

Original Article

LACO-GAI: A Laplace Copula Generative AI Network Model for Accurate Stroke Disease Prediction

R. Sujitha¹, S. Sivakumar²

^{1,2}Research Department of Computer Science, Cardamom Planters' Association College- Affiliated to Madurai Kamaraj University, Bodinayakanur, Theni, Tamil Nadu, India.

¹Corresponding Author : sujitha.r@cpacollege.org

Received: 21 April 2026

Revised: 08 August 2026

Accepted: 20 August 2026

Published: 30 September 2026

Abstract - Stroke is major wellbeing condition which creates serious hazard to human health and cause of bereavement and disability universal. Several risk factors contribute to its development, including hypertension, diabetes, tobacco use, raised cholesterol levels, and obesity. Conventional methods have been used to predict stroke, however it frequently face demands in achieving enhanced accuracy as well as minimizing errors. To overcome the demands mentioned above, a new model named the Laplace Copula Generative Artificial Intelligence (LACO-GAI) is introduced to deliver precise stroke predictive outcomes with minimal time and error rate. Initially, patient data samples are gathered from relevant datasets during the data acquisition stage. Input information is first transferred to input level, then forwarded to first hidden layer, where pre-processing takes place. This includes handling missing values through Multiple Imputations by Chained Equations (MICE) technique and removing anomalies or outliers using High Contrast Subspace method to compromise model performance. In second hidden level, is executed by Laplace Copula Multivariate feature ranking method, which identifies and retains the most relevant features. Subsequently, the third hidden layer handles the stroke prediction phase by employing Logistic Regression as a binary classification model to accurately forecast the presence or absence of stroke based on a selected set of relevant features. A probability is then used to classify the patient as either stroke or not. To optimize model performance during training, error backpropagation is employed for weight adjustment, and hyperparameters are fine-tuned using the Adaptive Moment Estimation optimizer. Experimental assessment of LACO-GAI model is conducted using various assessment parameters. Quantitatively examined outcomes expose LACO-GAI obtains superior accuracy in stroke disease prediction by lesser time consumption compared to DL techniques.

Keywords - High contrast subspace method, Laplace Copula Multivariate feature ranking method, Logistic Regression, Adaptive moment estimation optimizer, Multiple Imputations by Chained Equations (MICE), Stroke disease prediction, Generative Artificial Intelligence.

1. Introduction

Stroke is hazardous, severe illness which mainly involves people, however harmful diet is to increase of strokes in adolescence. Timely and accurate prediction of stroke diseases is pivotal in providing the treatment plans that lessen severe effects. The development of robust, accurate, and efficient stroke prediction systems has become a necessary focus in modern healthcare analytics. A Gated Recurrent Units (GRU) model was developed in [1] for achieving high stroke prediction accuracy and minimal time. The designed GRU model failed to effectively combine GRU by convolutional layers to improve its prognostic abilities. However, it did not improve applicability or strong performance during testing and validation across diverse datasets.

BiLSTM was introduced [2], aimed to achieve high efficiency, with a testing accuracy for stroke disease

prediction. However, BiLSTMs struggled when trained on large datasets due to high model complexity and large number of parameters Explainable artificial intelligence method was designed [3] for accurate stroke illness prediction based on DL and ML models. A modified Gated Recurrent Unit (GRU) model was introduced in [4] for accurate stroke prediction by removing the irrelevant values. ML-based methods were designed in [5] to predict disease and reduce severity of ischemic stroke. A hybridization of optimal FS as well as advanced classification methods were presented [6] to enhance result of stroke forecast.

Ensemble CNN model was designed [7] for identifying stroke patients with high testing accuracy. Random forest model was developed in [8] for forecasting stroke based on FS. Hybrid deep learning framework was developed in [9] by utilizing the Vision Transformer (ViT) for accurate and



interpretable stroke detection. An intelligent learning system was developed in [10] with the aim of stroke prediction that integrates an autoencoder through LDA. An AI-based framework was introduced in [11] for stroke diagnosis. ML method was designed in [12] to detect stroke patients with acute neurological signs. A weighted voting ensemble ML was developed [13] to optimize the result of brain stroke detection. Dense Stacking Ensemble (DSE) model was designed in [14] for premature stroke recognition to enhance patient results.

DL network method was introduced [15] to predict stroke-optimistic instance data through elevated reliability that enhances simplification of stroke prediction.

1.1. Novelty and Contributions of the Article

- Novel system called the LACO-GAI model, into the Generative Artificial Intelligence model is designed to enhance stroke disease prediction.
- Novelty of data pre-processing and select significant features by employing Laplace Copula
- A multivariate feature ranking method is used to enhance stroke disease prediction. This process helps in reducing the time consumption of stroke disease prediction.
- A novel method of Logistic regression functions is to classify stroke no stroke patients for improving stroke disease prediction accuracy. The gradient moment optimization algorithm is employed for tuning the hyperparameter of the model to further reduces error and increase accuracy in of stroke forecast.
- At last, a comprehensive assessment using several performance metrics is carried out to show the superior effectiveness of LACO-GAI model compared to conventional deep learning techniques.

2. Related Works

ML approach was designed [16] for stroke hazard forecast. A new meta-learning framework was designed in [17] to enhance stroke disease predictive performance that hybridizes advanced hybrid resampling techniques and XAI. A hybrid DL model was developed [18] to improve stroke prediction patient outcomes with lesser healthcare costs.

An ESPESM was developed [19] by measuring feature engineering as well as ensemble categorization. A novel hybrid system was developed in [20] for efficient brain stroke prediction by selecting the significant feature and classifier. A novel CBDA-ResNet50 DL method was introduced in [21] to improve forecast accuracy of stroke and increase patient outcomes with minimal healthcare costs.

A new ensemble method was developed in [22] for predicting stroke. It proved to effectual device for stroke prediction by elevated accuracy. Statistical analysis and ML classification methods were developed in [23] to identify key stroke predictors. Pair of RF and DT algorithms was developed in [24] using a hard voting technique for brain stroke forecast. Neural networks model was introduced in [25] for predicting stroke diseases with minimal time consumption and enhancing patient outcomes. A projection statistical FE with ML algorithm was developed in [26] to recognize stroke patients as well as hazard factors.

Transfer learning model was developed in [27] to detect stroke patients with higher accuracy by minimize FNR and FPR. However, advanced DL was not employed to improve effect of stroke forecast. ML methods were presented [28] to forecast stroke severity. A stacking ensemble method was introduced in [29] by implementing the Principal Component Analysis (PCA) for predicting stroke occurrences. Scalable machine learning technique was developed in [30] for efficient stroke disease prediction by combining big data analytics concepts.

2.1. Research Gap

Stroke is hazardous, severe sickness that involves people, but harmful diet is to improve of strokes in adolescence. A conventional Gated Recurrent Units (GRU) model, BiLSTM, provides insufficient accuracy. Optimization technique not achieve higher efficiency. In every existing method is unable to focus on computation cost minimization and time minimization. To concern these issues, a new technique called LACO-GAI is developed for delivering precise stroke predictive outcomes with minimum time.

Table 1. Comparison table

S.No	Method	Contribution	Merits	Demerits
1.	Gated Recurrent Units (GRU) model	A Gated Recurrent Units (GRU) model was designed for achieving high stroke prediction accuracy and minimal time.	Accuracy was improved	Gated Recurrent Units (GRU) model did not improve applicability or strong performance during testing and validation across diverse datasets.
2.	Bidirectional Long Short-Term Memory (BiLSTM)	BiLSTM was used to achieve high efficiency, with a testing accuracy for stroke disease prediction.	Time was minimized by BiLSTM.	BiLSTMs struggled when trained on large datasets due to high model complexity and large number of parameters.

3.	Explainable artificial intelligence model	Explainable artificial intelligence model was employed for accurate stroke disease prediction based on deep learning and machine learning models.	Precision was higher	Failed to consider either expanding the sample size or including data from multiple provinces
4.	Modified Gated Recurrent Unit (GRU) model	A modified Gated Recurrent Unit (GRU) model was used for accurate stroke prediction by removing the irrelevant values.	Accuracy was improved by a designed method	GRUs was not efficient, especially for very complex and long sequences
5.	Machine learning-based models	Machine learning-based models were designed to predict the disease and reduce the severity of ischemic stroke.	Recall was higher	Error rate reduction in the disease prediction remained unaddressed
6.	Hybridization of optimal feature selection and advanced classification techniques	A hybridization of optimal feature selection and advanced classification techniques were used to enhance the performance of stroke prediction.	Time was reduced by designed techniques.	Deep learning algorithms and post-processing algorithms were not employed to perform efficient diagnosis of different neurological diseases.
7.	Ensemble Convolutional Neural Network Model	Ensemble Convolutional neural network model was used for identifying stroke patients with high testing accuracy.	Accuracy was enhanced by a designed method.	Models did not achieve better results, especially with larger and more diverse datasets.
8.	Random forest model	Random forest model was used for predicting stroke on the healthcare stroke dataset based on feature selection algorithms.	Error rate was lesser.	Random forest model was unable to predict stroke occurrences within an adequate time.
9.	Hybrid deep learning framework	Hybrid deep learning framework was utilizing the Vision Transformer (ViT) for accurate and interpretable stroke detection.	Accuracy was improved by designed method.	Computational complexity of the model was not reduced.
10.	Intelligent learning system	An intelligent learning system was used for stroke prediction that integrates an autoencoder with a Linear Discriminant Analysis (LDA)	Recall was higher	Error handling process of the model was not enhanced

3. Proposal Methodology

Brain stroke is severe problem which directs to injuring intellectual cells, owing to blood stream to brain abruptly stop.

To address these issues, a novel model called LACO-GAI is introduced for achieving high precision and speed in classifying the normal cases and stroke patients with minimal

error. It utilizes advanced AI and optimized feature selection to examine and predict cerebral caress incidence at entity.

Figure 1, given above, illustrates overall flow processing diagram of LACO-GAI model for accurate stroke prediction from the given dataset.

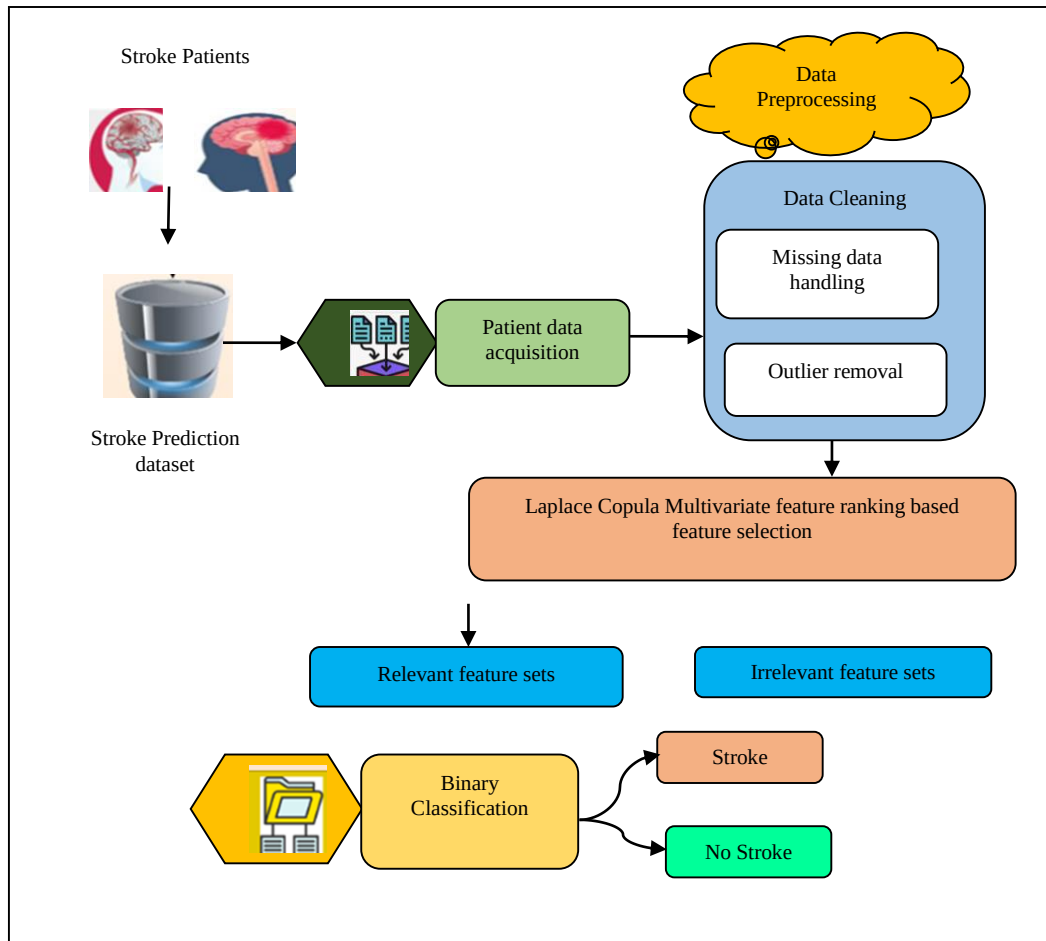


Fig. 1 Overall structural design of proposed LACO-GAI model

3.1. Data Acquisition

The primary step of LACO-GAI model is to perform the data collection, which involves gathering the raw patient data samples from the Kaggle public dataset. In the proposed LACO-GAI model, numbers of raw data samples are gathered from the Stroke Forecast Database at <https://www.kaggle.com/datasets/teamincirbo/stroke-prediction>. A database includes 15,000 records and 22 features of stroke prediction dataset. It includes a wide range of features, from biometric data like glucose levels and BMI, and so on. The main aim of dataset is to enhance as well as interpret ML or DL models capable of predicting stroke events.

3.2. Generative Artificial Intelligence Model Enabled Stroke Prediction

The second process of proposed LACO-GAI model is the accurate stroke prediction, which utilizes the Generative Artificial Intelligence model called Convolutional Deep Belief Networks. This is types of deep learning classifier to achieve fast and highly accurate stroke prediction with lesser computational overhead and error rate. The proposed deep learning classifier hybridizing the capability of DBNs and

strength of CNNs, the Convolutional Deep Belief Networks, is employed as a Hierarchical Discriminative AI model to efficiently enhance both feature analysis and stroke prediction.

As shown in the Figure 2, the overall structure have $\{D_i, Y_k\}$ D denotes training data samples $D_i = \{D_1, D_2, \dots, D_n\}$ gathered as of database and output ' Y_k ' symbolizing its output for stroke prediction. Through layer-wise pre-training, every level of network is trained independently to learn the input data. First hidden layer obtains the input data samples from the previous visible layer. The weight element of the matrix is constructed by means of the connection among visible as well as hidden unit h_j . Probabilistic max-pooling layer in RBM used to minimize dimensionality while preserving significant information. In the fine-tuning phase, entire network is optimized jointly using the backpropagation algorithm. This stage changes the model parameters and hyperparameters to reduce prediction error, thereby improving overall accuracy and robustness in stroke disease prediction tasks. Let us consider the input information matrix is provided as input to the input layer of a deep learning architecture. It is formulated as,

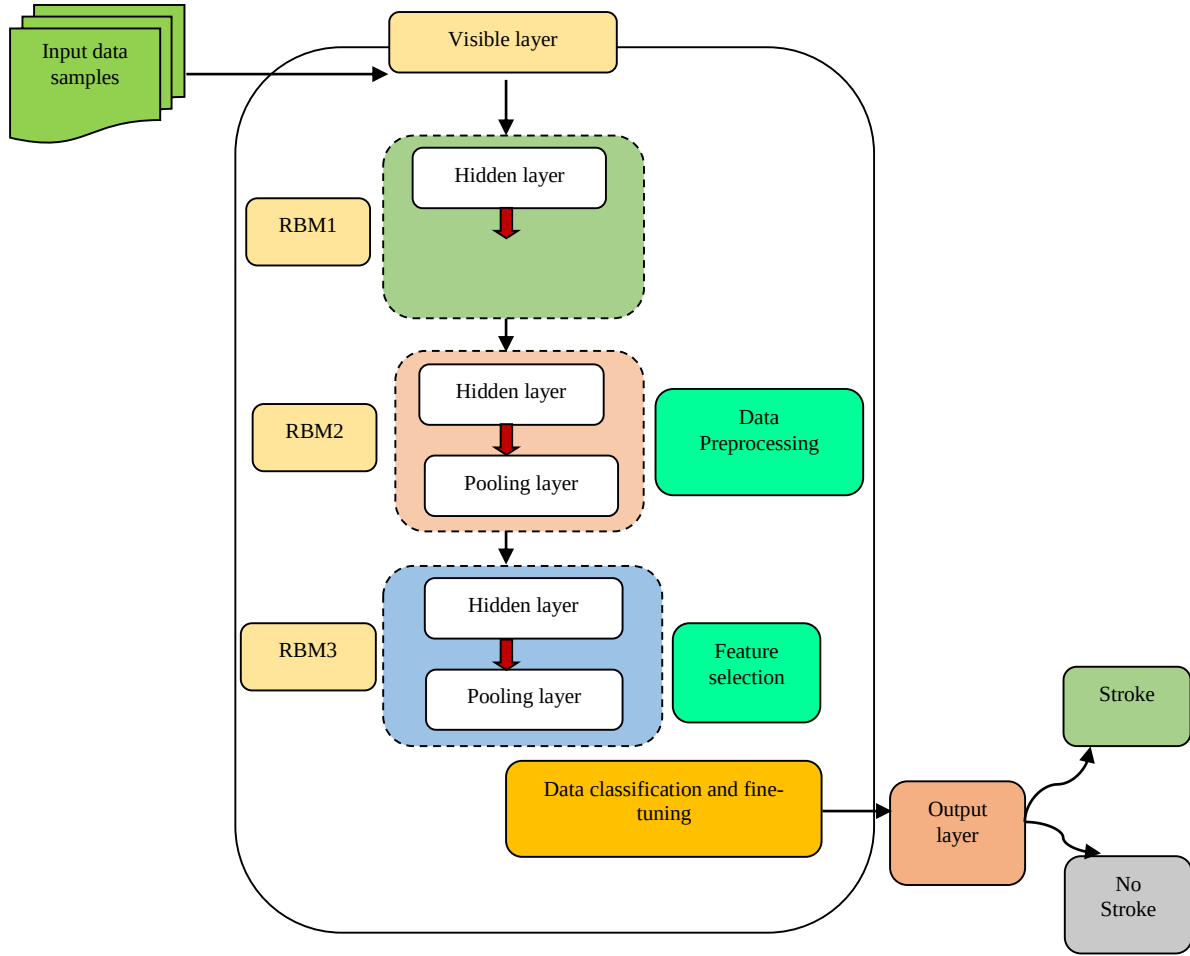


Fig. 2 Structure diagram of a Convolutional Deep Belief Network (CDBN)

$$Q = \begin{bmatrix} f_1 & f_2 & \dots & f_m \\ D_{11} & D_{12} & \dots & D_{1n} \\ D_{21} & D_{22} & \dots & D_{2n} \\ \vdots & \vdots & \dots & \vdots \\ D_{m1} & D_{m2} & \dots & D_{mn} \end{bmatrix} m = \text{rows}, n = \text{column} \quad (1)$$

Where Q represents a data matrix where every column denotes $f = \{f_1, f_2, \dots, f_m\}$, every row represents number of records $D = \{D_1, D_2, \dots, D_n\}$. This input information is transferred into hidden level of first RBM, where the neuron activation probability is measured. It plays a vital role that helps to find how quickly a neuron responds across all input data samples. Therefore, the activation probability of a hidden unit is calculated depends on weighted total of input information samples collected as of the previous visible level and a bias. Therefore, the neuron activation probability in the hidden layer ' Pr_{hid} ' is expressed as follows,

$$Pr_{hid} = A \sum Q * \beta_{vh} + b \quad (2)$$

Pr_{hid} represents a neuron activation probability at hidden, Q represents input matrix connected to a weight between visible and hidden layer ' β_{vh} ' and included with bias ' b ' helps

to transfer the input data samples left to right, giving it more flexibility, A indicates a sigmoid activation function. The output of the activation probability Pr_{hid} becomes 1 indicates that the neuron is always active state for processing the given input data samples. After that, data pre-processing task is executed to handle the missing data and outlier data in the given dataset. In pre-processing, the input raw dataset is transferred into a more suitable form for accurate stroke disease prediction. First, identify where the missing values (i.e., none or blanks are positioned in the given input dataset. After identifying the missing values, the MICE technique is employed to handle missing data in a given dataset for enhancing the accuracy and reliability of statistical analyses. This proposed imputation approach handling different data types and complex relationships. First, fill every missing value in dataset based on mean of previous known data samples.

$$D_M = \frac{1}{n} \sum_{i=1}^n D_i \quad (3)$$

Where, D_M denotes a missing value, D_i represents previously obtainable information samples in database, n denotes a total number of data samples. These values are

imputed into the corresponding blank places. In this way, all missing values are initially replaced with mean to create a total preliminary dataset. Then the linear Regression is employed to estimate missing values by measuring the relationships between variables in a dataset. In this approach, the variable with missing values is treated as the dependent variable, while the other observed variables serve as independent variables.

$$Z = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \dots \alpha_n D_n + \epsilon \quad (4)$$

Where Z indicates a dependent variable, $D_1, D_2, \dots, D_n$ indicates an independent variable, $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n$ indicates a regression coefficient, ϵ indicates the error term. The predicted value from the regression model is to fill the missing

values for a particular variable or features. This process is repeated for all features in the dataset that has missing values and achieves convergence. In this way, MICE-based missing data handling is efficiently performed in the given dataset. After handling missing data, outlier data points are detected and removed by applying the Extreme normalized residual test to discover outlier data points, which considerably diverge from the rest of the other data points in the given dataset. The normalized residual test is a statistical. Followed by, outlier data is handled by employing the high contrast subspace method. A subspace is a local dimensional reduction method to map subsets of the dataset to lower-dimensional spaces that determines outlier. The high contrast subspaces are used to measure the conditional dependence among the data samples in the dataset.

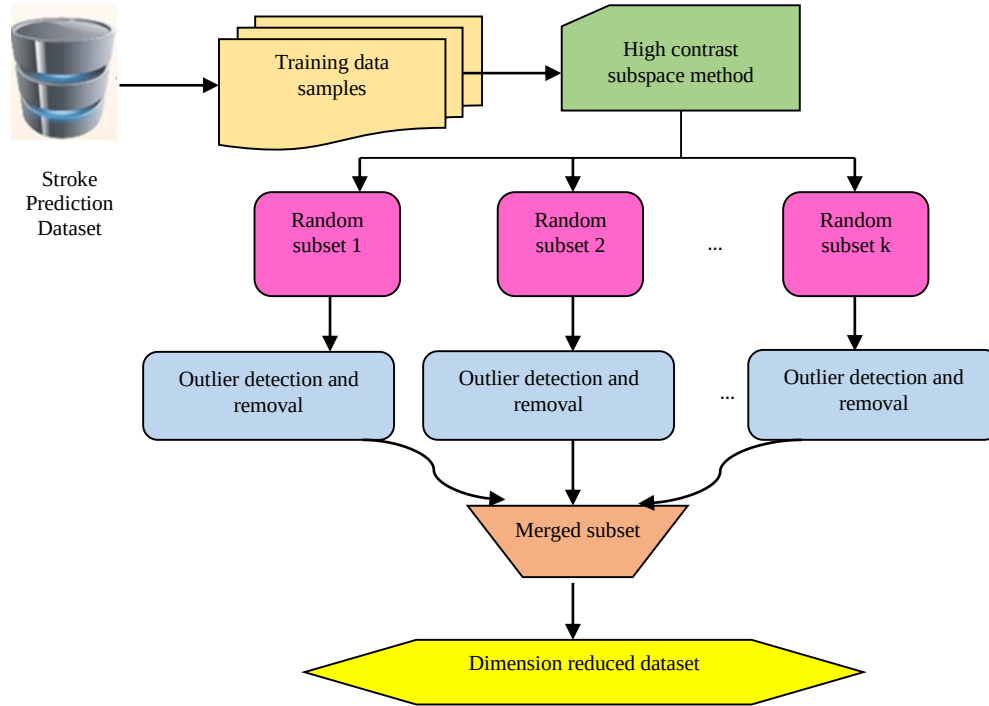


Fig. 3 Illustrates the different process of outlier detection

As shown in Figure 3, the method processes input data samples. $D_1, D_2, \dots, D_n$ collected from the datasets. In this framework, the subspace method partitions the total data samples into 'b'.

$$D_n \rightarrow \{S_1(D), S_2(D), \dots S_b(D)\} \quad (5)$$

After partitioning the subset, a normalized G-statistical test is employed to discover outlier data points, which significantly deviate from the rest of the other data points in the specified dataset. The normalized G-statistical test is expressed as follows,

$$GST = \arg \max \left[\frac{|D - \mu|}{\sqrt{\text{var}(D, \mu)}} \right] \quad (6)$$

$$\mu = \frac{1}{n} \sum_{i=1}^n D_i \quad (7)$$

$$\text{var}(D, \mu) = \sqrt{\frac{1}{n} \sum_{i=1}^n (D_i - \mu)^2} \quad (8)$$

Where, GST denotes a G-statistical test statistic, μ denotes a mean, var designates a deviation or variance, n indicates a number of input data samples, $\arg \max$ represents an argument of maximum function. In this approach, maximum statistical test statistics deviated from the mean, and then a data point is measured an outlier if it shows the largest absolute deviation from the mean. This detection process is performed iteratively, identifying and removing outliers until all significant anomalies in the dataset are addressed. After

finding and removing the outlier, the partitioned subsets are merged into whole set of data samples,

$$D_n \leftarrow \{S_1(D) \cup S_2(D) \cup \dots S_b(D)\} \quad (9)$$

After removing these outlier samples, the dataset is cleaned, and the pooling layer, followed by the hidden layer, minimizes its dimensionality, reducing the impact of remaining abnormality and preparing the data to further processing in the subsequent layers.

3.2.1. Feature Selection

The LACO-GAI model performs the feature selection process to reduce the size of input pre-processed dataset for a precise stroke forecast by lesser time utilization. The main objective to preserve the more significant features for accurate analysis and modelling by using Laplace Copula Multivariate feature ranking method. The selected less number of important features helps to minimize the computational performance in stroke prediction. The proposed LACO-GAI model utilizes the Laplace Copula Multivariate feature ranking method for accurate feature selection. First, each feature and the target variables are transformed into normal distribution space. Then, calculate the similarity between each transformed feature and the target variable using Laplace function. After that, the features are ranked based on the absolute value of their similarity coefficients. Finally, top features with the highest similarity values are selected and removed the other features. Copulas are used to describe the dependence (inter-correlation) between random variables. Let us consider the set of features in the dataset.

$$f_j = \{f_1, f_2, \dots, f_m\} \text{ where } j = 1, 2, 3 \dots m \quad (10)$$

The similarity relationship among T as well as input aspects is computed through Laplace kernel. The Laplace kernel is used to measure similarity between T as well as input aspects as follows,

$$LK = \exp \left[\sum_{j=1}^m - \left(\frac{|f_j - T|}{DV} \right) \right] \quad (11)$$

Where, LK represents a Laplace kernel, T indicates a target features, f_j indicates a features, D symbolizes deviation. The Laplace kernel function returns the output from 0 to 1. Rank the features based on similarity score values in descending order,

$$\{f_1 > f_2 > \dots > f_m\} \quad (12)$$

At last, top k features with the highest similarity score are selected and removed the other features.

$$Z = \begin{cases} \arg \max LK, & \text{select top } k \text{ features} \\ \text{Otherwise, remove irrelevant features} \end{cases} \quad (13)$$

Z denotes outcomes of feature selection, features with maximum similarity are considered relevant, while those lessening are considered irrelevant.

3.2.2. Classification

The final process of LACO-GAI model is to perform the classification in third RBM layer of Deep Learning network architecture with the selected relevant features for a precise stroke forecast. In context of classification, LACO-GAI model utilizes the Logistic Regression, also called as Logit Regression, is employed as a binary model to precisely forecast presence or absence of stroke depend on a selected set of relevant features.

These features, identified through a prior feature selection process. It helps to measure association among these aspects as well as target through probability. Association among dependent as well as independent variables is measured as follows,

$$L = \frac{1}{1 + \exp - (\vartheta_0 + \vartheta_1 \cdot f_1 + \vartheta_2 \cdot f_2 + \dots + \vartheta_k \cdot f_k)} \quad (14)$$

Where L denotes a Logistic Regression outcome, $\vartheta_0, \vartheta_1, \vartheta_2 \dots \vartheta_k$ represents a regression coefficient for the respective features, $f_1, f_2, \dots, f_k$. The regression outcomes provide outcomes from 0 to 1.

$$L = \begin{cases} 1 & ; \text{Stroke presence} \\ 0 & ; \text{No stroke} \end{cases} \quad (15)$$

The Logistic Regression model processed the input selected features according to target variable and provides the predicted likelihood strictly equivalent to the observed stroke outcomes. After predicting the stroke disease presence, prediction error ' PDE ' is computed using following expression.

$$PDE = \frac{\text{Misclassified data}}{\text{Total number of data}} \quad (16)$$

In the context of fine-tuning phase of proposed LACO-GAI model, the hyperparameters of convolutional deep belief network learning model are revised by means of stochastic adaptive moment algorithm to minimize the prediction errors and increase the accuracy of stroke prediction. This algorithm helps to iteratively change the weights among the layers of a deep learning model for quicker and more efficient convergence.

$$\beta_{new} = \beta_t - \eta \frac{G_t}{\sqrt{H_t + \epsilon}} \quad (17)$$

Where,

$$G_t = \tau_1 G_{t-1} + (1 - \tau_1) \left[\frac{\partial PDE}{\partial \beta_t} \right] \quad (18)$$

$$H_t = \tau_2 H_{t-1} + (1 - \tau_2) \left[\frac{\partial PDE}{\partial \beta_t} \right]^2 \quad (19)$$

Where, β_{new} represents a revised weights among the layers, β_t indicates present weight, η indicates ($\eta < 1$), $\frac{\partial PDE}{\partial \beta_t}$ denotes partial derivative of 'PDE' regarding current weight ' β_t ', τ denotes Momentum parameter (0.9), G_t denotes first moment, H_t denotes a second moment.. From updating procedure, numerous weights are updated depends on the computed gradients. After updating weights in a deep learning model, it is then propagated backward through the entire

network to accurately study provided input information samples. This procedure is repeated iteratively till error obtain reduced. Finally, output of stroke prediction results is examined at output level by A_{sig} .

$$Y = A_{sig}(h_t) \quad (20)$$

Y represents a last stroke prediction output, h_t indicates output of hidden, ' A_{sig} ' at output layer for generating binary class output. Stroke disease prediction is achieved to enhance the accuracy through minimizing the error.

//Algorithm 1: Laplace Copula Generative Artificial Intelligence based stroke prediction	
Input:	Dataset 'DS', features $f_1, f_2, \dots, f_m$ and data samples $D_i = \{D_1, D_2, \dots, D_n\}$
Output:	Improve the stroke disease prediction accuracy
Begin	
Step 1:	Collect number of data samples $D_1, D_2, \dots, D_n$ and features $f_1, f_2, \dots, f_m$ from the dataset// Data acquisition
Step 2:	Formulate input matrix using (1) given to visible layer // data processing
Step 3:	Transform input to RBM layer 1
Step 4:	For each data samples D_i do-----Hidden layer 1
Step 5:	Compute the neuron activation probability using (2)
Step 6:	End For
Step 7:	End For
Step 8:	For each data samples D_i do// data pre-processing
Step 9:	Perform missing data imputation using (3) and (4)
Step 10:	End for
Step 11:	partitions the total data samples into 'b' subsets using (5)
Step 12:	Compute the normalized G-statistical test 'GST' using (6)
Step 13:	If ($\arg \max GST$) then
Step 14:	Detect outlier data samples
Step 15:	else
Step 16:	Detect normal data samples
Step 17:	End if
Step 18:	Remove outlier data samples
Step 19:	Merge subsets using (9)
Step 20:	For each preprocessed data samples do // feature selection ----- RBM layer 2
Step 21:	Measure Laplace kernel using (11)
Step 22:	Rank the features based on similarity in descending order using (12)
Step 23:	Select top k features using (13)
Step 24:	End for
Step 25:	For each selected feature with training data samples // classification----- RBM layer 3
Step 26:	Perform multivariate regression analysis 'L' using (14)
Step 27:	if ($L = 1$) then
Step 28:	Data samples classified into disease presence
Step 29:	else
Step 30:	Data samples classified into disease absence
Step 31:	End if
Step 32:	End for
Step 33:	For each disease prediction result // Fine tuning
Step 34:	Compute the error 'PDE' using (16)
Step 35:	Update the weight using (17)
Step 36:	if ($\arg \min PDE$) then
Step 37:	Obtain final disease prediction results using (20) ----- (output layer)
Step 38:	else
Step 39:	Go to step 37

```

Step 40: End if
Step 41: End for
End

```

4. Experimental Analysis

Experimental evaluation of LACO-GAI, existing GRU [1], and BiLSTM [2] are implemented in Python with Stroke Prediction datasets. For experimental purposes, number of information samples is varied incrementally from 1500 to 15000.

5. Performance Evaluation of Models

Performance metric analyzes of the LACO-GAI model and existing [1, 2] are evaluated with various metrics across a dissimilar number of information samples.

5.1. Performance Assessment of Stroke Prediction Accuracy

It refers to the ratio of accurate stroke predictions with number of information samples to n . Number of information samples considered as input. Stroke prediction accuracy is mathematically given below,

$$SPA = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (21)$$

Where, 'SPA' represents the stroke prediction accuracy, true positive 'TP' indicates a model correctly predicts stroke presence, true negative 'TN' indicates correctly predicts no stroke data samples, false positive 'FP' represents a no stroke data samples incorrectly predicated as stroke, 'FN' denotes stroke samples incorrectly predicted as no stroke.

Table 2 presents a chart demonstrating the stroke prediction accuracy over LACO-GAI model and existing GRU [1] and BiLSTM [2]. Among the three different deep learning methods, the performance results of the LACO-GAI model demonstrate significant development compared to the existing deep learning methods.

For example, while considering 1500 data samples in the first run, the LACO-GAI model attained Stroke prediction accuracy of 96.66%. In contrast, the stroke prediction accuracy of the existing models [1] and [2] were observed to be 93.33% and 92%, respectively.

Table 2. Comparison of SPA versus data samples

Number of data samples	Stroke prediction accuracy (%)		
	Proposed LACO-GAI	Existing GRU [1]	Existing BiLSTM [2]
1500	96.66	93.33	92
3000	96.45	93.22	91.96
4500	96.38	93.18	91.87
6000	96.28	93.34	91.72
7500	96.15	93.21	91.63
9000	96.08	93.06	91.66
10500	96.17	93.12	91.52
12000	96.22	93.06	91.32
13500	96.07	93.22	91.45
15000	96.23	93.07	91.63

In a similar manner, ten various sets of results were observed for each deep learning method according to the dissimilar number of data samples.

Generally, observed results of the LACO-GAI model were then compared to those of the existing methods. The comparative analysis of the ten consequences demonstrated that the LACO-GAI model achieve high model accuracy, improved by 3% and 5% over [1] and [2], respectively.

This improvement is achieve due to the application of logic regression, which analyzes the training and testing data samples in the hidden layer of the deep learning model, resulting it increases the accuracy. Based on regression outcomes, predicts the stroke disease with better accuracy.

Besides, fine-tuning process of the proposed LACO-GAI model utilizes movement optimization to revise disease prediction error by enhance the true positive and true negative rates, while reducing the false positives and false negatives in the more accurate stroke prediction.

5.2. Performance Assessment of Precision

It is ratio of number of TP stroke forecasts made through algorithm, as well as whole number of data samples model forecasted as positive. It is computed as,

$$Precision = \left(\frac{TP}{TP+FP} \right) \quad (22)$$

TP indicates the true positive, FP denotes false positive.

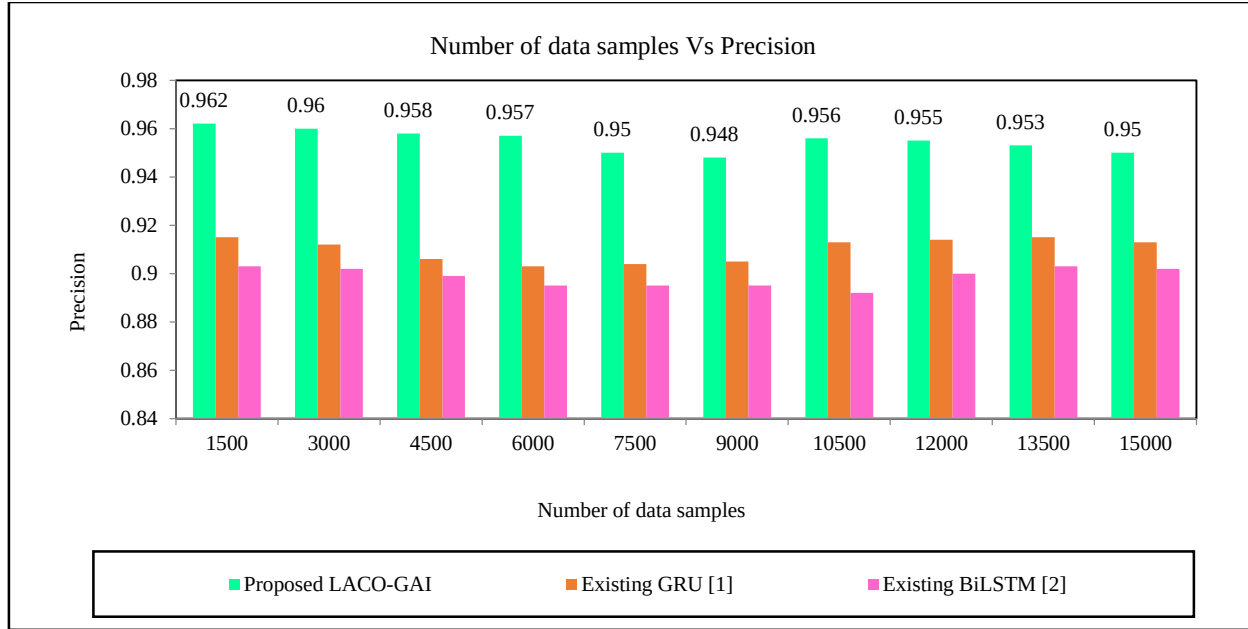


Fig. 4 Comparison chart of precision

Figure 4 shows the model comparison of precision about the stroke disease prediction number of data samples collected from the publically available dataset ranges from 1500 to 15000. The horizontal axis represents the number of input data samples, while the perpendicular axis of the chart designates the overall precision performance in stroke prediction. The experimental assessment of precision is done by employing three different deep learning models LACO-GAI model and existing GRU [1], and BiLSTM [2]. For example, 1500 samples are considered for computing the precision in the first run. In order to apply the LACO-GAI model achieved a precision of 0.962, while the existing methods [1] and [2] observed to be 0.915 and 91.22%, respectively. Similarly, dissimilar precision results were observed about various counts of input data counts. In the end, the observed results of LACO-GAI model are compared to the existing methods [1] and [2]. Based on the average of ten qualified evaluations, the LACO-GAI model shows an approximate 5% improvement in precision over method [1] and about a 6% improvement over

method [2]. The improved performance of the LACO-GAI model is achieved by integrating the convolutional deep belief network with the logistic regression for feature-based data analysis. In addition, the fine-tuning of the network layers is refined by means of the gradient momentum Optimization algorithm, which efficiently reduces both training and validation errors. This integrated approach considerably increases the performance of precision by enhancing the true positive rate while concurrently minimizing false positives in stroke prediction.

5.3. Performance Assessment of Recall

It is computed as ratio of number of TP through total numeral of actual positives, that is sum of TP and FN.

$$Recall = \left(\frac{TP}{TP+FN} \right) \quad (23)$$

TP indicates true positive, FN symbolizes false negative.

Table 3. Comparison of recall versus data samples

Number of data samples	Recall		
	Proposed LACO-GAI	Existing GRU [1]	Existing BiLSTM [2]
1500	0.975	0.962	0.949
3000	0.974	0.96	0.948
4500	0.973	0.958	0.947
6000	0.972	0.956	0.945
7500	0.973	0.957	0.943
9000	0.974	0.956	0.945
10500	0.97	0.955	0.943
12000	0.972	0.96	0.946
13500	0.973	0.961	0.944
15000	0.974	0.959	0.945

Table 3 presents a comparative examination of recall performance regarding varying numbers of data samples extracted from the dataset. The LACO-GAI model consistently achieves superior recall values in stroke disease prediction when compared with existing methods [1] and [2].

For instance, when 1,500 samples are used, LACO-GAI model attains a recall of 0.975, whereas methods [1] and [2] record recall values of 0.962 and 0.949, respectively.

Comparable trends are examined across dissimilar sample sizes. Overall investigational results specify that the average recall of the LACO-GAI model improves by approximately 2% over method [1] and by about 3% over method [2]. This enhanced recall performance is mainly achieved due to the

fine-tuning approach employed within the deep belief network architecture. Specifically, the convolutional deep belief network reduces stroke prediction errors by optimally adjusting network weights through the optimization algorithm. This iterative optimization process continues until error convergence is attained, directing to a reduction in false negatives and an increase in true positives, thereby facilitating more accurate stroke prediction in the healthcare domain.

5.4. F1 score

It is called as F-measure, computed based on harmonic average of precision, recall in the stroke prediction. It is expressed as follows

$$F1 - Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad (24)$$

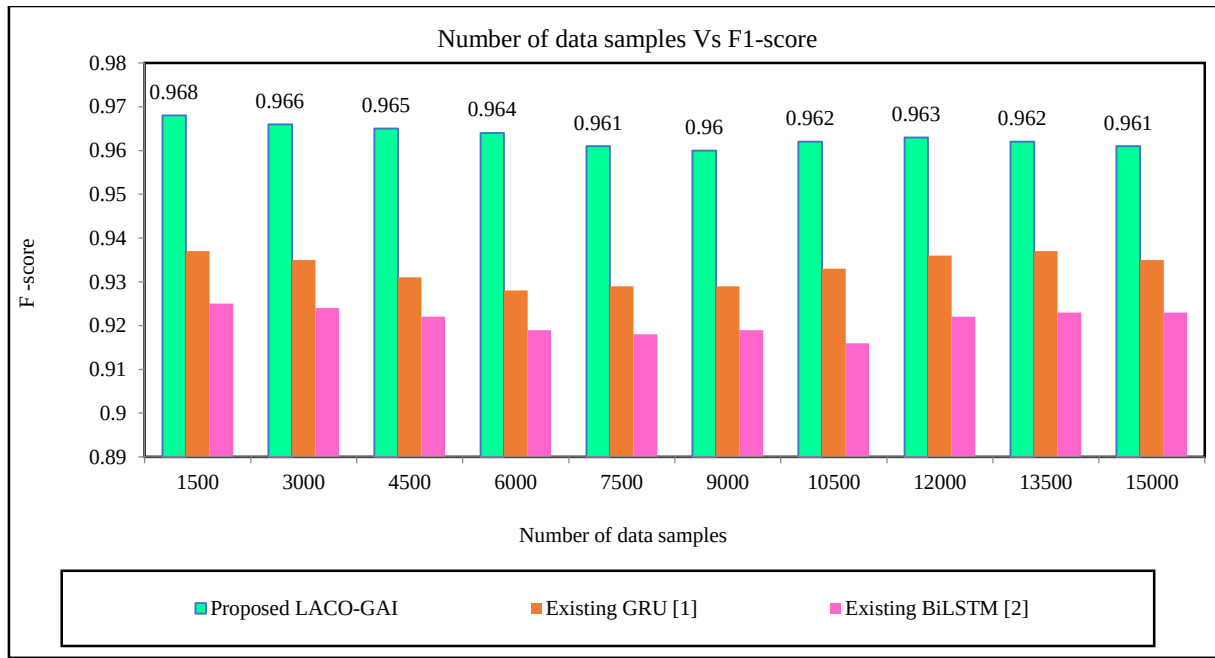


Fig. 5 Comparison charts of F1 score

Figure 5 provides a visual assessment of F1 score results by means of different quantities of data samples, ranging from 1500 to 15000. The results designate that the LACO-GAI model continuously achieves higher F1 scores when compared to existing deep learning approaches. For each method, ten separate experimental runs are observed according to various data samples. The average results illustrate that the proposed LACO-GAI model demonstrated a development of approximately 3% over method [1] and 5% over method [2]. This improvement highlights the model's ability to improve both precision and recall in stroke prediction. The superior performance of LACO-GAI model is mainly achieved due to its effective utilization of a convolutional deep belief neural network, which enables better feature analysis and prediction across different sample sizes. Finally, the overall comparative analysis indicates that the LACO-GAI model achieves an F1-

score improvement of approximately 3% and 5% when compared to [1] and [2], demonstrating its robustness and consistency in stroke disease prediction.

5.5. RMSE

It refers to fraction of variation between the actual disease prediction outcomes and observed disease prediction outcomes by the deep learning model. It is computed as,

$$RMSE = \left[\sqrt{\frac{(Y_{actual} - Y_{prediction})^2}{n}} \right] \quad (25)$$

Where, $RMSE$ symbolizes a root mean square error, Y_{actual} symbolizes the actual prediction results $Y_{prediction}$ represents a predicted output.

Table 4. Comparison of RMSE versus data samples

Number of data samples	RMSE		
	Proposed LACO-GAI	Existing GRU [1]	Existing BiLSTM [2]
1500	0.086	0.172	0.206
3000	0.064	0.123	0.146
4500	0.053	0.101	0.121
6000	0.048	0.085	0.106
7500	0.044	0.078	0.096
9000	0.041	0.073	0.087
10500	0.037	0.067	0.082
12000	0.034	0.063	0.079
13500	0.033	0.058	0.073
15000	0.030	0.056	0.068

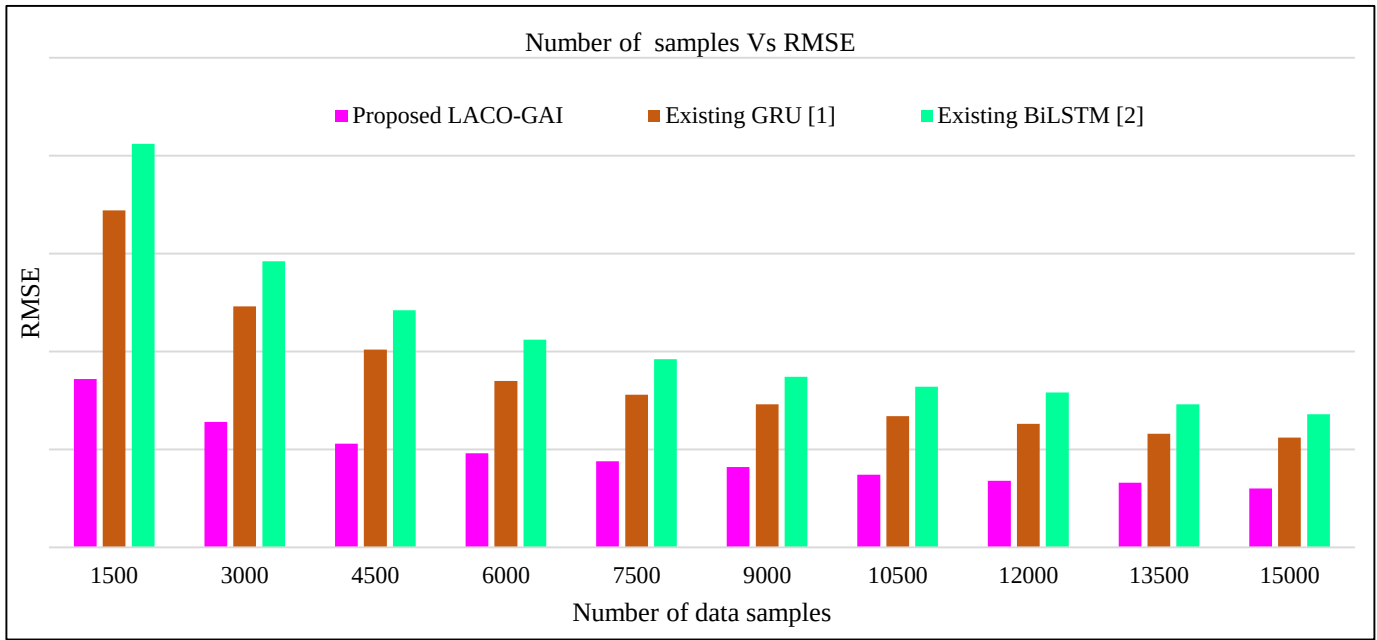


Fig. 6 Comparison charts of RMSE

Table 4 depicts the experimental evaluation of RMSE versus the number of data samples, ranging from 1500 to 15000. Three methods, namely the LACO-GAI model and existing GRU [1] and BiLSTM [2], are applied to estimate RMSE with a varying number of data samples. As shown in Figure 6, the horizontal axis of the graphical chart designates the number of data samples, while the vertical axis indicates the RMSE. The overall obtained performance results depict that the LACO-GAI model achieved lesser RMSE compared to existing methods [1, 2]. Consider a setting in which the dataset contains 1500 samples. In these settings, LACO-GAI model achieves an RMSE value of 0.086, whereas methods [1, 2] observed RMSE values of 0.172 and 0.206, respectively. Similar performance results were observed across varying sample sizes for all three deep learning methods. Overall, the experimental results illustrate that the LACO-GAI model considerably minimizes the RMSE, achieving approximately 46% reduction over method [1] and an 55% reduction over method [2]. This superior

performance is achieved due to the LACO-GAI model's ability to accurately determine stroke and no stroke samples at the output layer. Following classification, a fine-tuning phase iteratively updates the network weights until convergence to a minimum error is achieved. A lower RMSE result indicates higher stroke prediction accuracy, thereby increasing the reliability of a predictive model.

5.6. Performance Assessment of Stroke Prediction Time

It is the amount of time taken through the method for predicting stroke diseases from the input patient data samples. The stroke prediction time is computed as,

$$SPT = \sum_{i=1}^n D_i * TM(SP)] \quad (26)$$

Where 'SPT' indicates a stroke prediction time, $TM(SP)$ represents time for predicting strokes with one data sample. It is measured in ms.

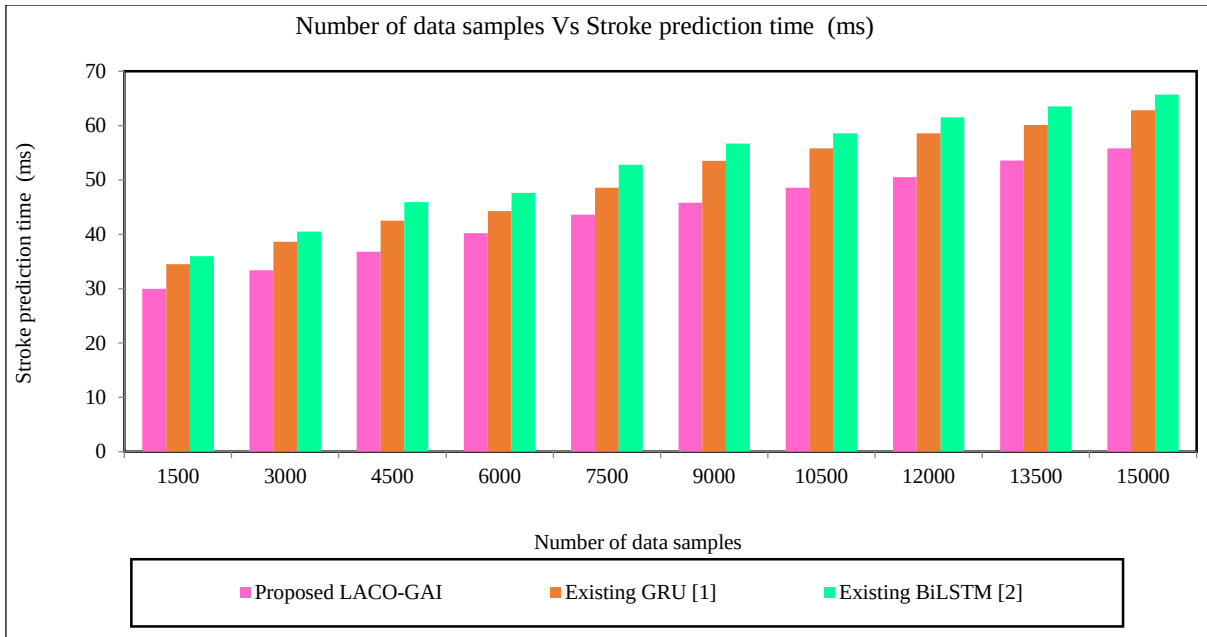


Fig. 7 Comparison charts of SPT

Figure 7 presents the overall performance of SPT using LACO-GAI model and existing GRU [1] and BiLSTM [2]. All models were estimated over ten experimental runs using a dataset of 15,000 distinct samples.

The figure demonstrates that as the number of samples rises, the stroke prediction time gets increased for all three deep learning models. In a first trial with 1500 samples, LACO-GAI model consumed 30ms, whereas [1, 2] required 34.5ms and 36ms, respectively.

This corresponds to an approximate 12% reduction in stroke prediction time compared to [1] and 17% reduction compared to [2]. This improved efficiency of LACO-GAI model is achieved due to its integration of data pre-processing and feature selection approach into the proposed deep learning model architecture.

In data pre-processing, High Contrast Subspace method is employed to remove the outliers from the given dataset, thus reducing the sizes of the original data.

By employing the Laplace Copula Multivariate feature ranking method, the model identifies and retains the most significant features while eliminating less relevant ones. With the selected less number of features, stroke prediction is performed, resulting it minimizes the time consumption.

6. Conclusion

A novel AI technique called LACO-GAI model is introduced for early stroke detection. The pre-processing stage of proposed AI model includes missing data handling as well as outlier data elimination. Afterwards, important features are chosen from the dataset for accurate stroke disease prediction. These two processes are employed to minimize the time consumption of early stroke detection.

Finally, logistic regression provides the stroke prediction outcomes by analyzing the selected relevant features. The hyperparameter optimization of DL model further minimizes the error as well as improves accuracy in the stroke forecast process. Quantitatively analyzed results designates that LACO-GAI model increases the stroke prediction accuracy, precision, recall while minimizing prediction time and RMSE than the conventional DL techniques.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

- [1] Afeez A. Soladoye et al., "Optimizing Stroke Prediction using Gated Recurrent unit and Feature Selection in Sub-Saharan Africa," *Clinical Neurology and Neurosurgery*, vol. 249, pp. 1-10, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Saeed Mohsen, and Ahmed Gomaa, "A Novel Intelligent Healthcare System based on Efficient deep Learning Models for Brain Stroke Predictio," *Multimedia Tools and Applications*, vol. 84, no. 34, pp. 42475-42506, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [3] Khadijeh Moulaei et al., "Explainable Artificial Intelligence for Stroke Prediction Through Comparison of deep Learning and Machine Learning Models," *Scientific Reports*, vol. 14, no. 1, pp. 1-16, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Afeez A. Soladoye et al., "Stroke Risk Prediction: A Deep Learning Approach for Identifying High-Risk Patients," *Discover Data*, vol. 3, no. 1, pp. 1-19, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Sapia Sakri et al., "An Improved Concatenation of Deep Learning Models for Predicting and Interpreting Ischemic Stroke," *IEEE Access*, vol. 12, pp. 53189-53204, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Mohammad Amin et al., "A Hybrid Brain Stroke Prediction Framework: Integrating Feature Selection, Classification, and Hyperparameter Optimization," *Engineering Reports*, vol. 7, no. 7, pp. 1-16, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Shilpa Bajaj, Manju Bala, and Mohit Angurala, "Machine Learning Models for Enhanced Stroke Detection and Prediction," *Egyptian Informatics Journal*, vol. 30, pp. 1-14, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Sheela Lavanya J M, and Subbulakshmi P, "Unveiling the Potential of Machine Learning Approaches in Predicting the Emergence of Stroke at its Onset: A Predicting Framework," *Scientific Reports*, vol. 14, no. 1, pp. 1-21, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Maha Samir Elsayed, "An Effective Brain Stroke Diagnosis Strategy based on Feature Extraction and Hybrid Classifier," *Scientific Reports*, vol. 15, no. 1, pp. 1-21, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Muhammad Asim Saleem et al., "An Intelligent Learning System based on Electronic Health Records for Unbiased Stroke Prediction," *Scientific Reports*, vol. 14, no. 1, pp. 1-14, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Marta Narigina et al., "AI-Powered Stroke Diagnosis System: Methodological Framework and Implementation," *Future Internet*, vol. 17, no. 5, pp. 1-26, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Jae-Young Maeng et al., "Machine Learning-based Diagnostic Model for Stroke in Non-Neurological Intensive care unit Patients with Acute Neurological Manifestations," *Scientific Reports*, vol. 14, no. 1, pp. 1-14, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Reeta Samuel, and Thanapal Pandi, Optimizing Brain Stroke Detection with a Weighted Voting Ensemble Machine Learning Model," *Scientific Reports*, vol. 15, no. 1, pp. 1-15, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Ahmad Hassan et al., "Predictive Modelling and Identification of key risk Factors for Stroke using Machine Learning," *Scientific Reports*, vol. 14, no. 1, pp. 1-23, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Chunhua Gao, and Hui Wang, "Intelligent Stroke Disease Prediction Model using Deep Learning Approaches," *Stroke Research and Treatment*, vol. 2024, no. 1, pp. 1-10, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Yunxia Duan, "Development and Validation of a Stroke risk Prediction model using Regional Healthcare big data and Machine Learning," *International Journal of Nursing Sciences*, vol. 12, no. 6, pp. 558-565, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Inam Abousaber, "A Novel Explainable Attention-based Meta-Learning Framework for Imbalanced Brain Stroke Prediction," *Sensors*, vol. 25, no. 6, pp. 1-32, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Gayatri Thakre, "A Hybrid Deep Learning Approach for Improved Detection and Prediction of Brain Stroke," *Applied Sciences*, vol. 15, no. 9, pp. 1-28, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] D. Ushasree, Praveen Krishna A.V, and Mallikarjuna Rao Ch, "Enhanced Stroke Prediction using Stacking Methodology (ESPESM) in Intelligent Sensors for Aiding Preemptive Clinical Diagnosis of Brain Stroke," *Measurement: Sensors*, vol. 33, pp. 1-7, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Priyanka Bathla, and Rajneesh Kumar, "A Hybrid System to Predict Brain Stroke using a Combined Feature Selection and Classifier," *Intelligent Medicine*, vol. 4, no. 2, pp. 75-82, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Muhammad Asim Saleem et al., "Enhancing Stroke risk Prediction Through class Balancing and data Augmentation with CBDA-ResNet50," *Scientific Reports*, vol. 15, no. 1, pp. 1-19, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Md. Maruf Hossain et al., "Optimizing Stroke Risk Prediction: A Primary Dataset-Driven Ensemble Classifier with Explainable Artificial Intelligence," *Health Science Reports*, vol. 8, no. 5, pp. 1-15, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Muhammed Sutcu et al., "Predicting Stroke Risk using Machine Learning: A Data-Driven Approach to Early Detection and Prevention," *Stroke Research and Treatment*, vol. 2025, no. 1, pp. 1-28, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Ravnoor Singh et al., "Optimizing Early Diagnosis by Integrating Multiple Classifiers for Predicting Brain Stroke and Critical Diseases," *Scientific Reports*, vol. 14, no. 1, pp. 1-17, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Aakanshi Gupta et al., "Predicting Stroke risk: An Effective Stroke Prediction model based on Neural Networks," *Journal of Neurorestoratology*, vol. 13, no. 1, pp. 1-8, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [26] Saad Sahriar et al., "Unlocking Stroke Prediction: Harnessing Projection-based Statistical Feature Extraction with ML Algorithms," *Heliyon*, vol. 10, no. 5, pp. 1-15, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [27] Dong-Her Shih et al., "Stroke Prediction using Deep Learning and Transfer Learning Approaches," *IEEE Access*, vol. 12, pp. 130091 - 130104, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [28] Amir Sorayaie Azar et al., "Predicting Stroke Severity of Patients using Interpretable Machine Learning Algorithms," *European Journal of Medical Research*, vol. 29, no. 1, pp. 1-23, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [29] Pritam Chakraborty et al., "Predicting Stroke Occurrences: A Stacked Machine Learning Approach with Feature Selection and data Preprocessing," *BMC Bioinformatics*, vol. 25, no. 1, pp. 1-23, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [30] Assefa Senbato Genale, and Tsion Ayalew Dessalegn, "Developing a Predictive Model for Stroke Disease Detection using a Scalable Machine Learning Approach," *Applied Computational Intelligence and Soft Computing*, vol. 2025, no. 1, pp. 1-19, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]