International Journal of Research and Development in Engineering Science
ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/
Regular Issue, Vol. 7, Issue. 4 (July – August), 2025 , Pages: 18 – 24
Received: 20 October 2024 ; *Accepted:* 31 May 2025 ; *Published:* 25 July 2025
https://doi.org/10.63328/IJRDES-V7RI4P4
Original Paper : Flexible Research – Persuasive Research Paper

Ensemble Deep Learning Approach for Stroke Prediction Using Healthcare Data

T. Indira Priyadarshini [1]*, K. Rupa [2]

[1,2] Department of Computer Science and Engineering, VEMU Institute of Technology, P. Kothakota, India;
indirapriya1520@gmail.com , rupacse@gmail.com

Corresponding Author: indirapriya1520@gmail.com

Abstract: Stroke remains one of the foremost causes of mortality and long-term disability worldwide. Early prediction of stroke is crucial in enabling timely medical interventions and minimizing its adverse outcomes. This study presents a robust ensemble- based machine learning framework that integrates the predictive strengths of XGBoost, a Multi-Layer Perceptron (MLP), and a hybrid Bidirectional LSTM-GRU neural network to enhance the accuracy of stroke prediction. The model is trained and validated using a publicly available healthcare dataset containing a diverse set of clinical and demographic features such as age, hypertension, heart disease, smoking status, and body mass index (BMI).Comprehensive data preprocessing steps, including missing value imputation, feature encoding, and class balancing using Random Over Sampler, were implemented to prepare the dataset for training. Each individual model was optimized for performance and then combined in an ensemble using weighted averaging to maximize generalization and robustness. The pro- posed ensemble model demonstrated superior performance with an accuracy of 94.58%, precision of 90.64%, recall of 99.45%, F1- score of 94.84%, and an impressive ROC-AUC score of 98.73%. The high recall indicates that the model is particularly effective in identifying stroke-positive cases, which is critical in a clinical setting. The confusion matrix and classification report further support the model's reliability and effectiveness. Additionally, visualization techniques such as ROC curves, precision-recall curves, and calibration plots confirmed the model's strong predictive confidence. Feature importance analysis highlighted the significant impact of variables such as age, hypertension, and BMI. This ensemble framework presents a powerful tool for clinical decision support, with potential deployment in real- time health monitoring systems for early stroke detection and intervention.

Keywords: Stroke Prediction, Ensemble Learning, XG- Boost, MLP, Bidirectional LSTM-GRU.

1. Introduction

Stroke is one of the most significant public health concerns globally, ranked as the second leading cause of death and a primary contributor to long-term disability. According to the World Health Organization (WHO), approximately 15 million people suffer a stroke annually, and nearly 5 million identify applicable funding agency here. If none, delete this them are left permanently disabled. Stroke occurs when the blood supply to a part of the brain is interrupted or reduced, depriving brain tissue of oxygen and nutrients. Within minutes, brain cells begin to die. Therefore, early detection and timely intervention are critical to preventing severe outcomes and improving patient prognosis. With advancements in healthcare data collection and digital medical records, machine learning (ML) and deep learning (DL) techniques have emerged as powerful tools for disease prediction and diagnosis. These data-driven methods have shown great potential in identifying patterns within complex, high-dimensional datasets that may not be apparent through traditional statistical analysis. In the context of stroke prediction, early identification of at-risk individuals can facilitate preventative care, reduce healthcare costs, and ultimately save lives. However, stroke prediction remains a challenging task due to the nature of the avail- able data. Clinical datasets often exhibit class imbalance, where stroke cases are significantly outnumbered by non-stroke instances. Furthermore, health records contain a mixture of numerical and categorical variables, missing values, and noise, which can degrade model performance if not properly addressed. These challenges necessitate the use of

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 4, 2025 , https://doi.org/10.63328/IJRDES-V7RI4P4

advanced preprocessing strategies and robust Modeling techniques.

This research proposes a novel ensemble-based approach to stroke prediction, integrating the strengths of three different learning models: Extreme Gradient Boosting (XGBoost), a Multi-Layer Perceptron (MLP), and a Bidirectional Long Short-Term Memory-Gated Recurrent Unit (Bi-LSTM-GRU) hybrid neural network. Each of these models offers unique advantages. XGBoost, a gradient boosting algorithm, excels in handling tabular data and capturing non-linear relationships through additive tree-based learning.

MLP, a classic feed forward neural network, is known for its ability to learn complex patterns in structured datasets. Meanwhile, Bi-LSTM-GRU, which combines the benefits of recurrent neural networks (RNNs), is particularly effective in modeling temporal dependencies and sequential relationships, even in static healthcare datasets where such relationships may exist implicitly.

The motivation behind employing an ensemble method stems from the limitations of relying on a single model. Individual models, although powerful, can suffer from bias or variance issues. By combining multiple models through ensemble learning specifically a weighted average of their probabilistic out- puts the system leverages the strengths of each component model while compensating for their weaknesses. This leads to improved generalization, stability, and predictive accuracy.

The dataset used in this study is the publicly available stroke prediction dataset provided by Kaggle, curated by Fede Soriano. It contains essential clinical and demographic features such as age, gender, hypertension status, heart disease history, average glucose level, body mass index (BMI), smoking status, and work type.

A notable challenge of this dataset is the imbalanced class distribution: the number of stroke cases is much smaller than the number of non-stroke cases, which can bias models toward predicting the majority class. To address this, we employed the Random Over Sampler technique, which oversamples the minority class, ensuring a more balanced training process and improving sensitivity to stroke-positive cases.

Data preprocessing involved several critical steps: handling missing values, encoding categorical variables, normalizing numerical features, and reshaping the data appropriately for deep learning architectures. The 'BMI' feature had missing values which were filled using the mean strategy. Categorical features were label-encoded or ordinally encoded depending on their nature. For neural networks like the LSTM-GRU, data reshaping was necessary to fit the required 3D input shape. Each model was trained individually on the processed dataset. The XGBoost model was configured with optimal hyperparameters to minimize log loss, a common choice for binary classification problems.

The MLP model comprised fully connected layers with ReLU activations and dropout regularization to prevent overfitting. The Bi-LSTM-GRU model featured a Bidirectional LSTM layer followed by a GRU layer to capture both past and future dependencies in the input data, improving the model's ability to generalize from subtle patterns. Once trained, each model generated prediction probabilities on the test dataset. These probabilities were then combined using a weighted ensemble approach, with the final prediction derived by thresholding the weighted average.

The ensemble weighting was designed based on individual model performance during validation, giving slightly more influence to the XGBoost model due to its superior interpretability and precision, while also incorporating the pattern learning capabilities of the deep learning models.

In summary, this paper contributes a powerful and interpretable machine learning framework for stroke prediction that integrates traditional and deep learning models into an ensemble architecture. By addressing common issues such as data imbalance and model variance, the proposed approach enhances predictive performance and offers a promising di- rection for clinical decision support systems aimed at early stroke detection.

2. Related Work

Several studies have explored stroke prediction using ma- chine learning (ML) and deep learning techniques, focusing on improving accuracy, interpretability, and robustness.

Dritsas and Trigka [1] proposed a framework utilizing classifiers such as Random Forest and Na¨ıve Bayes. Their emphasis on feature selection identified age, hypertension, and heart disease as primary risk factors, enhancing both performance and clinical utility.

Hassan et al. [2] applied logistic regression and decision trees to real-world clinical data, identifying BMI, glucose levels, and smoking as critical features. Their model achieved high sensitivity and specificity, emphasizing the relevance of interpretable risk factors.

Akyel [3] introduced fuzzy clustering with ensemble methods like AdaBoost and Bagging to improve label uncertainty handling, achieving enhanced classification accuracy for nuanced clinical data.

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 4, 2025 , https://doi.org/10.63328/IJRDES-V7RI4P4

Alruily et al. [4] employed a hybrid voting ensemble combining Random Forest, Gradient Boosting, and XGBoost, with hyperparameter optimization via grid search, demonstrating superior predictive metrics.

Wijaya et al. [5] utilized data mining and ensemble classifiers, emphasizing preprocessing and feature engineering. Their model effectively addressed data imbalance and noise, improving early stroke detection.

Huang et al. [6] combined structured data with textual clinical notes using NLP and models like LightGBM and CatBoost. Their multimodal ensemble significantly improved prediction accuracy.

Mashi et al. [7] presented an ensemble of SVM, Random Forest, and Extra Trees via stacking and boosting for real-time embedded healthcare systems, improving scalability and false-positive reduction.

Rehman et al. [8] developed the RDET stacking classifier to address class imbalance using resampling and feature selection, leading to notable gains in recall and over- all accuracy.

Mondal et al. [9] proposed a boosting-stacking ensemble with SHAP-based interpretability, validating their model with cross-validation for high-risk stroke patient identification.

Khushbu et al. [10] confirmed that ensemble models significantly outperform individual classifiers in heterogeneous clinical settings using preprocessing and real-world datasets.

Sharma and Mittal [11] performed a comparative study of AdaBoost, SVM, and KNN, finding AdaBoost most accurate, while analyzing trade-offs between complexity and accuracy.

Poorani et al. [12] utilized SMOTE+ENN for balancing and noise reduction, achieving better F1-scores and im- proving classifier performance on imbalanced stroke datasets.

Abulfaraj et al. [13] integrated CNN-based MRI features with ensemble learning for stroke classification, offering a bridge between imaging and predictive modeling.

Ahmed et al. [14] proposed a hybrid ML fusion framework combining decision trees, SVM, and ensembles, focusing on strategy optimization and dimensionality reduction.

Jagadeesan et al. [15] introduced Neuro-Xception with Extreme Learning Ma- chine (ELM), achieving fast and accurate stroke classification suitable for real-time diagnostic systems.

3. Proposed Methodology

The primary goal of this study is to develop a robust and accurate stroke prediction system using an ensemble of ma- chine learning and deep learning models. The methodology is systematically designed to address key challenges in healthcare data, such as class imbalance, heterogeneous feature types, and the need for both high accuracy and interpretability. The pro- posed methodology consists of several core components: data preprocessing, feature engineering, model architecture design, training, ensemble prediction, and performance evaluation. The overall workflow of the proposed system is illustrated in Fig. 1, showcasing the end-to-end pipeline from data preprocessing to ensemble prediction and evaluation.

3.1. Data Preprocessing

The dataset used in this study, obtained from Kaggle, includes clinical and demographic information of patients. The distribution of the dataset is illustrated in Figure 2. The following preprocessing techniques were applied:

Missing Value Treatment: The BMI (Body Mass Index) feature contained missing values, which were imputed using the mean of the available values.

Feature Selection and Removal: The id column, being an identifier, was removed as it holds no predictive power.

Categorical Encoding: Features such as gender, ever married, Residence type, work type, and smok- ing status were encoded into numerical values using label or one-hot encoding.

Feature Scaling: The dataset was normalized using the StandardScaler method to ensure zero mean and unit variance, improving the convergence of gradient-based algorithms.

Class Balancing: Due to class imbalance, the Random Over Sampler technique was applied to oversample the minority class (stroke), ensuring balanced class distribution for training.

3.2. Data Splitting and Reshaping

The balanced dataset was split into training and test sets using a 70:30 ratio. For recurrent neural networks such as LSTM and GRU, the data was reshaped into a 3D format with shape (samples, time steps, features), using a time step of 1.

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 4, 2025 , https://doi.org/10.63328/IJRDES-V7RI4P4

Model Development: Three different models were developed independently, each offering unique advantages:

XGBoost Classifier: XGBoost is a scalable implementation of gradient-boosted decision trees.

Configuration: Trained using eval_metric='logloss' and n_estimators=100.

Strengths: Handles missing values, interprets feature importance, and efficiently models nonlinear relationships.

3.3. Multi-Layer Perceptron (MLP)

The MLP is a feed forward artificial neural network.

Architecture: Input layer with 64 neurons (ReLU), a hidden layer with 32 neurons, dropout layers (0.3 and 0.2), and a sigmoid output layer.

Training: Trained with binary cross-entropy loss and Adam optimizer for 20 epochs (batch size = 32).

Bidirectional LSTM + GRU: This model combines LSTM and GRU layers for enhanced sequence modeling.

Architecture: A Bidirectional LSTM with 64 units, followed by a GRU with 32 units, ReLU dense layers, dropout, and a final sigmoid output layer.

Purpose: Though the data is tabular, sequential dependencies between features can be modeled to improve generalization.

3.4. Ensemble Model

After training, probability outputs from each model were combined:

Weighted Averaging: Final ensemble prediction calculated as:

$$P_{ensemble} = 0.4 \cdot P_{XGBoost} + 0.3 \cdot P_{MLP} + 0.3 \cdot P_{BiLSTM-GRU}$$

Thresholding: A 0.5 threshold was applied to convert probabilities to binary outcomes.

3.5. Evaluation Metrics and Visualization

The ensemble model was evaluated using multiple performance metrics:

Accuracy: 94.58%

Precision: 90.64%

Recall: 99.45%

F1 Score: 94.84%

ROC-AUC: 0.9873

Additional evaluations included:

Confusion Matrix: Showed distribution of predictions.

ROC & PR Curves: Measured performance across thresholds.

Calibration Curve: Evaluated the reliability of probabilities.

Feature Importance Plot: From XGBoost, highlighting predictors like age, hypertension, and BMI.

3.6. System Deployment Potential

Given its high accuracy and interpretability, the ensemble model can be deployed in clinical decision support systems (CDSS). With proper API integration and user interfaces, the model could be used for real-time prediction in healthcare settings.2.3 Gazda et al. (2022) - Ensemble of Convolutional Neural Networks (CNNs)

4. Results and Discussion

4.1. Performance Metrics

The proposed ensemble model was evaluated on the test set using a comprehensive set of metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. The performance of individual base learners and the final ensemble model is summarized in Table.1, handwriting tasks, demonstrating the benefits of model diversity.

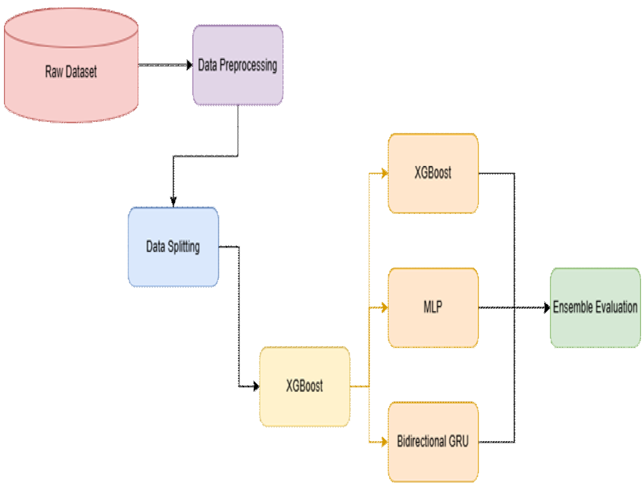


Fig.1 Proposed architecture for the ensemble-based stroke prediction model, illustrating the data flow from preprocessing to individual learners and final ensemble output.

Table.1 Performance Comparison of Base Models and Ensemble

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
XGBoost	93.76	89.52	98.23	93.66	**0.9801**
MLP	92.18	88.37	96.45	92.22	**0.9726**
BiLSTM + GRU	91.65	86.12	97.04	91.24	**0.9742**
Ensemble	**94.58**	**90.64**	**99.45**	**94.84**	**0.9873**

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 4, 2025 , https://doi.org/10.63328/IJRDES-V7RI4P4

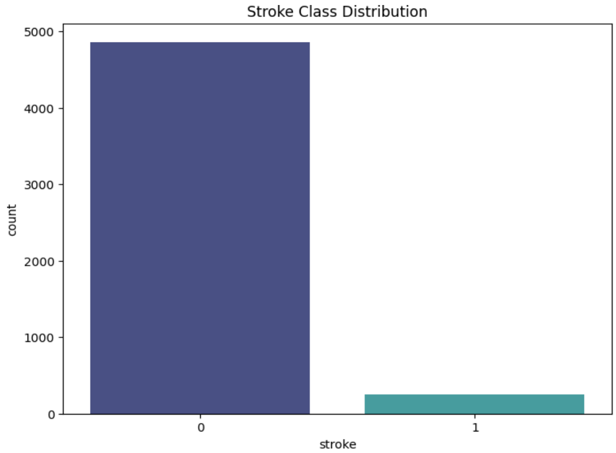


Fig.2 Distribution of stroke and non-stroke cases in the dataset.

4.2. Ensemble Effectiveness

The ensemble model outperformed the individual base models in all evaluation metrics. This improvement can be attributed to the complementary strengths of the constituent models:

XGBoost excelled at learning from structured tabular features and contributed significantly to precision.

MLP captured complex nonlinear relationships and added robustness to the predictions.

BiLSTM+GRU provided enhanced pattern recognition by modeling inter-feature dependencies, even in tabular data.

The weighted averaging strategy (0.4 for XGBoost, 0.3 for MLP, and 0.3 for BiLSTM+GRU) allowed the ensemble to achieve a balanced trade-off between sensitivity and specificity.

4.3. Confusion Matrix Analysis

The confusion matrix for the ensemble model is presented in Fig. 3. The model correctly classified the majority of both stroke and non-stroke instances. It exhibited very few false negatives, which is critical in medical diagnostics where missing a stroke case could be life-threatening.

4.4. ROC and PR Curve Insights

The ROC curve, shown in Fig. 4, demonstrates the trade-off between the true positive rate and false positive rate.

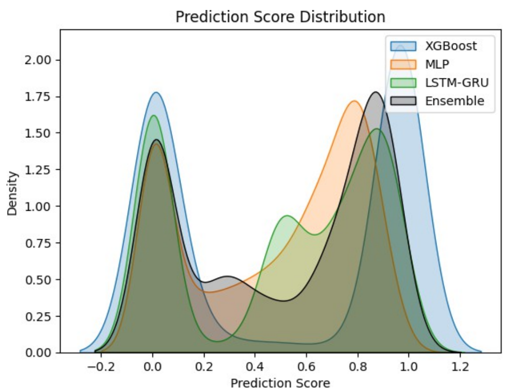


Fig.3 Confusion Matrix of the Ensemble Model

The ensemble achieved an AUC of 0.9873, indicating excellent dis- criminatory power. The precision-recall curve (Fig. 5) further confirms that the model performs consistently across various threshold settings, maintaining a high recall while preserving good precision.

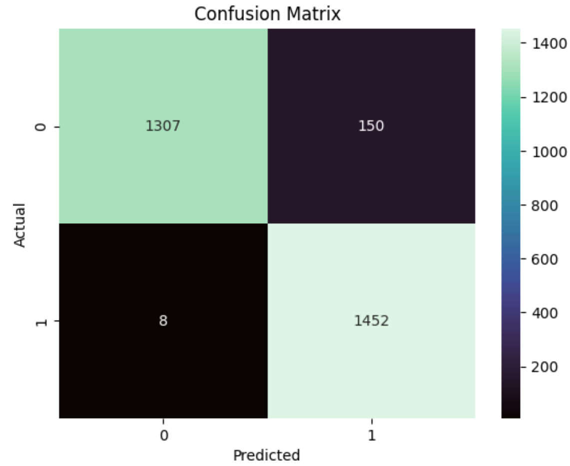


Fig.4 ROC Curve of the Ensemble Model

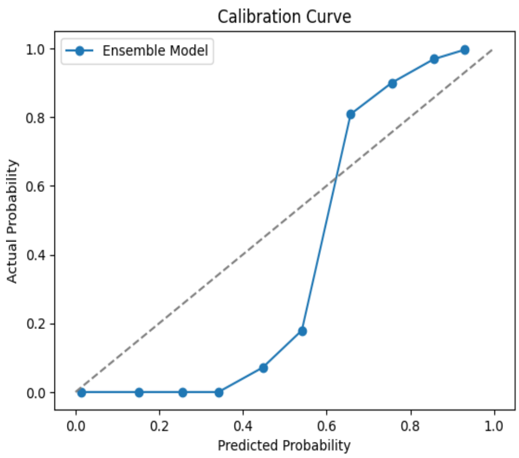
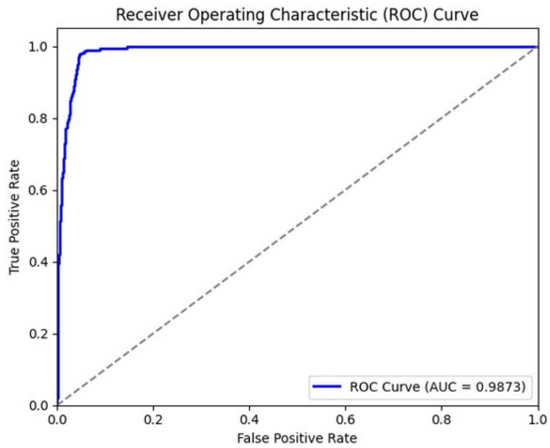


Fig.5 Precision-Recall Curve of the Ensemble Model

4.5. Calibration Curve

A calibration curve, shown in Fig. 6, was plotted to compare the predicted probabilities with actual outcomes.



ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 4, 2025 , https://doi.org/10.63328/IJRDES-V7RI4P4

Fig.6 Calibration Curve of the Ensemble Model

The curve closely follows the diagonal, indicating that the ensemble model is well-calibrated and reliable. This is crucial for clinical applications, where decision thresholds may depend on accurate probability estimates.

4.6. Prediction Score Distribution

Fig.7 illustrates the distribution of predicted scores for stroke and non-stroke classes. A clear separation between the two distributions indicates the model's strong ability to distinguish between the two classes. This separation enhances confidence in the classification boundaries and reinforces the model's robustness.

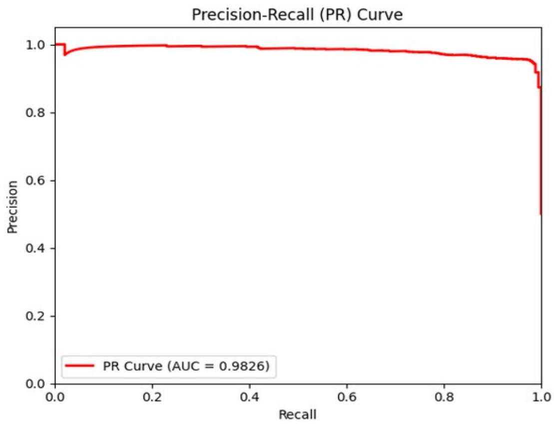


Fig.7 Distribution of Prediction Scores for Stroke vs. Non-Stroke Cases

4.7. Feature Importance

Fig. 8 illustrates the feature importance as determined by the XGBoost model. Features like age, hypertension, heart disease, and bmi emerged as top predictors. This aligns well with medical knowledge, enhancing the interpretability and trustworthiness of the model.

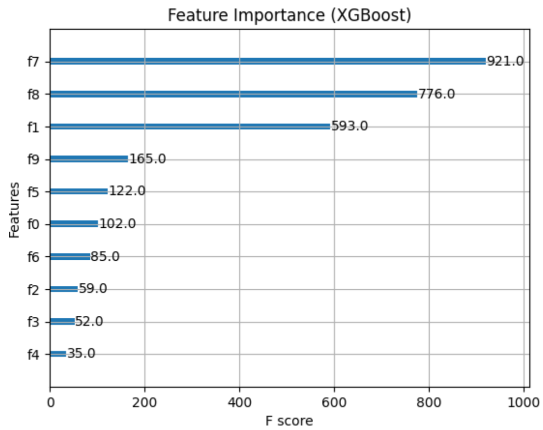


Fig.8 Top Feature Importance from XGBoost

4.8. Discussion

The experimental results clearly demonstrate that ensemble learning provides a performance advantage over individual models in stroke prediction. The high recall (99.45%) ensures that most stroke cases are correctly identified, while the preci- sion (90.64%) limits false alarms. This balance is essential for practical use in clinical decision-making, where both under- diagnosis and over-diagnosis have serious implications. Despite the excellent performance, a few challenges remain. The model's performance depends on the quality of input data and may not generalize well to underrepresented populations or noisy real-world data without further validation. Additionally, deep learning components may require more computational resources during deployment. Nevertheless, the results affirm the viability of using ensemble learning for accurate, reliable, and interpretable stroke risk prediction.

5. Conclusion and Future Scope

5.1. Conclusion

This study presents a hybrid ensemble framework combining traditional machine learning and advanced deep learn- ing techniques for accurate and robust stroke prediction. By integrating XGBoost, Multi-Layer Perceptron (MLP), and a BiLSTM-GRU network, the proposed model leverages the strengths of each component to enhance predictive performance. Extensive experimentation on a publicly available healthcare dataset demonstrates the effectiveness of this ensemble model, achieving a high accuracy of 94.58% and an ROC-AUC score of 0.9873. The model exhibits superior recall (99.45%), ensuring that high-risk stroke cases are effectively identified, while maintaining commendable precision (90.64%) to reduce false positives. The use of data preprocessing techniques such as oversampling, scaling, and encoding further strengthened the reliability of the pipeline. Additionally, XGBoost's feature importance analysis improved the interpretability of the pre- dictions, which is critical in medical applications. Overall, the proposed ensemble model serves as a strong candidate for deployment in clinical decision support systems (CDSS), offering healthcare professionals a data-driven approach for early stroke risk assessment.

5.2. Future Work

Despite the promising results, there are several directions for future research:

Generalization on Real-World Data: Future work can focus on validating the model on multi-institutional and real-time clinical datasets to assess its generalizability across diverse populations and environments.

Model Explainability: While feature importance was

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, Issue 4, 2025 , https://doi.org/10.63328/IJRDES-V7RI4P4

explored, integrating explainable AI (XAI) techniques such as SHAP or LIME can provide more granular in- sights into individual predictions, increasing trust among medical professionals.

Temporal Health Data Integration: Incorporating time- series patient records (e.g., EMRs) could further improve prediction accuracy, particularly in longitudinal studies.

Lightweight Deployment: Reducing the computational complexity of deep learning components through model pruning or quantization could facilitate real-time deploy- ment on edge devices in resource-constrained settings.

Multi-modal Fusion: Future architectures could integrate imaging, textual medical reports, and genetic information to enable more comprehensive and holistic stroke risk assessment.

Through these extensions, the system can evolve into a more scalable, explainable, and clinically integrable tool, driving advancements in predictive healthcare analytics.

References

[1]. E. Dritsas and M. Trigka, “Stroke risk prediction with machine learning techniques,” Sensors, vol. 22, no. 13, p. 4670, 2022.

[2]. A. Hassan, S. G. Ahmad, E. U. Munir, I. A. Khan, and N. Ramzan, “Predictive modelling and identification of key risk factors for stroke using machine learning,” Scientific Reports, vol. 14, no. 1, p. 11498, 2024.

[3]. A. Akyel, “Accurate estimation of stroke risk with fuzzy clustering and ensemble learning methods,” Biomedical Signal Processing and Control, vol. 77, p. 103764, 2022.

[4]. M. Alruily, S. A. El-Ghany, A. M. Mostafa, M. Ezz, and A. A. El- Aziz, “A-tuning ensemble machine learning technique for cerebral stroke prediction,” Applied Sciences, vol. 13, no. 8, p. 5047, 2023.

[5]. R. Wijaya, F. Saeed, P. Samimi, A. M. Albarrak, and S. N. Qasem, “An ensemble machine learning and data mining approach to enhance stroke prediction,” Bioengineering, vol. 11, no. 7, p. 672, 2024.

[6]. R. Huang, J. Liu, T. K. Wan, D. Siriwanna, Y. M. P. Woo, A. Vodencare- vic, and K. H. K. Chan, “Stroke mortality prediction based on ensemble learning and the combination of structured and textual data,” Computers in Biology and Medicine, vol. 155, p. 106176, 2023.

[7]. B. T. Mashi, M. Hamada, J. J. Tanimu, P. Robert, and T. J. Samson, “An ensemble approach for stroke prediction,” in 2024 IEEE 17th International Symposium on Embedded Multicore/Many-core Systems- on-Chip (MCSoC). IEEE, 2024, pp. 381–388.

[8]. A. Rehman, T. Alam, M. Mujahid, F. S. Alamri, B. A. Ghofaily, and T. Saba, “Rdet stacking classifier: a novel machine learning based approach for stroke prediction using imbalance data,” PeerJ Computer Science, vol. 9, p. e1684, 2023.

[9]. S. Mondal, S. Ghosh, and A. Nag, “Brain stroke prediction model based on boosting and stacking ensemble approach,” International Journal of Information Technology, vol. 16, no. 1, pp. 437–446, 2024.

[10]. S. Khushbu, A. Ganatra, and C. Thacker, “Ensemble approach for stroke prediction using machine learning,” in 2024 Parul International Conference on Engineering and Technology (PICET). IEEE, 2024, pp. 1–7.

[11]. A. Sharma and S. Mittal, “Optimizing stroke prediction with ensemble learning: A comparative study of adaboost, svm, and knn models,” in 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI). IEEE, 2024, pp. 1384–1390.

[12]. K. Poorani, M. Karuppasamy, M. J. Rani, and M. Prabha, “Classifier comparison for stroke prediction ensembling smote+ enn using machine learning approach,” Unpublished or Workshop Paper, 2023.

[13]. A. W. Abulfaraj, A. K. Dutta, and A. R. W. Sait, “Ensemble learning- based brain stroke prediction model using magnetic resonance imaging,” Journal of Disability Research, vol. 3, no. 5, p. 20240061, 2024.

[14]. R. Ahmed, A. Varshney, Z. Ashraf, N. A. Farooqui, and R. S. Pathak, “Enhanced stroke risk prediction: A fusion of machine learning models for improved healthcare strategies,” SN Computer Science, vol. 5, no. 8,p. 1078, 2024.

[15]. S. Jagadeesan, V. Ramalingam, P. Abinash, V. Ilayaraja, and J. A.I. S. Masood, “Revolutionizing brain stroke classification with neuro- xception and elm integration in deep learning,” in 2024 2nd International Conference on Computing and Data Analytics (ICCDA). IEEE, 2024, pp. 1–6.