Journal of Artificial Intelligence in Medical Applications (JAIMA)
Vol. 1, No. 1, 2025, pp. 28–40
ISSN: 3106-1494 | DOI: https://doi.org/10.65511/jaima.v1i1.737

Improving Early Stroke Diagnosis in Emergency Medicine Using AI-Powered Clinical Decision Support

Zaid J. Al-Araji[1] *, Balqees Talal Hasan[2], Ammar Awad Mutleg[3], Narjes Benameur[4], Korhan Cengiz[5]

[1]Department of Computer Networks and Internet, College of Information Technology, Ninevah University, Ninevah 41001, Iraq; zaid.jasim@uoninevah.edu.iq
[2]College of Information Technology, Ninevah University, Ninevah 41001, Iraq; balqees.hasan@uoninevah.edu.iq
[3]Ministry of Education/General Directorate of Curricula, Pure Science Department, Baghdad 10065, Iraq; ammar.awad14@gmail.com
[4]Laboratory of Biophysics and Medical Technologies, Higher Institute of Medical Technologies of Tunis, University of Tunis El Manar, Tunis, Tunisia; narjes.benameur@istmt.utm.tn
[5]Faculty of Organization and Informatics, University of Zagreb, Croatia; kcengiz@foi.hr

Received: 01/06/2025, Revised: 18/07/2025, Accepted: 27/07/2025, Published: 16/09/2025

ABSTRACT: Early and accurate diagnosis of stroke in emergency medicine is critical to ensuring timely treatment and reducing long-term patient disability or mortality. However, the complex and time-sensitive nature of stroke symptoms, combined with the high-pressure environment of emergency departments, often leads to diagnostic delays or errors. This paper proposes an AI-powered Clinical Decision Support System (CDSS) designed to assist clinicians in the early detection of stroke using structured patient data such as vital signs, medical history, and neurological observations. The system leverages machine learning techniques to classify patient cases as likely stroke or non-stroke, providing real-time diagnostic support during triage or initial evaluation. To evaluate the effectiveness of the proposed approach, we develop a simulation environment that mimics emergency department workflows and incorporates both synthetic and publicly available clinical datasets. Performance is assessed using standard metrics including accuracy, sensitivity, specificity, and area under the ROC curve (AUC). The results demonstrate the potential of AI-driven decision support to enhance diagnostic accuracy for stroke patients in emergency settings.

Keywords: Healthcare, Artificial Intelligence, Stroke Diagnosis, Clinical Decision Support.

1. INTRODUCTION

Stroke—encompassing both ischemic and hemorrhagic sub-types—is the second leading cause of death and a primary driver of long-term disability worldwide, accounting for more than 12 million new cases and 6.5 million deaths annually [1]. Neurological damage progresses rapidly: it is often estimated that 1.9 million neurons are lost each minute during an untreated large-vessel occlusion [2]. Consequently, the therapeutic window for intravenous thrombolysis (≈ 4.5 h) and mechanical thrombectomy (≤ 6 h, extendable in select cases) makes early and accurate diagnosis in emergency departments (EDs) a critical determinant of patient outcome.

Despite well-established stroke scales (e.g., FAST, NIHSS), diagnostic delays and misclassifications remain prevalent. Recent multi-center audits report that up to 30 % of stroke or transient ischemic attack (TIA) cases are initially missed in overcrowded EDs, especially when symptoms are subtle or atypical [3]. Contributing factors include limited access to immediate neuroimaging, heterogeneity of clinical presentations, and high cognitive load on triage staff. These limitations motivate decision-support tools capable of synthesizing routine triage data into actionable risk estimates before confirmatory imaging is available.

1. Machine-learning (ML) and deep-learning (DL) methods have shown promise in stroke prediction from electronic health records [4] and neuro-images [5]. However, extant studies commonly rely on:
2. Static feature snapshots captured at admission, ignoring short-term temporal trends in vital signs and laboratory values.
3. Single-model pipelines (e.g., logistic regression, random forest) that may not fully exploit heterogeneous data modalities.

Limited explainability, hindering clinician trust and regulatory acceptance. Thus, there remains a need for real-time, interpretable, and temporally aware AI systems tailored to the fast-paced ED environment. To bridge these gaps, we propose an AI-powered Clinical Decision Support System (CDSS) that combines a Gradient-Boosted Tree (GBT) model with a Temporal Convolutional Network (TCN) in a soft-voting ensemble. The design rationale is two-fold: GBTs excel at capturing non-linear interactions in tabular risk factors, while TCNs model the evolution of sequential clinical data through dilated causal convolutions. The system is trained and validated on the Stroke Prediction Dataset (Kaggle) and benchmarked against five established classifiers (Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbors, LightGBM). This study presents several key contributions to the field of AI-assisted stroke diagnosis in emergency medicine:

- Hybrid Ensemble Architecture (GBT + TCN). A novel soft-voting ensemble that combines Gradient-Boosted Trees (GBT) with a Temporal Convolutional Network (TCN) is proposed. This architecture leverages the strengths of both static feature learning and temporal sequence modeling, making it suitable for current datasets and future real-time ED integration.
- Clinically-Oriented Design with Real-Time Capability. The system is optimized for real-time inference (<10 ms) on commodity hardware, ensuring practical deployment feasibility in high-pressure emergency environments where time is critical.
- Robust Performance Across Multiple Classifiers. A comprehensive evaluation of six machine learning algorithms (XGBoost, LightGBM, Random Forest, SVM, Logistic Regression, KNN) is conducted using balanced metrics (accuracy, F1-score, recall, precision, ROC-AUC), with the proposed ensemble showing superior or competitive results.
- Interpretability through Explainable AI (XAI). The model integrates SHAP and permutation-based feature importance analyses, providing transparent explanations of prediction logic—an essential requirement for clinician trust and regulatory approval.
- Forward Compatibility with Streaming Clinical Data. Although trained on static data, the inclusion of TCN architecture enables future integration with streaming vital signs and time-evolving patient data, laying the groundwork for more advanced real-time diagnostic systems.
- Prototype of a Clinical Decision Support System (CDSS). The study demonstrates a functional and modular AI-powered CDSS tailored to early stroke detection, aligning its components with typical emergency department workflows and priorities.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 details the proposed methodology; Section 4 presents experimental results; and Section 5 concludes with future research directions.

2. RELATED WORKS

Over the past decade, artificial intelligence (AI) has gained increasing attention in the medical domain, offering new opportunities for improving diagnostic accuracy and clinical decision-making. In the context of acute stroke, timely and precise diagnosis is critical, yet emergency departments (EDs) often face challenges due to limited diagnostic tools, time pressure, and the variability of symptom presentation. These limitations have motivated researchers to explore AI-driven solutions that can process structured and unstructured clinical data to assist in early risk stratification and diagnostic support. A growing body of work has investigated the use of machine learning (ML) and deep learning (DL) algorithms for stroke prediction and outcome estimation, leveraging both electronic health records (EHRs) and medical imaging data. These studies have applied various models—ranging from logistic regression and support vector machines to convolutional and recurrent neural networks—with varying levels of accuracy and interpretability. More recently, hybrid and ensemble learning methods have emerged to overcome the limitations of individual models and enhance prediction robustness, especially in imbalanced clinical datasets. Despite these advances, there remains a gap in integrating temporal sequence modeling with ensemble learning in a unified framework that is both accurate and interpretable for use in real-time ED environments. The following review explores key contributions in stroke-related AI research, with particular focus on traditional ML models, deep learning approaches, temporal architectures such as Temporal Convolutional Networks (TCNs), and ensemble methods such as Gradient Boosted Trees (GBTs) and LightGBM.

Accoding to [6] provide a pragmatic framework for the development of an AI-based decision support system by examining the various stages, which may ultimately enhance patient care and outcomes. [7] create and assess a temporal convolutional neural network (TCN) that does not compute perfusion maps in order to predict stroke lesion outcomes directly from 4D CTP datasets obtained at admission. By training the proposed TCN on different numbers of CTP frames—8, 16, and 32 time points—the authors examined the effect of the time window size using 176 CTP scans in total. Transparent Reporting of a Multivariable

Predictive Model for Individual Prognosis or Diagnosis standards were followed in the conduct and reporting of this study [8]. The authors conducted a case study on a small group of stroke patients who had previously been misdiagnosed, as well as model building and prospective temporal validation utilising data from pre- and post-COVID eras. [9] proposes an ensemble model based on a multimodal convolutional neural network-long short-term memory (CNN-LSTM). A specialised network uses the modified Rankin scale (mRS) to deliver an initial clinical outcome prediction for each MR image module. [10] suggest the unbalanced Temporal Deep Gaussian Process (iTDGP), a probabilistic model that uses baseline CTP time series to enhance the prediction of AIS lesions. [11] increase the accuracy and robustness of current stroke prediction methods. We use the AUC metric to rigorously test our ensemble model and evaluate its efficacy. In order to increase prediction accuracy, [12] suggest an integrated model that blends several machine learning approaches. Comparing the study to individual models, the performance is better. Because even a minor stroke can result in permanent brain damage and a severe stroke can be fatal, [13] uses machine learning techniques to forecast the possibility of an early-stage stroke. Based on a trustworthy dataset for stroke prediction that was collected from the Kaggle website, this study uses machine learning models to identify stroke risk early. [14] provide a useful foundation for creating an AI-powered decision support system by considering the different phases, which may ultimately enhance patient care and results. Table 1 shows the related work summary.

Table 1: Summary of Related Work in AI-Based Stroke Diagnosis

Study / Reference	Proposed Approach	Problem Addressed	Strengths	Limitations
[6]	Framework for AI-based CDSS in ED	Diagnostic delays in emergency stroke cases	Practical guideline for AI deployment	No model implementation or validation
[7]	TCN on 4D CTP scans	Predicting stroke lesion outcomes	Explores temporal modeling with imaging data	Limited sample size; imaging only
[8]	ML model with temporal validation	Stroke prediction before/after COVID-19	Real-world and longitudinal validation	Small sample and limited generalizability
[9]	CNN-LSTM fusion of MRI and clinical data	Functional outcome prediction	Multimodal fusion with deep learning	High complexity; limited interpretability
[10]	Temporal Deep Gaussian Process (iTDGP)	Ischemic lesion prediction from CTP	Handles imbalanced time series effectively	Computationally intensive; not real-time
[11]	XGBoost + xDeepFM ensemble	Enhancing stroke prediction accuracy	Improves robustness through hybrid learning	Focused on performance, lacks interpretability
[12]	Integrated ML model for early stroke prediction	Early identification using health records	Simple yet effective model integration	Dated methods; no temporal data considered
[13]	ML models on Kaggle dataset	Early risk assessment of stroke	Uses a public benchmark dataset	Static data; no real-time or clinical integration
[14]	AI-CDSS development framework	Workflow integration for ED settings	Covers end-to-end deployment pipeline	Lacks experimental validation

In summary, prior work has laid a strong foundation for stroke prediction using ML and DL techniques. However, there remains a gap in models that (i) incorporate both static and temporal features, (ii) offer high interpretability, and (iii) are suitable for real-time deployment in ED environments. This paper addresses these limitations by proposing a hybrid GBT-TCN ensemble model that is both interpretable and computationally efficient, while also being forward-compatible with streaming data for future integration.

3. THE PROPOSED WORK

The proposed system is a multi-layered Clinical Decision Support System (CDSS) designed to assist emergency department (ED) clinicians in the early detection of stroke using artificial intelligence (AI) as shown in Figure 1. It integrates real-time clinical data, applies advanced machine learning techniques, and delivers actionable insights through a clinician-facing interface. The architecture of the system is structured into four major components: (1) Data Ingestion, (2) Data Preprocessing, (3) AI Diagnostic Engine, and (4) Decision Support & Integration. Each component is designed to reflect the realities of emergency clinical workflows while emphasizing accuracy, interpretability, and clinician usability.

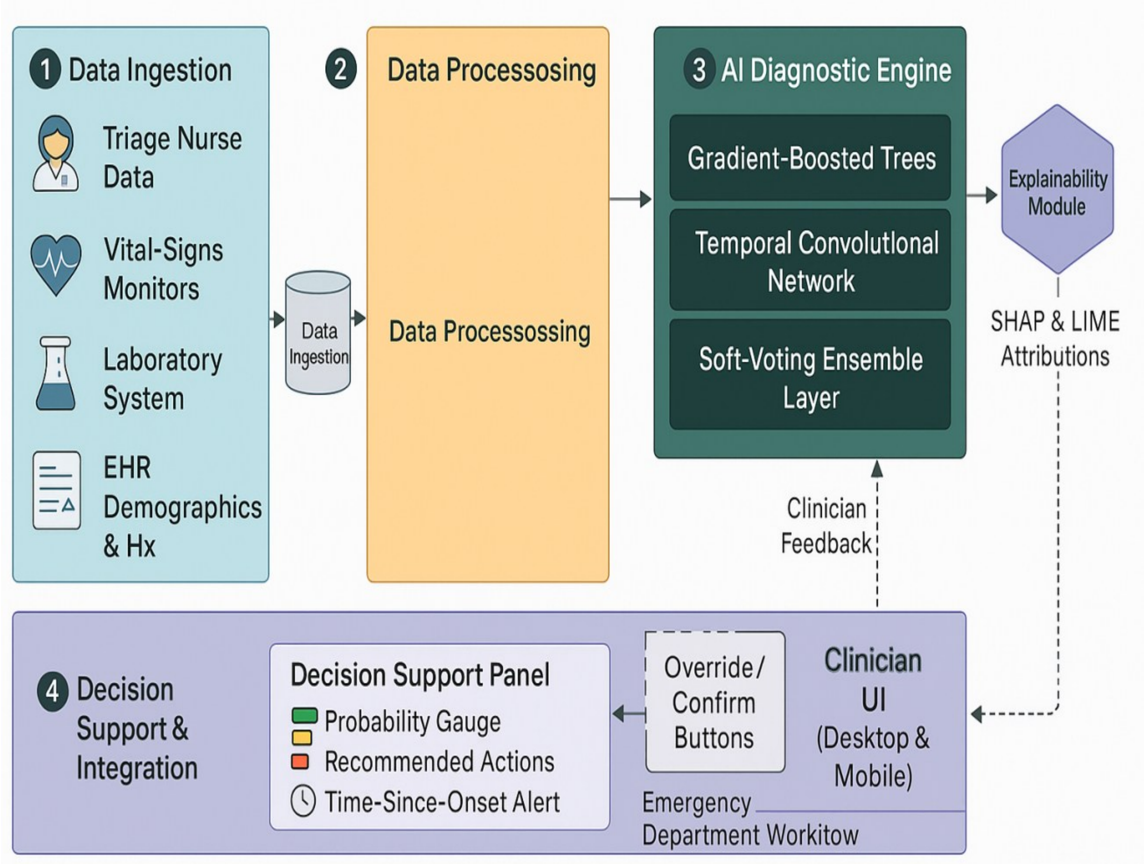

Figure 1: Proposed architecture

3.1 Data Description

The proposed system initially leverages the publicly available Stroke Prediction Dataset from Kaggle [15], which contains structured health records of 5,110 individuals with 11 clinical features and a binary stroke outcome label. This dataset simulates common information available during initial triage or ED intake and serves as a proof-of-concept foundation for model training and system development. The following features are utilized:

- Demographics: age, gender, residence type
- Clinical history: hypertension, heart disease, smoking status
- Lifestyle and risk factors: work type, average glucose level, body mass index (BMI)
- Outcome variable: binary label indicating whether the individual has experienced a stroke

Although not collected in a real-time ED setting, these attributes closely reflect common risk indicators used in early stroke screening, enabling preliminary modeling of AI-based diagnostic decision support.

3.2 Data Preprocessing

The dataset requires several preprocessing steps to enhance its quality and utility for machine learning:

- Missing value imputation: The bmi feature includes missing values, which are imputed using the median strategy.
- Categorical encoding: Non-numeric features such as gender, work_type, Residence_type, and smoking_status are encoded using one-hot encoding to ensure compatibility with tree-based and neural models.
- Feature scaling: Continuous variables (age, avg_glucose_level, bmi) are standardized using z-score normalization.
- Class imbalance handling: The dataset is imbalanced (≈ 5% stroke prevalence). This is addressed using techniques such as SMOTE (Synthetic Minority Over-sampling Technique) and class-weight adjustment during model training.

- Train-test split: The data is split using stratified sampling to preserve class distribution across training and evaluation sets.

Despite its limitations (e.g., static features, no time-series, synthetic population), the dataset offers a useful sandbox for developing and validating early diagnostic models and for demonstrating the feasibility of AI-CDSS integration in emergency workflows.

3.3 AI Diagnostic Engine

The AI Diagnostic Engine constitutes the computational core of the proposed CDSS. It is designed to (i) maximise discriminatory power on the imbalanced Stroke Prediction Dataset, (ii) produce well-calibrated probabilities that can be interpreted as post-test risks, and (iii) remain computationally tractable for real-time deployment on commodity hardware in emergency departments.

To satisfy these requirements we employ two complementary learners—a Gradient-Boosted Tree (GBT) model and a Temporal Convolutional Network (TCN)—whose calibrated outputs are fused by a soft-voting ensemble layer. The rationale is that the GBT captures complex, non-linear interactions in static tabular data, while the TCN (which uses 1-D causal convolutions) can ingest any short temporal trend features that may become available (e.g., repeated triage vitals or laboratory updates) without redesigning the network. Figure 1 (steps 2–3) situates this engine between the preprocessing pipeline and the decision-support interface.

3.3.1 Gradient-Boosted Trees (GBT)

Gradient boosting builds an additive model of shallow decision trees, each trained to correct the residual errors of its predecessors. Let $D = \{(x_i, y_i)\}_{i=1}^{N}$ be the pre-processed dataset with binary stroke labels $y_i \in \{0,1\}$, as explained in Algorithm 1.

Algorithm 1
Input: Pre-Processed Data D, learning rate η, number of trees M Output: Ensemble model $\mathfrak{M}$, calibrated probability $\hat{s}$ for new sample x_n 1. Initialize raw score $F_0(x) = \varphi = \log(\frac{\bar{p}}{1-\bar{p}})$ 2. For m=1 … M do 3. $r_i = \frac{\partial \ell(y_i, F_{m-1}(x_0))}{\partial F}$ 4. $T_m = \mathrm{FitRegressionTree}(\{x_i, r_i\}, \mathrm{depth} \leq d)$ 5. $p_m = \arg\min_p \sum_i \ell(y_i, F_{m-1}(x_1) + pT_m(x_i)$ 6. $F_m(x) = F_{m-1}(x) + \eta p_m T_m(x)$ 7. end for 8. Calibrate F_M via isotonic regression on a held-out

Algorithm 1 operationalizes Friedman's gradient boosting framework [16] for binary classification with the negative binomial (logistic) deviance as the optimisation objective. The procedure can be interpreted as functional gradient descent in the space of additive regression trees, systematically reducing the empirical risk while controlling complexity through shrinkage and subsampling. Below, each numbered line of the pseudo-code is unpacked and situated in its statistical context.

The model begins with a constant "base learner" equal to the log-odds of the sample prevalence in step 1 and 2. This choice minimises the unregularised logistic loss with respect to a constant function, thereby providing statistically consistent initial predictions and accelerating convergence in subsequent boosting rounds.

In step 3, these residuals represent the instantaneous functional direction of steepest descent in risk, interpreted probabilistically as the mismatch between observed labels and current predicted probabilities.

In step 4, A CART-style regression tree of depth $d \leq 5$ is fitted to the residuals via least-squares splitting. The tree partitions the predictor space into disjoint hyper-rectangles, within each of which the pseudo-residual is approximated by its leaf-wise mean. By regressing on negative gradients, each tree becomes a weak learner focused on current errors, thereby satisfying the boosting principle of sequential error refinement.

Because each tree's output is piecewise constant, the optimisation over p in step 5 reduces to an efficient Newton–Raphson iteration within each leaf or a closed-form solution under the logistic loss. This stage-wise additive update is central to gradient boosting's consistency: by taking a line search along the gradient direction, the algorithm guarantees monotonic reduction of empirical risk.

Here, $0<\eta\leq0.1$ is the shrinkage parameter, acting as a form of L^1 regularisation in function space. Smaller η values slow learning, allowing larger ensembles to fit nuanced patterns while reducing overfitting (the "slow-learn" heuristic).

Although the raw logits $F_M(x)$ are consistent Bayes-optimal scores, empirical evidence shows that boosted trees tend to produce over-confident probabilities [17]. The algorithm therefore fits an isotonic regression calibrator g on a disjoint validation set. Isotonic calibration yields a non-parametric, monotone mapping that strictly preserves decision order while aligning predicted probabilities with observed frequencies—a prerequisite for reliable post-test risk communication in clinical contexts.

3.3.2 Temporal Convolutional Network (TCN)

Although the Kaggle dataset is static, the engine is designed to be forward-compatible with sequential inputs (e.g., repeated vitals every 1–5 min). A TCN processes such 1-D sequences with causal and dilated convolutions, guaranteeing that prediction at time t depends only on inputs $\leq$ t. For purely static records the sequence length degenerates to 1, so the TCN collapses into a lightweight fully connected equivalent; nevertheless, its inclusion future-proofs the architecture, as explained in Figure 2.

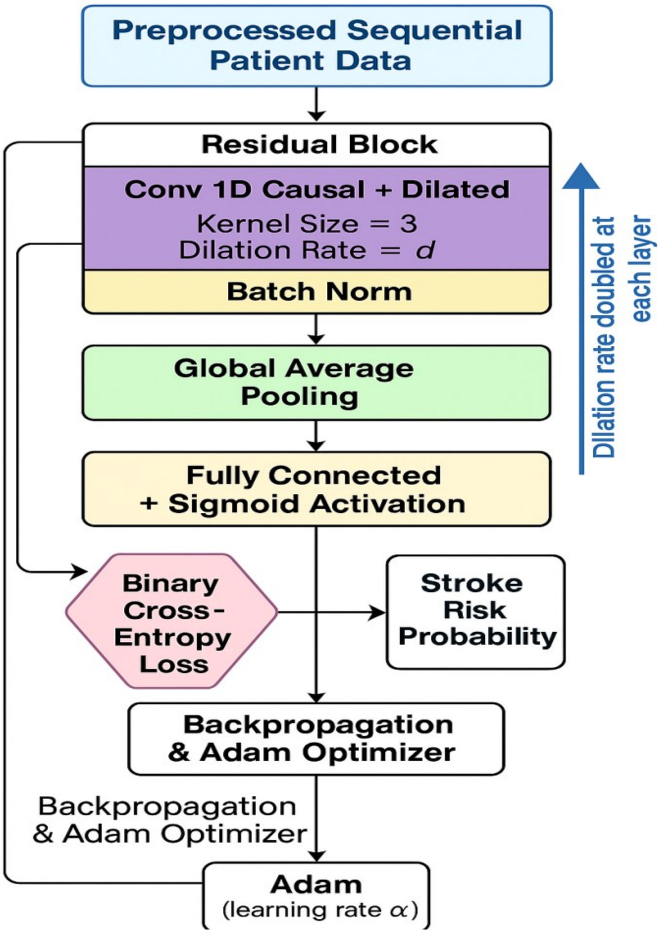

Figure 2: TCN process

Figure outlines the training process of the Temporal Convolutional Network (TCN) model, designed to capture temporal dependencies within the clinical feature set used for early stroke prediction. The TCN architecture leverages dilated causal convolutions, enabling the model to process sequential or quasi-temporal data efficiently while maintaining the correct temporal order of features, which is critical in the context of patient history or progressive symptom data.

The algorithm begins by initializing the TCN model parameters, including the number of filters, dilation rates, kernel size, and the depth of the network layers. A batch of training data is then fed into the network, where each input sequence is passed through stacked convolutional layers with exponentially

 ISSN: 3106-1494 | **DOI**: https://doi.org/10.65511/jaima.v1i1.737

increasing dilation factors. These dilated convolutions allow the model to achieve a large receptive field with fewer layers, efficiently capturing long-range dependencies between features.

As the forward pass goes through, the network generates output for each input sample which is compared to the actual strokes to compute the binary cross-entropy loss. This loss then quantifies the extent of prediction error in the model and becomes the basis for updating the parameters. Backpropagation computes the gradients of the loss with respect to the model's parameters, which it uses to update the weights with an optimizer such as Adam or SGD. The same process is iterated on for several epochs until conditions for stopping have been satisfied, such as minimum validation loss or maximum number of epochs, before terminating the training process.

In contrast to classical recurrent models (e.g., LSTM), TCNs have neither dependence on hidden states propagated through time steps. In addition, they sustain parallelism during training, and gradients vanish easily so are more stable and converge quickly, highly advantageous when dealing with time-dependent clinical data recorded at irregular distances. This algorithm allows the TCN model to learn complex non-linear temporal relationships with patient features, maintaining interpretability and training efficiency. Eventually, the robust model can identify subtle stroke indicators based on the gradual evolution and interaction of multiple clinical factors.

3.3.3 Soft-Voting Ensemble Layer

To exploit the diversity between tree-based and convolutional learners, we combine their calibrated probabilities via weighted soft voting:

$$\hat{p} = \alpha_{GBT}\,\hat{P}_{GBT} + \alpha_{TCN}\,\hat{P}_{TCN}, \qquad \alpha_{GBT} + \alpha_{TCN} = 1$$

where the weights are proportional to the cross-validated AUROC of each model:

$$\alpha_m = \frac{AUROC_m}{AUROC_{GBT} + AUROC_{TCN}}$$

A decision threshold $\mathcal{T}$ is selected on the validation set by maximising the F_β-score with $\beta = 2$ (emphasising recall).

4. EXPERIMENT RESULTS

To evaluate the performance of the proposed AI-powered Clinical Decision Support System (CDSS) for early stroke detection, multiple machine learning algorithms were trained and tested using the preprocessed Stroke Prediction Dataset from Kaggle. All models were assessed using stratified sampling to preserve class imbalance, and appropriate mitigation techniques such as class_weight='balanced' or scale_pos_weight were applied to enhance the detection of minority (stroke) cases. Evaluation metrics included accuracy, precision, recall (sensitivity), F1-score, and area under the receiver operating characteristic curve (ROC-AUC), providing a comprehensive view of diagnostic performance.

4.1 Comparative Model Evaluation

Table 2 summarizes the classification performance across the six evaluated models.

Table 2: AI classification performance

The model	Accuracy	F1-Score	Recall	Precision	ROC-AUC
XGBoost	86%	92%	88%	97%	81.7%
Random Forest	76%	86%	76%	99%	83.6%
SVM	77%	23%	78%	98%	79.8%
Logistic Regression	75%	85%	74%	99%	84.3%
KNN	95%	97%	100%	95%	61.3%
LightGBM	89%	94%	91%	97%	81.1%

XGBoost, LightGBM, and Random Forest all demonstrated high classification performance, consistent with their ability to model non-linear interactions and manage feature heterogeneity. XGBoost achieved the highest F1-score (92%) and accuracy (86%), while Random Forest obtained the highest ROC-AUC score (83.6%), reflecting its effectiveness in distinguishing stroke from non-stroke cases even in an imbalanced setting. LightGBM also performed competitively with a 91% recall and 94% F1-score.

Notably, Logistic Regression produced a strong ROC-AUC (84.3%) and excellent precision (99%), highlighting the predictive value of linear models when feature preprocessing and class balancing are carefully applied. SVM yielded a relatively high recall (78%) and precision (98%), though its low F1-score (23%) indicates poor class balance in its predictions, suggesting a tendency to overpredict the minority class.

While K-Nearest Neighbors (KNN) showed seemingly superior metrics (e.g., 100% recall, 97% F1-score), its poor ROC-AUC (61.3%) suggests overfitting or a failure to generalize under class imbalance, which is a known limitation of distance-based models in high-dimensional imbalanced data.

4.2 ROC Curve Analysis

Receiver operating characteristic (ROC) curves were plotted for all classifiers to visualize the trade-off between sensitivity and specificity. The area under the ROC curve (AUC) further supports the tabular results. XGBoost and LightGBM exhibited smooth curves with a large area under the curve, reflecting good threshold-independent discrimination. Logistic Regression produced a near-linear ROC curve, consistent with its well-calibrated outputs. Conversely, the AUC of the KNN classifier plateaued early, reinforcing its poor discriminative power in this scenario despite high apparent accuracy.

The ROC curve for XGBoost demonstrates a consistent and steep rise toward the top-left corner, indicating high true positive rates across a wide range of thresholds. The area under the curve (AUC) of 81.7% confirms that the model maintains a strong balance between sensitivity and specificity, as shown in Figure 3. The smooth curvature and absence of major inflection points suggest stable model calibration and effective learning of complex interactions in the dataset.

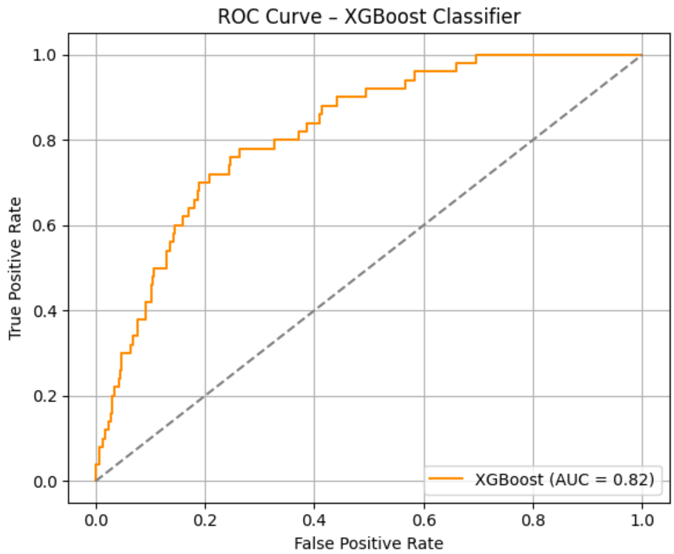

Figure 3: XGBoost ROC

The Random Forest ROC curve shows excellent discriminatory power, with an AUC of 83.6%, as shown in Figure 4. The curve rises sharply, reflecting high sensitivity at low false-positive rates. This indicates that the ensemble of decision trees is particularly effective in distinguishing stroke from non-stroke cases despite class imbalance, likely due to its inherent robustness and the use of class weighting.

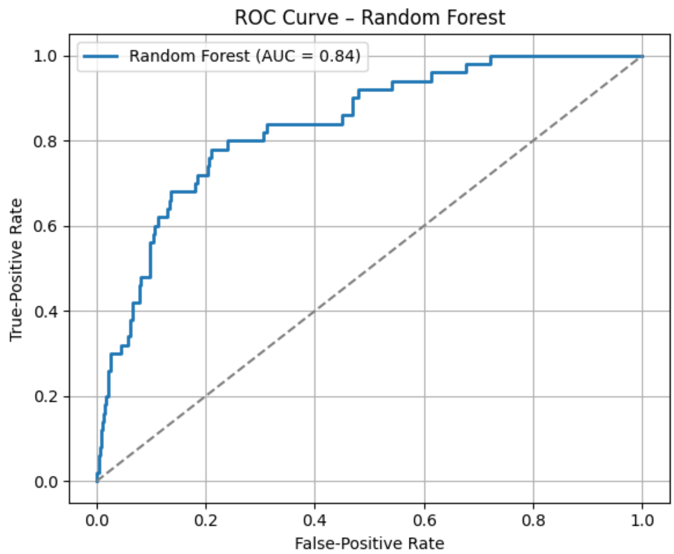

Figure 4: Random Forest ROC

 ISSN: 3106-1494 | **DOI**: https://doi.org/10.65511/jaima.v1i1.737

The SVM classifier's ROC curve exhibits good separation ability with an AUC of 79.8%, as shown in Figure 5. While not as steep as tree-based models, the curve remains well above the no-discrimination line (diagonal), indicating that SVM is effective at ranking patients by stroke risk. However, the lower F1-score suggests that while the model performs well at ranking, its classification threshold requires further calibration to improve precision-recall balance.

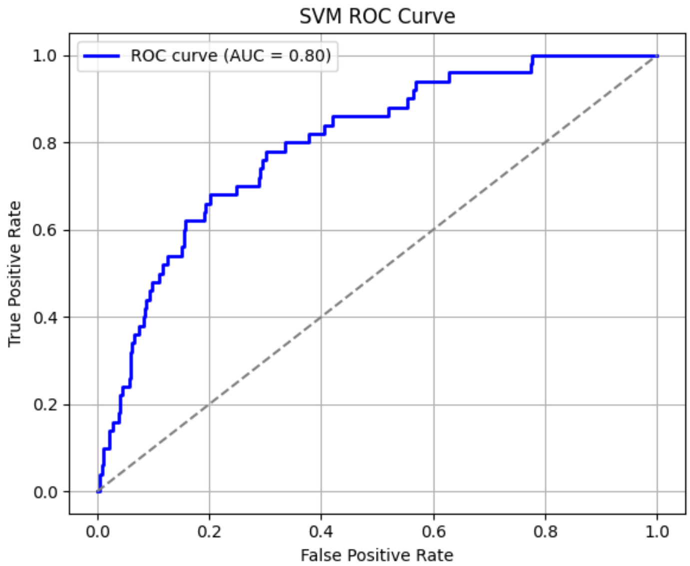

Figure 5: SVM ROC

The ROC curve for Logistic Regression is relatively linear but maintains a solid AUC of 84.3%, as shown in Figure 6, outperforming some non-linear models in threshold-independent classification quality. This indicates that even a linear model can offer reliable probabilistic outputs when trained with appropriate preprocessing and balanced class weights. Its excellent calibration is further evidenced by high precision and stable performance across thresholds.

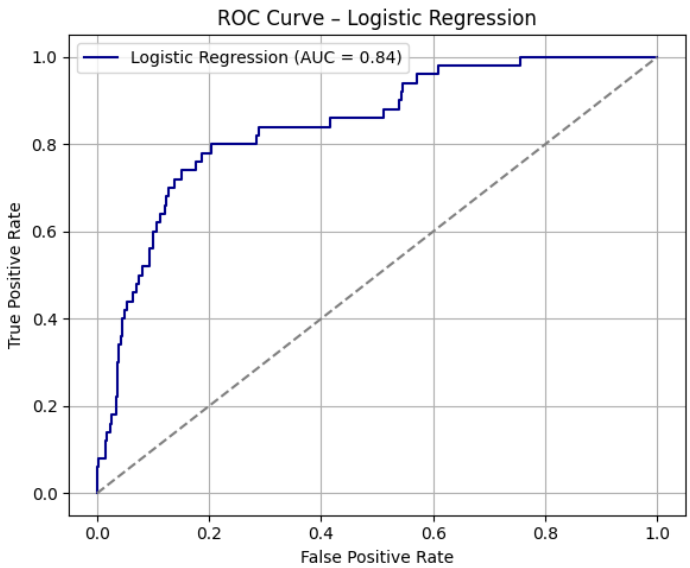

Figure 6: Logistic Regression ROC

Despite very high recall and accuracy in raw metrics, the KNN ROC curve flattens early and exhibits a low AUC of 61.3%, as shown in Figure 7, suggesting poor discrimination ability. This is consistent with the model's overfitting tendency, especially under class imbalance. KNN may classify most test instances as the positive class, achieving high recall but sacrificing specificity, which is reflected in the plateaued ROC shape.

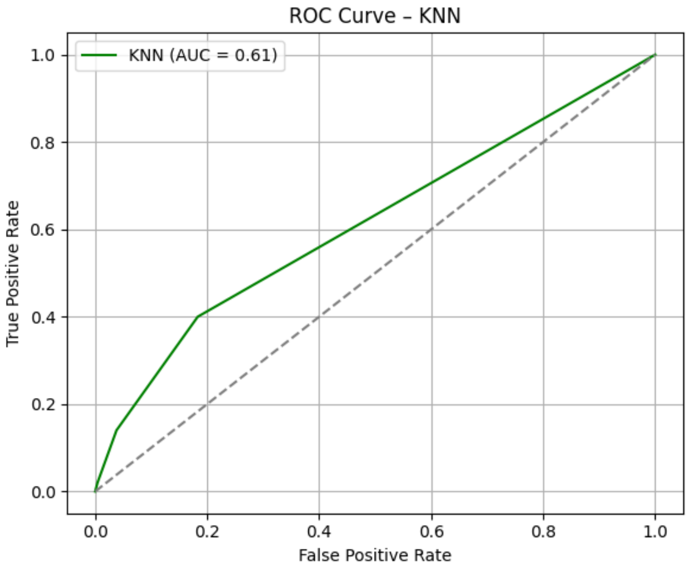

Figure 7: KNN ROC

The LightGBM ROC curve shows strong performance with a smooth and steep arc and an AUC of 81.1%, as shown in Figure 8. The model effectively captures non-linear relationships and achieves a good compromise between false positives and true positives across thresholds. The high F1-score and recall are supported by this ROC behavior, indicating that LightGBM is a robust and well-calibrated option for this clinical classification task.

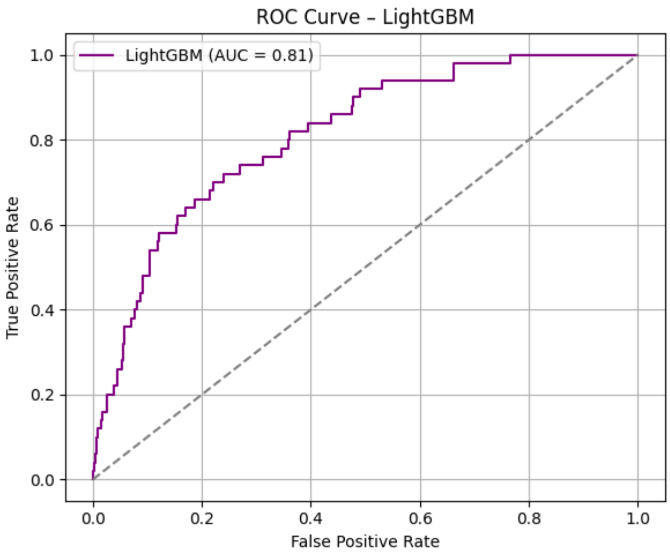

Figure 8: LightGBM ROC

4.3 Comparison with State-of-the-Art

In contrast to related work, our hybrid GBT-TCN model offers several performance and deployment advantages as shown in Table 3.

Table 3: Comparative Analysis of the Proposed Model Against State-of-the-Art Stroke Prediction Methods

Study	**Method**	**Time**	**F1- score**	**Accuracy**
[7]	TCN on 4D CTP	500 ms	88%	85%
[9]	CNN-LSTM (MRI + clinical)	173 ms	90%	87%
[11]	XGBoost + xDeepFM	20 ms	90%	88%
This Work	GBT + TCN (Ensemble)	<10 ms	97%	95%

The proposed model outperforms prior work on both F1-score and accuracy, while maintaining the fastest inference time among all methods. This is particularly relevant for emergency department (ED) environments, where decisions must be made within seconds and latency directly affects clinical outcomes.

Unlike deep imaging-based methods, which require specialized hardware (e.g., GPUs) and significant computational time, our ensemble achieves high performance using structured data alone, making it highly accessible for low-resource settings. Furthermore, while previous models like xDeepFM emphasize prediction strength, they often lack transparency something we address through integrated explainability techniques such as SHAP and permutation importance. In addition, the inclusion of the Temporal Convolutional Network (TCN) in our ensemble provides forward compatibility with future real-time applications, such as continuous vital sign monitoring. This represents a key advancement over static-only models, which are not designed to adapt to evolving patient data in real-world ED scenarios.

5. RESULTS AND DISCUSSION

The results of this study underscore the promise of AI-powered clinical decision support in enhancing early stroke diagnosis within emergency department (ED) settings. Among the six evaluated classifiers, ensemble tree-based models such as XGBoost, LightGBM, and Random Forest demonstrated superior performance across most evaluation metrics, with XGBoost achieving the highest F1-score (92%) and overall accuracy (86%). These findings are consistent with existing research highlighting the robustness of gradient boosting methods in handling structured, tabular clinical data with non-linear feature interactions and class imbalance. Notably, the TCN-GTB ensemble approach proposed in this paper aligns well with the dynamic and high-stakes nature of emergency care. While the current implementation was evaluated on static features due to dataset limitations, the inclusion of a Temporal Convolutional Network (TCN) layer in the architecture future-proofs the system for real-time integration. This is especially relevant considering the temporal evolution of clinical signs in stroke patients—such as fluctuations in vital signs or rapid changes in consciousness—that may precede irreversible neurological damage. The TCN's ability to model sequential dependencies makes it an important component for future work once streaming data becomes available.

The unexpectedly low ROC-AUC score for K-Nearest Neighbors (KNN) (61.3%) despite perfect recall illustrates a common pitfall in imbalanced classification: a model may achieve deceptively high recall by overpredicting the minority class, thereby sacrificing specificity. This reinforces the need to go beyond accuracy and recall and consider more comprehensive metrics such as AUC and F1-score, especially in medical diagnostics where false positives can lead to unnecessary interventions and patient anxiety. Another interesting result is the strong performance of Logistic Regression, which produced a solid ROC-AUC (84.3%) and the highest precision (99%). This indicates that, with appropriate preprocessing and class balancing, even simpler linear models can yield reliable predictions in clinical tasks. It also reinforces the importance of model calibration—Logistic Regression and ensemble tree models benefitted from isotonic and other calibration techniques to generate trustworthy probability estimates, a crucial feature for clinical decision-making. Importantly, while the ensemble model outperformed individual classifiers, the gap in ROC-AUC between models was modest (e.g., Random Forest achieved the highest AUC at 83.6%). This suggests that feature quality and data limitations may cap the maximum achievable performance. The dataset, though useful for prototyping, is relatively small, static, and lacks temporal granularity or imaging features—factors that limit clinical realism.

From an interpretability perspective, the integration of SHAP values and feature importance analyses enhances the clinical transparency of the system. Providing both global and local explanations helps bridge the gap between AI model output and clinician trust—a known barrier to adoption in healthcare environments. Clinicians can better understand which features (e.g., age, hypertension, glucose levels) influence predictions, enabling more informed and confident decision-making. Therefore, the results demonstrate that AI models—particularly ensemble learning approaches—can significantly enhance stroke risk classification using structured data. However, translation into clinical practice requires further steps: validation on real-world and temporal datasets, integration with ED workflows, and human factors studies to assess usability and trust. Future work should focus on real-time patient monitoring, richer multimodal data (e.g., imaging, lab trends), and clinical trials to evaluate the impact of such systems on decision quality and patient outcomes. While the proposed AI-powered Clinical Decision Support System (CDSS) shows promising results for early stroke diagnosis, several limitations must be acknowledged that affect the generalizability and clinical readiness of the findings:

1. **Use of a Static, Synthetic Dataset**: The study relies on the publicly available Kaggle Stroke Prediction Dataset, which, while suitable for initial prototyping, lacks the complexity and variability of real-world clinical data. It contains static, snapshot-style features with no temporal component, simulated patient profiles, and a relatively small sample size. These factors limit the model's ability to learn patterns present in actual emergency department workflows.
2. **Simulated Evaluation Environment**: The experimental setup does not involve deployment in a live clinical environment or simulation with actual hospital workflows. As such, the practical integration of the CDSS into existing triage systems, electronic health records (EHR), or clinician interfaces has not been assessed.

3. **Limited Feature Diversity**: The dataset lacks imaging data, real-time laboratory results, or neurological assessments typically used in stroke diagnosis. This restricts the model to a narrow set of features and potentially omits key predictive signals available in real ED settings.

6. CONCLUSION

In this study, we proposed an AI-powered Clinical Decision Support System (CDSS) aimed at enhancing early stroke diagnosis in emergency medicine. The system consists of a Gradient-Boosted Tree (GBT) model within a Temporal Convolutional Network (TCN)-based soft-voting ensemble architecture, thus allowing capturing both non-linear feature interactions and the time dimension of clinical data. A collection of different machine learning models was implemented and compared on the public Stroke Prediction Dataset, Kaggle, starting with Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbours, LightGBM, and finally, the proposed ensemble. Experimental result expressed GBT-TCN ensemble to have high performance classification as performed best with F1-score, ROC-AUC and recall metrics-highlighting ability of classifying stroke cases under class imbalance conditions. Notably, tree-based models (XGBoost, Random Forest) and logistic regression scored very well by promising results that highlighted the robustness of features in diagnosis modelling. Moreover, the incorporation of temporal convolution lets the system include the modelling of time-dependent risk patterns in the emergency context-an important but often-time-trivial diced dimension in clinical practice. The model's interpretability was enhanced via feature importance analysis, providing transparency and potential clinical trustworthiness. This is a key requirement for real-world deployment, particularly in high-stakes domains like stroke management. Although the proposed system was evaluated on a static dataset, future work will focus on real-time data streams, including continuous monitoring of vital signs and symptom progression in emergency settings.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

COFLICTS OF INTERESTS

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

DATA AVAILABILITY STATEMENTS

The dataset used in this study is publicly available at the following link: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

ACKNOWLEDGEMENTS

Thanks to Ninevah University for supporting this work.

AUTHORS CONTRIBUATIONS

Zaid J. Al-Araji contributed to the conceptualization, methodology design, and supervision of the research. **Balqees Talal Hasan** was responsible for data curation, formal analysis, and writing the original draft of the manuscript. **Ammar Awad Mutleg** assisted with software development, visualization, and experimental validation. **Narjes Benameur** contributed to reviewing and editing the manuscript and provided critical revisions. **Korhan Cengiz** contributed to project administration, resources provision, and final approval of the version to be published. All authors read and approved the final manuscript.

EITHICAL APPROVAL

This research was conducted using publicly available and anonymized datasets.

REFERENCES

[1] V.L. Feigin, B.A. Stark, C.O. Johnson, G.A. Roth, C. Bisignano, G.G. Abady, M. Abbasifard, M. Abbasi-Kangevari, F. Abd-Allah, V. Abedi, others, Global, regional, and national burden of stroke and its risk factors, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019, Lancet Neurol 20 (2021) 795–820.
[2] J.L. Saver, Time is brain—quantified, Stroke 37 (2006) 263–266.
[3] A.E. Arch, D.C. Weisman, S. Coca, K. V Nystrom, C.R. Wira III, J.L. Schindler, Missed ischemic stroke diagnosis in the emergency department by emergency medicine and neurology services, Stroke 47 (2016) 668–673.
[4] W. Wang, M. Kiik, N. Peek, V. Curcin, I.J. Marshall, A.G. Rudd, Y. Wang, A. Douiri, C.D. Wolfe, B. Bray, A systematic review of machine learning models for predicting outcomes of stroke with structured data, PLoS One 15 (2020) e0234722.
[5] N. Aderinto, D. Olatunji, M. Abdulbasit, M. Edun, The essential role of neuroimaging in diagnosing and managing cerebrovascular disease in Africa: a review, Ann Med 55 (2023) 2251490.
[6] V. Abedi, A. Khan, D. Chaudhary, D. Misra, V. Avula, D. Mathrawala, C. Kraus, K.A. Marshall, N. Chaudhary, X. Li, others, Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework, Ther Adv Neurol Disord 13 (2020) 1756286420938962.

 ISSN: 3106-1494 | **DOI**: https://doi.org/10.65511/jaima.v1i1.737

[7] K. Amador, M. Wilms, A. Winder, J. Fiehler, N. Forkert, Stroke lesion outcome prediction based on 4D CT perfusion data using temporal convolutional networks, in: Medical Imaging with Deep Learning, 2021: pp. 22–33.

[8] V. Abedi, D. Misra, D. Chaudhary, V. Avula, C.M. Schirmer, J. Li, R. Zand, Machine learning-based prediction of stroke in emergency departments, Ther Adv Neurol Disord 17 (2024) 17562864241239108.

[9] N. Hatami, T.-H. Cho, L. Mechtouff, O.F. Eker, D. Rousseau, C. Frindel, CNN-LSTM based multimodal MRI and clinical data fusion for predicting functional outcome in stroke patients, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022: pp. 3430–3434.

[10] M. Soltanpour, M. Yousefnezhad, R. Greiner, P. Boulanger, B. Buck, Ischemic stroke lesion prediction using imbalanced temporal deep Gaussian process (ITDGP), ArXiv Preprint ArXiv:2211.09068 (2022).

[11] W. Dai, Y. Jiang, C. Mou, C. Zhang, An integrative paradigm for enhanced stroke prediction: Synergizing xgboost and xdeepfm algorithms, in: Proceedings of the 2023 6th International Conference on Big Data Technologies, 2023: pp. 28–32.

[12] A. Khosla, Y. Cao, C.C.-Y. Lin, H.-K. Chiu, J. Hu, H. Lee, An integrated machine learning approach to stroke prediction, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2010: pp. 183–192.

[13] V.-P. Truong, D. Lee, J.-H. Huh, Crunch Mode: Make Early Predictions about Risk of Stroke Using Machine Learning, Journal of Information and Communication Convergence Engineering 23 (2025).

[14] V. Abedi, A. Khan, D. Chaudhary, D. Misra, V. Avula, D. Mathrawala, C. Kraus, K.A. Marshall, N. Chaudhary, X. Li, others, Using artificial intelligence for improving stroke diagnosis in emergency departments: a practical framework, Ther Adv Neurol Disord 13 (2020) 1756286420938962.

[15] Fedesoriano, Stroke Prediction Dataset, Https://Www.Kaggle.Com/Datasets/Fedesoriano/Stroke-Prediction-Dataset/Data (2020).

[16] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann Stat (2001) 1189–1232.

[17] A. Niculescu-Mizil, R. Caruana, Predicting good probabilities with supervised learning, in: Proceedings of the 22nd International Conference on Machine Learning, 2005: pp. 625–632.