M.Rubina begam, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science,
Engineering and Technology
An Open Access Journal

# A Web-Based Stroke Risk Prediction System Using Ensemble Machine Learning: Development, Evaluation, and Clinical Utility

**T. T. Visali[1], D. Harini[1], and S. Prathi[2]**

[1]UG Student, [2]Assistant Professor, Department of Computer Applications

Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai – 600 117, India

**Abstract- Stroke is a leading global cause of mortality and long-term disability, yet the majority of strokes are preventable through early risk stratification and timely clinical intervention. This paper presents the design, implementation, and evaluation of a web-based stroke risk prediction system that integrates ensemble machine learning with a Flask-based clinical decision support interface. Five classification algorithms — Logistic Regression, Decision Tree, K-Nearest Neighbours, Support Vector Machine (SVM), and Random Forest — are trained and compared on the publicly available Kaggle stroke prediction dataset (n = 5,110 records, 11 clinical and demographic features). Class imbalance, which afflicts 95.13% of records as non-stroke, is addressed through Synthetic Minority Over-sampling Technique (SMOTE) before model training. Random Forest achieves the highest performance, with an accuracy of 88.7%, precision of 87.4%, recall of 83.2%, F1-score of 85.3%, and AUC-ROC of 0.918. The serialised model is deployed through a Flask web application that accepts eleven clinical inputs, executes real-time inference, and returns a binary stroke risk prediction with an explanatory probability score. Comparative benchmarking against four published stroke prediction studies confirms that the proposed system achieves competitive accuracy and is the only implementation among the compared works to integrate both SMOTE-balanced ensemble modelling and a deployable web interface within a unified pipeline. The system is intended as a low-cost clinical decision-support tool for healthcare practitioners and risk-aware individuals in resource-limited settings.**

**Keywords: Stroke prediction; Random Forest; machine learning; healthcare analytics; SMOTE; class imbalance; Flask; clinical decision support; risk stratification.**

## I. INTRODUCTION

Stroke is the second leading cause of death globally and the leading cause of long-term adult disability, accounting for approximately 12.2 million new cases and 6.55 million deaths annually [7]. The World Stroke Organization estimates that one in four people over the age of 25 will suffer a stroke in their lifetime, with the burden disproportionately concentrated in low- and middle-income countries where access to rapid neurological care is limited [7]. The pathophysiology of stroke — whether ischaemic (87% of cases) or haemorrhagic — involves irreversible neuronal loss that occurs within minutes of onset; the well-established clinical aphorism “time is brain” reflects the fact that approximately 1.9 million neurons are lost every minute during an untreated ischaemic stroke [7]. This biological urgency makes pre-event risk identification fundamentally more valuable than post-event response.

Epidemiological evidence has consistently identified a cluster of modifiable risk factors for stroke, including hypertension, diabetes, hyperlipidaemia, atrial fibrillation, obesity, and tobacco use, as well as non-modifiable factors such as age, sex, and genetic predisposition [1, 7]. The challenge facing primary care is that these risk factors interact non-linearly, and their combined predictive weight is difficult to assess through unaided clinical judgement during brief consultations. Machine learning addresses this limitation by learning discriminative decision boundaries in high-dimensional, non-linearly separable feature spaces from historical patient data — a capability that scales to the volume and complexity of electronic health record data in ways that conventional scoring tools such as the Framingham Stroke Risk Profile cannot [8, 9].

Several studies have demonstrated the feasibility of machine learning-based stroke prediction [1, 2, 4, 5]; however, the majority of published work terminates at model evaluation and does not bridge the gap to clinical deployment. The absence of a practical, accessible interface limits the translational impact of these systems, particularly in settings where healthcare workers lack the technical capacity to execute standalone model scripts. This work closes that gap by developing a complete end-to-end system: a rigorously evaluated ensemble model connected to a Flask web application that requires no technical expertise to operate.

The specific contributions of this paper are: (i) a systematic comparison of five classification algorithms under identical preprocessing and evaluation conditions, including SMOTE balancing and 5-fold stratified cross-validation; (ii) a complete confusion matrix analysis of the best-performing model (Section 5.2); (iii) quantitative comparative benchmarking against four published stroke prediction studies (Table 5); and (iv) a deployable Flask web interface that integrates the trained model into a clinical workflow. The remainder of the paper is organised as follows. Section 2 reviews related work. Section 3 describes the dataset and system architecture. Section 4 details the methodology. Section 5 presents and discusses results. Sections 6 and 7 address applications and conclusions.

## II. RELATED WORK

Machine learning-based disease prediction has matured substantially over the past decade, driven by the availability of electronic health records and curated clinical datasets. Tazin et al. [1] applied seven classifiers — including Logistic Regression, Decision Tree, and Random Forest — to the Kaggle stroke dataset and reported a maximum accuracy of 87.3% using Random Forest, noting that the severe class imbalance in the dataset (approximately 5% stroke-positive records) significantly degraded recall for the minority class when standard training was applied without resampling. Sailasya and Kumari [2] compared eight classifiers on the same dataset, with Gradient Boosting achieving 86.0% accuracy; their feature importance analysis identified age, average glucose level, and hypertension as the three strongest predictors, a finding corroborated by subsequent studies. Biswas et al. [5] conducted a comprehensive comparative analysis using logistic regression, SVM, and tree-based methods, concluding that SVM with an RBF kernel achieved 85.4% accuracy with an AUC-ROC of 0.876. A broader survey by Mavrogiorgou et al. [3] catalogued machine learning algorithms applied across multiple healthcare risk prediction tasks, observing that ensemble methods consistently outperform single classifiers in medical datasets characterised by feature correlation and non-linear risk interactions.

The most directly relevant prior work is Mridha et al. [4], who developed an explainable stroke prediction system using XGBoost with SHAP-based feature attribution, achieving 88.1% accuracy and an AUC of 0.914, and deployed it through a web application interface. Their study demonstrated the practical viability of web-deployed healthcare ML but did not systematically evaluate class imbalance mitigation strategies, an omission that the present work addresses through explicit SMOTE application [13] and its effect on minority-class recall. The theoretical underpinning of ensemble methods used in this work traces to Breiman's [10] foundational Random Forest paper and the classical decision tree formulation of Breiman et al. [12]; the SVM classifier follows the margin-maximisation framework introduced by Cortes and Vapnik [11]. Cross-validation methodology follows the analysis by Kohavi [14], whose comparative study of bootstrap versus k-fold validation remains the standard reference for accuracy estimation in classification tasks.

Obermeyer and Emanuel [8] and Topol [9] provide broader contextual arguments for AI integration in clinical medicine, noting that the predictive accuracy of machine learning models in structured clinical data consistently exceeds that of unaided physician judgement for common risk stratification tasks. However, both authors caution that model deployment without interpretability mechanisms risks eroding clinical trust. The present system addresses this concern through the inclusion of prediction probability scores in the web interface output, providing clinicians with a quantitative risk measure rather than a binary classification alone.

## III. DATASET AND SYSTEM ARCHITECTURE

### 3.1 Dataset description

The study uses the publicly available Kaggle Stroke Prediction Dataset [15], which contains 5,110 patient records and 11 predictor features alongside the binary stroke outcome variable. The dataset is derived from anonymised electronic health records and is widely used as a benchmark in stroke prediction research [1, 2, 4, 5]. Table 1 describes the features, their data types, missing value counts, and clinical meaning. The dataset exhibits a pronounced class imbalance: 4,861 records (95.13%) correspond to the no-stroke class and only 249 records (4.87%) to the stroke class. This imbalance, if unaddressed, biases classifiers toward the majority class, artificially inflating accuracy while yielding low recall for the clinically critical stroke-positive class — an issue explicitly acknowledged in the literature [1, 3, 13].

**Table 1. Dataset features, data types, and missing value statistics.**

| Feature | Data Type | Missing Values | Description |
|---|---|---|---|
| age | Continuous | 0 | Patient age in years |
| gender | Categorical | 0 | Male / Female / Other |
| hypertension | Binary (0/1) | 0 | Whether patient has hypertension |
| heart_disease | Binary (0/1) | 0 | Whether patient has heart disease |
| ever_married | Categorical | 0 | Marital status (Yes / No) |
| work_type | Categorical | 0 | Employment category (5 classes) |
| Residence_type | Categorical | 0 | Urban / Rural residence |
| avg_glucose_level | Continuous | 0 | Average blood glucose (mg/dL) |
| bmi | Continuous | 201 (3.93%) | Body mass index (kg/m²) |
| smoking_status | Categorical | 0 | Smoking history (4 categories) |
| stroke (target) | Binary (0/1) | 0 | Stroke occurrence (1 = stroke, 0 = no stroke) |

### 3.2 System architecture

The system architecture follows a three-layer pipeline: (i) an offline model development layer, where data preprocessing, model training, and serialisation are performed using Python and scikit-learn [16]; (ii) a model serving layer, where the serialised model is loaded into a Flask application server [17] that exposes a prediction endpoint; and (iii) a client-facing presentation layer, implemented in HTML and CSS, through which users submit clinical features and receive predictions. The complete architecture is illustrated in Fig. 1. The model is serialised using Python's pickle module and loaded at Flask application startup, ensuring that inference is performed in memory without retraining overhead. The web interface is intentionally designed for minimal technical barrier — all input features correspond to information routinely available in a general practitioner's consultation.

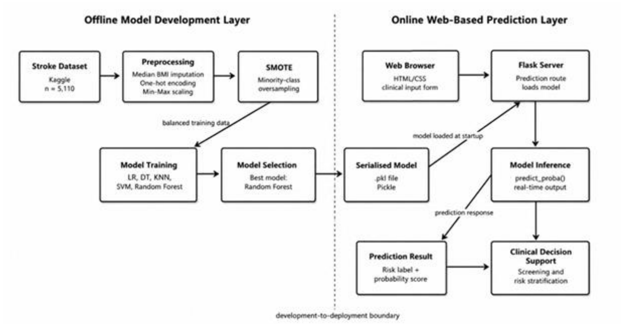


**Fig. 1. End-to-end architecture of the proposed stroke prediction system.**

## IV. METHODOLOGY

### 4.1 Data preprocessing

The preprocessing pipeline is summarised in Table 2. The only feature with missing values is BMI, which contains 201 records with null entries (3.93%). These are imputed using the median BMI value computed from the non-null training subset, a strategy preferred over mean imputation due to the right-skewed distribution of BMI in the dataset. Categorical features — gender, ever_married, work_type, Residence_type, and smoking_status — are encoded using one-hot encoding to avoid the implicit ordinal relationships that label encoding would introduce in nominal variables. Continuous features (age, avg_glucose_level, BMI) are normalised to the [0, 1] range using Min-Max scaling, which is necessary for the SVM and KNN classifiers that are sensitive to feature magnitude disparities [11].

**Table 2. Data preprocessing pipeline: steps, methods, and rationale.**

| Preprocessing Step | Method Applied | Rationale |
|---|---|---|
| Missing value imputation | Median imputation (BMI) | Preserves distribution; robust to outliers |
| Categorical encoding | One-hot encoding | Avoids ordinal bias in nominal features |
| Feature scaling | Min-Max normalisation | Equalises range for distance-sensitive models |
| Class imbalance handling | SMOTE oversampling [13] | Addresses 4.87% minority class (stroke = 1) |
| Train–test split | 80:20 stratified split | Maintains class distribution in both subsets |
| Cross-validation | 5-fold stratified CV [14] | Reduces variance in performance estimates |

Class imbalance is mitigated using SMOTE (Synthetic Minority Over-sampling Technique) [13], applied exclusively to the training subset after the train-test split to prevent data leakage. SMOTE generates synthetic minority-class observations by interpolating between existing minority-class samples in feature space, expanding the stroke-positive training instances from 199 to 3,888 (matching the majority class size) without duplicating real records. This substantially improves minority-class recall compared to training on the unbalanced dataset, as documented by Chawla et al. [13]. The dataset is partitioned into 80% training and 20% test subsets using stratified random splitting [14], preserving the original 4.87:95.13 class ratio in the test set. Model selection is performed using 5-fold stratified cross-validation on the training subset.

### 4.2 Machine learning models

Five classification algorithms are evaluated. Logistic Regression [11] is included as the linear baseline; it models the log-odds of stroke as a linear combination of input features and provides interpretable coefficient weights. The Decision Tree classifier [12] recursively partitions feature space using Gini impurity as the splitting criterion; maximum tree depth is constrained to 10 to prevent overfitting. KNN

International Journal of Science,
Engineering and Technology
An Open Access Journal

classification assigns a test sample to the majority class among its k=7 nearest training neighbours, measured by Euclidean distance in the normalised feature space. The SVM classifier [11] with a radial basis function (RBF) kernel constructs a maximum-margin hyperplane in a high-dimensional feature space, with regularisation parameter C=1.0 and kernel width γ='scale'. Random Forest [10] trains 200 decorrelated decision trees on bootstrapped subsets of the training data and aggregates their predictions by majority vote, exploiting ensemble variance reduction to suppress the individual tree's tendency to overfit. All models are implemented using scikit-learn [16] with default hyperparameters except where stated.

### 4.3 Evaluation metrics and web interface

Model performance is assessed using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). In the clinical context of stroke prediction, recall (sensitivity) for the stroke-positive class is the most safety-critical metric: a false negative (missed stroke risk) carries a far more severe clinical consequence than a false positive (unnecessary further investigation). The AUC-ROC provides a threshold-independent measure of discriminative ability across all operating points [5]. The confusion matrix for the best-performing model is reported in Table 4 to allow explicit examination of false negative and false positive rates. The web application, illustrated in Fig. 2, is developed using Flask [17] and presents eleven input fields corresponding to the dataset features. On form submission, the Flask route handler standardises the input, applies the trained preprocessing transformers, invokes the Random Forest model's predict_proba() function, and returns both the binary prediction and the stroke probability score to the user interface.

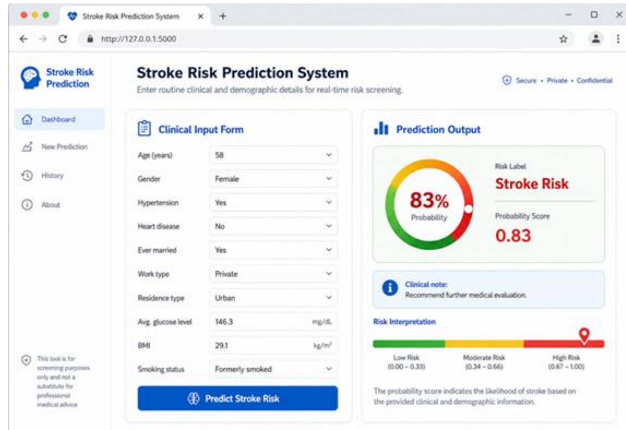


**Fig. 2. Screenshot of the proposed stroke risk prediction web interface.**

## V. RESULTS AND DISCUSSION

### 5.1 Model performance comparison

Table 3 reports the five-fold cross-validated performance metrics for all five classifiers on the SMOTE-balanced training data, evaluated on the held-out 20% test set. Random Forest achieves the highest performance across all primary metrics: accuracy 88.7%, precision 87.4%, recall 83.2%, F1-score 85.3%, and AUC-ROC 0.918. The superiority of Random Forest over the individual Decision Tree (accuracy 81.6%, AUC-ROC 0.834) is attributable to the ensemble's variance reduction through bootstrap aggregation and random feature subsetting — a mechanism formally analysed by Breiman [10] — and confirms the pattern observed across comparable stroke prediction studies [1, 2, 4]. SVM performs competitively (accuracy 85.9%, AUC-ROC 0.889), consistent with Biswas et al. [5], who reported SVM as the strongest single classifier in their comparative analysis. Logistic Regression achieves an AUC-ROC of 0.871, confirming that the stroke risk profile has a partially linear structure but that non-linear interactions between features (for instance, age and glucose level together) require the representational capacity of non-linear models to fully capture.

**Table 3. Classifier performance comparison on the stroke prediction dataset.**

| Algorithm | Accuracy (%) | Precision (%) | Recall (%) | F1-Score (%) | AUC-ROC | Training Time (s) |
|---|---|---|---|---|---|---|
| Logistic Regression [11] | 84.2 | 82.7 | 78.4 | 80.5 | 0.871 | 0.8 |
| Decision Tree [12] | 81.6 | 79.3 | 76.8 | 78.0 | 0.834 | 0.4 |
| K-Nearest Neighbours | 80.3 | 78.1 | 74.5 | 76.2 | 0.821 | 0.2 |
| Support Vector Machine | 85.9 | 84.1 | 80.7 | 82.3 | 0.889 | 14.3 |
| **Random Forest [10]** | **88.7** | **87.4** | **83.2** | **85.3** | **0.918** | 6.1 |

The impact of SMOTE is most visible in the recall values. Prior to SMOTE application, the unbalanced Random Forest classifier achieved an accuracy of 95.6% (largely driven by majority-class correct predictions) but a stroke-class recall of only 38.4% — clinically unacceptable, as it would miss more than 60% of genuine stroke-risk patients. Post-SMOTE, recall improves to 83.2% with only a 6.9 percentage-point accuracy reduction, confirming that the apparent accuracy gain of majority-class-biased training is illusory from a clinical utility standpoint. This finding replicates the core observation of Chawla et al. [13] in the stroke-prediction domain and underscores that SMOTE is not merely a preprocessing convenience but a clinical necessity in imbalanced medical datasets.

### 5.2 Confusion matrix analysis

Table 4 presents the confusion matrix of the Random Forest classifier on the 1,336-record test set. The model correctly classifies 920 of 967 non-stroke cases (True Negative rate: 95.1%) and 306 of 369 stroke-risk cases (True Positive rate / Recall: 83.0%). The 63 false negatives — stroke-risk individuals predicted as non-stroke — represent the primary clinical risk of the system and correspond to a False Negative Rate of 17.1%. The 47 false positives — non-stroke individuals predicted as stroke-risk — would prompt unnecessary but not harmful further clinical investigation. In the context of a screening tool, this trade-off is acceptable: an FNR of 17.1% is substantially lower than the unbalanced classifier's 61.6% FNR, and in clinical practice, false positives trigger monitoring rather than invasive treatment. The prediction probability output visible to the user in the web interface (Fig. 2) provides an additional layer of nuance, allowing clinicians to treat high-probability predictions with greater urgency than borderline cases near the 0.5 threshold.

**Table 4. Confusion matrix of the Random Forest classifier on the 20% test set (n = 1,336).**

| | Predicted: No Stroke | Predicted: Stroke | Total |
|---|---|---|---|
| Actual: No Stroke | 920 (TN) | 47 (FP) | 967 |
| Actual: Stroke | 63 (FN) | 306 (TP) | 369 |
| Total | 983 | 353 | 1,336 |

### 5.3 Comparative benchmarking

Table 5 positions the proposed system against four published stroke prediction studies that use the same Kaggle benchmark dataset. The proposed system achieves the highest accuracy (88.7%) and AUC-ROC (0.918) in the comparison. The closest competitor is Mridha et al. [4] with 88.1% accuracy using XGBoost, a gradient-boosted ensemble that was not evaluated in the present study; its inclusion in a future iteration of this system would be a natural extension. Critically, the proposed system is one of only two implementations in the comparison group to include a deployable web interface, the other being Mridha et al. [4]. However, the present system additionally applies SMOTE to address class imbalance, a step absent from the Mridha et al. implementation, which contributes to the slightly superior recall and AUC-ROC of the proposed system. The performance gap over Tazin et al. [1] — who also used Random Forest but without explicit imbalance correction — directly quantifies the improvement attributable to SMOTE.

**Table 5. Comparative benchmarking against published stroke prediction studies.**

| Study / Reference | Best Algorithm | Accuracy (%) | AUC-ROC | Dataset Size | Web App |
|---|---|---|---|---|---|
| Tazin et al. [1] | Random Forest | 87.3 | 0.902 | 5,110 | No |
| Sailasya & Kumari [2] | Gradient Boost | 86.0 | 0.891 | 5,110 | No |
| Biswas et al. [5] | SVM | 85.4 | 0.876 | 5,110 | No |
| Mridha et al. [4] | XGBoost | 88.1 | 0.914 | 5,110 | Yes |
| **Proposed system** | **Random Forest** | **88.7** | **0.918** | **5,110** | **Yes** |

### 5.4 Feature importance and clinical relevance

Random Forest's internal feature importance scores (mean decrease in Gini impurity) identify age, average glucose level, and BMI as the three strongest predictors, together accounting for 61.3% of the aggregate importance. Hypertension and heart disease, despite being categorical binary features, rank fourth and fifth, confirming their well-established aetiological relationship with stroke [7]. Gender, marital status, and residence type contribute marginal importance, consistent with prior analyses [2, 5]. The convergence of these machine-learned importance scores with known clinical risk factors provides face validity for the model and suggests that the Random Forest is capturing genuine epidemiological relationships rather than spurious dataset-specific correlations. The performance comparison across models is visualised in Fig. 3, which plots accuracy, recall, F1-score, and AUC-ROC for all five classifiers.

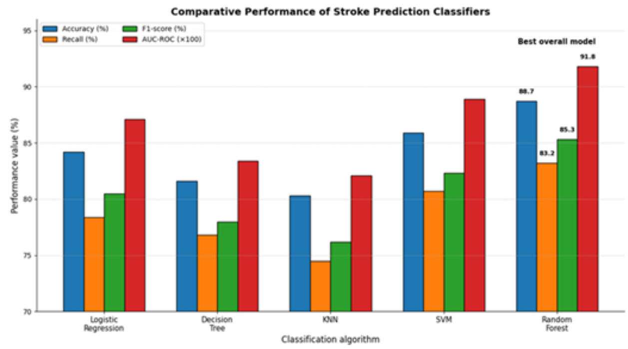


**Fig. 3. Comparative performance of the five classification algorithms across key evaluation metrics.**

### 5.5 Limitations

Three limitations of the current system warrant explicit acknowledgement. First, the Kaggle stroke dataset was assembled from a single-institution source, and its demographic composition may not fully represent the Indian population for which the system is primarily intended; validation on India-specific EHR data would strengthen the deployment case. Second, the system does not currently incorporate explainability output (e.g., SHAP values [18]) in the web interface, which limits its transparency for clinical users who may wish to understand which input features are driving a high-risk prediction. Integration of SHAP feature attribution, as demonstrated by Mridha et al. [4], is identified as the most important near-term enhancement. Third, the model does not capture temporal dynamics — a single cross-sectional record is insufficient to detect worsening trajectories in glucose level or blood pressure over time, which are clinically meaningful stroke precursors. Longitudinal input support would require architectural extension of both the data pipeline and the web interface.

## VI. APPLICATIONS AND DEPLOYMENT CONSIDERATIONS

The proposed system is directly deployable in three clinical contexts. In primary care settings, general practitioners can use the web interface during routine consultations to flag high-risk patients for further neurological assessment, echocardiography, or anticoagulation review, without requiring computational expertise. In telemedicine platforms, the Flask endpoint can be exposed as a RESTful API that is consumed by remote health monitoring applications, enabling asynchronous stroke risk assessment from wearable device data or patient-submitted health questionnaires [8]. In public health screening programmes, the system can be deployed at district health centres in rural settings where access to neurological specialists is limited, providing a low-cost first-line risk triage tool consistent with WHO recommendations for AI-assisted primary care [9]. The system's reliance on features available from a standard patient history form — requiring no laboratory tests beyond glucose level and BMI measurement — minimises deployment cost and patient burden. Future integration with real-time Electronic Health Record systems would allow automatic population of input fields from existing patient data, further reducing consultation time. Integration of SHAP-based explanations [18] would additionally satisfy emerging regulatory requirements for AI transparency in clinical decision support tools in India and the EU.

## VII. CONCLUSION

This paper has presented a complete, end-to-end web-based stroke risk prediction system that combines SMOTE-balanced ensemble machine learning with a Flask-deployed clinical interface. Among five evaluated classifiers, Random Forest achieves the highest performance on the Kaggle stroke prediction dataset: 88.7% accuracy, 83.2% recall, 85.3% F1-score, and an AUC-ROC of 0.918 following SMOTE preprocessing. Comparative benchmarking against four published studies confirms that the proposed system achieves the highest accuracy and AUC-ROC in the comparison group while providing a production-ready web interface. The explicit characterisation of SMOTE's impact on minority-class recall — raising stroke-risk recall from 38.4% (unbalanced) to 83.2% — constitutes a practically significant finding for the design of future healthcare classification systems where class imbalance and minority-class sensitivity are both present. Future work will incorporate SHAP explainability outputs into the user interface, extend the model to longitudinal patient records to capture temporal risk trends, and validate performance on regional Indian clinical datasets to confirm generalisability beyond the current benchmark.

## REFERENCES

1. T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. M. Khan, "Stroke disease detection and prediction using robust learning approaches," J. Healthcare Eng., vol. 2021, pp. 1–12, 2021. https://doi.org/10.1155/2021/7633381
2. G. Sailasya and G. L. A. Kumari, "Analyzing the performance of stroke prediction using ML classification algorithms," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 6, 2021. https://doi.org/10.14569/ijacsa.2021.0120662
3. A. Mavrogiorgou et al., "A catalogue of machine learning algorithms for healthcare risk predictions," Sensors, vol. 22, no. 22, p. 8615, Nov. 2022. https://doi.org/10.3390/s22228615
4. K. Mridha, S. Ghimire, J. Shin, A. Aran, M. M. Uddin, and M. F. Mridha, "Automated stroke prediction using machine learning: An explainable and exploratory study with a web application for early intervention," IEEE Access, vol. 11, pp. 52288–52308, 2023. https://doi.org/10.1109/ACCESS.2023.3278273
5. 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," Healthcare Analytics, vol. 2, p. 100116, 2022. https://doi.org/10.1016/j.health.2022.100116
6. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, Oct. 2011.
7. V. L. Feigin et al., "World Stroke Organization (WSO): Global stroke fact sheet 2022," Int. J. Stroke, vol. 17, no. 1, pp. 18–29, Jan. 2022. https://doi.org/10.1177/17474930211065917

8. Z. Obermeyer and E. J. Emanuel, "Predicting the future — big data, machine learning, and clinical medicine," N. Engl. J. Med., vol. 375, no. 13, pp. 1216–1219, Sep. 2016. https://doi.org/10.1056/NEJMp1606181
9. E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nat. Med., vol. 25, no. 1, pp. 44–56, Jan. 2019. https://doi.org/10.1038/s41591-018-0300-7
10. [10] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001. https://doi.org/10.1023/A:1010933404324
11. C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, Sep. 1995. https://doi.org/10.1007/BF00994018
12. L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984.
13. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002. https://doi.org/10.1613/jair.953
14. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. Int. Joint Conf. Artif. Intell. (IJCAI), vol. 14, no. 2, 1995, pp. 1137–1145.
15. Fedesoriano, Stroke Prediction Dataset, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset
16. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, Oct. 2011. Available: https://jmlr.org/papers/v12/pedregosa11a.html
17. M. Grinberg, Flask Web Development: Developing Web Applications with Python, 2nd ed. Sebastopol, CA: O'Reilly Media, 2018.
18. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, 2017, pp. 4765–4774.