted composite score is then defined to select the top 10 features. Based on this compact and informative feature set, a LightGBM classifier is trained and evaluated. The results demonstrate that the proposed framework achieves 100% Recall in identifying stroke patients while maintaining high Precision (97.12%) and a suitable F1-score, providing a favorable balance between minimizing missed stroke cases and controlling false alarms.

2 | Data Preprocessing

2.1 | Dataset Preparation and Initial Cleaning

The experiments were conducted on the Healthcare Stroke Dataset. After loading the dataset, the following preprocessing steps were applied:

I. The identifier column id was removed because it does not carry any analytical or predictive information.

II. Records with gender = 'Other' were removed to ensure a homogeneous definition of the gender variable.

After preprocessing, the final dataset included:

$$n = 5110$$

samples.

2.2 | Missing Value Imputation for BMI

The BMI attribute contained missing values. Instead of discarding these records, BMI was imputed using the median BMI within subgroups defined by gender and work_type.

Let BMI_i denote the BMI value of record i, which belongs to a subgroup with gender $g$and work type w. Define:

I. The index set of records in the same subgroup with observed BMI:

$S_{g,w}=\{j|gender_j=g, work\_type_j=w, BMI_j \text{ is observed}\}$.

II. The group-wise median BMI for this subgroup in *Eq. (1)*:

$$\widetilde{BMI}_{g,w} = \text{median}(\{BMI_j \mid j \in S_{g,w}\}). \quad \textbf{(1)}$$

Then missing BMI values are imputed according to:

BMIi={BMIi if BMI is observed,\[4pt]BMI~g,w,if BMI is missing.

This conditional median imputation leverages demographic and occupational structure to provide robust estimates for BMI while preserving all records.

2.3 | Categorical Variable Encoding

Several predictors in the dataset are categorical and cannot be directly used by tree-based or gradient-boosting models in raw form. Therefore, the following categorical variables were transformed using One-Hot Encoding:

- *gender*
- *ever_married*
- *work_type*
- *Residence_type*
- *smoking_status*

For a generic categorical variable X with categories $\{c_1, c_2, \dots, c_K\}$, One-Hot Encoding creates K binary variables $X_1, \dots, X_K$ such that in *Eq. (2)*:

$$X_k = \begin{cases} 1, & \text{if } X = c_k, \\ 0, & \text{otherwise,} \end{cases} \quad k = 1, \dots, K. \quad \textbf{(2)}$$

This transformation yields a purely numerical feature space suitable for the subsequent modeling and feature selection steps.

2.4 | Handling Class Imbalance

The target variable stroke is highly imbalanced, with the stroke class (positive cases) being much smaller than the non-stroke class (negative cases). To mitigate this issue and avoid bias toward the majority class, an oversampling strategy was applied to increase the number of minority-class samples.

Let N_{major} and N_{minor} denote the number of samples in the majority and minority classes, respectively. Oversampling was used to obtain a new minority class size $N_{minor}^{(new)}$ such that in *Eq. (3)*:

$$N_{minor}^{(new)} \approx N_{major}. \quad \textbf{(3)}$$

The focus of this step is to improve the model's ability to correctly identify stroke cases by providing a more balanced training distribution.

2.5 | Cross-Validation Strategy

Model evaluation and feature scoring were performed using Stratified 5-Fold Cross Validation. The full dataset D was split into five approximately equal folds in *Eq (4)*:

$$D = D_1 \cup D_2 \cup \cdots \cup D_5, \quad (4)$$

such that the class distribution in each fold D_k is similar to the overall distribution (stratification). For the k-th fold ($k = 1, \ldots, 5$) in *Eq. (5)*:

$$D_{train}^{(k)} = D \setminus D_k, D_{test}^{(k)} = D_k. \quad (5)$$

For each fold, a model (here, XGBoost for feature scoring) was trained on $D_{train}^{(k)}$ and evaluated on $D_{test}^{(k)}$. The stratified design ensures that performance estimates and feature scores are robust across different subsets of the data, especially under class imbalance.

2.6 | Feature Selection Framework

A hybrid feature selection framework was used to identify the most informative predictors for stroke. The final feature score for each variable combines three complementary criteria:

I. SHAP-based importance (from XGBoost).

II. Absolute correlation with the target stroke.

III. Stability of the feature across CV folds.

The goal is to select the top 10 features that are simultaneously important, predictive, and stable.

2.6.1 | SHAP-based importance

For each fold k, an XGBoost model was trained on the corresponding training set $D_{train}^{(k)}$. Using SHAP TreeExplainer, SHAP values were computed for all features [9], [10].

Let $\phi_{i,f}^{(k)}$ denote the SHAP value of feature f for sample i in fold k. The fold-wise SHAP importance of feature f in fold k is defined as the mean absolute SHAP value in *Eq. (6)*:

$$SHAP\_imp_f^{(k)} = \frac{1}{n_k} \sum_{i \in D_{train}^{(k)}} | \phi_{i,f}^{(k)} |, \quad (6)$$

where $n_k = | D_{train}^{(k)} |$.

The overall SHAP score for feature f is then obtained by averaging over all 5 folds in *Eq. (7)*:

$$SHAP\_Score_f = \frac{1}{5} \sum_{k=1}^{5} SHAP\_imp_f^{(k)}. \quad (7)$$

This score reflects how strongly feature f contributes, on average, to the model's predictions across different training splits.

Correlation with the target

For each fold k, the Pearson correlation coefficient between each feature f and the binary target stroke was computed. Let $X_f^{(k)}$ be the vector of values of feature f and $Y^{(k)}$ be the corresponding target vector in fold k. The Pearson correlation is [7] in *Eq. (8)*:

$$r_f^{(k)} = \mathrm{corr}(X_f^{(k)}, Y^{(k)}). \quad (8)$$

To obtain a fold-invariant correlation measure, the absolute correlation coefficients were averaged over the 5 folds in *Eq. (9)*:

$$\mathrm{Corr\_Score}_f = \frac{1}{5}\sum_{k=1}^{5} | r_f^{(k)} |. \quad (9)$$

This score captures the linear association between each feature and the stroke outcome, aggregated across cross-validation splits.

To better understand the relationships among numeric variables and to assess potential multicollinearity, we computed the Pearson correlation matrix for age, hypertension, heart disease, average glucose level, BMI, and the stroke label.

As shown in *Fig. 1*, none of the pairwise correlations exceeds 0.4, indicating that strong linear dependencies are absent in this subset of features. BMI shows the highest correlation with age, while all predictors exhibit only mild correlations with the stroke outcome. This pattern suggests that stroke risk is driven by the joint interaction of several factors rather than by a single dominant variable.

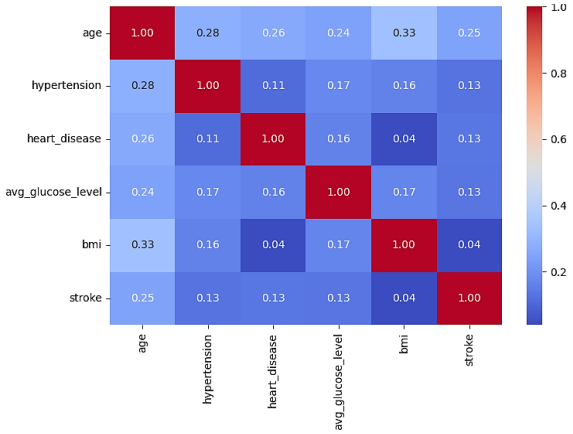

Fig. 1. Pearson correlation matrix for numeric features (age, hypertension, heart disease, average glucose level, BMI, and stroke).

Stability across folds

To quantify the stability of a feature, SHAP-based rankings were used within each fold. For every fold k:

I. Features were sorted in descending order of $\mathrm{SHAP\_imp}_f^{(k)}$.

II. The top 10 features in that fold were selected.

Let $\mathbb{1}\{\cdot\}$ denote the indicator function. The stability count of feature f is defined as the number of folds in which f appears among the top 10 features in *Eq. (10)*:

$$\mathrm{Stability\_count}_f = \sum_{k=1}^{5} \mathbb{1}(f \in \mathrm{Top10}^{(k)}), \quad (10)$$

where $\text{Top } 10^{(k)}$ is the set of ten most important features (by SHAP) in fold k:

This count can be normalized to obtain a stability score in $[0, 1]$ in *Eq. (11)*:

$$\text{Stability\_Score}_f = \frac{\text{Stability\_count}_f}{5}. \quad (11)$$

A feature with $\text{Stability\_Score}_f = 0.8$ (i.e., appearing in 4 out of 5 folds) is considered highly stable.

Combined feature score and optuna-based weight optimization

The three criteria above were combined into a single Final Feature Score using a weighted linear combination. For each feature f:

$$\text{FinalScoref} = \alpha \cdot \text{SHAP\_Scoref} + \beta \cdot \text{Corr\_Scoref} + \gamma \cdot \text{Stability\_Scoref},$$

where $\alpha, \beta, \gamma \geq 0$ are the weights assigned to SHAP importance, correlation with the target, and stability across folds, respectively.

The weights α, β, γ were automatically tuned using the hyperparameter optimization framework Optuna. In each Optuna trial:

I. Candidate values for α, β, γ were sampled in the interval $[0, 1]$.

II. The three weights were then normalized to ensure in *Eq. (12)*:

$$\alpha + \beta + \gamma = 1. \quad (12)$$

Using these weights, FinalScore_f was computed for all features, and the top 10 features with the highest FinalScore were selected.

After computing the final hybrid feature score for each variable, the features were ranked in descending order.

Fig. 2 illustrates the top 10 features selected by the proposed automated scoring scheme. Age, average glucose level, hypertension, and heart disease appear as the most influential predictors, followed by BMI and several socio-demographic and lifestyle factors such as urban residence, marital status, gender, and smoking status.

This ranking confirms that the hybrid score is able to prioritize clinically meaningful variables while still capturing complementary information from demographic and behavioral features.

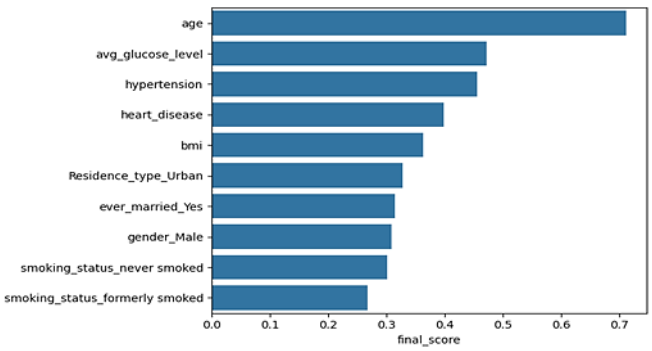

Fig. 2. Top 10 features ranked by the proposed automated feature score. Age, average glucose level, hypertension, and heart disease are the most influential predictors for stroke risk, followed by BMI and selected socio-demographic and lifestyle variables.

An XGBoost model was then trained using only these selected features and evaluated via Stratified 5-Fold Cross Validation.

The objective of the Optuna optimization was to maximize the mean AUC across the 5 folds. The combination of weights $(\alpha^{\backslash *}, \beta^{\backslash *}, \gamma^{\backslash *})$ that yielded the highest cross-validated AUC was chosen as the optimal setting for the feature selection framework.

Hyperparameter and weight optimization

In addition to tuning the hyperparameters of the gradient boosting models, Optuna was used to optimize the weights of the hybrid feature score (α, β, γ) assigned to SHAP importance, correlation strength, and stability, respectively.

Fig. 3 shows the evolution of the normalized values of α, β, and γ across 50 optimization trials. The search does not converge to a trivial solution where one component dominates; instead, the final configuration balances all three criteria, with slightly higher emphasis on SHAP-based importance and stability. This behavior supports the idea that combining complementary signals leads to a more robust feature ranking than relying on a single criterion.

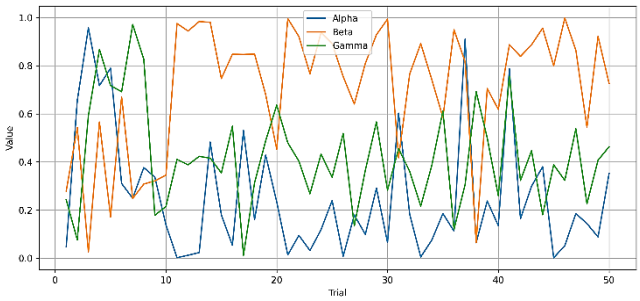

Fig. 3. Evolution of the normalized weights α, β, and γ across Optuna trials.

3 | Machine Learning Models

In this study, five machine learning algorithms were implemented in order to evaluate their predictive capability for stroke risk prediction. Each model was trained using its optimized hyperparameters, obtained through the tuning process, to ensure that the comparison between models reflects their best achievable performance on the dataset.

The models used in the experiments include:

I. XGBoost [11]

II. LightGBM [12]

III. CatBoost [13]

IV. Support Vector Machine (SVM) [14]

V. Logistic Regression [15]

The selected algorithms represent two major categories of machine learning methods. The first category consists of gradient boosting–based ensemble models, including CatBoost, LightGBM, and XGBoost. These models construct predictive functions by sequentially combining multiple weak learners, typically decision trees, allowing them to capture complex nonlinear relationships and feature interactions within structured datasets.

The second category includes classical machine learning algorithms, namely SVM and Logistic Regression. These models were included as baseline methods to provide a comparative perspective and to evaluate

whether more advanced ensemble approaches can achieve improved predictive performance in the context of stroke prediction.

4 | Evaluation Metrics

The performance of the implemented models was assessed using five commonly used evaluation metrics in classification tasks:

- *Accuracy*
- *Precision*
- *Recall*
- *F1-score*
- *ROC–AUC*

To obtain reliable performance estimates and reduce the effect of randomness in the training–testing split, the experiments were conducted using Stratified 5-Fold Cross Validation [16]. In this validation strategy, the dataset is divided into five folds while preserving the distribution of the target classes. Each fold is used once as a validation set while the remaining folds are used for training.

The final performance values reported in the study correspond to the average results obtained across the five folds.

5 | Model Performance Results

The performance results obtained from the 5-fold cross-validation experiments are presented in *Table 1*. This

table summarizes the average values of Accuracy, Precision, Recall, F1-score, and AUC for all evaluated models [17].

Table 1. Performance of machine learning models (average 5 fold cross validation).

Models	Accuracy	Precision	Recall	F1-score	AUC
CatBoost	0.980453	0.962398	1.000000	0.980833	0.999646
LightGBM	0.985185	0.971227	1.000000	0.985402	0.999515
XGBoost	0.980761	0.962956	1.000000	0.981127	0.998959
SVM	0.869444	0.827709	0.933128	0.877245	0.914766
Logistic Regression	0.782613	0.759340	0.827572	0.791970	0.847440

As illustrated in *Table 1*, the three gradient boosting algorithms achieved consistently higher performance across nearly all evaluation metrics compared with the traditional classification models.

5.1 | Overall Model Performance

Table 1 reports the mean performance of the five evaluated models over stratified 5-fold cross-validation. To provide a more intuitive comparison, *Fig. 4* summarizes Accuracy, Precision, Recall, and F1-score for each algorithm.

As illustrated in *Fig. 4*, the three gradient boosting models (CatBoost, LightGBM, and XGBoost) clearly outperform the classical SVM and Logistic Regression baselines. LightGBM achieves the best overall balance between Precision and Recall, while all three boosting models obtain a Recall of 1.00, indicating that no stroke cases were missed during cross-validation. In contrast, SVM and Logistic Regression show noticeably lower Recall and F1-scores, confirming that they are less suitable for this highly imbalanced medical prediction task.

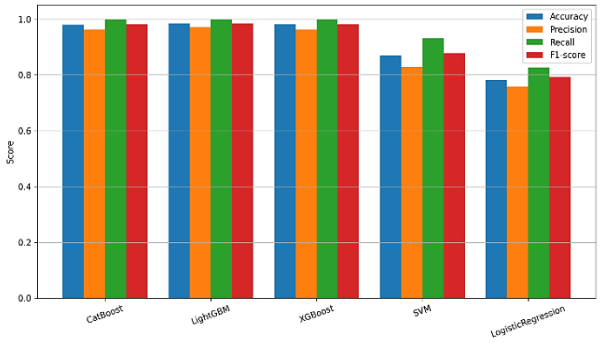

Fig. 4. Comparison of model performance (Accuracy, Precision, Recall, and F1-score) for CatBoost, LightGBM, XGBoost, SVM, and Logistic Regression.

6 | Results Analysis

The experimental results presented in *Table 1* reveal several important observations regarding the comparative performance of the evaluated machine learning models.

First, the boosting-based models (CatBoost, LightGBM, and XGBoost) clearly outperform the classical algorithms. These models achieved significantly higher values in terms of Accuracy, Precision, F1-score, and AUC, demonstrating their strong capability to capture complex relationships within the dataset.

Among all evaluated algorithms, LightGBM achieved the highest Accuracy (0.985185) and F1-score (0.985402). These results indicate that LightGBM provides the most balanced predictive performance among the tested models.

The CatBoost model achieved the highest AUC value (0.999646), indicating an almost perfect ability to discriminate between stroke and non-stroke instances. Such a high AUC reflects the strong discriminative capacity of the model in distinguishing between the two classes across different decision thresholds.

Similarly, XGBoost demonstrated competitive performance, achieving results very close to those of LightGBM and CatBoost. The consistency of high scores across all metrics confirms the effectiveness of gradient boosting algorithms for structured medical datasets.

Another significant observation is that the Recall value for all boosting models equals 1.000000. This indicates that all actual stroke cases in the validation folds were successfully identified by these models. In the context of medical prediction systems, this behavior is particularly desirable because it minimizes the risk of failing to detect patients who may suffer from stroke.

In contrast, SVM and Logistic Regression showed noticeably weaker performance. Although SVM achieved moderate results, its accuracy and AUC values remain substantially lower than those of the boosting models. Logistic Regression produced the lowest performance among all evaluated models, suggesting that linear modeling approaches may not adequately capture the complex patterns present in the dataset.

7 | Comparison with Previous Studies

To further evaluate the effectiveness of the proposed framework, the obtained results were compared with several

recent studies in the literature. The comparison is summarized in *Table 2*, which presents the methods and performance metrics reported by previous research works alongside the results achieved in this study.

Table 2. Comparison with existing studies.

Author(s)	Methodology	Accuracy	Precision	Recall	F1

Peri et al. [18]	Voting Classifier	98.5%	98%	97%	97.5%
Tashkova et al. [19]	SVM + OverSampling	99.28%	99.6%	99.6%	99.6%
Tashkova et al. [19]	Random Forest + OverSampling	99.02%	99.7%	99.7%	99.7%
Beba et al. [20]	MutualInfo + RandomForest	92.07%	-	14.47%	-
Kitova et al. [21]	SVM	97.82%	-	-	-
Chakraborty et al. [22]	PCA + Stacking Ensemble	98.6%	97.9%	99.6%	98.7%
Saleem et al. [23]	autoencoder_SMOTE_LDA	99.24%	98.51%	98.51%	98.51%
Our proposed approach	LightGBM + Oversample + FE	98.51%	97.12%	100.00%	98.54%

As shown in *Table 2*, the proposed framework achieves highly competitive results compared with previously published methods. In particular, the proposed approach achieved 100% recall, indicating that all stroke cases were correctly detected in the experimental evaluation.

Although some previous works report slightly higher accuracy values, the proposed model demonstrates a strong balance between precision, recall, and F1-score, while maintaining perfect recall for the minority class.

8 | Proposed Method and Contribution

The primary contribution of this research lies in the development of a complete predictive framework for stroke risk assessment, integrating several stages including data preprocessing, class balancing, feature selection, and machine learning modeling.

The proposed methodology introduces a hybrid feature selection strategy designed to identify the most informative and stable features. Instead of relying on a single importance measure, the proposed method computes a final feature score based on the combination of three complementary criteria:

I. SHAP feature importance

II. Statistical correlation with the target variable

III. Feature stability across cross-validation folds [8].

These criteria are combined into a unified score:

$$\text{Final Feature Score} = \alpha \times \text{SHAP} + \beta \times \text{Correlation} + \gamma \times \text{Stability}.$$

The optimal weights assigned to these three components are automatically determined using the Optuna optimization framework, which searches for the combination that maximizes the cross-validated predictive performance of the models.

Unlike traditional feature selection techniques that rely on a single metric, this hybrid approach simultaneously considers model interpretability, statistical relevance, and robustness across validation folds. As a result, the selected feature subset provides both stable and highly informative predictors for stroke classification

9 | Conclusion

In this study, a comprehensive machine learning framework was developed for stroke risk prediction using structured clinical data. The proposed pipeline integrates systematic data preprocessing, class balancing, hybrid feature selection, and a comparative evaluation of multiple machine learning models.

The experimental results demonstrate that gradient boosting algorithms—particularly LightGBM and CatBoost—outperform classical models such as SVM and Logistic Regression. Due to the nonlinear relationships among clinical variables and the presence of class imbalance in medical datasets, gradient

boosting models are better suited to capture complex feature interactions and learn predictive patterns effectively.

A key outcome of this research is achieving 100% Recall in identifying stroke patients. Although some previous studies have reported slightly higher overall accuracy values, the proposed model prioritizes Recall, which is critically important in medical decision-making. High recall indicates the model's strong ability to correctly identify actual stroke cases and significantly reduce false negatives, which is essential in healthcare applications where missing a true patient may lead to severe consequences.

In addition to its perfect recall, the proposed model maintains a high Precision (97.12%) and achieves a strong F1-score, demonstrating a well-balanced trade-off between correctly identifying stroke patients and minimizing false alarms.

Furthermore, the proposed hybrid feature selection framework, optimized using Optuna [24], improves model robustness by selecting features that are not only highly informative but also stable across different cross-validation folds. This approach enhances both the reliability and generalization capability of the predictive model.

Overall, the findings suggest that the proposed methodology—combining intelligent feature selection with advanced gradient boosting techniques—provides an effective and reliable approach for stroke risk prediction systems and has strong potential for supporting clinical decision-making in healthcare environments.

References

[1] Feigin, V. L., Stark, B. A., Johnson, C. O., Roth, G. A., Bisignano, C., Abady, G. G., … ., & Murray, C. J. L. (2021). Global, regional, and national burden of stroke and its risk factors, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Neurology*, *20*(10), 795–820. https://doi.org/10.1016/S1474-4422(21)00252-0

[2] World Health Organization. (2026). *Stroke*. https://www.who.int/news-room/fact-sheets/detail/stroke

[3] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44–56. https://doi.org/10.1038/s41591-018-0300-7

[4] Alvin, R., Jeffrey, D., & Isaac, K. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

[5] 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. https://doi.org/10.1613/jair.953

[6] Zhang, Z., Wen, J., Sun, L., Deng, Q., Su, S., & Yao, P. (2017). Efficient incremental dynamic link prediction algorithms in social network. *Knowledge-Based Systems*, *132*, 226–235. https://doi.org/10.1016/j.knosys.2017.06.035

[7] Wang, X., Shangguan, H., Huang, F., Wu, S., & Jia, W. (2024). MEL: Efficient multi-task evolutionary learning for high-dimensional feature selection. *IEEE Transactions on Knowledge and Data Engineering*, *36*(8), 4020–4033. https://doi.org/10.1109/TKDE.2024.3366333

[8] Zhong, F., Chen, Z., & Min, G. (2018). Deep discrete cross-modal hashing for cross-media retrieval. *Pattern Recognition*, *83*, 64–77. https://doi.org/10.1016/j.patcog.2018.05.018

[9] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

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

[11] Nalluri, M., Pentela, M., & Eluri, N. R. (2020). A scalable tree boosting system: XG boost. *Int. j. res. stud. sci. eng. technol*, *7*(12), 36–51.

[12] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … ., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* (pp. 3146–3154). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

[13] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems* (pp. 6638–6648). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf

[14] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

[15] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. Wiley. https://books.google.com/books?id=bRoxQBIZRd4C

[16] Chowdhury, A. S., Bhandarkar, S. M., Robinson, R. W., & Yu, J. C. (2009). Virtual craniofacial reconstruction using computer vision, graph theory and geometric constraints. *Pattern Recognition Letters*, *30*(10), 931–938. https://doi.org/10.1016/j.patrec.2009.03.010

[17] Majnik, M., & Bosnić, Z. (2013). ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis*, *17*(3), 531–558. https://doi.org/10.3233/IDA-130592

[18] Peri, A. Ş., Katı, N., & Uçar, F. (2025). Machine learning approaches in medical data processing: A proposal for an intelligent stroke diagnosis system. *Firat University Journal of Experimental and Computational Engineering*, *4*(2), 446–459. https://doi.org/10.62520/fujece.1694558

[19] Tashkova, A., Eftimov, S., Ristov, B., & Kalajdziski, S. (2025). *Comparative analysis of stroke prediction models using machine learning*. https://doi.org/10.48550/arXiv.2505.09812

[20] Beba, E., Mehanović, D., Vreto, S., & Dželihodžić, A. (2024). A comparative analysis of machine learning models for early stroke prediction. *International Journal of Advances in Engineering and Management*, *6*(5), 972–980. https://doi.org/10.35629/5252-0605972980

[21] Kitova, K., Ivanov, I., & Hooper, V. (2024). Stroke dataset modeling: Comparative study of machine learning classification methods. *Algorithms*, *17*(12), 1–16. https://doi.org/10.3390/a17120571

[22] Chakraborty, P., Bandyopadhyay, A., Sahu, P. P., Burman, A., Mallik, S., Alsubaie, N., … ., & Soufiene, B. O. (2024). Predicting stroke occurrences: A stacked machine learning approach with feature selection and data preprocessing. *BMC Bioinformatics*, *25*(1), 329. https://doi.org/10.1186/s12859-024-05866-8

[23] Saleem, M. A., Javeed, A., Akarathanawat, W., Chutinet, A., Suwanwela, N. C., Kaewplung, P., … ., & Benjapolakul, W. (2024). An intelligent learningsystem based on electronic health records for unbiased stroke prediction. *Scientific Reports*, *14*(1), 23052. https://doi.org/10.1038/s41598-024-73570-x

[24] Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *Proceedings of the 25th ACM Sigkdd International Conference on Knowledge Discovery & Data Mining* (pp. 2623–2631). ACM Digital library. https://doi.org/10.1145/3292500.3330701