



OPEN

## Explainable artificial intelligence for stroke prediction through comparison of deep learning and machine learning models

Khadijeh Moulaei<sup>1,2</sup>, Lida Afshari<sup>3</sup>, Reza Moulaei<sup>4</sup>, Babak Sabet<sup>2,5</sup>,  
Seyed Mohammad Mousavi<sup>6</sup> & Mohammad Reza Afrash<sup>2,7✉</sup>

Failure to predict stroke promptly may lead to delayed treatment, causing severe consequences like permanent neurological damage or death. Early detection using deep learning (DL) and machine learning (ML) models can enhance patient outcomes and mitigate the long-term effects of strokes. The aim of this study is to compare these models, exploring their efficacy in predicting stroke. This study analyzed a dataset comprising 663 records from patients hospitalized at Hazrat Rasool Akram Hospital in Tehran, Iran, including 401 healthy individuals and 262 stroke patients. A total of eight established ML (SVM, XGB, KNN, RF) and DL (DNN, FNN, LSTM, CNN) models were utilized to predict stroke. Techniques such as 10-fold cross-validation and hyperparameter tuning were implemented to prevent overfitting. The study also focused on interpretability through Shapley Additive Explanations (SHAP). The evaluation of model's performance was based on accuracy, specificity, sensitivity, F1-score, and ROC curve metrics. Among DL models, LSTM showed superior sensitivity at 96.15%, while FNN exhibited better specificity (96.0%), accuracy (96.0%), F1-score (95.0%), and ROC (98.0%) among DL models. For ML models, RF displayed higher sensitivity (99.9%), accuracy (99.0%), specificity (100%), F1-score (99.0%), and ROC (99.9%). Overall, RF outperformed all models, while DL models surpassed ML models in most metrics except for RF. DL models (CNN, LSTM, DNN, FNN) achieved sensitivities from 93.0 to 96.15%, specificities from 80.0 to 96.0%, accuracies from 92.0 to 96.0%, F1-scores from 87.34 to 95.0%, and ROC scores from 95.0 to 98.0%. In contrast, ML models (KNN, XGB, SVM) showed sensitivities between 29.0% and 94.0%, specificities between 89.47% and 96.0%, accuracies between 71.0% and 95.0%, F1-scores between 44.0% and 95.0%, and ROC scores between 64.0% and 95.0%. This study demonstrates the efficacy of DL and ML models in predicting stroke, with the RF models outperforming all others in key metrics. While DL models generally surpassed ML models, RF's exceptional performance highlights the potential of combining these technologies for early stroke detection, significantly improving patient outcomes by preventing severe consequences like permanent neurological damage or death.

**Keywords** deep learning, machine learning, predicting, stroke

Stroke poses a significant risk of mortality. Individuals who survive may face impairments such as vision and speech loss, paralysis, and cognitive disorientation. Each year, strokes affect 15 million individuals globally, resulting in 5 million deaths and leaving another 5 million permanently incapacitated, thereby exerting a considerable strain on families and communities. Strokes are rare in individuals younger than 40 years; however, when they do happen, high blood pressure is often the primary factor. Moreover, approximately 8% of children

<sup>1</sup>Health Management and Economics Research Center, Health Management Research Institute, Iran University of Medical Sciences, Tehran, Iran. <sup>2</sup>Artificial Intelligence in Medical Sciences Research Center, Smart University of Medical Sciences, Tehran, Iran. <sup>3</sup>Department of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran. <sup>4</sup>Department of Orthopedic and Trauma Surgery, Shariati Hospital and School of Medicine, Tehran University of Medical Sciences, Tehran, Iran. <sup>5</sup>Department of Surgery, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran. <sup>6</sup>Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. <sup>7</sup>Department of Artificial Intelligence in Medical Sciences Research Center, Smart University of Medical Sciences, Tehran, Iran. ✉email: M.afrash@sbmu.ac.ir; Mmdrza.lp19@gmail.com

with sickle cell disease experience strokes<sup>1</sup>. Therefore, failure to recognize a stroke in time can be perilous, as it delays crucial medical intervention<sup>2</sup>. Without prompt identification of stroke symptoms, individuals are at risk of suffering irreversible brain damage and enduring long-term disabilities. Timely recognition is paramount for optimizing treatment outcomes and minimizing the impact of this potentially life-threatening condition<sup>3</sup>.

So, a stroke, as a critical medical situation, demands urgent attention. Prompt recognition and effective treatment are essential to reduce additional harm to the brain's impacted region and prevent complications in other body parts<sup>4</sup>. As the cost of stroke hospitalization escalates, there is a heightened demand for cutting-edge technologies. Such technologies are crucial for supporting clinical diagnosis, treatment, forecasting of clinical incidents, advising on potential therapeutic measures, rehabilitation schemes, and so forth<sup>5</sup>. Artificial intelligence (AI) encompasses the application of computer systems to undertake tasks that might challenge human capabilities, frequently in manners that are elusive to specify. Machine Learning (ML) and Deep Learning (DL), branches of AI, can be utilized for the timely diagnosis of stroke<sup>6</sup>. These methods enhance the accuracy and speed of diagnosing strokes, enabling the analysis of vast datasets with unparalleled precision<sup>7,8</sup>. Feng et al.'s review<sup>7</sup> showed that DL is a potent tool for data-driven stroke management, especially in acute interventions and prognosis. They highlighted its effectiveness in delivering rapid and powerful results, making it essential for modern stroke specialists and paving the way for personalized care in ischemic stroke cases. The findings of Shirsat et al.'s study<sup>9</sup> also showed that ML can have wide applications in stroke management, helping in the timely prevention, diagnosis, treatment, and prognosis of strokes.

To our knowledge, no study has been conducted to compare ML and DL models for stroke prediction. The studies that have been conducted have only utilized ML models<sup>4,10,11</sup> or DL models<sup>12–14</sup> for stroke prediction and did not compare the performance of these two types of models, ML and DL, within the same study. So, the aim of this study was to compare deep learning models with machine learning models in stroke prediction.

## Methods

### Study design and setting

This study was conducted based on the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist (Appendix 1). It was carried out in 2024 using a dataset comprising 663 records of patients hospitalized at Hazrat Rasool Akram Hospital in Iran over the past five years, with data collected on December 28, 2022. This retrospective observational study aimed to analyze stroke prediction in patients. The dataset included 401 cases of healthy individuals and 262 cases of stroke patients admitted in hospital. Eight well-established ML and DL models were utilized for predicting stroke using a variety of clinical and demographic features. In order to address the risk of overfitting, the training of models incorporated a 10-fold cross-validation approach.

### Dataset

This phase involved two key steps. Initially, 53 important features influencing stroke in patients were identified by two neurologists. These features were organized into a questionnaire utilizing a five-point Likert scale to assess their significance. The questionnaire was distributed and collected among 11 neurologists in person, enabling us to pinpoint the most critical features for our study. Following the approval of these features by the neurologists, researchers extracted data from patients' medical records using a data extraction form. This process included details of any treatments received by the patients and how these treatments were managed during model development and evaluation.

It should be noted that, to ensure the integrity of the outcome assessment and to blind the evaluation process, we anonymized patient records by removing all identifiable information, including names and personal details. Each record was assigned a unique numerical ID to maintain confidentiality while allowing for systematic data management. The data extraction and analysis were conducted by a separate team of researchers who were not involved in the initial feature selection or questionnaire distribution. This team was responsible for ensuring that the outcome assessors remained unaware of the patients' identities and treatment details, thereby preventing any bias in interpreting outcomes. Additionally, all outcome assessors were required to have relevant qualifications in neurology, with a minimum of five years of clinical experience, ensuring that their evaluations were both objective and reliable. By implementing these measures, we aimed to minimize subjective interpretation and uphold the rigor of our study.

To arrive at the study size, we utilized a convenience sampling approach, incorporating all available data from patients admitted to Hazrat Rasool Akram Hospital within the past five years<sup>15</sup>. In total, 663 records were gathered, consisting of 401 healthy individuals and 262 patients with stroke. The inclusion criteria for selecting subjects in the study were as follows:

- Patients admitted to Hazrat Rasool Akram hospital.
- Patients whose medical records were documented within the past 5 years.
- Patients with stroke symptoms and sign.

The exclusion criteria involved patients whose information was inadequately documented in the electronic medical record.

### Pre-processing of the data

To preprocess the data for analysis, we first imported the dataset into Jupyter Notebook within the Anaconda environment, utilizing Python version 13.1. For efficient data preprocessing, we used the NumPy library, which is crucial for handling array operations and mathematical functions. NumPy facilitates various preprocessing tasks, including managing missing values and conducting statistical calculations, thereby streamlining data

manipulation and setting the stage for deeper analysis. Firstly, we conducted an exploratory analysis to assess the extent of missing data. Approximately 12% of the dataset contained missing values, with certain variables exhibiting a higher degree of missingness than others. Variables with missingness less than 30% were imputed using appropriate measures: the mean was used for continuous variables, and the mode for categorical variables. This imputation helped preserve the overall structure of the data without significantly affecting variability. For variables with missingness greater than 70%, the corresponding rows were removed to avoid introducing bias that might skew model performance. These thresholds were determined based on standard data cleaning practices, ensuring that enough data remained for meaningful analysis while reducing noise. Secondly, the nominal values of variables in the columns were replaced with numerical values using one-hot encoding or label encoding, depending on the variable type. This was done to allow machine learning algorithms to effectively interpret categorical data and improve model learning.

After addressing missingness, Min-max normalization was applied to the data. This process involved adjusting numerical data to a predefined range, typically [0, 1], while preserving the relative relationships between values. Min-max normalization ensured consistent feature scaling, thus preventing certain features from disproportionately influencing the model<sup>16</sup>. Additionally, outliers were identified and addressed through statistical methods, which helped mitigate their impact on model predictions. By implementing these preprocessing steps, we aimed to create a well-balanced and clean dataset, ensuring that the algorithms could learn effectively without being hindered by data quality issues.

It should be noted that the ratio of healthy individuals to stroke patients is approximately 1.53:1. While this may seem imbalanced, it is within an acceptable range for many medical studies, where having a larger number of healthy controls is often beneficial for establishing a robust baseline<sup>17,18</sup>. Moreover, existing literature suggests that a ratio of around 2:1 is often adequate for distinguishing between cases and controls<sup>19</sup>. Our dataset is close to this guideline, allowing the model to learn effectively from both groups without requiring aggressive balancing techniques.

### Machin learning and deep learning models

In total, eight widely recognized models from both deep learning (DL) and machine learning (ML) domains were employed to stroke prediction. Among the DL models utilized were the Deep Neural Network (DNN), feedforward neural network (FNN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). On the other hand, the ML models included Extreme Gradient Boosting (XGB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and an additional Random Forest (RF).

The chosen models offer a diverse set of methodologies suitable for stroke prediction, covering both deep learning and machine learning approaches. Deep learning models such as DNN, FNN, LSTM, and CNN are capable of capturing intricate patterns within the data, while ML models like XGB, SVM, DT, and RF provide robust and interpretable predictions. By leveraging a range of models, we aimed to maximize prediction accuracy and gain insights into the complex factors influencing stroke prediction.

It is important to acknowledge that although CNNs are primarily utilized for image data because of their capacity to capture spatial hierarchies, recent research<sup>20,21</sup> has shown that they can also be effective with tabular data. CNNs are capable of learning local dependencies and patterns in tabular datasets, much like they identify features in images. The study by Buturović et al.<sup>20</sup> demonstrated that CNNs can accurately predict diseases using tabular data. Moreover, although LSTMs are commonly used for time-series data, they can also be effective for tabular datasets where there is a need to capture sequential patterns, even if the data is not inherently temporal<sup>22</sup>. In medical datasets, certain features (e.g., patient history) may have inherent sequential relationships that an LSTM can exploit to improve prediction accuracy. Second, recent research has demonstrated the versatility of LSTMs in handling tabular data when combined with techniques like feature engineering<sup>23</sup>. Prabowo et al.<sup>24</sup>, also employed LSTM networks in conjunction with tabular data to improve the prediction of university students' GPA by utilizing both historical GPA as time-series data and various static features from tabular data. The LSTM branch captures temporal dependencies within the sequential nature of historical GPA values, enabling the model to discern how past academic performance influences future outcomes. By integrating this with the MLP branch for processing tabular data, the proposed MLP-LSTM architecture effectively leverages the strengths of both data types. These adaptations allow LSTMs to extract meaningful representations, particularly in domains like healthcare or education, where tabular data often includes implicit sequential information. This dual-input approach significantly enhances predictive accuracy, as demonstrated by superior performance metrics compared to models relying on a single data type. Consequently, the combination of LSTM with tabular data enriches the model's ability to understand complex relationships and offers a more robust framework for predictive modeling<sup>24</sup>.

### Cross-validation and hyperparameter tuning

To prevent overfitting, we implemented 10-fold cross-validation for training all suggested models. This technique involves dividing the dataset into 10 equally sized folds, where the model is trained on 9 folds and validated on the remaining fold in each iteration. This process is repeated 10 times to ensure thorough validation. The final performance metric is then derived by averaging the results from these iterations, providing a robust assessment of the model's performance<sup>25</sup>.

Additionally, to optimize each algorithm's performance, we conducted hyperparameter tuning. This process utilized a grid search methodology, where a comprehensive set of hyperparameter values was systematically evaluated. The aim was to pinpoint the parameter configurations that maximize the efficiency and accuracy of each model. Through this detailed and iterative exploration of the hyperparameter space, we were able to fine-tune the models effectively. This meticulous adjustment ensures that our models are precisely calibrated, significantly enhancing their ability to analyze and predict outcomes accurately with the dataset at hand<sup>26</sup>.

Output		Prediction value	
		Death (+)	Living (-)
Real value	Death (+)	TP	FN
	Living (-)	FP	TN

**Table 1.** Confusion matrix. True positive (TP): The number of deaths that the model has correctly identified. False positive (FP): The number of living people but the model has incorrectly identified them as dead. True negative (TN): The number of people who are living and the model correctly identified them as living. False negative (FN): The number of people who are dead but the model has identified them as living incorrectly.

Evaluation measures	Formula
Accuracy	$(TP + TN) / (TP + FP + TN + FN)$
Sensitivity/ Recall	$TP / (TP + FN)$
Specificity	$TN / (TN + FP)$
F1-measure	$2 * ((Precision * Sensitivity) / (Precision + Sensitivity))$

**Table 2.** The performance evaluation measures.

### Explanation and justification the output of ML and DL models

Both ML and DL methods are often considered “black box” models, meaning that their inner workings are complex and not easily interpretable<sup>27,28</sup>. This lack of interpretability can be a significant drawback, especially in critical applications like healthcare where understanding the reasoning behind predictions is crucial. To address this challenge, researchers have been focusing on developing methods to improve the interpretability of these models. One notable technique that has emerged in recent years is Shapley Additive Explanations (SHAP), which was introduced by Lundberg and Lee. SHAP is specifically designed to provide insights into the predicted outcomes of ML models<sup>29</sup>. It builds upon the concept of Shapley values from cooperative game theory, which have gained traction in various fields, including clinical studies<sup>30,31</sup>.

The core idea behind SHAP is to assign contribution values to individual features in the dataset, indicating how much each feature influences the predicted outcomes. These contribution values are calculated by comparing predictions made with and without specific features included in the model. By considering all possible combinations of features, SHAP offers a comprehensive understanding of how each feature contributes to the final predictions. This holistic approach allows researchers to identify which features have the most significant impact on outcomes and whether their influence is positive or negative<sup>30</sup>.

Overall, SHAP provides a valuable tool for interpreting the predictions made by ML and DL models, thereby increasing transparency and trust in these models’ decision-making processes. Its application in healthcare settings, for example, can help clinicians better understand and validate the predictions made by AI models, leading to more informed and effective decision-making<sup>32</sup>. As in other studies, employing SHAP in this study yields invaluable insights into the influence of features on the predicted outcomes of both ML and DL models. Consequently, SHAP diagrams were generated for the model exhibiting the best performance across sensitivity, specificity, accuracy, ROC, and F1-score indices<sup>33,34</sup>.

It should be noted that while the importance of simpler algorithms like LASSO or Ridge for interpretability is acknowledged, our study aims to balance both accuracy and explainability. Complex models such as deep learning, when coupled with SHAP, maintain high predictive performance while providing interpretability through feature attribution. Simplicity has its advantages, particularly in practical medical settings, but the nuanced insights from SHAP enhance the understanding of the model’s decision-making process, which can be beneficial in complex cases like stroke prediction<sup>35</sup>.

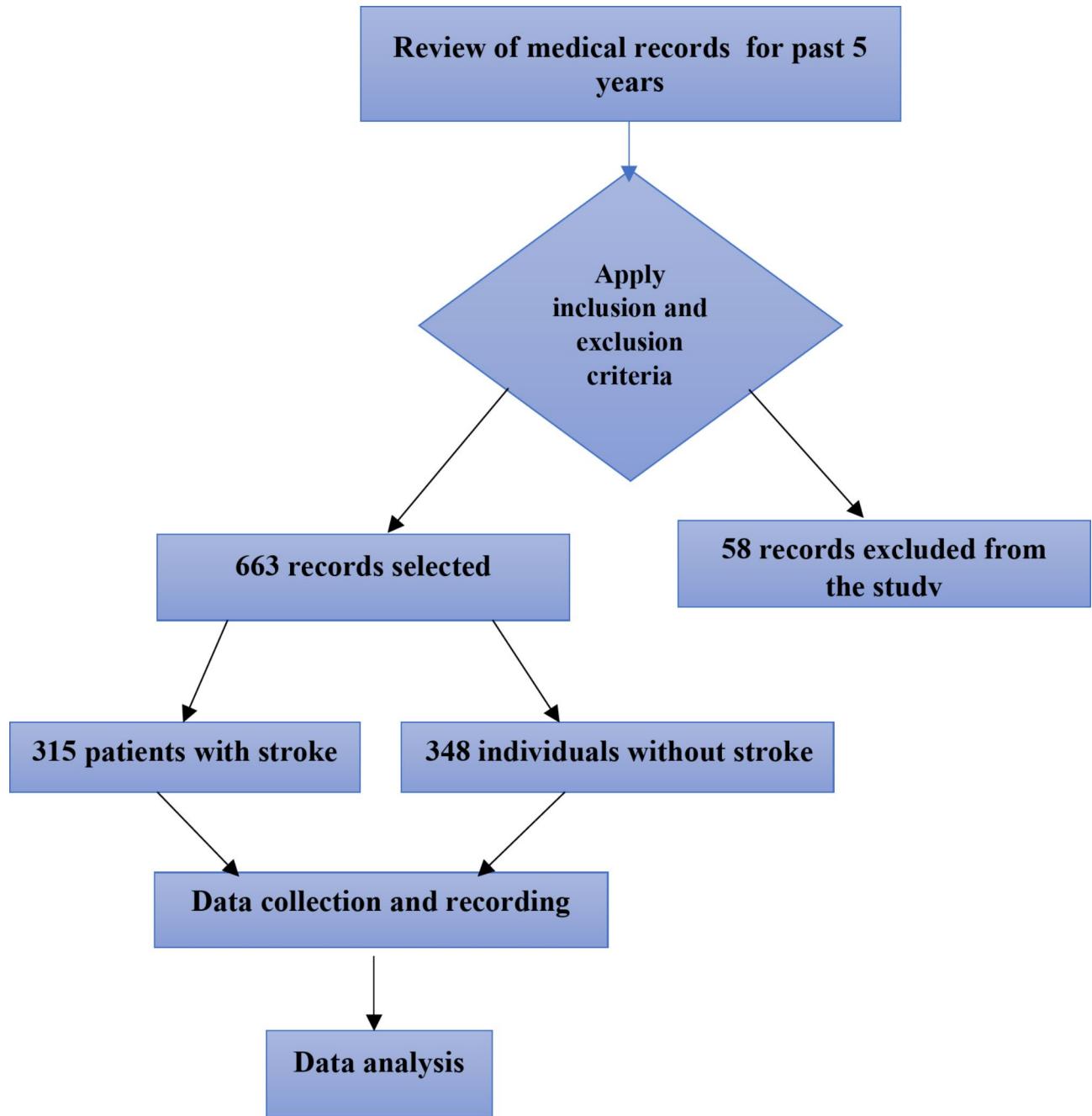
### Performance evaluation of models

The effectiveness of both ML and DL models was thoroughly assessed through performance metrics derived from the confusion matrix, as detailed in Table 1. Several key metrics, such as accuracy, specificity, sensitivity, F1-score, and the receiver operating characteristic (ROC) curve, were employed to evaluate model performance comprehensively (Table 2). It should be noted that in similar studies focused on stroke prediction, commonly used performance metrics include accuracy, sensitivity, specificity, F1-score, and ROC curve, which we have also employed<sup>35–38</sup>.

Accuracy provided an overall measure of the correctly predicted instances, while specificity and sensitivity gave insights into the model’s ability to identify true negatives and true positives, respectively. The F1-score balanced precision and recall, offering a single metric to evaluate performance, especially when dealing with imbalanced datasets. The ROC curve further illustrated the trade-off between sensitivity and specificity across different thresholds, enabling a visual comparison of model performance<sup>39,40</sup>.

## Results

Figure 1 illustrates the data extraction and analysis process.

**Fig. 1.** Study flow diagram.

#### Selected features for stroke prediction based on physicians' opinions

Based on physicians' opinions, 53 features influencing stroke in patients were identified (Table 3). The most important features identified were: Platelet levels (Mean = 224.37, SD = 85.29), Alkaline Phosphatase (ALP) (Mean = 199.56, SD = 102.47), and Cholesterol (Mean = 153.22, SD = 24.81).

#### Optimal parameters using hyperparameter tuning

Table 4 displays different models along with their parameters, showcasing the optimal parameters derived from hyperparameter tuning and majority voting.

#### Performance evaluation of selected models

Table 5 shows the performance evaluation of chosen models. Among all DL models, LSTM performed better in terms of sensitivity (96.15%). Additionally, among all DL models, FNN demonstrated superior performance in terms of specificity (96.0%), accuracy (96.0%), F1-score (95.0%), and ROC (98.0).

Features	Mean	std
Platelet	224.37	85.29
Alkaline Phosphatase (ALP)	199.56	102.47
Cholesterol	153.22	24.81
Serum Na	140.79	15.20
Serum K	140.79	15.20
Spontaneous Bacterial Peritonitis (SBP)	133.49	19.58
Fasting blood sugar (FBS)	104.45	37.97
Triglyceride	99.83	41.61
Oxygen saturation (SpO <sub>2</sub> )	95.63	3.57
Creatine phosphokinase (CPK)	89.93	128.62
PT Activity	87.29	16.89
Mean corpuscular volume (MCV)	86.20	6.47
Dibutyl phthalate (DBP)	80.30	12.52
PR	80.25	6.27
Troponin	74.76	1557.03
LDL	64.47	19.46
Age	43.88	22.26
Hematocrit (HCT)	39.89	5.99
HDL	36.27	16.82
Mean corpuscular hemoglobin concentration (MCHC)	34.00	1.79
Partial Thromboplastin Time (PTT)	31.82	8.42
Mean corpuscular hemoglobin (MCH)	29.39	2.84
Serum Glutamic Oxaloacetic Transaminase (SGOT)	20.55	12.47
Platelet large cell ratio (P-LCR)	20.30	8.01
Erythrocyte sedimentation rate (ESR)	18.27	13.87
Creatine phosphokinase-MB (CPK-MB)	17.31	17.31
Serum glutamic pyruvic transaminase (SGPT)	16.95	14.67
Blood urea nitrogen (BUN)	15.13	9.03
Prothrombin Time (PT)	14.48	6.45
RDW-CV	13.80	6.68
Hemoglobin (Hb)	13.76	2.84
Platelet distribution width (PDW)	11.91	4.18
W.B.C	11.20	8.60
mean platelet volume (MPV)	10.59	30.92
C-reactive protein (CRP)	10.55	15.76
Serum Ca	9.52	3.20
Hemoglobin A1C (HbA1C)	5.77	0.71
R.B.C	4.71	1.00
Inorganic P	3.45	1.63
Serum Mg	2.07	0.36
Creatinine	1.61	5.17
International normalised ratio (INR)	1.55	5.70
Gender	1.36	0.48
Hemiparesis	1.02	0.51
Hemiparesis	1.02	0.51
Headache	0.92	0.28
Total Bili	0.82	0.80
Aphasia	0.81	0.39
Marital status	0.68	0.47
Direct Bili	0.27	0.44
High blood pressure (hypertension)	0.23	0.42
DM	0.12	0.32
Cerebra Vascular Accident (CVA)	0.10	0.30

**Table 3.** Selected features for stroke prediction.

Num	ML & DP models	Hyper-Parameters
1	CNN	Conv1D (filters = 64, kernel_size = 5, MaxPoolingD (pool_size = 3), Dense (2)
2	LSTM	LSTM Units = 64, Activation Function = 'sigmoid', Optimizer = 'adam', Loss Function = 'binary_crossentropy', Epochs = 20, Batch Size = 128.
3	DNN	Layers and Neurons = 84, 42, 1, Activation Functions = 'relu', 'relu', 'sigmoid'; Optimizer = 'adam', Loss Function = 'binary_crossentropy', Epochs = 20, Batch Size = 128.
4	FNN	Hidden Layer Size = (128,), Max Iterations = 2000, Random State = 42, Activation Function = 'relu', Solver = 'adam', Learning Rate = 'constant', Batch Size = '128'.
5	RF	n_estimators = 400, criterion = 'entropy', max_depth = None, min_samples_split = 2, min_samples_leaf = 1, max_features = 'auto', bootstrap = True, oob_score = False, random_state = 42.
6	KNN	Test_size = 0.2, Random_state: 42, CV = 10
7	XGB	learning_rate = 0.05, n_estimators = 500, max_depth = 10, min_child_weight = 5, subsample = 0.9, colsample_bytree = 0.8, gamma = 0, reg_lambda = 0.
8	SVM	Probability = 'True', C = 5.0, kernel = 'rbf', degree = 0, gamma = 0.1

**Table 4.** Optimal parameters using hyperparameter tuning.

Models	Sensitivity (%)	Specificity (%)	Accuracy (%)	F1-score (%)	ROC (%)
Performance of each deep learning models					
CNN	96.0	87.0	92.0	91.0	96.0
LSTM	96.15	80.0	89.47	87.34	97.0
DNN	93.0	93.0	95.0	93.0	95.0
FNN	94.0	96.0	96.0	95.0	98.0
Performance of each machine learning models					
RF	99.9	100	99.0	99.0	99.0
KNN	70.0	94.0	86.0	80.0	83.0
XGB	94.0	95.0	95.0	95.0	95.0
SVM	29.0	94.0	71.0	44.0	64.0

**Table 5.** Performance evaluation of selected models.

Among ML models, RF performed better in terms of sensitivity (99.9%), accuracy (99.0%), specificity (100%), F1-score (99.0%), and ROC (99.9%). After the RF model, XGB performed better in terms of sensitivity (94.0%), accuracy (95.0%), specificity (95.0%), F1-score (95.0%), and ROC (95.0%).

In general, the performance of RF was better than that of all other models. However, the performance of DL models was better than that of ML models in all performance measures except for RF. Deep learning models (CNN, LSTM, DNN, and FNN) achieved sensitivities between 93.0% and 96.15%, while ML models (KNN, XGB, SVM) achieved sensitivities ranging from 29.0 to 94%. DL models achieved specificities ranging from 80.0 to 96.0%, while ML models achieved specificities between 89.47% and 96.0%. DL models achieved accuracies ranging from 92.0 to 96.0%, while ML models achieved accuracies between 71.0% and 95.0%. DL models achieved F1-scores ranging from 87.34 to 95%, while ML models achieved F1-scores between 44.0% and 95.0%. Additionally, DL models achieved ROC scores between 95.0% and 98.0%, while ML models achieved ROC scores between 64.0% and 95.0%.

Figure 2 illustrates the performance of selected MA and DL models in predicting stroke, whereas Fig. 3 contrasts the ROC curves of these models. Additionally, Fig. 4 presents the calibration plots comparing the ML and DL models for stroke prediction.

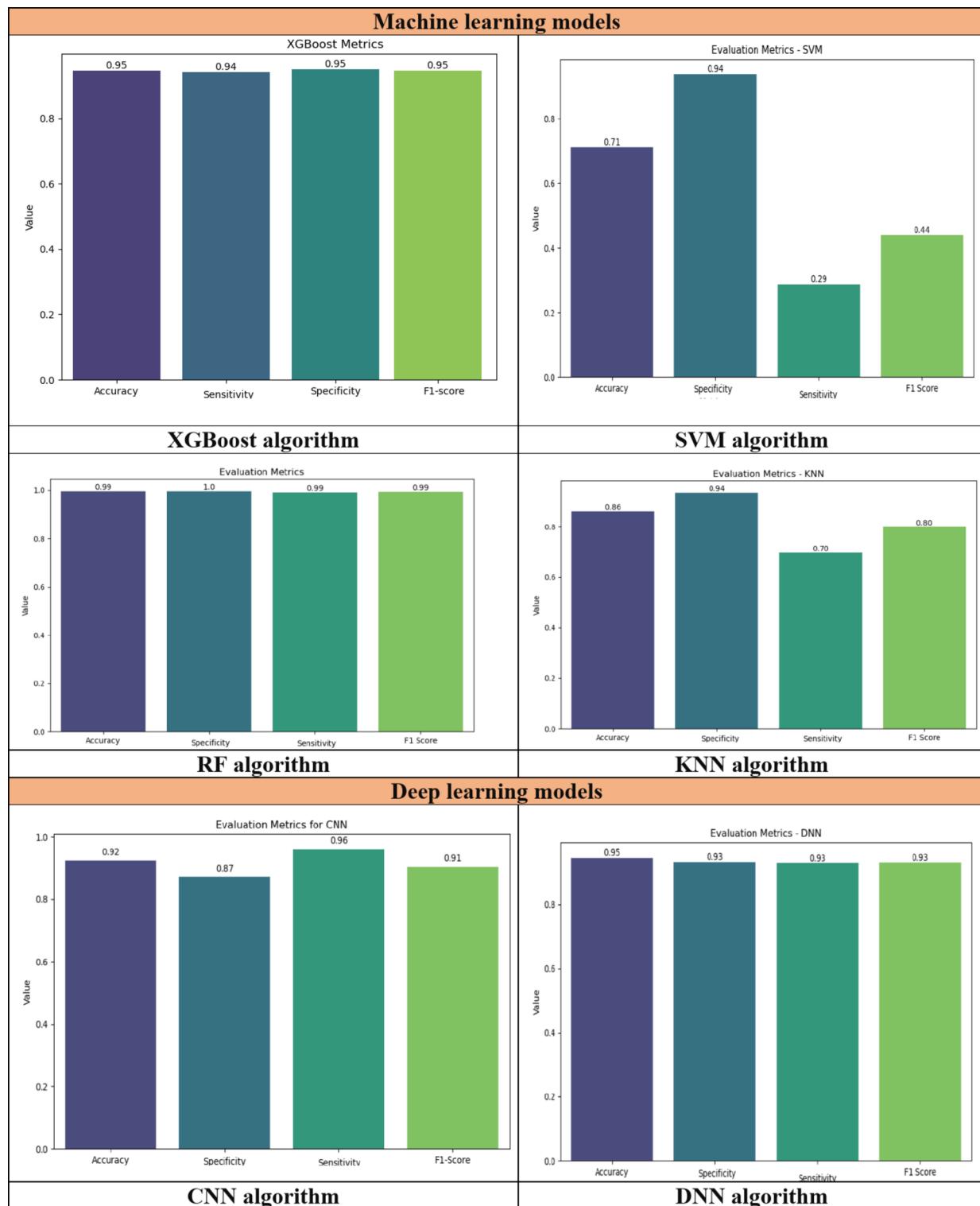
### SHAP summary plot

As noted previously, the RF algorithm outperformed all others in terms of sensitivity, specificity, accuracy, F1-score, and ROC, as illustrated in Fig. 5. Consequently, a SHAP plot was specifically generated for this algorithm. As can be seen in the figure below, age, triglyceride, and aphasia were the most important descriptor of SVM in predicting the stroke.

### Discussion

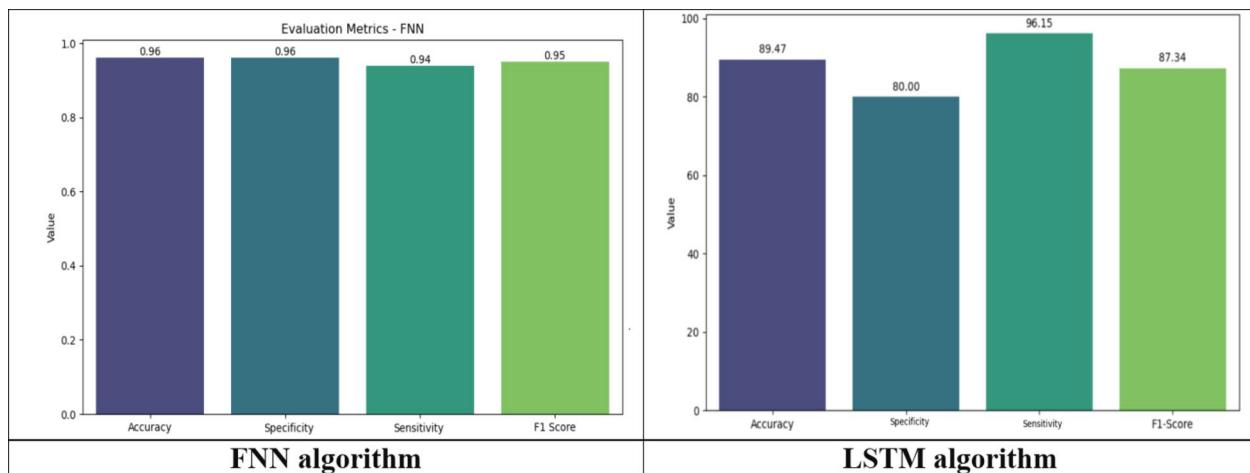
In this study, the performance of various DL and ML models in predicting strokes was assessed and compared. Among the DL models, LSTM emerged as the standout performer, demonstrating superior sensitivity at 96.15%. Furthermore, the FNN excelled in terms of specificity, accuracy, F1-score, and ROC among all DL models. On the ML side, RF exhibited remarkable performance, outshining other models with the highest sensitivity, accuracy, specificity, F1-score, and ROC. Following RF, XGB showcased commendable results in sensitivity, accuracy, specificity, F1-score, and ROC. Despite RF's overall superiority, DL models, including CNN, LSTM, DNN, and FNN, outperformed ML models in all performance measures except for RF.

Our study findings reveal the remarkable performance of LSTM and FNN models within the domain of DL models for stroke prediction. The LSTM model emerged as the superior DL model, demonstrating exceptional sensitivity at 96.15%. This high sensitivity indicates LSTM's proficiency in identifying individuals at risk of stroke,



**Fig. 2.** Performance evaluation of ML and DL models for stroke prediction.

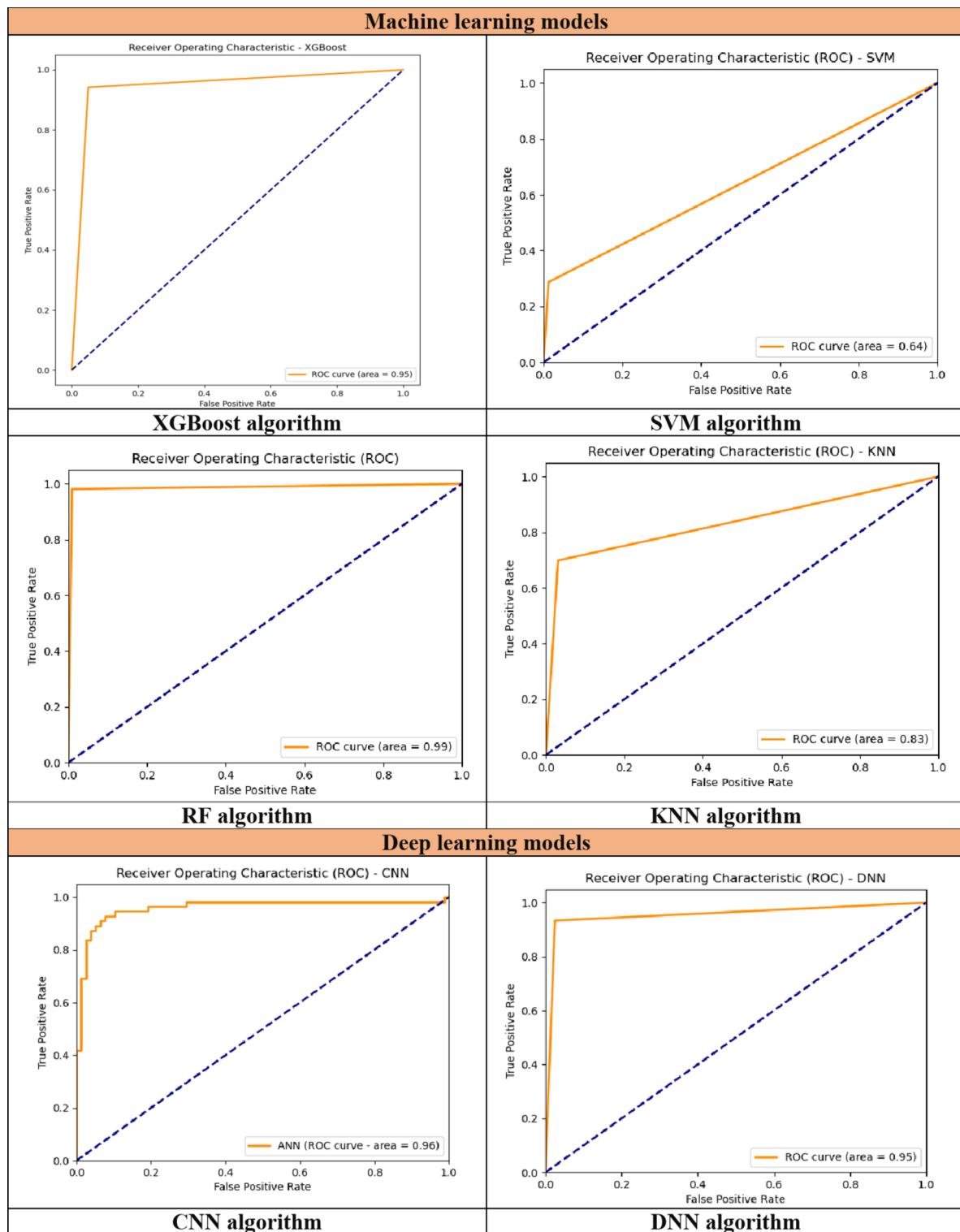
a critical aspect for implementing preventive healthcare measures effectively. The ability of LSTM to accurately identify such patients underscores its potential to contribute significantly to early intervention strategies aimed at reducing stroke incidence and associated morbidity and mortality rates. Choi et al.'s study<sup>41</sup> reveals that the LSTM model applied to raw EEG data achieved a 94.0% accuracy in predicting stroke, with low FPR (6.0%) and FNR (5.7%), highlighting the efficacy of non-invasive LSTM methods for real-time monitoring and prediction of stroke. This approach is noted for its potential to significantly advance early stroke detection, providing a cost-effective and more comfortable alternative to conventional imaging methods. Yu et al.'s study<sup>42</sup> also

**Figure 2.** (continued)

demonstrates the LSTM model's effectiveness in predicting Hemorrhagic Transformation (HT) in acute stroke patients, achieving an AUC-ROC of 89.4% and surpassing current machine learning benchmarks. These results highlight the LSTM model's clinical potential to improve management and therapy customization for acute ischemic stroke patients. So, according to the remarkable performance of the LSTM model, this underscores its potential contribution to early intervention strategies and its clinical utility in identifying individuals at risk of stroke. Furthermore, the effectiveness of the LSTM in predicting stroke and hemorrhagic transformation, as evidenced by Choi et al.<sup>41</sup>, and Yu et al.<sup>42</sup>, highlights its significant potential for improving management and therapy customization in acute ischemic stroke patients. The findings from the study by Yu et al.<sup>43</sup>, indicate that both the RF algorithm, a machine learning technique, and LSTM, a deep learning approach, are effective in detecting and predicting strokes using real-time EMG bio-signals. Therefore, it can be said that the potential of LSTM to enhance early detection and intervention strategies highlights its significance in clinical settings, paving the way for improved management and personalized treatment for acute stroke patients. Overall, these results advocate for the integration of advanced deep learning models like LSTM into routine clinical practice to optimize patient outcomes and reduce the burden of stroke-related morbidity and mortality.

In our analysis, among the DL models examined, the FNN model stands out due to its exemplary performance across a range of critical metrics. Notably, the FNN model demonstrated robustness in identifying true negatives, a key factor in reducing the number of false alarms in clinical diagnoses. Its high scores in specificity, accuracy, F1-score, and ROC curve evaluations underscore its precision and reliability as a predictive tool. The FNN's comprehensive efficacy across these metrics not only attests to its capability in accurately predicting stroke risk but also emphasizes its potential applicability in real-world clinical settings. The ability to identify individuals at risk of stroke with high confidence and accuracy is paramount in preventative healthcare strategies. By leveraging the strengths of the FNN model, healthcare providers can potentially improve early intervention measures, tailor individualized prevention plans, and ultimately, contribute to reducing the incidence and impact of stroke within populations. In Sanderson et al.'s study<sup>44</sup>, the optimal FNN model (AUC: 83.52) surpassed logistic regression (AUC: 81.79), indicating improved predictive accuracy for quantifying suicide risk. Additionally, the optimal FNN model showed higher sensitivity (69.96) compared to logistic regression (65.31), with similar specificity, PPV, and NPV. Despite varying learning rates and epoch combinations, each FNN neuron configuration achieved the same maximum AUC, suggesting adaptable model optimization. These results highlighted the potential of FNN models to enhance suicide risk assessment using available healthcare data<sup>44</sup>. Raziani et al.<sup>45</sup>, also mentioned that FNNs, primarily due to their straightforward architecture, allow for efficient training and prediction. Their layered structure enables the modeling of complex relationships between inputs and outputs, making them particularly adept at tasks involving classification, regression, and pattern recognition. Additionally, their capacity to generalize from provided data helps in making accurate predictions on unseen data, making FNNs invaluable for tasks requiring high levels of precision and reliability<sup>46</sup>. The analysis highlights the FNN model as a standout among deep learning models due to its exceptional performance in critical metrics for predicting stroke and suicide risks. Its robustness in identifying true negatives significantly reduces false alarms in clinical diagnoses, making it a reliable predictive tool. With high scores in specificity, accuracy, and F1-score, the FNN model demonstrates its effectiveness in real-world clinical applications.

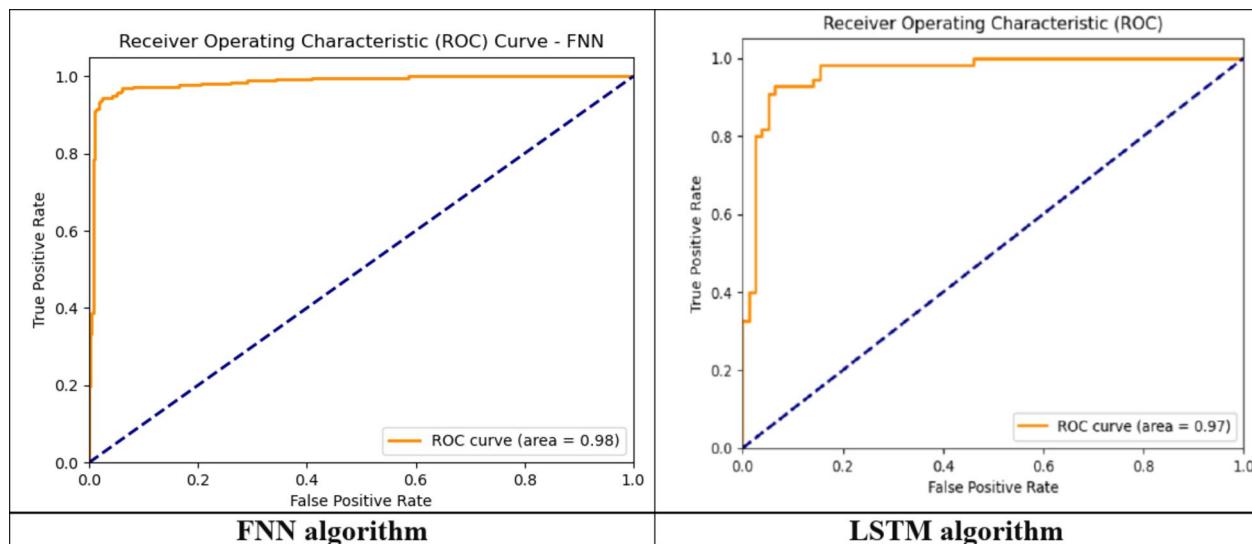
In total, these findings contribute valuable insights into the application of DL models, particularly LSTM and FNN, in stroke prediction. By leveraging the strengths of these models, healthcare practitioners can potentially enhance risk stratification strategies and implement targeted interventions for individuals deemed at high risk of stroke. Furthermore, the superior performance of LSTM and FNN underscores the importance of leveraging advanced computational techniques in healthcare decision-making processes, particularly in the realm of preventive medicine. However, further research is warranted to explore the generalizability of these findings across diverse populations and healthcare settings. Additionally, efforts to optimize the performance and interpretability of DL models, as well as their integration into existing clinical workflows, remain crucial



**Fig. 3.** ROC curves comparing ML and DL models for stroke prediction.

for their effective adoption in real-world healthcare scenarios. Overall, the findings of our study highlight the promising role of DL models, particularly LSTM and FNN, in revolutionizing stroke prediction and preventive care practices.

According to our study findings, when examining traditional ML models, the RF model outperformed all others, including DL models, in nearly all performance metrics. RF's sensitivity, accuracy, specificity, F1-score, and ROC were exceptionally high, signifying its strong predictive power and reliability in stroke prediction. Following RF, the XGB model also showed commendable performance, although it did not match the RF's exemplary

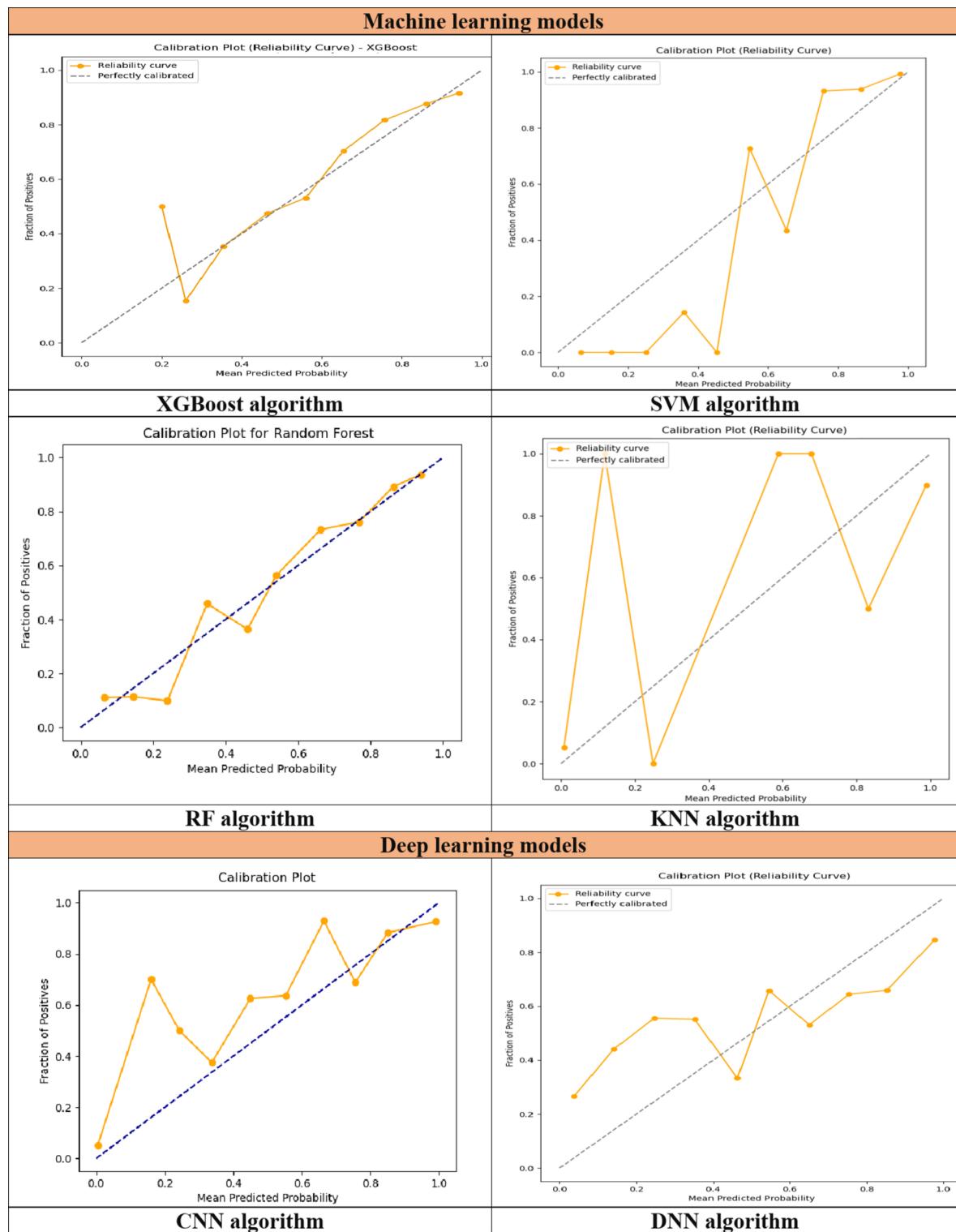
**Figure 3.** (continued)

results. This comparative analysis highlights a pivotal finding: while DL models generally surpassed traditional ML models in stroke prediction, the RF model stood as a notable exception, surpassing all other models in performance. This suggests that while DL technologies hold significant promise for advancing predictive analytics in healthcare, certain ML models like RF maintain a competitive edge, especially in tasks requiring high sensitivity and specificity. The overall superior performance of DL models over ML models, with the exception of RF, underscores the potential of DL in healthcare applications. These findings align with previous studies<sup>47–51</sup> that identified the RF model as an effective off-the-shelf solution in the medical field, where understanding feature importance plays a crucial role in patient interoperability. Consequently, the random forest algorithm was recognized as the most reliable predictor for stroke incidence. Fernandez-Lozano et al.<sup>52</sup>, also utilized the RF model to predict stroke outcomes. Their findings demonstrated that RF can effectively predict long-term mortality and complications in stroke patients. They noted that RF, an ensemble learning method, combines multiple predictions to enhance output robustness and accuracy, uniquely providing an internal measure for evaluating feature importance. This model excels across diverse problems, even with unbalanced or incomplete data. Importantly, RF analyzes variable significance using the Gini importance index, which assesses each variable's contribution to model purity during the selection process, offering a weighted evaluation rather than a mere summation of values across all trees<sup>52</sup>. In an other study Islam et al.<sup>16</sup>, found that among the Logistic Regression, Decision Tree Classifier (DTC), K-NN, and RF models, the RF model achieved the highest accuracy across all performance metrics at 96%. The K-NN model secured third place with a performance metric of 90%, while the DTC came in second with an accuracy of 93%.

This analysis indicates that while deep learning models generally excel in predictive analytics, the RF model remains a significant contender, particularly in tasks requiring high sensitivity and specificity. Previous research supports these findings, recognizing RF as an effective tool in medical applications, especially for understanding feature importance and patient interoperability. Additionally, studies have shown that RF can effectively predict long-term outcomes in stroke patients, further establishing its reliability as a predictive model in healthcare.

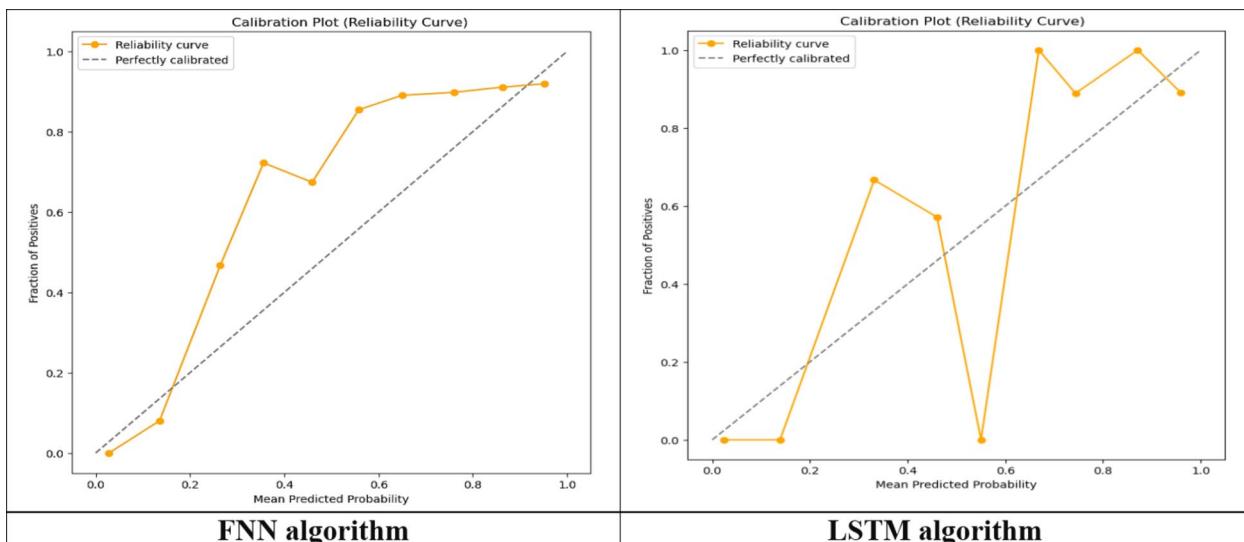
In our study we observed a notable distinction in performance between DL models and traditional ML models in predicting stroke outcomes. Despite RF's overall superiority, DL models (CNN, LSTM, DNN, and FNN) outperformed ML models in all performance measures except for RF. Different studies<sup>53,54</sup> show that DL excels in predicting diseases compared to traditional machine learning methods due to its ability to automatically discover intricate patterns and relationships within vast and complex datasets. Unlike conventional machine learning models that often rely on handcrafted features and predetermined rules, DL models, particularly neural networks, can autonomously extract relevant features directly from the raw input data<sup>55</sup>. This feature extraction process enables DL models to capture subtle nuances and hidden patterns that may not be apparent to human experts or may be difficult to encode manually<sup>56</sup>. Moreover, the scalability and adaptability of DL architectures contribute to their superior performance in disease prediction tasks<sup>57</sup>.

DL models can effectively handle high-dimensional data, such as medical images<sup>58</sup>, genomic sequences<sup>59</sup>, and electronic health records<sup>60</sup>, without the need for extensive feature engineering or dimensionality reduction. Additionally, DL models can learn from large-scale datasets, benefiting from the abundance of labeled and unlabeled data available in healthcare domains<sup>61</sup>. The ability to learn from vast amounts of data allows DL models to generalize well to unseen examples and adapt to diverse patient populations, thereby improving the robustness and reliability of disease prediction systems. Overall, the data-driven nature, feature learning capabilities, and scalability of DL make it a powerful approach for predicting diseases with high accuracy and precision, paving the way for more effective healthcare interventions and personalized treatment strategies.

**Fig. 4.** Calibration plots comparing ML and DL models for stroke prediction.

### Study limitations

Given the challenges associated with collecting data from various hospitals across Iran, our study exclusively relied on data from a single hospital (Hazrat Rasool Akram Hospital). To enhance the generalizability of findings, future research endeavors should consider either expanding the sample size or including data from multiple provinces. Furthermore, our study utilized only eight models, encompassing four DL and four machine learning approaches. To gain a more comprehensive understanding, it is advisable for future studies to explore a broader array of models.

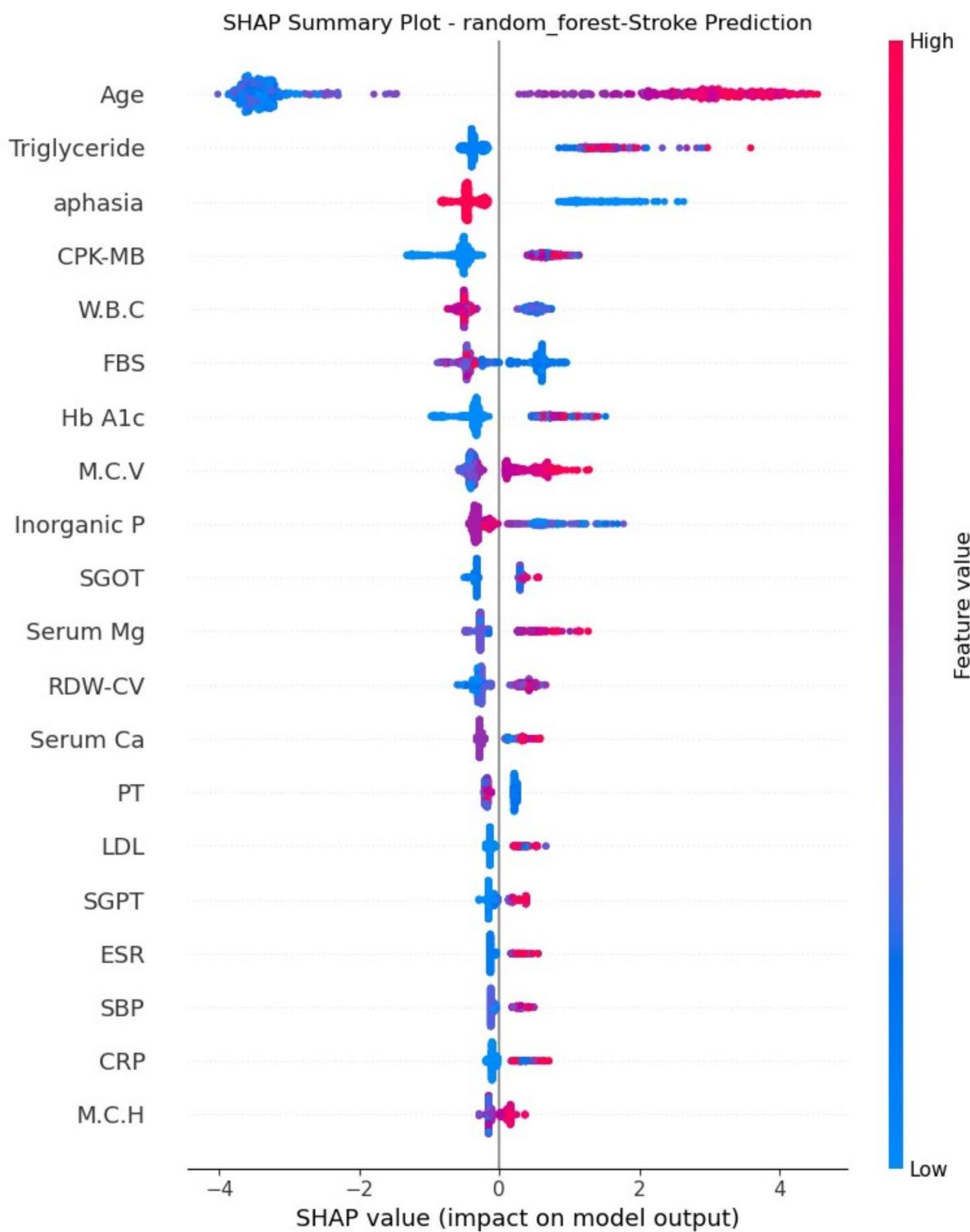
**Figure 4.** (continued)

Additionally, it is important to acknowledge that our findings may be influenced by the structure of missing data. In this study, we did not explicitly analyze or differentiate the missing data mechanism, such as missing at random (MAR) or missing not at random (MNAR). If the missingness mechanism was not completely random, it could introduce bias in the models' predictions. Future studies should consider investigating the nature of missingness and applying suitable imputation or correction techniques to minimize potential bias.

## Conclusion

In this study, the efficacy of DL and ML models was assessed in the early detection of strokes, aiming to improve patient outcomes by reducing the risk of severe consequences such as permanent neurological damage or death. By comparing various DL and ML models, the study sought to identify the most effective models for predicting stroke based on a dataset of patients and healthy individuals.

The findings revealed that among DL models, LSTM displayed the highest sensitivity, while FNN showed superior performance in specificity, accuracy, F1-score, and ROC metrics. Notably, the RF model outperformed all other models, including DL models, in almost all performance measures. This suggests that while DL models generally offer significant advantages over traditional ML models in predicting stroke, the exceptional performance of RF highlights its potential as a highly effective tool in clinical settings. Incorporating RF alongside DL models could enhance the predictive accuracy and reliability of stroke detection systems, ultimately improving patient care and outcomes.



**Fig. 5.** SHAP summary plot for RF model.

#### Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Received: 7 July 2024; Accepted: 10 December 2024

Published online: 28 December 2024

## References

1. Stroke Cerebrovascular accident ( World Health Organization (WHO), 2024).
2. Lacy, C. R., Suh, D. C., Bueno, M. & Kostis, J. B. Delay in presentation and evaluation for acute stroke: Stroke Time Registry for Outcomes Knowledge and Epidemiology (STROKE). *Stroke* **32**, 63–69 (2001).
3. Hayes, M. et al. Tales from the trips: a qualitative study of timely recognition, treatment, and transfer of emergency department patients with acute ischemic stroke. *J. Stroke Cerebrovasc. Dis.* **28**, 1219–1228 (2019).
4. Emon, M. U. et al. Performance analysis of machine learning approaches in stroke prediction. *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*1464–1469 (IEEE, 2020).
5. Sirsat, M. S., Fermé, E. & Câmara, J. Machine learning for brain stroke: A review. *J. Stroke Cerebrovasc. Dis.* **29**, 105162 (2020).
6. Mouridsen, K., Thurner, P. & Zaharchuk, G. Artificial intelligence applications in stroke. *Stroke* **51**, 2573–2579 (2020).
7. Feng, R., Badgeley, M., Mocco, J. & Oermann, E. K. Deep learning guided stroke management: a review of clinical applications. *J. neurointerventional Surg.* **10**, 358–362 (2018).
8. Heo, J. et al. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke* **50**, 1263–1265 (2019).
9. Sirsat, M. S., Fermé, E. & Câmara, J. Machine learning for brain stroke: a review. *J. Stroke Cerebrovasc. Dis.* **29**, 105162 (2020).
10. Rajora, M., Rathod, M. & Naik, N. S. Stroke prediction using machine learning in a distributed environment. *Distributed Computing and Internet Technology: 17th International Conference, ICDCIT 2021, Bhubaneswar, India, January 7–10, 2021, Proceedings* 17 238–252. (Springer, 2021).
11. Lee, H. et al. Machine learning approach to identify stroke within 4.5 hours. *Stroke* **51**, 860–866 (2020).
12. Cheon, S., Kim, J. & Lim, J. The use of deep learning to predict stroke patient mortality. *Int. J. Environ. Res. Public Health* **16**, 1876 (2019).
13. Yu, Y. et al. Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. *JAMA Netw. Open* **3**, e200772–e200772 (2020).
14. Nielsen, A., Hansen, M. B., Tietze, A. & Mouridsen, K. Prediction of tissue outcome and assessment of treatment effect in acute ischemic stroke using deep learning. *Stroke* **49**, 1394–1401 (2018).
15. Golzar, J., Noor, S. & Tajik, O. Convenience sampling. *Int. J. Educ. Lang. Stud.* **1**, 72–77 (2022).
16. Islam, M. M. et al. Stroke prediction analysis using machine learning classifiers and feature technique. *Int. J. Electron. Commun. Syst.* **1**, 17–22 (2021).
17. El-Melegy, M. et al. Prostate cancer diagnosis via visual representation of tabular data and deep transfer learning. *Bioengineering* **11** 635 (2024).
18. Borsos, B., Allaart, C. G. & van Halteren, A. Predicting stroke outcome: a case for multimodal deep learning methods with tabular and CT perfusion data. *Artif. Intell. Med.* **147**, 102719 (2024).
19. Liang, S. & Yu, H. Revealing new therapeutic opportunities through drug target prediction: a class imbalance-tolerant machine learning approach. *Bioinformatics* **36**, 4490–4497 (2020).
20. Buturović, L. & Miljković, D. A novel method for classification of tabular data using convolutional neural networks. *BioRxiv*, 2020.2005.2002.074203 (2020).
21. Zhu, Y. et al. Converting tabular data into images for deep learning with convolutional neural networks. *Sci. Rep.* **11**, 11325 (2021).
22. Gicic, A., Donko, D. & Subasi, A. Time sequence deep learning model for ubiquitous tabular data with unique 3D tensors manipulation. *Entropy* **26**, 783 (2024).
23. Alvi, R. H., Rahman, M. H., Khan, A. A. S. & Rahman, R. M. Deep learning approach on tabular data to predict early-onset neonatal sepsis. *J. Inform. Telecommunication.* **5**, 226–246 (2021).
24. Prabowo, H. et al. Aggregating time series and tabular data in deep learning model for university students' gpa prediction. *IEEE Access*. **9**, 87370–87377 (2021).
25. Malakouti, S. M., Menhaj, M. B. & Suratgar, A. A. The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction. *Clean. Eng. Technol.* **15**, 100664 (2023).
26. Bardehnet, R., Brendel, M., Kégl, B. & Sebag, M. Collaborative hyperparameter tuning. *Int. Conf. Mach. Learn. PMLR*, 199–207 (2013).
27. Papernot, N. et al. Practical black-box attacks against machine learning. Proceedings of the 2017 ACM on Asia conference on computer and communications security 506–519 (2017).
28. Sudmann, A. On the media-political dimension of artificial intelligence: Deep learning as a black box and OpenAI. *Digit. Cult. Soc.* **4**, 181–200 (2018).
29. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural. Inf. Process. Syst.* **30** (2017).
30. Sun, J., Sun, C. K., Tang, Y. X., Liu, T. C. & Lu, C. J. Application of SHAP for explainable machine learning on age-based subgrouping mammography questionnaire data for positive mammography prediction and risk factor identification. *Healthcare (Basel)* **11** (2023).
31. Wang, F. et al. Potential of the non-contrast-enhanced chest CT radiomics to distinguish molecular subtypes of breast cancer: A retrospective study. *Front. Oncol.* **12**, 848726 (2022).
32. Hu, C. et al. Interpretable machine learning for early prediction of prognosis in sepsis: a discovery and validation study. *Infect. Dis. Ther.* **11**, 1117–1132 (2022).
33. Moulaei, K. et al. Explainable artificial intelligence (XAI) for predicting the need for intubation in methanol-poisoned patients: a study comparing deep and machine learning models. *Sci. Rep.* **14**, 15751 (2024).
34. Zhai, Y., Lin, X., Wei, Q., Pu, Y. & Pang, Y. *Interpretable prediction of cardiopulmonary complications after non-small cell lung cancer surgery based on machine learning and SHapley additive exPlanations* 9 (Heliyon, 2023).
35. Chandramohan, R. M. *Stroke Detection and Prediction Using Deep Learning Techniques and Machine Learning Algorithms* (National College of Ireland, 2022).
36. Rehman, A. et al. RDET stacking classifier: a novel machine learning based approach for stroke prediction using imbalance data. *PeerJ Comput. Sci.* **9** (2023).
37. Al-Zubaidi, H., Dweik, M. & Al-Mousa, A. Stroke prediction using machine learning classification methods. *2022 international Arab conference on information technology (ACIT)* 1–8 (IEEE, 2022).
38. Abedi, V. et al. Prediction of long-term stroke recurrence using machine learning models. *J. Clin. Med.* **10**, 1286 (2021).
39. Le, N. Q. K., Li, W. & Cao, Y. Sequence-based prediction model of protein crystallization propensity using machine learning and two-level feature selection. *Brief. Bioinform.* **24** (2023).
40. Kha, Q. H., Le, V. H., Hung, T. N. K., Nguyen, N. T. K. & Le, N. Q. K. Development and validation of an explainable machine learning-based prediction model for drug-food interactions from chemical structures. *Sensors (Basel)* **23** (2023).
41. Choi, Y. A. et al. Deep learning-based stroke disease prediction system using real-time bio signals. *Sensors* **21**, 4269 (2021).
42. Yu, Y. et al. LSTM network for prediction of hemorrhagic transformation in acute stroke. *Medical Image Computing and Computer Assisted Intervention-MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV* 22 177–185 (Springer, 2019).
43. Yu, J. et al. AI-based stroke disease prediction system using real-time electromyography signals. *Appl. Sci.* **10**, 6791 (2020).
44. Sanderson, M., Bulloch, A. G., Wang, J., Williamson, T. & Patten, S. B. Predicting death by suicide using administrative health care system data: can feedforward neural network models improve upon logistic regression models? *J. Affect. Disord.* **257**, 741–747 (2019).

45. Raziani, S., Ahmadian, S., Jalali, S. M. J. & Chalechale, A. An efficient hybrid model based on modified whale optimization algorithm and multilayer perceptron neural network for medical classification problems. *J. Bionic Eng.* **19**, 1504–1521 (2022).
46. Jamett, M. & Acuña, G. An interval approach for weight's initialization of feedforward neural networks. *Mexican International Conference on Artificial Intelligence* 305–315 (Springer, 2006).
47. Charbuty, B. & Abdulazeez, A. Classification based on decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends.* **2**, 20–28 (2021).
48. Biau, G. & Scornet, E. A random forest guided tour. *Test* **25**, 197–227 (2016).
49. Sarica, A., Cerasa, A. & Quattrone, A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Front. Aging Neurosci.* **9**, 329 (2017).
50. Shanthakumari, R., Nalini, C., Vinothkumar, S., Roopadevi, E. & Govindaraj, B. Multi disease prediction system using random forest algorithm in healthcare system. *2022 International Mobile and Embedded Technology Conference (MECON)* 242–247 (IEEE, 2022).
51. Belgiu, M. & Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. photogrammetry remote Sens.* **114**, 24–31 (2016).
52. Fernandez-Lozano, C. et al. Random forest-based prediction of stroke outcome. *Sci. Rep.* **11**, 10071 (2021).
53. Zafar, I. et al. Reviewing methods of deep learning for intelligent healthcare systems in genomics and biomedicine. *Biomed. Signal Process. Control.* **86**, 105263 (2023).
54. Zhao, J., Han, X., Ouyang, M. & Burke, A. F. Specialized deep neural networks for battery health prognostics: Opportunities and challenges. *J. Energy Chem.* (2023).
55. Lee, S. H., Chan, C. S., Mayo, S. J. & Remagnino, P. How deep learning extracts and learns leaf features for plant classification. *Pattern Recogn.* **71**, 1–13 (2017).
56. Yu, Z., Wang, K., Wan, Z., Xie, S. & Lv, Z. Popular deep learning algorithms for disease prediction: a review. *Clust. Comput.* **26**, 1231–1251 (2023).
57. Loukil, Z., Mirza, Q. K. A., Sayers, W. & Awan, I. A Deep Learning based scalable and adaptive feature extraction framework for medical images. *Inform. Syst. Front.* 1–27 (2023).
58. Shen, D., Wu, G. & Suk, H. I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).
59. Koumakis, L. Deep learning models in genomics; are we there yet? *Comput. Struct. Biotechnol. J.* **18**, 1466–1473 (2020).
60. Rajkomar, A. et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.* **1**, 18 (2018).
61. Leclerc, S. et al. Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Trans. Med. Imaging* **38**, 2198–2210 (2019).

## Author contributions

Conceptualization: Khadijeh Moulaei, Mohammad Reza Afrash, Seyed Mohammad Mousavi, Lida Afshri, Reza Moulaei, Babak Sabet. Data curation: Lida Afshri, Reza Moulaei. Funding: Khadijeh Moulaei, Mohammad Reza Afrash. Project administration: Mohammad Reza Afrash, Khadijeh Moulaei. Design and Modeling: Mohammad Reza Afrash, Khadijeh Moulaei. Resources: Mohammad Reza Afrash, Khadijeh Moulaei. Supervision: Maryam Ahmadi, Mohammad Reza Afrash, Khadijeh Moulaei. Writing-original draft: Mohammad Reza Afrash, Khadijeh Moulaei. Writing-review & editing: Mohammad Reza Afrash, Khadijeh Moulaei.

## Declarations

### Competing interests

The authors declare no competing interests.

### Ethical approval

The ethics committee of Iran University of Medical Sciences approved the study, as indicated by reference number R.IUMS.REC.1398.234. All procedures followed the relevant guidelines and regulations set by the committee of Iran University of Medical Sciences. When participants were unable to give consent, their families provided informed consent on their behalf. Additionally, the informed consent collected at our institution included permission for potential future retrospective analyses.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-82931-5>.

**Correspondence** and requests for materials should be addressed to M.R.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024