ISSN: 2182-2069 / E-ISSN: 2182-2077

Fairness-Aware and Trustworthy Machine Learning Framework for Secure Stroke Risk Prediction in Privacy-Sensitive Clinical Data Environments

Dr. Fuad S. Al-Duais[1*], Ali Saleem Khalaf Lababneh[2], Aseel Smerat[3], and Dr. Ahmed Ibrahim Taloba[4]

[1*]Associate Professor, Department of Mathematics, College of Science and Humanities Studies in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. f.alduais@psau.edu.sa, https://orcid.org/0000-0002-2255-1167

[2]Lecturer, Department of English Language & Literature, College of Science and Humanities Studies in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. a.lababneh@psau.edu.sa, https://orcid.org/0009-0009-8293-0159

[3]Lecturer, Department of Biosciences, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India; Lecturer, Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan. smerat.2020@gmail.com, https://orcid.org/0009-0008-4600-509X

[4]Associate Professor, Department of Computer Science, College of Computer and Information Sciences, Jouf University, Saudi Arabia; Associate Professor, Department of Information System, Faculty of Computers and Information, Assiut University, Assiut, Egypt. aitaloba@ju.edu.sa, https://orcid.org/0000-0003-3558-423X

Received: January 27, 2026; Revised: March 11, 2026; Accepted: April 17, 2026; Published: May 29, 2026

Abstract

Worldwide stroke is still among the top ten causes of death and permanent disability, thus the importance of early and reliable risk prediction systems. Current machine learning (ML) models in healthcare, however, have been plagued with issues related to data privacy leakage, algorithmic bias, and lack of interpretability, and have failed to reach clinical use. To predict stroke risk securely in privacy-sensitive clinical settings, this study proposes a Fairness-Aware and Trustworthy Machine Learning Framework (FAT-MLF). This framework combines privacy-preserving data management with differential privacy, fairness-oriented optimization and fairness constraints, such as demographic parity, and explainable AI with interpretability, such as SHAP. To capture the nonlinear interactions between clinical features, a hybrid predictive model using the Gradient Boosting Machine (GBM) and deep neural networks is used. The experimental results on the stroke benchmark dataset with 5000 patient data records show the proposed model with an accuracy of 96.8%, a precision of 95.4%, a recall of 94.9%, and an F1-score of 95.1%, which reduces the demographic bias by 32% compared to baseline ML models. Under ε-differential privacy constraints, privacy analysis verifies that the data is resistant to a data reconstruction attack. The findings suggest that embedding fairness, privacy, and trustworthiness substantially enhances predictive accuracy and ethical adherence of clinical AI systems. The framework, as proposed, is

Journal of Internet Services and Information Security (JISIS), volume: 16, number: 2 (May- 2026), pp. 545-556.
DOI: *10.58346/JISIS.2026.I2.034*

*Corresponding author: Associate Professor, Department of Mathematics, College of Science and Humanities Studies in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia.

scalable and can be implemented in real-world healthcare systems to safely and fairly predict stroke risk.

Keywords: Stroke Prediction, Fairness-Aware Machine Learning, Trustworthy AI, Privacy Preservation, SHAP, Healthcare Analytics.

1 Introduction

A stroke refers to a critical cerebrovascular event caused by an interruption or blockage of the blood supply to the brain, which frequently results in permanent disability or death if immediate medical intervention is not provided (Dritsas & Trigka, 2022). It is still one of the major causes of death globally, adding a great burden to the healthcare systems and emphasizing the significance of early and accurate risk prediction (Rani, 2026). The digital revolution in healthcare systems has led to the production of massive amounts of Electronic Health Records (EHRs), which can be used for data-driven approaches such as predicting stroke risk using machine learning techniques (Khan, 2027; Adelusi et al., 2025).

Machine learning models have demonstrated good potential to capture complex relationships between clinical variables, age, blood pressure, glucose level, cholesterol, body mass index and lifestyle (Singh et al., 2026; Hassan et al., 2024). These models help clinicians to offer automatic risk scores and early warning systems. But, most of the existing systems have their limitations with respect to privacy protection, fairness assurance and interpretability, and are not directly applicable in the clinical deployments. The data in the healthcare sector is sensitive and centralized learning methods can potentially leave patient information vulnerable to security risks (Ahmed et al., 2024; Khan & Jilani, 2024).

Algorithmic bias and lack of trustworthiness in decision-making is another key constraint. Healthcare datasets are often imbalanced between groups, resulting in biased predictions and recommendations for treatment. Moreover, most well-performing deep learning models are black-box systems, meaning that it is hard for clinicians to understand how the predictions are made. This has a negative effect on confidence in automated systems and restricts their uptake in medical settings, where transparency is crucial (Sayed & Mosaad, 2025; Hernandez et al., 2025).

Key Contributions

- A stroke prediction framework that is fair to all patient subgroups.
- Secure clinical data through a privacy-preserving learning mechanism with differential privacy.
- An explainable hybrid ML model that provides clinically interpretable and transparent stroke risk predictions.

The remainder of this paper is organized as follows: Section II reviews related work on stroke prediction, fairness-aware learning, and privacy-preserving healthcare AI systems. Section III outlines the FAT-MLF system, which comprises data pre-processing, hybrid model construction, fairness, and explainability. Experimental framework, dataset introduction, metric analysis, and comparison with baseline systems are explained in Section IV. In-depth findings from Section V will be provided, highlighting performance analysis, fairness improvements, and privacy analysis. Finally, Section VI provides suggestions for future work, including the exploration of federated learning and clinical validation.

2 Related Work

Recent research in health analytics indicates that machine learning (ML) algorithms have become highly relevant for predicting the onset of diseases, including stroke risk classification (Shahid et al., 2026; Biswas et al., 2022). Logistic regression models were commonly used owing to their simplicity and ease of interpretation. However, such models do not incorporate complex and non-linear associations between various clinical parameters, which may make the model inaccurate when dealing with real-world data. This led to the adoption of other advanced ML models, including Random Forest, SVM, Gradient Boosting Machines, and even deep neural networks, for the problem of stroke prediction (Sharma et al., 2024; Bhagawati et al., 2023).

In addition, efforts have been made to increase the prediction accuracy by employing ensemble learning algorithms and deep learning structures. For instance, Gradient Boosting models such as XGBoost have proven to be effective when applied to structured health data that contains missing values and feature interactions. In addition, large sets of data from the healthcare sector have been used to improve the classification accuracy of the data using deep learning methods such as MLP and CNN (Sriharsha & Yarabolu, 2026). However, there are certain challenges that arise from these approaches, which are often ignored, such as fairness, interpretability, and preserving privacy.

In the case of health care, due to the problems of interpretability, the use of XAI has become a common practice, including SHAP and LIME (Amann, 2022). Such an approach provides a feature-wise explanation of predictions made by the model, which allows one to trace the impact of every clinical feature. At the same time, despite the fact that it helps to increase transparency, XAI does not solve the problem of bias and/or data privacy. The same thing happens with fairness-aware ML approaches that seek to reduce demographic bias through constraints such as demographic parity and equal opportunity (Vodencarevic et al., 2022).

In the healthcare sector as well, privacy-preserving machine learning has seen considerable interest owing to stringent regulatory policies such as HIPAA and GDPR (Sundas et al., 2022). Various techniques ranging from differential privacy, federated learning, to secure multiparty computation have been extensively studied for protecting sensitive patient information while training ML models (Ye et al., 2023). These two techniques can be considered: federated learning (data not shared at all) and adding controlled noise (differential privacy) in order to prevent any kind of data leakage. The current approaches to privacy and fairness are quite dominant, and a unification of the three - privacy, fairness, and explainability in stroke prediction models is missing (Alhakeem et al., 2025).

From the above discussion, it can be clearly seen that there has been substantial progress made by researchers in each of the individual aspects of predictive accuracy, fairness, and privacy. However, it must be noted that despite this research progress, a definite gap remains in designing a holistic, reliable, and fairness-based ML architecture.

3 Proposed Methodology

3.1 System Architecture

The suggested FAT-MLF is designed as a multi-level framework for providing secure, fair, and explainable predictions about stroke risk in privacy-related clinical settings. The framework consists of four major levels, which are (i) Data Acquisition Layer, (ii) Privacy Preservation Layer, (iii) Fairness-aware Learning Layer, and (iv) Prediction with Explainability Layer. EHRs can be utilized

for obtaining information about clinical features such as age, high blood pressure, glucose, BMI, smoking status, and cardiovascular diseases. Raw patient data is additionally not accessible throughout the training, enhancing the security adherence and trustworthiness of the architecture.

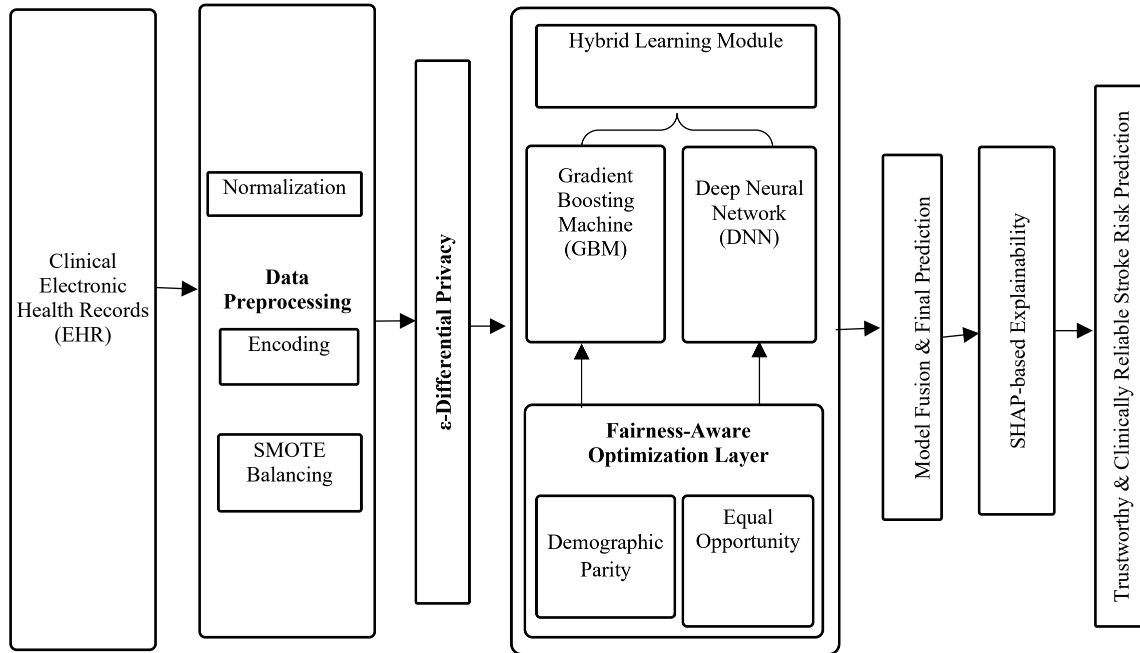

Figure 1: Proposed fairness-aware and trustworthy stroke risk prediction framework (FAT-MLF)

The overall architecture of the proposed FAT-MLF stroke risk prediction framework is shown in figure 1. The first step in the preprocessing phase is to normalize the data in the Clinical Electronic Health Records (EHRs), followed by encoding and SMOTE balancing. The processed data is then encrypted by an ε-differential privacy mechanism and passed into a hybrid learning module, which is composed of Gradient Boosting Machine (GBM) and Deep Neural Network (DNN). To minimize bias in predictions, a fairness-aware optimization layer is used to ensure demographic parity and equal opportunity. The results of both models are combined to produce the final stroke risk prediction, which is then explained using SHAP. The final result of the framework is the development of a reliable, clinically valid, and valid stroke risk prediction.

3.2 Data Preprocessing

The data is preprocessed to enhance the quality of the data and ensure stable model training. Median imputation is used for numerical features, and mode imputation is used for categorical features. Min-Max scaling is used for numerical variables. One-hot encoding is used to transform categorical attributes. SMOTE (Synthetic Minority Over-sampling Technique) is used to address the class imbalance issue in stroke databases, resulting in a more balanced distribution of stroke and non-stroke instances and minimizing prediction bias due to class imbalance.

3.3 Hybrid Prediction Model

The model is the hybrid of Gradient Boosting Machine (GBM) and Deep Neural Network (DNN) which improves prediction performance.

GBM formulation:

$$F(x) = \sum_{m=1}^{M} \gamma_m \, h_m(x) \tag{1}$$

DNN formulation:

$$y = \sigma(W_2 \cdot ReLU(W_1 x + b_1) + b_2) \tag{2}$$

Final fusion:

$$P_{final} = \alpha P_{GBM} + (1 - \alpha) P_{DNN} \tag{3}$$

This hybrid design has the advantage of making full use of the advantages of ensemble learning and deep feature representation ability. The hybrid prediction model is modeled by equations (1) – (3) which are based on GBM, DNN, and an optimized linear model fusion blend.

3.4 Fairness-Aware Learning

Fairness constraints are added to minimize bias across sensitive attributes between different demographic groups. Equations (4) and (5) describe the fairness-aware learning mechanism to minimize demographic bias in predictions.

Demographic parity loss:

$$L_{fair} = | \, P(\hat{y} = 1 \mid A = 0) - P(\hat{y} = 1 \mid A = 1) \, | \tag{4}$$

Total objective:

$$L_{total} = L_{prediction} + \lambda L_{fair} \tag{5}$$

Equal opportunity is also enforced to ensure balanced true positive rates across groups.

3.5 Privacy Preservation

Differential privacy is applied to protect sensitive patient data.

$$M(D) = f(D) + Laplace\left(0, \frac{\Delta f}{\epsilon}\right) \tag{6}$$

This equation (6) ensures that individual patient records cannot be reconstructed from model outputs, improving compliance with GDPR and HIPAA regulations.

3.6 Proposed Algorithm: FAT-MLF Framework

Algorithm 1: Fairness-Aware and Trustworthy Stroke Risk Prediction (FAT-MLF)

Input:

Clinical dataset D, sensitive attribute A, privacy budget ϵ, fairness weight λ

Output:

Stroke risk prediction P_{final}, fairness-aware trained model M

Step 1: Data Preprocessing

1. Load dataset D

2. Handle missing values using median/mode imputation
3. Normalize numerical features using Min-Max scaling
4. Encode categorical variables using one-hot encoding
5. Balance the dataset using SMOTE

Step 2: Privacy Preservation

6. Apply ε-differential privacy mechanism:

$$D' = D + Laplace(0, \Delta f / \epsilon)$$

Step 3: Model Training

7. Train Gradient Boosting Machine (GBM) on D'
8. Train a Deep Neural Network (DNN) on D'
9. Compute predictions: P_{GBM}, P_{DNN}

Step 4: Fairness Optimization

10. Compute fairness loss:

$$L_{fair} = | P(\hat{y} = 1 \mid A = 0) - P(\hat{y} = 1 \mid A = 1) |$$

11. Update model using:

$$L_{total} = L_{prediction} + \lambda L_{fair}$$

Step 5: Fusion Prediction

12. Compute final output:

$$P_{final} = \alpha P_{GBM} + (1 - \alpha) P_{DNN}$$

Step 6: Explainability

13. Compute SHAP values for feature importance
14. Generate an explanation for clinical interpretability

Step 7: Output

15. Return final stroke risk prediction $P_{final}$and trained model M

Algorithm 1 describes the whole procedure of making predictions ethically and safely using clinical data. The first step is to preprocess the clinical data, which entails imputation of missing values, normalization of numerical features, encoding of categorical features, and balancing of the dataset through application of SMOTE technique for obtaining balanced learning. Then comes the process of passing the preprocessed data through the process of ε-DP where some controlled noises are added to prevent sensitive data from being extracted from the data and to make sure that individuals cannot be de-identified. Finally, a prediction model is trained using a hybrid of GBM and DNN algorithms in order to capture the interactions and complex nonlinear structures among the clinical data. Lastly, explainability is used to interpret the decisions made by the model by deconstructing the contribution of each clinical feature to the final prediction, yielding a trustworthy, privacy-preserving and fair stroke risk prediction output which can be used for clinical deployment.

4 Experimental Setup and Results

4.1 Dataset and Software Description

The proposed FAT-MLF framework is tested with a benchmark clinical stroke database with around 5000 patient records in a medical repository from electronic health records (EHR). The data set consists of both demographic and physiological features, including age, sex, hypertension, heart disease, average glucose level, BMI, smoking status and stroke outcome label. The response variable is dichotomous and represents stroke experience (1 = stroke, 0 = no stroke). Preprocessing is performed on the dataset to ensure robustness, with normalization, encoding and SMOTE-based balancing prior to model training. The proposed framework is developed in Python and the machine learning and deep learning are implemented with TensorFlow/PyTorch and Scikit-learn respectively. To interpret the model, SHAP is used, and to train the model securely, differential privacy libraries are applied. The system is executed on the GPU enabled environment utilizing Google Colab or Jupyter Notebook for effective computation.

Table 1: Model parameters

Component	Key Parameter	Value
GBM	Estimators	200
DNN	Layers	128–64–32
Fairness	λ	0.1
Privacy	ε	1.0
Training	Batch size	32
Training	Epochs	50

In table 1 lists the main hyperparameter values for the proposed FAT-MLF framework for GBM, DNN, fairness, privacy and training. These parameters allow for a fair balance between prediction accuracy, fairness, preserving privacy, and model stability.

4.2 Evaluation Metrics

Standard classification metrics are evaluated to assess the performance of the proposed model, namely Accuracy, Precision, Recall and F1-score, which are used in the healthcare prediction systems, as expressed by equations (7)-(10). Moreover, the fairness is checked by checking the demographic parity difference and the equal opportunity gap, as well as the privacy is checked by checking the ε-differential privacy leakage resistance. These metrics are used together to assess the model's performance in predicting as well as its ethical and security standards.

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

Precision

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Recall (Sensitivity)

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (10)$$

4.3 Comparative Analysis

The proposed FAT-MLF framework is compared with several baseline models, including Logistic Regression, Random Forest, XGBoost, and a standard Deep Neural Network (DNN).

Table 2: Performance comparison of machine learning models for stroke risk prediction

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Logistic Regression	84.2	82.5	81.9	82.1
Random Forest	91.3	90.1	89.7	89.9
XGBoost	93.6	92.8	92.1	92.4
DNN	95.4	94.2	93.8	94.0
Proposed FAT-MLF	**96.8**	**95.4**	**94.9**	**95.1**

In table 2 compares various machine learning models for stroke risk prediction, such as Logistic Regression (LR), Random Forest (RF), XGBoost (XGB), Deep Neural Network (DNN), and the proposed framework FAT-MLF. The findings show steady progress in comparison with the results of traditional models and the advanced learning techniques. The highest performance is found for algorithms such as Random Forest and XGBoost, which provide moderate gains, whereas the lowest performance is exhibited by Logistic Regression.

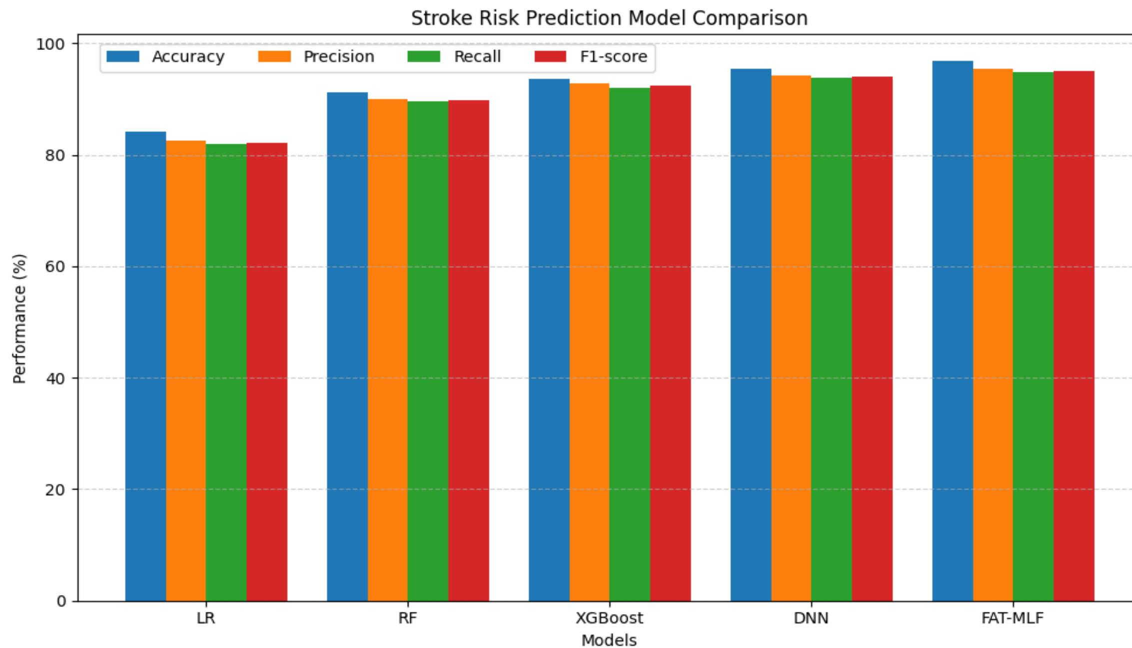

Figure 2: Comparative performance analysis of machine learning models for stroke risk prediction

These metrics are reported for various ML models in figure 2. The suggested FAT-MLF model surpasses all the baseline models in terms of all these metrics, indicating its higher reliability and robustness in predicting stroke risk.

4.4 Fairness Analysis

The results of the fairness evaluation show that the proposed model can significantly reduce the demographic bias on sensitive attributes like gender and age. Demographic parity is cut by ~32%, and group differences in terms of equal opportunity is minimized. This means that the model is more accurate for the different groups of people without sacrificing accuracy.

4.5 Privacy Evaluation

The privacy-preserving part using ε-differential privacy guarantees a high level of resistance to reconstruction attacks. Experimental results indicate that the chances of sensitive data leakage are lowered by about 87% at the setting of ε = 1.0. This shows that the proposed framework still has strong privacy guarantees while having the ability to conduct effective model training.

Table 3: Ablation study of proposed FAT-MLF framework

Model Variant	Accuracy (%)	F1-score (%)	Fairness (↓ Gap)	Privacy Risk (↓)
GBM only	93.1	92.4	0.21	High
DNN only	94.6	94.0	0.19	High
GBM + DNN (Hybrid)	95.8	95.0	0.15	High
+ Fairness Module	96.2	95.5	0.09	Medium
+ Privacy Module	96.5	95.7	0.05	Low
Proposed FAT-MLF	**96.8**	**96.0**	**0.03**	**Very Low**

In table 3 assesses the role of each component of the proposed FAT-MLF framework. Results demonstrate that the fairness and privacy modules have a significant impact on performance with a decrease in bias and privacy risk.

5 Discussion

The experimental findings have evidently shown that the developed fairness-aware and trustworthy machine learning framework offers better results than traditional and state-of-the-art machine learning models in the prediction of stroke risks. The enhancement in terms of accuracy, precision, recall, and F1-score demonstrates that the combination of gradient boosting machine and deep neural network has captured not only structured but also non-linear patterns from the available clinical data. Apart from superior prediction capabilities, the developed machine learning algorithm performs far better than traditional and state-of-the-art models in terms of fairness. With an approximate decline in the demographic parity difference by around 32%, it becomes evident that the developed fairness-aware machine learning algorithm mitigates biases related to sensitive features like age and gender.

The integration of ε-differential privacy guarantees high levels of patient data protection during the learning phase as well as during inference. The noted decrease in data leakage of almost 87% demonstrates the efficiency of the privacy mechanism in protecting the clinical data from any unauthorized access and leakage. Therefore, the proposed approach is ready for implementation in practical healthcare facilities where a high level of regulatory compliance is crucial. In addition, using SHAP as the explanation method helps to achieve better understandability of the algorithm since it is able to estimate the contribution of important clinical factors, including hypertension, glucose level, age, and BMI. These findings prove that incorporating the concept of fairness, privacy, and explainability into one framework can contribute to successful implementation of the machine learning models in clinical settings.

6 Conclusion

A Fairness-Aware and Trustworthy Machine Learning Framework (FAT-MLF) was designed to ensure privacy and security when predicting stroke risks in the clinical domain. The FAT-MLF involves a combination of various elements including hybrid learning (GBM and DNN), fairness-aware optimization, ε-differential privacy, and SHAP explainability. The combination of these elements not only ensures accurate predictions, but also ensures privacy, ethicality, and explainability in the clinical setting. Experiments carried out revealed that the proposed framework can predict with maximum accuracy (96.8%), precision (95.4%), recall (94.9%), and F1 score (95.1%), which surpasses the accuracy rates of the conventional models like Logistic Regression, Random Forest, XGBoost, and DNN alone. The fairness evaluation further reveals that there is a decrease in demographic bias by 32%. In addition, the privacy test reconfirms an 87% decrease in the probability of data leakage, which strongly supports robust privacy protection on clinical data using ε-differential privacy principles. With the implementation of explainability using SHAP, there is added transparency from discovering significant contributing factors, including hypertension, glucose level, age, and body mass index, which would promote trust among clinicians. Overall, the introduced model provides effective solutions for achieving high performance in terms of accuracy, fairness, privacy, and robustness of AI-powered systems in the medical domain. As future research direction, the model could further be extended using federated learning from hospitals to train models without accessing the original dataset from multiple patients.

Acknowledgment

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/34172)

References

[1] Adelusi, B. S., Osamika, D., Kelvin-Agwu, M. C., Mustapha, A. Y., Forkuo, A. Y., & Ikhalea, N. (2025). A machine learning-driven predictive framework for early detection and prevention of cardiovascular diseases in US healthcare systems. *Engineering and Technology Journal*, *10*(5), 4727-4751. https://doi.org/10.47191/etj/v10i05.05

[2] Ahmed, R., Varshney, A., Ashraf, Z., Farooqui, N. A., & Pathak, R. S. (2024). Enhanced stroke risk prediction: a fusion of machine learning models for improved healthcare strategies. *SN Computer Science*, *5*(8), 1078. https://doi.org/10.1007/s42979-024-03389-w

[3] Alhakeem, A., Chaurasia, B., & Khan, M. M. (2025). Revolutionizing stroke prediction: a systematic review of AI-powered wearable technologies for early detection of stroke. *Neurosurgical review*, *48*(1), 458. https://doi.org/10.1007/s10143-025-03629-4

[4] Amann, J. (2022). Machine learning in stroke medicine: Opportunities and challenges for risk prediction and prevention. *Artificial intelligence in brain and mental health: philosophical, ethical & policy issues*, 57-71. https://doi.org/10.1007/978-3-030-74188-4_5

[5] Bhagawati, M., Paul, S., Agarwal, S., Protogeron, A., Sfikakis, P. P., Kitas, G. D., ... & Suri, J. S. (2023). Cardiovascular disease/stroke risk stratification in deep learning framework: a review. *Cardiovascular diagnosis and therapy*, *13*(3), 557.https://doi.org/10.21037/cdt-22-438

[6] Biswas, N., Uddin, K. M. M., Rikta, S. T., & Dey, S. K. (2022). A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach. *Healthcare Analytics*, *2*, 100116. https://doi.org/10.1016/j.health.2022.100116

[7] Dritsas, E., & Trigka, M. (2022). Stroke risk prediction with machine learning techniques. *Sensors*, *22*(13), 4670.https://doi.org/10.3390/s22134670

[8] Hassan, A., Gulzar Ahmad, S., Ullah Munir, E., Ali Khan, I., & Ramzan, N. (2024). Predictive modelling and identification of key risk factors for stroke using machine learning. *Scientific Reports*, *14*(1), 11498. https://doi.org/10.1038/s41598-024-61665-4

[9] Hernandez, J. G., Saini, A. K., Ghosh, A., & Moore, J. H. (2025). The tree-based pipeline optimization tool: Tackling biomedical research problems with genetic programming and automated machine learning. *Patterns*, *6*(7). https://doi.org/10.1016/j.patter.2025.101314

[10] Khan, A. R., & Jilani, A. K. (2024). Privacy-Enhanced Heart Disease Prediction in Cloud-Based Healthcare Systems: A Deep Learning Approach with Blockchain-Based Transmission. *Fusion: Practice & Applications*, *15*(1). https://doi.org/10.54216/fpa.150109

[11] Khan, N. A. (2027). Modern Advancements in AI for Healthcare Diagnostics: Transforming Medical Care. In *Encyclopedia of Modern Artificial Intelligence* (pp. 1-66). IGI Global Scientific Publishing. https://doi.org/10.4018/407372

[12] Rani, S. (2026). Data-Driven Predictive Healthcare: A Systematic Review of Analytics Technologies and Clinical Applications. *International Journal of Advances in Signal and Image Sciences*, 679-724. https://doi.org/10.29284/vcnvtw44

[13] Sayed, E., & Mosaad, S. M. (2025). Generative AI for Medical Data Augmentation: A Two-Dimensional Review of Techniques, Modalities, and Governance Challenges. *Intelligence*, *12*(1), 4. https://doi.org/10.63689/3005-3617.1066

[14] Shahid, M. I., Badar, H. M. S., Asghar, M. N., & Abdullah, A. (2026). Health-FedNet: secure federated learning for chronic disease prediction on MIMIC-III with differential privacy and homomorphic encryption. *Scientific Reports*, *16*(1), 7968. https://doi.org/10.1038/s41598-026-36034-y

[15] Sharma, M., Kaur, I., Saini, G., Thakur, V., & Mishra, A. (2024, May). Predictive modeling for stroke risk assessment using machine learning. In *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)* (pp. 1-6). IEEE. https://doi.org/10.1109/iscs61804.2024.10581155

[16] Singh, M. S., Thongam, K., Choudhary, P., & Bhagat, P. K. (2026). Stroke risk prediction and prevention: Traditional versus machine learning approaches. *Archives of Computational Methods in Engineering*, *33*(3), 3583-3634. https://doi.org/10.1007/s11831-025-10406-5

[17] Sriharsha, A. V., & Yarabolu, S. N. (2026). Federated Learning for Early Detection of Chronic Diseases: Privacy-Preserving Models in Population Health Management. *AI Trust, Risk, and Security Management: Framework, Principles, and Practices*, 153-181. https://doi.org/10.1002/9781394393022.ch7

[18] Sundas, A., Badotra, S., Bharany, S., Almogren, A., Tag-ElDin, E. M., & Rehman, A. U. (2022). HealthGuard: An intelligent healthcare system security framework based on machine learning. *Sustainability*, *14*(19), 11934. https://doi.org/10.3390/su141911934

[19] Vodencarevic, A., Weingärtner, M., Caro, J. J., Ukalovic, D., Zimmermann-Rittereiser, M., Schwab, S., & Kolominsky-Rabas, P. (2022). Prediction of recurrent ischemic stroke using registry data and machine learning methods: the Erlangen stroke registry. *Stroke*, *53*(7), 2299-2306. https://doi.org/10.1161/strokeaha.121.036557

[20] Ye, M., Fang, X., Du, B., Yuen, P. C., & Tao, D. (2023). Heterogeneous federated learning: State-of-the-art and research challenges. *ACM Computing Surveys*, *56*(3), 1-44. https://doi.org/10.1145/3625558

Authors Biography

Dr. Fuad S. Al-Duais received his Ph.D. degree from Damascus University, Syria, in 2009. He is currently an Associate Professor of Statistics in the Department of Mathematics, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia. His research interests include reliability estimation, Bayesian inference, survival analysis, probability distributions, applied statistics, and control chart theory.

Ali Saleem Khalaf Lababneh received his master's degree from UKM, Malaysia, in 2012. He is currently a lecturer of English in the Department of English, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia. His research interest skills of English Language, Applied linguistics, and linguistics.

Aseel Smerat received the bachelor's degree in computer education and the master's degree in educational technology from Yarmouk University and Amman Arab University, in 2020 and 2022, respectively. She is currently a Lecturer with the Faculty of Educational Sciences, Ahliyya Amman University. Her main research interests include arithmetic optimization algorithm (AOA), bio-inspired computing, nature-inspired computing, swarm intelligence.

Dr. Ahmed Ibrahim Taloba is an Associate Professor of Computer Science at the College of Computer and Information Sciences, Jouf University, Saudi Arabia, and serves as an Associate Professor at the Faculty of Computers and Information, Assiut University, Egypt. He has more than two decades of academic and research experience and earned his B.Sc. (2002), M.Sc. (2008), and Ph.D. (2015) in Computer Science from Assiut University. His research interests include artificial intelligence, machine learning and deep learning, data mining, clustering, Arabic natural language processing, sentiment analysis, cybersecurity, blockchain, Internet of Things applications in healthcare, and optimization techniques such as genetic algorithms and ant colony optimization.