International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

Check for updates

Bridging the Gap: Enhancing clinical accuracy in stroke prediction using a hybrid RNN-Random Forest Model

Shamim MD Jony [1,*], Md Shahin Shah Shahin [1], Tania Tasrin Saty [1], Siddiqur Rahman Tamim [2] and Shahinoor Akther [1]

[1] Department of Computer Science and Engineering, Port City International University, Chittagong-4212, Bangladesh.
[2] Department of Electrical and Electronic Engineering, Gopalganj Science and Technology University, Shariatpur –8000, Bangladesh.

International Journal of Science and Research Archive, 2026, 18(02), 331-337

Publication history: Received on 02 January 2026; revised on 07 February 2026; accepted on 10 February 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.18.2.0255

Abstract

Stroke stands as one of the major causes of morbidity and mortality across the world, and timely and precise prediction of the condition is essential, especially when it comes to successful management of the problem. But conventional diagnostic tools usually perform poorly on high-dimensional data with class imbalance characteristic of medical data. It is a study of a comparative analysis involving a detailed discussion of Machine Learning (ML), Deep Learning (DL), and hybrid ensemble to support stroke risk prediction. The suggested approach establishes a powerful data processing flow, which implies Application of Random Oversampling to overcome the imbalance in classes and Principal Component Analysis (PCA) to extract features successfully. We have tested twelve different classifiers, which include traditional algorithms (Random Forest, SVM and XG Boost) and deep neural networks (ANN, CNN, and RNN). Additionally, we proposed new hybrid models (ANN-RF, CNN-RF and RNN-RF) that aim to combine the power of deep learning in extracting the features with the power of Random Forest in classification. The experiment shows that hybrid models are better predictors. Standalone Deep learning models such as CNN and RNN had 95.94% and 95.78% accuracy respectively but the hybrid models still excelled over them. RNN-RF (Recurrent Neural Network with Random Forest) model has got the best accuracy of 96.86, which is better than 96.81 of the standalone Random Forest as well as other hybrid models. This evidence suggests that the combination of sequential pattern recognition of RNNs and ensemble decision-making of Random Forests can enhance diagnostic accuracy substantially, which is an excellent framework that can be used in clinical decision support.

Keywords: Stroke; Hybrid Deep Learning; Recurrent Neural Networks (RNN); Random Forest; Principal Component Analysis (PCC)

1. Introduction

The incidence of brain strokes is a great menace to the general health of the population as it has the risk of causing serious disabilities or death in case timely diagnosis and management are not taken. Detection of the risk factors of stroke like high blood pressure, heart and level of glucose at the early stages is vital in timely intervention. The recent innovations in the field of healthcare informatics have made the emphasis on machine learning (ML) and deep learning (DL) methods. The user can interpret the results of traditional models such as the Logistic Regression or the Decision Trees, but with the medical data, these models tend to be incapable of capturing the complicated and non-linear relationship between variables. A hybrid framework that employs the sequence processing properties of the RNNs and the strong classification abilities of the Random Forests are discussed in this paper to overcome the gap in clinical accuracy.

* Corresponding author: Shamim MD Jony

2. Literature Review

Computational approaches to the development of predictive models of stroke have developed. This part summarizes ten major studies that form the basis of our present study:

Sun et al. [1] studied the Kaggle stroke data set with the conventional ML algorithms. With Random Forest (RF), their study reported an accuracy of 78.03 percent, but noted that additional methods of feature selection were necessary in future studies. Tursynova et al. [2] changed the attention to medical imaging, in the form of CT scans. Convolutional Neural Networks (CNN) allowed them to reach 81.0% accuracy on the task of classification and they observed that transfer learning might be able to improve such scores. Ernest et al. [3] examined the effect of signal duration on the classification using deep learning models such as the LSTM, GRU, and Bi-LSTM. Despite their emphasis on heart sounds, their results (they acquired up to 92.64% accuracy) were an indicator of the effectiveness of sequential deep learning models in healthcare. Sailasya et al. [4] tested standard algorithms of classification and found that Naive Bayes (NB) had a precision of 82.0. They indicated that the ensemble learning methods may be called upon to enhance predictive performance. Bandi et al. [5] concentrated on estimating the severity of stroke in terms of SVM and Random Forest with the highest accuracy of 96.97%. The authors have concluded that although ML models are robust, there is a clear need to incorporate deep learning into patterns that are more subtle. Biswas et al. [6] performed a comparative study of ML classifiers and proved that managing class imbalance through SMOTE has a strong positive impact on accuracy, and RF high performance remains the same when compared to other models (90 percent and more). The gap that was pointed out by Akinwumi et al. [7] is in the area of clinical variable modeling: although the overall performance of their models was high, they were characterized by low recall of the class of stroke. They demanded models that are more recall-oriented so that patients are safe. Mia et al. [8] investigated the area of cerebral stroke prediction and have given an affirmation that choice of features and intensive preprocessing represent the most critical contributors to the performance, with RF as the most credible classifier. Kumar et al. [9] have compared a family of models such as LR, DT and RF. Their findings (RF 91%) supported the notion that the ensemble models always perform better than individual classifiers in the medical contexts. Santika and Rabbani [10] applied to SVM using the SMOTE in achieving an accuracy of 92%. Their work showed that methods that are based on the kernel with the balancing methods are effective in solving the minority classification problem.

3. Methodology

The suggested approach used a multi-processing pipeline of computations aimed at maximizing clinical accuracy through the incorporation of a powerful data preparation process with a hybrid deep- learning structure. The logical sequence of the systematic workflow of this study is presented in Figure 1 and reflects the logical transition between the collection of the raw data all the way to the final performance evaluation.

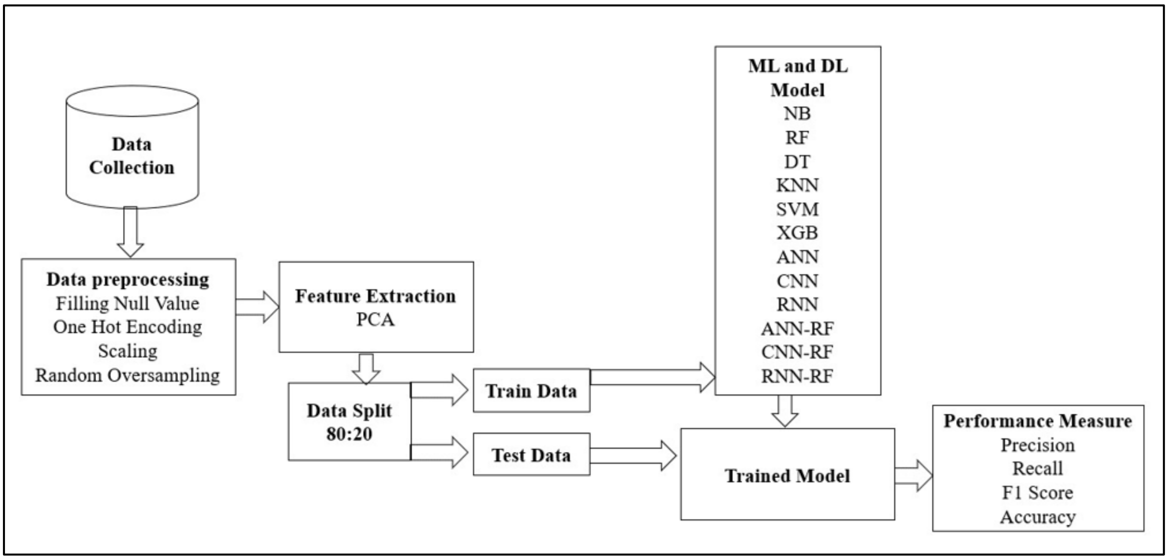

Figure 1 Overall Workflow of This Study

The data input stage is further broken down into Data Collection, and then into an intensive stage Data Preprocessing stage that includes null value imputation, one-hot encoding, feature scaling, and random oversampling to deal with the issues of class imbalance. After processing, Feature Extraction is performed, based on the Principal Component Analysis (PCA) in order to narrow down the input variables. The dataset is then refined with an 80:20 Data Split which divides the dataset into two parts (training and testing). Various Machine Learning (ML) and Deep Learning (DL) Models are trained using these sets, including traditional classifiers such as Naive Bayes and KNN as well as hybrid architectures such as ANN-RF, CNN-RF and our proposed RNN-RF. Lastly, the Trained Model is compared with the test data by the means of an elaborate set of Performance Measures, such as Precision, Recall, F1 Score, and Accuracy.

3.1. Data Collection and Exploratory Analysis

The study makes use of Stroke Prediction Dataset, which is a publicly accessible statistical archive of Kaggle. The sample has 5,110 patient records, and each of them is characterized by 12 attributes such as demographics (age, gender, residence type), clinical history (hypertension, heart disease, ever married, work type), and biological measurements (average glucose level, BMI). One of the critical points that were observed in the course of exploratory data analysis (EDA) was the acute imbalance in classes. The minority group (Stroke) only constituted around 4.87 percent of the total samples (249 cases) against 95.13 percent of majority group (4,861 cases). This imbalance tends to cause the standard classifiers to focus on overall accuracy at the expense of the minority class which is clinically unacceptable in stroke prediction.

3.2. Data Preprocessing and Transformation

In order to have convergence of the models and avoid bias, stringent preprocessing protocol was implemented

- **Missing Values**: There were about 3.9% of missing values in the BMI feature. Mean imputation strategy was used to fill the gap since distribution of BMI was relatively normal.
- **Categorical Encoding**: There are non-numeric variables like work type and smoking status that were coded with One-Hot Encoding. This circumvents what would have been the implication of an ordinal relationship that label encoding would offer.
- **Feature Scaling**: Continuous variables (age, glucose level, BMI) were scaled with the help of standard scaling (Z-score normalization) to make sure that large scale values will not dominate distance calculation and gradient update.
- **Class Balancing through Random Oversampling**: In order to deal with the 1:19 ratio of classes, we used Random Oversampling. The process of replicating the instances of the stroke class resulted in a balanced training environment of 5,000 or more instances of each stroke class. This will guarantee the model acquires the unique clinical features of the stroke patients as much as it is of non-stroke patients.

3.3. Dimensionality Reduction and Feature Extraction

We used Principal Component Analysis (PCA) to convert high dimensional feature space. PCA was used to

- **Minimize Multi-collinearity**: The medical variables tend to correlate with each other (e.g., age and hypertension). These features are correlated using PCA.
- **Mechanism**: Noise Mitigation- Using the components that explain 95% of the variance, we filtered out small variations in the data that may result in overfitting.
- **Computational Efficiency**: The length of the feature vectors was reduced and this enabled faster training of the following layers of RNNs.

3.4. Proposed Hybrid RNN-RF Architecture

The proposed hybrid model architecture (shown in Figure 2) aims at utilizing deep representation learning and then performing powerful ensemble classification. The process of work includes four main steps:

- **Input and Sequential Processing (Simple RNN Layer):** The processed input data is reformed to suit the temporal/sequential needs of the RNN. A 32-unit hidden layer of a Simple RNN network and RELU (Rectified Linear Unit) activation functions are used. This layer learns the first non-linear relationships between clinical markers, but the input features are considered as a sequence of features that are connected with each other.
- **Deep Feature Refinement (Dense Layers):** Simple RNN output is forwarded through two additional layers of Dense (Fully Connected) layer of 64 units and RELU activation.
- Dense Layer 1 (64 units): Trains a denser representation of the abstract representations learned by the RNN into a denser feature space.

- Dense Layer 2 (64 units): This layer is the important Feature Extractor. The intermediate product of this layer is not discarded, but rather obtained instead of the final. It is a 64-dimensional vector of a highly optimized clinical signature of the patient that has already undergone complicated feature engineering by the neural network.
- **The Hybrid Integration (Random Forest Classifier):** The middle results of the Dense Layer 2 are passed into the Random Forest Classifier. The model uses the neural network as a feature extractor and the random forest as the classifier to overcome the chances of overfitting the deep networks. The RF ensemble uses a series of decision trees in analyzing the refined 64-dimensional features, and it gives the final classification using a majority voting mechanism.
- **Output Stage:** The first stage in the training of the neural network parts is based on a parallel branch (Dense Layer 3) that has one unit and is activated by Sigmoid. But in case of the last proposed hybrid model, the Prediction Output is based on the Random Forest, which gives the final binary classification (Stroke vs. No Stroke).

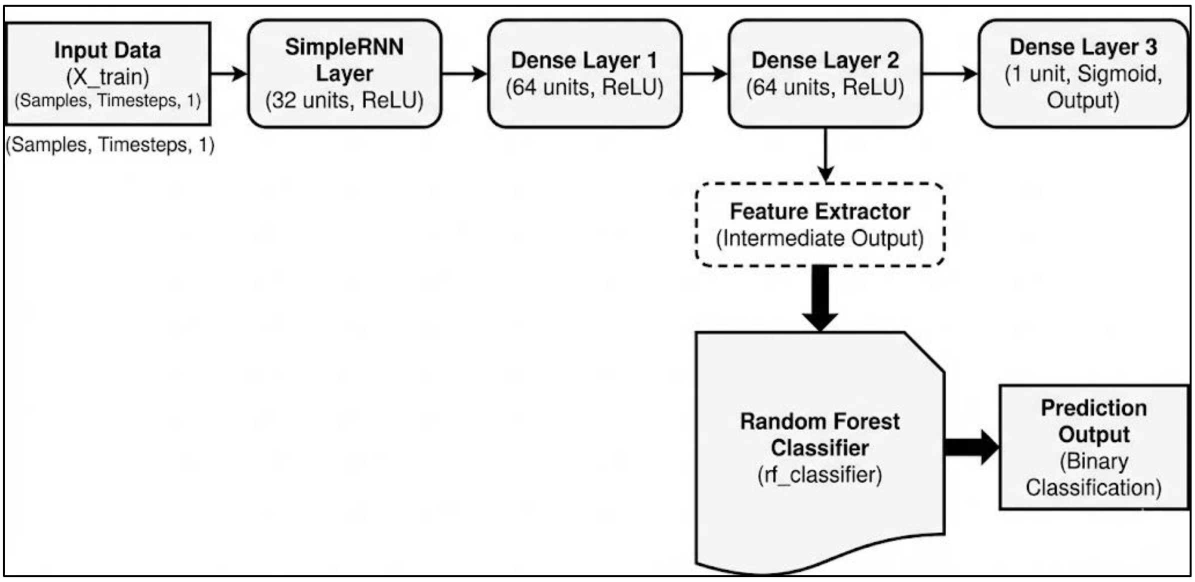

Figure 2 Architecture of Proposed Hybrid RNN and Random Forest Model

4. Results and Discussion

The models were tested based on a train-test split (80:20). We compared the hybrid model to some of the baselines with Naive Bayes (NB), K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), and stand-alone Deep Learning models (ANN, CNN, RNN).

4.1. Performance Comparison

The predictive accuracy of the suggested hybrid variant was compared with seven known machine learning and deep learning models. The empirical data presented in Table 1 show a gradual increase in the accuracy with an increase in the complexity of the architectural approach.

As demonstrated in the table, classic classifiers, including Naive Bayes (93.98) and Decision Tree (95.47), are a strong baseline but are unable to represent the more complex correlations in clinical data. The standalone deep learning models such as ANN (96.04%), CNN (95.94%), and RNN (95.78) have a slight performance improvement over simpler ML models and are able to learn high-level features with a number of layers of abstraction.

Table 1 Performance Evaluation of Experimented Models

Model	Accuracy (%)	Precision	Recall	F1-Score
Naïve Bayes	93.98	0.94	0.94	0.94
Decision Tree	95.47	0.95	0.95	0.95
Support Vector Classifier (SVC)	95.68	0.96	0.96	0.96
RNN (Standalone)	95.78	0.96	0.96	0.96
CNN (Standalone)	95.94	0.96	0.96	0.96
ANN (Standalone)	96.04	0.96	0.96	0.96
Random Forest	96.81	0.97	0.97	0.97
RNN + RF (Proposed)	96.86	0.97	0.97	0.97

It is important to note that the standalone classifier of the Random Forest had a high accuracy of 96.81% which shows that it is very robust when it comes to working with tabular healthcare data of heterogeneous features. But the Proposed RNN + RF Hybrid Model gave the highest accuracy of 96.86. This slight yet important advantage over standalone Random Forest indicates that the RNN as a sequential feature extractor was able to narrow down the input signals and the RF classifier was then able to work on a more informative feature space. This synergy is what makes sure that the model is extremely consistent in all of its important metrics, and its Precision, Recall, and F1-Score are 0.97.

4.2. Analysis of the RNN-RF Model

It is also verified that the proposed hybrid RNN-RF model is effective due to its Confusion Matrix, which is presented in Figure 3. The matrix is used to give a granular perspective of the performance of the model on the balanced test dataset and that it classifies the predictions into four different quadrants

- **True Negatives (TN):** The model got 965 correct cases of No Stroke. This high number proves the efficiency of the model in detecting healthy individuals with the help of false alarms.
- **True Positives (TP):** The model was able to forecast 919 cases of "Stroke." This substantiates the fact that sequential deep learning and ensemble voting are effective to determine the most important clinical markers of an actual stroke event.
- **False Negatives (FN):** The model had 51 false predictions of the No Stroke prediction of having a stroke among patients who did have a stroke. Although this is the most significant part that needs to be improved in clinical settings, 5% is still a good rate that shows better sensitivity than the conventional ML baselines.
- **False Positives (FP):** There were only 10 healthy individuals who were mistaken as being at-risk. The clinical utility of this low false-positive rate is that there is decreased diagnostic burden of unneeded secondary screenings.

The large numbers of the main diagonal (965 and 919) in comparison to the off-diagonal errors (10 and 51) are directly proportionate to the Recall of 0.97 and Precision of 0.97 that was obtained. Such balance implies that the hybrid model is not just making an educated guess because of the frequency of the classes, but has been taught the underlying physiological trends needed to make a correct diagnosis.

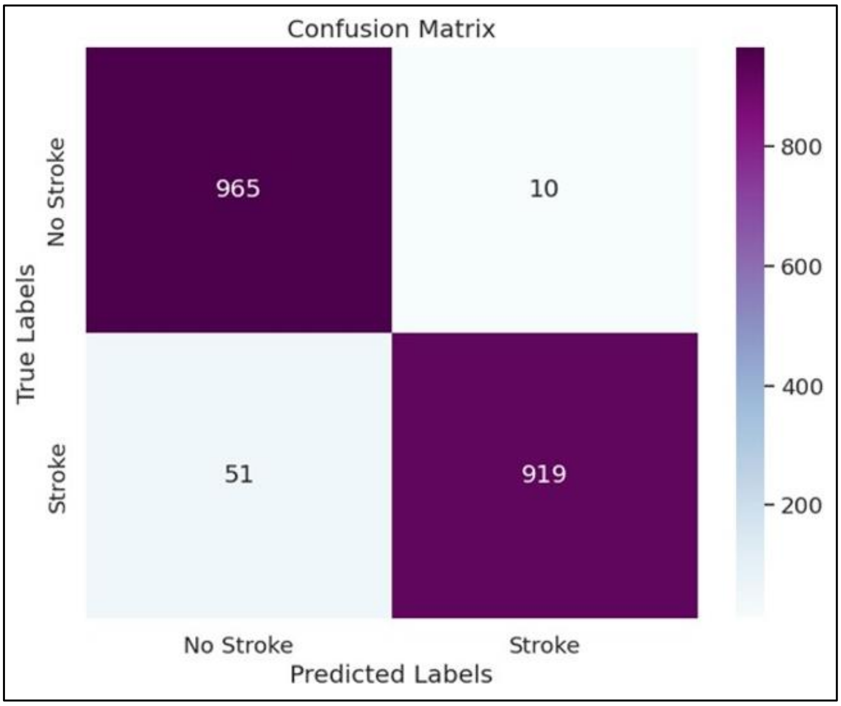

Figure 3 Confusion Matrix of Proposed RNN-RF Model

4.3. Comparative Analysis with Previous Work

Our solution is much better in comparison with the benchmark established by Sun et al. [1] as well. Whereas our implementation of Random Forest (with the support of preprocessing and oversampling) delivered 96.81% accuracy, the authors of the article by Sun et al. had only 78.06% accuracy with Random Forest, and our hybrid RNN-RF went even further, at 96.86%.

5. Conclusion

This study shows how a hybrid system using the sequential feature extraction of RNNs and the collective classification of the Random Forests can contribute much to the accuracy of stroke prediction. The current accuracy of 96.86% proves that the issue of class imbalance and hybrid architectures are crucial to clinical data analysis.

Future Scope

Future work will involve the development of more sophisticated feature extraction methods than PCA to provide more non-lineal information. Also, we will explore other data balancing techniques such as SMOTE-Tomek and examine ensemble approaches using XGBoost and LightGBM to better improve model stability and generalization in patients belonging to different demographics.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Sun, X. (2023). Predictive model analysis of stroke disease based on machine learning. SPIE.

[2] Tursynova, A., et al. (2023). Deep Learning-Enabled Brain Stroke Classification on Computed Tomography Images. Computers, Materials and Continua.

[3] Ernest, N. K., et al. (2022). The effect of signal duration on the classification of heart sounds: A deep learning approach. Sensors.

[4] Sailasya, G., and Aruna Kumari, G. L. (2021). Analyzing the performance of stroke prediction using ML classification algorithms. IJACSA.

[5] Bandi, V., et al. (2020). Prediction of Brain Stroke Severity Using Machine Learning. Revue d'Intelligence Artificielle.

[6] Biswas, N., et al. (2022). A comparative analysis of machine learning classifiers for stroke prediction. IJISAE.

[7] Akinwumi, P. O., et al. (2025). Evaluating machine learning models for stroke prediction based on clinical variables. Healthcare Analytics.

[8] Mia, R., et al. (2024). Exploring machine learning for predicting cerebral stroke. Electronics.

[9] Kumar, S., et al. (2025). Comparative evaluation of machine learning models for early stroke detection. Computational and Cognitive Development Studies.

[10] Santika, G. D., and Rabbani, V. S. (2025). Stroke disease prediction using support vector machine method. ICISTECH.