



Interpretable Incremental Tree-Based Learning for Brain Stroke Prediction with Rule Evolution

Ahmed A. AL-Shalabi , Ahmed S. Al-Hegami and Majid A. Aslan *

Department of Computer Science, Faculty of Computer and Information Technology, Sana'a University, Sana'a, Yemen

*Corresponding author: maj.aslan@su.edu.ye

ABSTRACT

Accurate brain stroke prediction requires models that can detect high-risk cases while adapting to newly arriving clinical data. This study proposes an interpretable incremental tree-based learning approach for brain stroke prediction using Decision Tree, Random Forest, and Gradient Boosting. Experiments were conducted using the Cerebral Stroke Dataset, a public imbalanced stroke prediction dataset with 43,400 records and 12 attributes. Since the dataset does not provide verified patient-level timestamps, the data were divided into three sequential simulated cohorts to evaluate model behavior under progressive data availability. Decision Tree was used as a refitted interpretable baseline, while Random Forest and Gradient Boosting were updated using a warm-start strategy by appending additional estimators at later stages. Rule extraction and rule classification were used to translate tree-based decision logic into readable rules and categorize them as conforming, generalized, specialized, unexpected, or novel, enabling analysis of rule evolution across cohorts. Experimental results showed that the incremental approach improved sensitivity-oriented performance and reduced update time compared with traditional retraining. At the final stage, incremental Gradient Boosting achieved 98.044% accuracy, 75.32% F1-score, 91.03% recall, and 94.66% ROC AUC, with only 40 false negatives. Therefore, incremental Gradient Boosting provided the most clinically suitable trade-off between stroke-case detection, update efficiency, and interpretable rule behavior.

ARTICLE INFO

Keywords:

Brain Stroke Prediction, Incremental Learning, Tree-Based Models, Decision Tree, Random Forest, Gradient Boosting, Streaming Data, Rule Evolution.

Article History:

Received: 30-May-2026,

Revised: 26-July-2026,

Accepted: 28-September-2026,

Published: 28 October 2026.

1. INTRODUCTION

Brain stroke is a major neurological disorder that occurs when the blood supply to the brain is interrupted or reduced, leading to acute neurological dysfunction and, in many cases, long-term disability or death. Recent global estimates indicate that stroke remains one of the leading causes of death and disability worldwide, which makes early risk recognition an important public health and clinical priority [1]. Clinically, stroke is commonly classified into ischemic and hemorrhagic types, and stroke risk is influenced by several modifiable and non-modifiable factors, including hypertension, diabetes, smoking, cardiovascular conditions, age, and lifestyle-related factors [2]. Because clinical outcomes are strongly influenced by timely intervention, early prediction of high-risk cases remains an important objective in preventive healthcare and clinical decision support. Machine learning has

been increasingly used for stroke prediction because it can learn predictive patterns from demographic, clinical, and lifestyle-related variables. Previous studies have applied several classification algorithms for stroke prediction, including decision trees, random forests, logistic regression, support vector machines, naïve Bayes, and ensemble-based methods, and have shown that predictive performance is strongly affected by preprocessing, feature selection, and class-imbalance handling [3]. However, high accuracy does not necessarily indicate clinical reliability in stroke prediction. Stroke datasets are often imbalanced, and a classifier may obtain high accuracy by correctly predicting the majority non-stroke class while missing a considerable number of actual stroke cases [4]. This issue makes false-negative reduction a central concern. In a clinical screening context, failing to identify a patient who is actually at risk may be more harmful

than producing a false alarm. For this reason, evaluation should give particular attention to recall, F1-score, ROC AUC, and sensitivity under high specificity rather than relying only on overall accuracy. A useful stroke prediction model should therefore provide not only acceptable global performance but also strong sensitivity to minority positive cases. Most existing stroke prediction studies follow a traditional batch-learning strategy, where the model is trained on a fixed dataset and retrained from scratch when new data become available. This strategy is simple and widely used; however, it is not fully aligned with real healthcare environments, where patient data may arrive continuously, and the underlying data distribution may change over time. In machine learning settings where data distributions change, concept drift is a known challenge, and adaptive learning methods are commonly used to maintain model relevance over time [5]. Incremental supervised learning addresses this issue by updating models progressively as new data arrive, instead of rebuilding the model entirely from the beginning [6]. Incremental learning offers a practical direction for stroke prediction because it allows the model to incorporate newly observed patient cases while retaining useful knowledge from previous training stages. This can reduce the update time and may improve sensitivity to emerging patterns without relying on complete retraining. In this study, incremental learning was used to simulate a setting in which clinical data become available across sequential simulated cohorts. Tree-based models are particularly suitable for this study because they provide a balance between predictive performance and interpretability. Decision Tree models produce explicit decision paths, Random Forest improves robustness through ensemble learning, and Gradient Boosting enhances prediction by sequentially correcting earlier errors. In addition, tree-based models allow predictive rules to be extracted and examined, which is valuable in applications where model behavior should be explainable and auditable [7]. Rule analysis also makes it possible to observe whether newly generated rules conform to known patterns or represent novel patterns learned from later data cohorts. This study proposes an incremental learning approach for brain stroke prediction using three tree-based machine learning models: decision tree, random forest, and gradient boosting. The proposed approach uses sequential simulated cohorts to simulate the arrival of new clinical data over time. It compares incremental learning with traditional retraining and evaluates the models using predictive performance, computational efficiency, and rule-based interpretability. The study particularly emphasizes sensitivity to stroke cases, model update time, and the evolution of extracted rules across sequential training stages. The main contributions of this study are summarized as follows. First, the study implements a sequential cohort-based evaluation protocol for stroke prediction using tree-based learning models. Sec-

ond, it examines the behavior of decision trees, random forests, and gradient boosting under progressive data availability, with emphasis on sensitivity-oriented evaluation rather than accuracy alone. Third, it provides a comparative assessment against traditional retraining in terms of predictive performance and update time. Fourth, it integrates rule extraction and rule classification to examine whether model updates preserve stable decision logic or introduce new predictive patterns. The remainder of this paper is organized as follows. Section 2 presents the motivation of the study. Section 3 defines the research problem. Section 4 reviews related work. Section 5 describes the proposed incremental learning approach. Section 6 presents the experimental results and discussion. Section 7 concludes the paper and outlines future research directions.

2. MOTIVATION

The motivation for this study is mainly practical and clinical. In healthcare environments, patient data may become available progressively, while repeatedly retraining prediction models from scratch can increase computational cost and delay model updating. Therefore, an adaptive learning strategy is needed to incorporate newly available cases more efficiently [5], [6]. In stroke prediction, the clinical cost of false-negative predictions is particularly high because missed high-risk cases may delay preventive intervention. For this reason, this study emphasizes recall, F1-score, ROC AUC, and false-negative reduction rather than relying only on overall accuracy. Tree-based models were selected because they combine predictive capability with interpretable decision structures. Their extracted rules allow the analysis of how prediction logic changes across sequential cohorts, which is useful for auditing model behavior in clinical decision-support settings.

3. PROBLEM STATEMENT

The main problem addressed in this study is the limited adaptability and interpretability of conventional stroke prediction models when new patient data become available. Most existing approaches are trained as static batch models and require full retraining when additional data are introduced. This limits their practical suitability for settings where data are processed progressively. A second problem is that stroke prediction is affected by class imbalance, where high accuracy may hide poor detection of actual stroke cases. Therefore, this study investigates how tree-based models can be evaluated under a sequential cohort-based setting, with particular attention to sensitivity-oriented performance, update time, and rule evolution. Therefore, this study asks how an interpretable incremental learning framework can improve stroke-case detection, reduce update cost com-

pared with traditional retraining, and support transparent analysis of changing decision rules.

4. RELATED WORKS

Previous studies related to this work can be grouped into three main directions: traditional machine learning for stroke prediction, incremental and continual learning approaches, and rule extraction for interpretable tree-based models. This organization reflects the main components of the proposed study: stroke-risk classification, model adaptation over sequential data cohorts, and rule-based interpretation of model behavior.

4.1. STROKE PREDICTION USING TRADITIONAL MACHINE LEARNING

Machine learning has been widely investigated for brain stroke prediction using static clinical and demographic datasets. Early studies applied conventional supervised learning algorithms, such as decision trees, random forests, logistic regression, support vector machines, K-nearest neighbors, Naïve Bayes, and ensemble-based classifiers. These studies generally showed that machine learning models can identify stroke-risk patterns from patient-related features; however, the reported performance varied according to preprocessing, feature selection, sampling strategy, and the evaluation metrics used [3], [8], [9]. Several comparative studies have reported that random forest and boosting-based models often achieve competitive performance in stroke prediction. For example, some studies have compared multiple classifiers and found that ensemble-based methods can improve classification stability compared with single learners, especially when the dataset contains heterogeneous clinical and lifestyle-related variables [10], [11], [12]. Other studies have investigated stroke-risk classification using national screening data or publicly available datasets and shown that model performance depends strongly on feature representation, dataset structure, and class distribution characteristics [13]. Despite these contributions, most existing stroke prediction studies are designed as static batch-learning pipelines. In such pipelines, the model is trained once on a fixed dataset and is later evaluated on test data drawn from the same or similar distribution. This design is useful for initial model development, but it does not fully reflect healthcare environments in which patient records may arrive continuously and risk patterns may evolve over time. Moreover, many studies still emphasize overall accuracy, even though stroke datasets are commonly imbalanced. In this setting, a classifier may achieve high accuracy by correctly predicting the majority non-stroke class while failing to detect a clinically important proportion of stroke cases [3], [4]. This limitation is particularly important in clinical screening, where false-negative predictions

are more critical than false positives. A missed high-risk case may delay preventive intervention, while a false alarm can usually be followed by additional clinical assessment. Therefore, sensitivity-oriented metrics, such as recall, F1-score, ROC AUC, and sensitivity under high specificity, are more informative than accuracy alone. However, relatively few stroke prediction studies explicitly combine sensitivity-focused evaluation with adaptive model updating.

Recent LSTM-based and Transformer-based models have been used in sequential clinical prediction, particularly when longitudinal records, time-stamped events, imaging sequences, or multimodal temporal data are available. However, the dataset used in this study is structured tabular data and does not provide genuine longitudinal patient sequences. Therefore, tree-based models were selected because they are more suitable for tabular clinical features, require less data than deep sequence models, and provide interpretable decision rules for rule-evolution analysis.

4.2. INCREMENTAL AND CONTINUAL LEARNING APPROACHES

Incremental learning has been proposed as an alternative to traditional batch learning in applications where data arrive sequentially or where retraining from scratch is computationally inefficient. Instead of rebuilding the model whenever new data become available, incremental learning updates the model using newly observed data while attempting to retain previously acquired knowledge. This makes it suitable for data-stream and time-evolving environments, where adaptability and computational efficiency are required [5, 6], [14]. A major challenge in evolving data environments is concept drift, which occurs when the statistical properties of incoming data differ from those observed during earlier training stages [15][15]. Concept drift can reduce model reliability if the model remains fixed after initial training. Adaptive learning methods and data-stream approaches have therefore been studied to maintain prediction performance under changing distributions [5], [14], [16]. These studies provide a general foundation for updating predictive models over time, but they are not specifically designed for brain stroke prediction. Continual learning research further highlights the challenge of retaining previous knowledge while incorporating new information. When a model is updated sequentially, it may suffer from catastrophic forgetting, where newly learned patterns overwrite useful knowledge acquired from earlier data. Several approaches have attempted to reduce this problem by preserving old knowledge, regularizing model updates, or adapting the learning process to new data without fully discarding prior information [17–19]. Although incremental and continual learning have been applied in different domains, their systematic use in stroke prediction re-

Table 1. Summary of Selected Stroke Prediction Studies

Study	Dataset / Do-main	Model(s) Used	Best Reported Result	Main Limitation
Gurjar et al.	Stroke warning symptoms dataset	Logistic Regression, Decision Tree, KNN, Random Forest, Naïve Bayes	Random Forest achieved about 95.4% accuracy	Static batch evaluation; no incremental updating or rule-evolution analysis
Sailasya and Kumari	Stroke prediction dataset	Logistic Regression, Decision Tree, Random Forest, KNN, SVM, Naïve Bayes	Naïve Bayes achieved about 82% accuracy	Focused mainly on conventional classification performance
Rahman et al.	Kaggle stroke prediction dataset	XGBoost, AdaBoost, LightGBM, Random Forest, Decision Tree, Logistic Regression, KNN, SVM, Naïve Bayes, ANN	Random Forest achieved 96% accuracy	No sequential cohort-based updating
Zhang	Stroke prediction dataset	Random Forest, SVM	Random Forest achieved 95% accuracy and 93% precision	High false-negative rate; no incremental learning
Kumar et al.	Kaggle stroke dataset	AdaBoost, Random Forest, XGBoost, and other ML classifiers	Random Forest achieved 97% accuracy	Static evaluation; no rule-evolution analysis
Al-Mansoori et al.	Stroke forecasting dataset	Five supervised ML classifiers	Random Forest achieved 94.7% accuracy	No adaptive updating or interpretability tracking
This study	Cerebral Stroke Dataset, 43,400 records	Decision Tree baseline, Random Forest, Gradient Boosting	Incremental GB achieved 98.044% accuracy, 91.03% recall, 75.32% F1-score, and 94.66% ROC AUC	Uses simulated sequential cohorts; external validation is still required

mains limited. Most stroke prediction studies focus on static model comparison rather than sequential model updating. This leaves an important gap: how to design a stroke prediction framework that can update models over sequential simulated cohorts, maintain sensitivity to stroke cases, reduce update time, and preserve useful knowledge from previous stages. Recent healthcare studies have further emphasized the need for adaptive and continual learning models in dynamic clinical environments, where patient populations, devices, and data distributions may change over time. Nevertheless, the use of incremental learning for stroke prediction on structured tabular data remains limited, which supports the relevance of the present study [20].

4.3. RULE EXTRACTION AND INTERPRETABLE TREE-BASED MODELS

Interpretability is an important requirement in medical prediction tasks because clinicians and decision-makers need to understand the reasoning behind model predictions. Tree-based models are useful in this context because their decision structures can be translated into explicit rules. Decision Trees provide direct rule paths, while Random Forest and Gradient Boosting can also be analyzed through feature importance, extracted decision logic, or ensemble interpretation methods [7], [21, 22]. Rule extraction has been studied as a way to improve transparency, support model auditing, and transform learned patterns into human-readable knowledge. Some studies focused on extracting rules from machine

learning models, while others used decision-tree-based techniques to derive business or domain rules from complex data. These approaches show that extracted rules can help users inspect how predictions are made, identify important decision conditions, and compare learned patterns across models [7], [22]. Incremental rule mining and interesting pattern discovery also provide useful foundations for analyzing how knowledge changes when new data arrive. Prior work on incremental association-rule mining and temporal rule discovery has shown that newly arriving data can produce conforming, generalized, specialized, unexpected, or novel patterns. Such rule-level analysis is useful because it goes beyond predictive performance and allows the researcher to observe whether the model is learning new patterns or merely repeating previously known ones [23, 24]. However, relatively few studies combine incremental learning with longitudinal rule analysis in medical prediction. In stroke prediction specifically, most studies evaluate classification performance but do not track how decision rules evolve across sequential cohorts. This limits the ability to assess whether model updates preserve clinically meaningful logic, introduce new risk patterns, or produce unstable decision behavior. Therefore, rule evolution analysis remains an underexplored component in adaptive stroke prediction.

4.4. RESEARCH GAP

Based on the reviewed literature, most stroke prediction studies rely on traditional batch learning and evaluate

models mainly on static datasets. Incremental learning has been studied in other domains, but its systematic use for brain stroke prediction remains limited. In addition, rule extraction is often treated as a post-hoc interpretability tool rather than as a longitudinal mechanism for tracking how decision logic changes over time. To address these gaps, this study proposes an incremental learning approach for brain stroke prediction using Decision Tree, Random Forest, and Gradient Boosting models. The proposed approach evaluates predictive performance over sequential simulated cohorts, compares incremental updating with traditional retraining, and analyzes extracted rules in terms of conformity, novelty, and evolution. This provides a more adaptive and interpretable framework for stroke prediction in dynamic healthcare environments.

5. THE PROPOSED APPROACH:

This study proposes an incremental learning approach for brain stroke prediction using three tree-based machine learning models: Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GB). The proposed approach is designed to update prediction models as new patient-data cohorts become available, while preserving previously acquired knowledge and maintaining interpretable decision logic through rule extraction and rule analysis. Unlike traditional batch learning, where the model is retrained from the beginning whenever new data are introduced, the proposed approach follows a sequential update strategy. The model is first trained on an initial cohort and then progressively updated using subsequent cohorts. This design reflects the nature of clinical environments, where patient records are continuously generated over time and predictive models may need to adapt to newly arriving data [5, 6].

5.1. GENERAL FRAMEWORK

The proposed framework consists of five main stages: data preprocessing, initial model construction, incremental model updating, rule extraction and classification, and model evaluation. The dataset is divided into three sequential simulated cohorts, D_1 , D_2 , and D_3 , and processed sequentially at T_1 , T_2 , and T_3 . The first cohort is used to train the initial model, while the later cohorts are used to update the model incrementally. The general workflow is as follows. First, each cohort is preprocessed using the same data preparation pipeline. Second, DT, RF, and GB models are trained on the initial cohort. Third, the trained models are updated incrementally when new cohorts arrive. Fourth, decision rules are extracted from each model after every training or update stage. Finally, the models are compared using predictive performance, update time, and rule-based interpretability. Figure 1 illustrates the workflow of the proposed framework. The figure distinguishes the initial training stage at

T_1 from the incremental update stages at T_2 and T_3 . In RF and GB, previously learned ensemble components are retained during warm-start updating, allowing later updates to benefit from earlier training stages, while DT is refitted as an interpretable baseline.

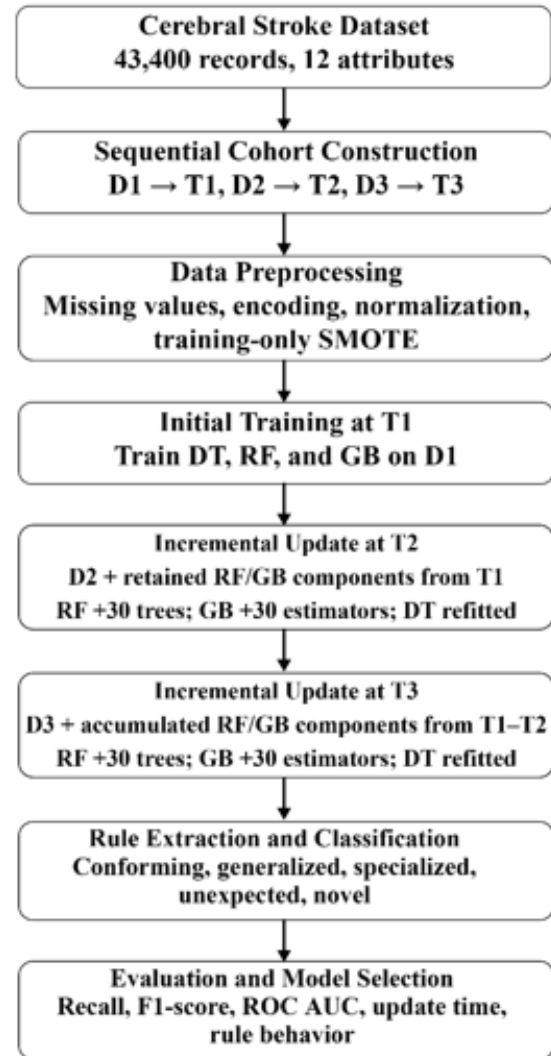


Figure 1. Publication-quality workflow of the proposed incremental tree-based learning framework for brain stroke prediction.

This framework allows the study to examine whether incremental learning improves stroke prediction performance, reduces update time compared with full retraining, and preserves interpretable rule behavior across sequential cohorts.

5.2. SEQUENTIAL COHORT REPRESENTATION

Let the complete sequential simulated cohorts be represented as:

$$D = \{D_1, D_2, \dots, D_T\} \quad (1)$$

where D_t denotes the data cohort arriving at time T_t , and T is the total number of sequential cohorts. In this study, the experimental setting used three cohorts: D_1, D_2 , and D_3 . Each cohort D_t contains a set of patient records represented as:

$$D_t = \left\{ \left(x_i^{(t)}, y_i^{(t)} \right) \right\}_{i=1}^{n_t}, \quad y_i^{(t)} \in \{0, 1\} \quad (2)$$

where $x_i^{(t)}$ is the feature vector of the i -th patient in cohort D_t , $y_i^{(t)}$ is the corresponding class label, and n_t is the number of instances in the cohort. The class label $y_i^{(t)} = 1$ represents a stroke case, while $y_i^{(t)} = 0$ represents a non-stroke case.

5.3. DATA PREPROCESSING

The dataset used in this study is the Cerebral Stroke Dataset, a public imbalanced stroke prediction dataset available on Kaggle. It contains patient demographic, clinical, and lifestyle-related attributes, including age, sex, hypertension, heart disease, marital status, work type, residence type, average glucose level, BMI, smoking status, and stroke target label. The dataset was originally associated with the work of Liu, Fan, and Wu on cerebral stroke prediction using an imbalanced medical dataset [25–27]. Since the dataset does not provide verified patient-level timestamps, the three cohorts were treated as sequential simulated cohorts rather than clinically verified temporal cohorts. This design was used to evaluate the model behavior when new batches of patient records become available progressively. Before model construction, each cohort is processed using a consistent preprocessing pipeline. This includes handling missing values, encoding categorical attributes, normalizing numerical attributes, splitting the data into training and testing parts, and addressing class imbalance. Consistency in preprocessing is important because the objective is to compare model behavior across sequential stages. If preprocessing differs between cohorts, changes in model performance may result from preprocessing variation rather than genuine incremental learning. The dataset contains both numerical and categorical attributes related to stroke risk. Categorical variables are transformed into numerical form, while numerical variables are normalized to reduce scale-related bias. Since stroke cases represent the minority class, imbalance handling is applied to improve the model's ability to detect positive stroke cases [4], [28]. In the practical implementation, class imbalance was handled using the Synthetic Minority Oversampling Technique (SMOTE). SMOTE was applied to the training data in each cohort to increase the representation of stroke cases while preserving the original testing distribution for unbiased evaluation. This procedure was applied consistently across D_1, D_2 , and D_3 , supporting a fair comparison of the model performance across the sequential stages. Table 2 provides a simplified description

of the dataset.

SMOTE was applied only to the training subset of each cohort and not to the validation or testing subsets. This was done to avoid data leakage and preserve an unbiased evaluation under the original class distribution. After SMOTE, the minority stroke class in the training subset was oversampled to match the majority non-stroke class.

5.4. INCREMENTAL MODEL UPDATING

At the first stage, an initial model is trained using the first cohort:

$$M_1 = \text{Train}(D_1) \quad (3)$$

where M_1 represents the initial prediction model. When a new cohort D_t becomes available, the model is updated incrementally instead of being retrained from scratch. The update process is defined as:

$$M_t = \text{Update}(M_{t-1}, D_t), \quad t = 2, 3, \dots, T \quad (4)$$

where M_t is the updated model at time T_t , M_{t-1} is the previously learned model, and D_t is the newly arriving data cohort. The model-specific implementation of this generic update is described below and summarized in Table 3. In the practical implementation, the sequential update protocol was applied over three nearly equal-sized cohorts, D_1, D_2 , and D_3 , containing 13,828, 13,827, and 13,827 records, respectively. The model was first trained at T_1 using D_1 . At T_2 , a new cohort D_2 was introduced, and the model was updated using the newly available data while retaining the knowledge learned from T_1 . At T_3 , the same procedure was repeated using D_3 , allowing the model to benefit from the accumulated knowledge learned across the previous stages. For Random Forest and Gradient Boosting, incremental updating was implemented using a warm-start strategy. Both ensemble models were initialized with 100 estimators at T_1 . At T_2 , 30 additional estimators were appended, increasing the ensemble size to 130. At T_3 , another 30 estimators were appended, resulting in 160 estimators. The previously learned ensemble components were retained during these updates, while the selected hyperparameter settings were kept fixed after the initial tuning stage to ensure a fair comparison and reduce retuning bias. For Gradient Boosting, the learning rate was kept unchanged during the incremental updates. For the decision tree, the standard implementation does not support native online or warm-start updating. Therefore, decision tree was used as a refitted interpretable baseline under the same sequential cohort protocol, rather than as a native incremental learner.

The purpose of this update is not only to learn from new data but also to retain useful knowledge from earlier

**Table 2.** Cohort Size and Class Distribution Before SMOTE

Cohort	Number Records	of Non-Stroke Cases	Stroke Cases	Stroke Ratio
D1 / T1	13,828	13,614	214	1.55%
D2 / T2	13,827	13,613	214	1.55%
D3 / T3	13,827	13,610	217	1.57%
Total	41,482	40,837	645	1.55%

Table 3. Model-Specific Update Strategy Used in the Sequential Protocol

Model	Role in This Study	T1	T2	T3	Hyperparameter Policy
Decision Tree	Refitted interpretable baseline	in- Trained on D1	Refit under the sequential protocol using D2	Refit under the sequential protocol using D3	Fixed after initial tuning
Random Forest	Warm-start incremental ensemble	100 trees trained on D1	D2 introduced; +30 trees appended, total = 130 trees	D3 introduced; +30 trees appended, total = 160 trees	Fixed after initial tuning
Gradient Boosting	Warm-start incremental boosting model	100 estimators trained on D1	D2 introduced; +30 estimators appended, total = 130 estimators	D3 introduced; +30 estimators appended, total = 160 estimators	Fixed after initial tuning; learning rate unchanged

cohorts. This follows the general principle of incremental and continual learning, where the model should incorporate new information while reducing the loss of previously acquired knowledge [6], [17]. Knowledge retention in this implementation was operationally achieved for RF and GB by retaining the previously fitted ensemble components through warm-start updating. No separate knowledge-retention regularization term was optimized in the experiments.

5.5. RULE EXTRACTION AND RULE EVOLUTION

Because the proposed approach uses tree-based models, decision rules can be extracted after each training or update stage. Let R_t denote the set of rules extracted from model M_t :

$$R_t = \text{ExtractRules}(M_t) \quad (5)$$

The extracted rules are compared across sequential stages to determine how the decision logic evolves over time. Rule extraction from tree-based models is useful because it provides an interpretable representation of model behavior and supports decision auditing [7], [22]. Newly discovered rules at time T_t are defined as:

$$R_t^{\text{novel}} = R_t \setminus R_{t-1} \quad (6)$$

where R_t^{novel} represents the rules that appear in the current stage but were not present in the previous stage. Conforming rules are defined as rules that remain

consistent with previously learned rules:

$$R_t^{\text{conforming}} = R_t \cap R_{t-1} \quad (7)$$

The novelty ratio is calculated as:

$$NR_t = \frac{|R_t^{\text{novel}}|}{|R_t|} \times 100 \quad (8)$$

where NR_t represents the percentage of newly discovered rules at time T_t .

Similarly, the conformity ratio is calculated as:

$$CR_t = \frac{|R_t^{\text{conforming}}|}{|R_t|} \times 100 \quad (9)$$

where CR_t represents the percentage of rules that conform to previously learned patterns. To measure the degree of change in rule structure between two consecutive stages, the rule evolution rate is defined as

$$RE_t = 1 - \frac{|R_t \cap R_{t-1}|}{|R_t \cup R_{t-1}|} \quad (10)$$

A higher value of RE_t indicates a greater change in the model's decision logic, whereas a lower value indicates higher rule stability. In the context of stroke prediction, both aspects are important: the model should be stable enough to preserve clinically meaningful knowledge but adaptive enough to learn new patterns from incoming data.

Table 4. Representative Examples of Extracted Rules and Their Classification

Model / Stage	Representative Extracted Rule	Rule Category	Interpretation
DT / T1	If age is high and hypertension = yes, then stroke risk increases.	Conforming	Consistent with established stroke-risk factors.
RF / T2	If age is high, average glucose level is elevated, and heart disease = yes, then stroke risk increases.	Specialized	Adds more specific clinical conditions to an existing risk pattern.
GB / T3	If age is high, average glucose level is elevated, BMI is high, and smoking status indicates risk, then stroke risk increases.	Novel	Represents a newly emerging combined pattern after incremental updating.
GB / T3	If younger age is combined with no hypertension and normal glucose level, then non-stroke prediction is more likely.	Conforming	Reflects a lower-risk pattern consistent with clinical expectations.

5.6. RULE CLASSIFICATION

The extracted rules are classified into five categories: conforming, generalized, specialized, unexpected, and novel rules. This classification is consistent with prior work in knowledge discovery, where novelty has been used as a measure of interestingness to distinguish newly discovered or unexpected rules from previously known patterns [29]. Conforming rules are consistent with previously learned or clinically expected patterns. Generalized rules describe broad patterns that remain valid across multiple cohorts. Specialized rules apply to narrower subsets of patients. Unexpected rules represent patterns that differ from previous knowledge and may require further analysis. Novel rules indicate new patterns that emerge after incremental updates. This classification allows the proposed framework to evaluate interpretability beyond simple rule extraction. It helps determine whether the model is merely repeating old patterns or learning meaningful new patterns from later data cohorts. This is particularly important in clinical decision support, where prediction performance alone is not sufficient; the model's reasoning should also remain understandable and auditable. To further illustrate the interpretability role of the proposed framework, Table 4 presents representative examples of extracted tree-based rules and their corresponding rule classifications. These examples show how the extracted decision logic can be inspected and related to clinically meaningful stroke-risk factors.

These representative rules suggest that the extracted decision logic is clinically plausible. Rules involving older age, hypertension, heart disease, elevated average glucose level, BMI, and smoking-related information are consistent with commonly recognized stroke-risk factors. Therefore, the rule-classification component supports interpretability by showing not only model predictions but also how decision patterns evolve across sequential cohorts.

5.7. PROPOSED ALGORITHM

Algorithm. Proposed Incremental Learning Algorithm for Brain Stroke Prediction

Input:

Sequential simulated cohorts D1, D2, and D3;

Tree-based models DT, RF, and GB.

Output:

Updated models, performance metrics, and classified rules.

1. For each cohort D_t , apply the same preprocessing pipeline:

missing-value handling, categorical encoding, normalization, and training-only SMOTE.

2. Train the initial models on D1 at T1:

DT1 = Train(DT, D1)

RF1 = Train(RF, D1)

GB1 = Train(GB, D1)

3. Extract and classify rules from DT1, RF1, and GB1.

4. For each subsequent cohort D_t , where $t = 2, 3$:

a. Refit DT as an interpretable baseline using D_t .

b. Update RF using warm-start by retaining previous trees and adding 30 new trees.

c. Update GB using warm-start by retaining previous estimators and adding 30 new estimators.

d. Evaluate predictive performance and update time.

e. Extract rules from the updated models.

f. Classify rules as conforming, generalized, specialized, unexpected, or novel.

5. Compare the incremental models with traditional retraining.

6. Select the final model based on recall, F1-score, ROC AUC, update time, and rule behavior.

5.8. EVALUATION METRICS

The proposed approach was evaluated from three perspectives. The first is predictive performance, measured using accuracy, precision, recall, F1-score, and ROC AUC. These metrics were selected because stroke prediction is an imbalanced classification problem in which

accuracy alone may not reflect the model's ability to detect actual stroke cases. Therefore, recall, F1-score, ROC AUC, and sensitivity under high specificity received particular attention, as they provide a more informative assessment of minority-class detection and reduce missed high-risk cases [3], [4]. The second perspective is computational efficiency, measured by comparing the model update time in the incremental approach with the training time in the traditional batch approach. This is necessary because one of the main objectives of incremental learning is to reduce the cost of repeated full retraining when new data become available [6]. The third perspective is interpretability, measured through rule extraction and rule classification across sequential cohorts, which supports the analysis of how the model's decision logic evolves over time [7]. The final model was selected based on its ability to provide balanced performance across these perspectives. A suitable model should detect stroke cases effectively, update efficiently when new data arrive, and maintain interpretable decision logic that can be inspected through extracted rules.

6. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental results of the proposed incremental learning approach for predicting brain stroke. The results are discussed from four perspectives: predictive performance, improvement over traditional retraining, computational efficiency, and rule-based interpretability. The comparison includes three tree-based models: decision tree (DT), random forest (RF), and gradient boosting (GB).

6.1. EXPERIMENTAL SETUP

The experiments were conducted using three sequential simulated cohorts extracted from the Cerebral Stroke Dataset, denoted as D_1 , D_2 , and D_3 , and processed sequentially at T_1 , T_2 , and T_3 . For each cohort, the data were divided into training and testing subsets using an 80:20 ratio. To preserve the class distribution between stroke and non-stroke cases, stratified K-fold partitioning was applied with $k = 5$. The same splitting procedure was applied consistently to D_1 , D_2 , and D_3 , ensuring comparable evaluation across the sequential stages and reducing the risk of biased performance estimates caused by class imbalance. To ensure a fair comparison, all models were tuned under the same validation-based procedure at the initial stage, T_1 . RandomizedSearchCV was used with $n_iter = 10$ for each model to search for model-specific hyperparameters. The selected hyperparameters were then kept fixed during T_2 and T_3 to avoid repeated retuning bias and to ensure that later performance changes were mainly attributed to the sequential update protocol rather than stage-specific hy-

perparameter optimization. The same preprocessing pipeline, SMOTE procedure, train-test split ratio, and evaluation metrics were applied consistently across all cohorts.

Each model was evaluated under two learning settings. The first setting was the traditional batch approach, in which models were trained without using the previous model state. The second setting was the proposed incremental approach, in which each model was updated progressively as new data cohorts became available. The final selected settings were retained unchanged after T_1 and applied consistently during T_2 and T_3 for all models. The evaluated models were DT, RF, and GB. The comparison used several performance indicators, including accuracy, precision, recall, F1-score, ROC AUC, and training time. Because stroke prediction is a clinically sensitive task, special attention was given to recall and false-negative reduction, because failing to detect actual stroke cases may delay preventive intervention. All experiments were conducted on a personal computer running Windows 11 (version 10.0.26100) with an Intel64 Family 6 Model 154 processor and 15.71 GB RAM. The implementation was performed using Python 3.13.1 and scikit-learn 1.6.0. All models were evaluated under the same hardware and software environment to ensure consistent timing comparisons.

6.2. INCREMENTAL LEARNING VERSUS TRADITIONAL RETRAINING

The experimental results show that the incremental approach consistently improved the predictive performance of the evaluated models compared with the traditional approach. The most important improvement was observed in the recall, which is particularly relevant for stroke prediction. Table 5 illustrates this.

At the final stage D_3 , all models showed improved predictive performance under the incremental approach compared with the traditional approach. Decision Tree achieved moderate gains, with recall increasing from 42.11 to 50.19 and F1-score increasing from 41.73 to 50.19. Random Forest showed stronger improvement, with recall increasing from 57.62 to 68.11 and ROC AUC increasing from 75.18 to 83.58. The largest improvement was observed in Gradient Boosting, where recall increased from 50.266 to 91.03, F1-score increased from 50.20 to 75.32, and ROC AUC increased from 81.70 to 94.66. The observed recall improvement for Gradient Boosting may be associated with the combined effect of training-only SMOTE, warm-start estimator expansion, and the sensitivity-oriented evaluation setting. In particular, later GB stages retained previously learned boosting estimators and appended additional estimators when new cohorts became available, which may have helped the model detect more minority stroke cases. However, this interpretation should not be considered a definitive

Table 5. Hyperparameter Tuning Policy

Model	Tuned ters	Hyperparame- ters	Tuning Method	Tuning Stage	Policy in T2 and T3
Decision Tree	max_depth, min_samples_split, min_samples_leaf		Randomize	T1 only	Fixed after initial tuning; model refitted as an inter- pretable baseline
Random Forest	n_estimators, max_depth, max_features, min_samples_split, min_samples_leaf		Randomized Search	T1 only CV	Fixed after initial tuning; +30 trees appended at each update
Gradient Boosting	n_estimators, learning_rate, min_samples_leaf	learn- ing_rate, max_depth,	Randomize	T1 only	Fixed after initial tuning; +30 estimators appended at each update; learning rate unchanged

Table 6. Performance Comparison between Traditional and Incremental Approaches at D_3

Model	Approach	Accuracy (%)	F1- Score	Precision	Recall	ROC AUC (%)	Training Time (s)
DT	Traditional	93.950	41.73	41.35	42.11	72.11	5.245
DT	Incremental	96.191	50.19	50.19	50.19	74.11	2.021
RF	Traditional	96.228	56.75	55.89	57.62	75.18	42.817
RF	Incremental	98.162	68.11	68.11	68.11	83.58	11.718
GB	Traditional	96.240	50.20	50.135	50.266	81.70	320.221
GB	Incremental	98.044	75.32	64.24	91.03	94.66	61.215

Table 7. Performance Improvement of Incremental Learning over Traditional Learning at D_3

Model	Δ Accuracy	Δ F1-Score	Δ Precision	Δ Recall	Δ ROC AUC	Δ Training Time
DT	+2.241	+8.46	+8.84	+8.08	+2.00	-3.224 s
RF	+1.934	+11.36	+12.22	+10.49	+8.40	-31.099 s
GB	+1.804	+25.12	+14.105	+40.764	+12.96	-259.006 s

causal conclusion. The observed behavior reflects the specific experimental setting, simulated cohort construction, and preprocessing protocol used in this study, and further validation on independent and genuinely time-stamped clinical datasets is required. These results indicate that the incremental approach was particularly effective for ensemble-based models, especially gradient boosting. This is illustrated in Table 7.

Table 7 summarizes the direct improvement achieved by the incremental approach over the traditional approach at the final cohort D_3 . The largest improvement was obtained using gradient boosting, where the recall increased from 50.266 to 91.03, representing a gain of 40.764 points. Its F1-score also increased by 25.12 points, and the ROC AUC improved by 12.96 points. Random Forest showed consistent gains across all predictive metrics, whereas Decision Tree achieved more modest improvements. This suggests that the incremental set-

ting was most useful when combined with ensemble models, particularly gradient boosting, where the reduction in false negatives is more important clinically than the small difference in overall accuracy. The differences were calculated as Incremental D_3 minus Traditional D_3 . Positive values indicate improvement in predictive performance, while negative values in training time indicate a reduction in computational cost. To provide a clearer visual interpretation of the model behavior across the sequential stages, Figure 2 shows the evolution of recall for DT, RF, and GB under the incremental protocol from T1 to T3.

As shown in Figure 2, gradient boosting achieved the strongest recall improvement at the final stage, increasing from 47.80% at T1 to 51.37% at T2 and 91.03% at T3. Random Forest also showed a steady improvement from 50.99% to 68.11%, while decision tree achieved a more limited increase from 39.47% to 50.19%. This visual

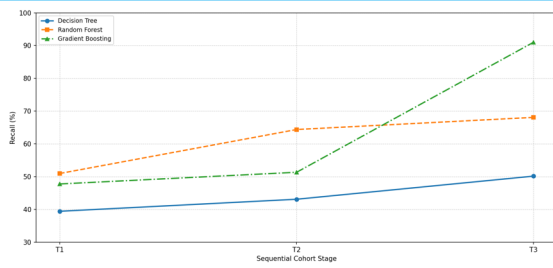


Figure 2. Recall evolution of the incremental models across T1–T3.

trend supports the numerical results reported in Tables 6 and 7 and further explains the sensitivity gain achieved by the incremental update protocol. Figure 3 presents the ROC curves and AUC values for the traditional and incremental models on the D3 test set.

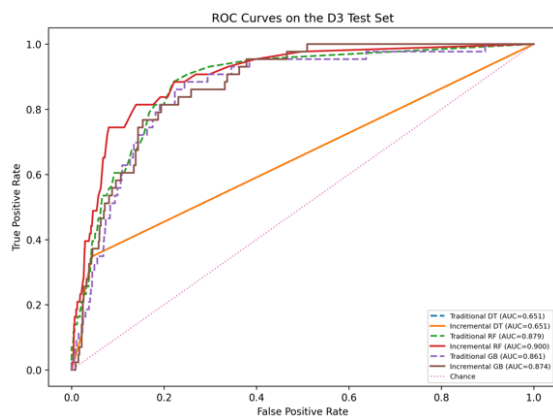


Figure 3. ROC curves and AUC values for the evaluated models on the D3 test set.

Additional statistical validation was also conducted on the D3 test set in response to the final reviewer's recommendation. The D3 data were evaluated using a stratified 80:20 test split, and SMOTE was applied only to the training subset. McNemar's exact test compared paired predictions from the traditional and incremental versions of each model on the same D3 test instances. The exact p-values were 1.0000 for DT, 0.0275 for RF, and 0.0020 for GB, indicating statistically significant paired prediction differences for RF and GB but not for DT. The corresponding traditional versus incremental ROC AUC values were 0.651 versus 0.651 for DT, 0.879 versus 0.900 for RF, and 0.861 versus 0.874 for GB. These results provide supplementary statistical evidence and should be interpreted together with the sensitivity-oriented metrics and false-negative behavior. The observed recall improvement for Gradient Boosting may be associated with the combined effect of training-only SMOTE, warm-start estimator expansion, and the sensitivity-oriented evaluation setting. In particular, later GB stages retained previously learned boosting estimators and appended additional estimators when new cohorts became available, which

may have helped the model detect more minority stroke cases. However, this interpretation should not be considered a definitive causal conclusion. The observed behavior reflects the specific experimental setting, simulated cohort construction, and preprocessing protocol used in this study, and further validation on independent and genuinely time-stamped clinical datasets is required.

6.3. COMPUTATIONAL EFFICIENCY

The incremental approach also reduced training time as the models progressed from D_1 to D_3 . The reduction was clear across all three models. Decision Tree achieved a total training-time reduction of 67.46% from D_1 to D_3 , while Random Forest achieved 73.23%. Gradient Boosting achieved the highest reduction, with a total decrease of 80.87%. The following table details this. The reduction in training time can be explained by the sequential update design. Instead of rebuilding the ensemble models from scratch at each stage, the warm-start strategy retained the previously learned estimators and appended only a fixed number of new estimators when a new cohort became available. As a result, the models incorporated new cohort information at a lower computational cost than full retraining.

Table 8. Training-Time Reduction within the Incremental Approach

Model	Reduction ($D_1 \rightarrow D_2$)	Reduction ($D_2 \rightarrow D_3$)	Total Reduction ($D_1 \rightarrow D_3$)
DT	38.23%	47.32%	67.46%
RF	49.05%	47.52%	73.23%
GB	63.60%	47.47%	80.87%

This result is important because GB had the highest training cost at the first stage, but the incremental approach substantially reduced its update time in later stages. Specifically, the GB training time decreased from approximately 320.063 s at D_1 to 61.215 s at D_3 . This indicates that incremental updating can make computationally expensive models more practical in dynamic healthcare environments. Although DT and RF remained faster in absolute training time at D_3 , their predictive behavior was less clinically favorable. DT was very fast but had low recall, whereas RF was efficient and accurate but produced more false negatives than GB. Therefore, GB provided the best trade-off between update efficiency and clinical sensitivity.

6.4. CONFUSION MATRIX ANALYSIS

The confusion matrix results provide a clearer explanation of why GB was the most suitable model. At D_3 , GB achieved 406 true positives and only 40 false negatives.

In contrast, RF achieved 267 true positives and 125 false negatives, while DT achieved 261 true positives and 259 false negatives. Table 8 presents these results.

Table 9. Confusion Matrix Results at D_3

Model	TP	TN	FP	FN
DT	261	12821	259	259
RF	267	13083	125	125
GB	406	12928	226	40

The RF confusion-matrix values were verified against the original experimental records. Although FP and FN are numerically equal for RF at the final stage, both values were recorded as 125 in the final output; therefore, the equality is coincidental and not a transcription error. Table 9 confirms that gradient boosting achieved the lowest number of false negatives at the final stage, which supports its selection as the most clinically suitable model despite random forest having slightly higher overall accuracy. This difference is clinically meaningful. A lower number of false negatives indicates fewer actual stroke-risk cases are missed. Although RF obtained slightly higher accuracy, GB detected substantially more true stroke cases. Therefore, from a clinical decision-support perspective, GB is preferable because it provides a stronger ability to identify high-risk patients. The confusion matrix also shows that DT had the weakest sensitivity at the final stage. Its false negatives remained high, indicating that the model failed to detect many stroke cases despite acceptable overall accuracy. RF improved over DT, but GB offered the most balanced result by combining high true-positive detection with a relatively low false-negative count.

6.5. RULE CLASSIFICATION AND INTERPRETABILITY

In addition to predictive performance, the study analyzed the extracted rules from DT, RF, and GB across the three-time stages. The rules were classified using the novelty measure into conforming, generalized, specialized, unexpected, and novel rules under thresholds from $\Phi = 0.1$ to $\Phi = 0.9$. This analysis was used to observe how the decision logic evolved during incremental learning. Table 10 summarizes the rule classification results at the 0.5 threshold.

Table 10 shows that Random Forest and Gradient Boosting achieved more than 90% conforming rules at $\Phi = 0.5$, indicating stable rule behavior after incremental updating. Decision Tree remained slightly below this level, with 88.70% conforming rules, although it produced the smallest and most interpretable rule set. For the decision tree, the number of rules was relatively small, which made the model highly interpretable. At D_3 , the total

number of rules reached 292. However, DT showed weaker clinical performance, especially in recall. This indicates that interpretability alone is not sufficient if the model fails to detect stroke cases effectively. Random Forest generated a much larger number of rules. At D_3 , the total number of RF rules exceeded 20,000, making interpretation difficult. Although RF achieved strong accuracy and acceptable recall, the large rule volume reduced practical interpretability. At $\Phi = 0.5$, RF showed more than 90% conforming rules, indicating stable rule behavior, but its high number of rules made clinical auditing less convenient. Gradient boosting provided a better balance. It generated a moderate number of rules compared with RF, ranging from approximately 5,208 rules at T_1 to 8,604 rules at T_3 . At $\Phi = 0.5$, GB achieved more than 90% conforming rules, showing that its decision logic remained largely stable after incremental updates. At lower thresholds, GB also produced a noticeable number of novel rules, especially at T_3 , which suggests that the model was able to learn new patterns from later cohorts. This rule behavior supports the main goal of the proposed approach. The model should not only achieve strong predictive results but also preserve interpretable and stable decision logic while learning from new data. Compared with DT and RF, GB provided the most appropriate compromise between predictive strength, rule conformity, and manageable interpretability.

6.6. OVERALL MODEL RANKING

The final model comparison shows that the selection of the best model depends on clinical priorities rather than overall accuracy alone. RF achieved the highest final accuracy, but GB achieved the best recall and the lowest false-negative count. DT was the most interpretable and fastest model, but its low recall made it unsuitable for stroke-risk screening. The following table presents these results.

Therefore, gradient boosting was ranked as the best model for the proposed framework. It achieved the highest recall (91.03), strong ROC AUC (94.66), acceptable training time after incremental updating (61.215 s at (D_3) , and more than 90% conforming rules at $\Phi = 0.5$. These results indicate that incremental GB provides the best balance between clinical sensitivity, computational efficiency, and rule-based interpretability. Overall, the findings confirm that the proposed incremental learning approach improves stroke prediction compared with traditional retraining. The improvement is especially clear in sensitivity-oriented metrics and training-time reduction. The rule analysis further shows that incremental learning can support interpretable model evolution, making it more suitable for adaptive clinical decision-support systems. Compared with previous static stroke prediction studies summarized in Table 1, the proposed incremental gradient boosting model achieved a competitive overall

**Table 10.** Rule Classification Summary at $\Phi=0.5$ and T_3

Model	Total Rules	Conforming Rules	Generalized Rules	Specialized Rules	Unexpected Rules	Novel Rules	Conforming Ratio
DT	292	259	20	11	2	0	88.70%
RF	20812	18866	1442	504	0	0	90.65%
GB	8604	7788	606	208	2	0	90.51%

Table 11. Performance and Rule Classification

Criterion	Gradient Boosting	Random Forest	Decision Tree
Accuracy at (D_3)	98.044%	98.162%	96.191%
Recall at (D_3)	91.03	68.11	50.19
FN at (D_3)	40	125	259
Rule Quality	> 90% conforming at $\Phi = 0.5$	> 90% conforming at $\Phi = 0.5$	< 90% conforming at all Φ
Interpretability	Moderate, 5k–8k rules	Poor, 20k+ rules	High, 292 rules
Training Time at (D_3)	61.215 s	11.718 s	2.021 s
Clinical Fit	Best	Risky due to higher FN	Unsuitable due to low recall

performance and stronger sensitivity-oriented behavior. For example, prior studies commonly reported high accuracy values, such as 96% for Random Forest in Rahman et al. and 95% for Random Forest in Zhang, while the present incremental GB model achieved 98.044% accuracy and 91.03% recall at the final stage. This comparison suggests that the proposed approach is particularly useful when the evaluation priority is not only overall accuracy but also the detection of true stroke cases under class imbalance. However, this comparison should be interpreted cautiously because datasets, preprocessing strategies, validation protocols, and reported metrics differ across studies.

7. CONCLUSION

This study presents an incremental learning approach for predicting brain stroke using decision tree, random forest, and gradient boosting models. The approach was tested on sequential simulated cohorts to examine whether model updating can improve prediction performance while reducing the need for repeated retraining from scratch. Rule extraction and rule classification were also used to observe how the decision behavior of the models changed after each update. The results showed that incremental updating improved the performance of all three models, although the level of improvement differed from one model to another. Gradient boosting achieved the strongest clinical performance at the final

stage, with a recall of 91.03, a ROC AUC of 94.66, and only 40 false negatives. This result is important because, in stroke-risk screening, missing a true high-risk case is more serious than a small difference in overall accuracy. Although Random Forest achieved slightly higher accuracy, it produced more false negatives, making gradient boosting the more suitable model from a clinical decision-support perspective. The results also showed that incremental updating reduced training time across the sequential stages. Gradient boosting achieved the largest reduction from (D_1) to (D_3), which suggests that incremental updating can make computationally expensive models more practical when new data arrive. In addition, the rule analysis showed that the extracted rules can be used to track changes in model logic over time. This adds an interpretability layer to the prediction process and helps examine whether the updated model preserves stable patterns while learning from new cohorts. This study has several limitations. First, the cohorts were constructed as sequential simulated cohorts because the dataset does not provide verified patient-level timestamps; therefore, the results do not fully validate performance under real clinical temporal ordering. Second, the evaluation was conducted on a single public dataset, which limits generalizability to other populations and healthcare institutions. Third, although SMOTE was applied only to training data to reduce class imbalance, the use of synthetic oversampling may still influence learned decision boundaries. Fourth, although an additional McNemar statistical validation was added to the D_3 test set, broader statistical robustness still requires repeated independent runs, external validation, and evaluation on genuinely time-stamped clinical datasets. Future work should therefore include repeated experiments, statistical testing across multiple datasets, and real concept-drift evaluation. Overall, the final results indicate that the incremental GB model was the most appropriate choice for the evaluated screening scenario because it minimized missed stroke cases while maintaining an acceptable update cost and stable rule behavior.

8. DECLARATIONS

8.1. ETHICAL APPROVAL:

This study did not involve any new experiments on humans or animals performed by the authors. The work

was conducted using an existing anonymized dataset for research purposes. The authors declare that the manuscript is original, has not been published previously, and is not currently under consideration for publication elsewhere. The order of authors listed in the manuscript has been approved by all authors.

8.2. CONFLICT OF INTEREST:

The authors declare that they have no conflict of interest.

8.3. COMPETING INTERESTS:

The authors declare that they have no financial or proprietary interests in any material discussed in this article.

8.4. AUTHORS' CONTRIBUTIONS:

All authors contributed to the study conception and design. Data preprocessing, model implementation, simulations, result analysis, and manuscript preparation were performed collaboratively by the authors. All authors reviewed and approved the final version of the manuscript.

8.5. FUNDING

No funding was received from any organization for this work.

8.6. AVAILABILITY OF DATA AND MATERIALS:

The dataset used in this study was used for research purposes. Data availability can be provided by the corresponding author upon reasonable request, subject to any applicable data-use restrictions.

8.7. CONSENT FOR PUBLICATION

All authors have read and approved the final manuscript and consent to its publication.

8.8. CONSENT TO PARTICIPATE

Not applicable, as this study did not involve the direct participation of human subjects by the authors.

8.9. DECLARATION OF AI-ASSISTED TECHNOLOGIES:

The authors declare that artificial intelligence (AI) tools were used only for limited purposes, such as language editing, grammar improvement, and formatting assistance. All scientific content, data analysis, interpretations, conclusions, and final manuscript revisions were reviewed and verified by the authors, who take full responsibility for the integrity and originality of the work.

REFERENCES

- [1] V. L. Feigin et al., "World Stroke Organization: Global Stroke Fact Sheet 2025," *Int. J. Stroke*, vol. 20, no. 2, pp. 132–144, Feb. 2025. DOI: [10.1177/17474930241308142](https://doi.org/10.1177/17474930241308142)
- [2] A. K. Boehme, C. Esenwa, and M. S. Elkind, "Stroke Risk Factors, Genetics, and Prevention," *Circ. Res.*, vol. 120, no. 3, pp. 472–495, Feb. 2017. DOI: [10.1161/CIRCRESAHA.116.308398](https://doi.org/10.1161/CIRCRESAHA.116.308398)
- [3] N. Biswas, K. M. M. Uddin, S. T. Rikta, and S. K. Dey, "A comparative analysis of machine learning classifiers for stroke prediction: A predictive analytics approach," *Healthc. Anal.*, vol. 2, p. 100116, 2022. DOI: [10.1016/j.health.2022.100116](https://doi.org/10.1016/j.health.2022.100116)
- [4] A. Fernández, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, 2018. DOI: [10.1613/jair.1.11192](https://doi.org/10.1613/jair.1.11192)
- [5] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surv.*, vol. 46, no. 4, pp. 1–37, 2014. DOI: [10.1145/2523813](https://doi.org/10.1145/2523813)
- [6] A. Chefrour, "Incremental supervised learning: Algorithms and applications in pattern recognition," *Evol. Intell.*, vol. 12, no. 2, pp. 97–112, 2019. DOI: [10.1007/s12065-019-00203-y](https://doi.org/10.1007/s12065-019-00203-y)
- [7] R. Bemthuis, W. Wang, M.-E. Iacob, and P. Havinga, "Business rule extraction using decision tree machine learning techniques: A case study into smart returnable transport items," *Procedia Comput. Sci.*, vol. 220, pp. 446–455, 2023. DOI: [10.1016/j.procs.2023.03.057](https://doi.org/10.1016/j.procs.2023.03.057)
- [8] M. U. Emon, M. S. Keya, T. I. Meghla, M. M. Rahman, M. S. Al Mamun, and M. S. Kaiser, "Performance analysis of machine learning approaches in stroke prediction," in *Proceedings of the 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, IEEE, 2021, pp. 1464–1469. DOI: [10.1109/ICECA49313.2020.9297525](https://doi.org/10.1109/ICECA49313.2020.9297525)
- [9] T. Tazin, M. N. Alam, N. N. Dola, M. S. Bari, S. Bourouis, and M. M. Khan, "Stroke disease detection and prediction using robust learning approaches," *J. Healthc. Eng.*, vol. 2021, no. 1, p. 7633381, 2021. DOI: [10.1155/2021/7633381](https://doi.org/10.1155/2021/7633381)
- [10] S. Rahman, M. Hasan, and A. K. Sarkar, "Prediction of brain stroke using machine learning algorithms and deep neural network techniques," *Eur. J. Electr. Eng. Comput. Sci.*, vol. 7, no. 1, pp. 23–30, 2023. DOI: [10.24018/ejece.2023.7.1.483](https://doi.org/10.24018/ejece.2023.7.1.483)
- [11] Y. Zhang, "Stroke prediction based on machine learning," in *ITM Web of Conferences*, vol. 70, EDP Sciences, 2025, p. 04029. DOI: [10.1051/itmconf/20257004029](https://doi.org/10.1051/itmconf/20257004029)
- [12] G. Sailasya and G. L. A. Kumari, "Analyzing the performance of stroke prediction using ml classification algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, 2021. DOI: [10.14569/IJACSA.2021.0120662](https://doi.org/10.14569/IJACSA.2021.0120662)
- [13] X. Li, D. Bian, J. Yu, M. Li, and D. Zhao, "Using machine learning models to improve stroke risk level classification methods of china national stroke screening," *BMC Med. Informatics Decis. Mak.*, vol. 19, no. 1, p. 261, 2019. DOI: [10.1186/s12911-019-0998-2](https://doi.org/10.1186/s12911-019-0998-2)
- [14] J. Read and I. Žliobaitė, "Learning from data streams: An overview and update," *ACM Comput. Surv.*, 2025, Preprint: arXiv:2212.14720. DOI: [10.1145/3737279](https://doi.org/10.1145/3737279)
- [15] A. A. Mustafa, A. A. Al-Khulidi, and A. S. Al-Hegami, "A proposed incremental distance based clustering of interesting patterns for big data," *Sana'a Univ. J. Appl. Sci. Technol.*, vol. 4, no. 4, pp. 1938–1949, 2026. DOI: [10.59628/jast.v4i4.2271](https://doi.org/10.59628/jast.v4i4.2271)



- [16] A. O. AlQabbany and A. M. Azmi, "Measuring the effectiveness of adaptive random forest for handling concept drift in big data streams," *Entropy*, vol. 23, no. 7, p. 859, Jul. 2021. DOI: [10.3390/e23070859](https://doi.org/10.3390/e23070859)
- [17] G. M. Van de Ven, N. Soures, and D. Kudithipudi, *Continual learning and catastrophic forgetting*, 2024. DOI: [10.48550/arXiv.2403.05175](https://doi.org/10.48550/arXiv.2403.05175) arXiv: [2403.05175](https://arxiv.org/abs/2403.05175).
- [18] H.-G. Doan, H.-Q. Luong, T.-O. Ha, and T. T. T. Pham, "An efficient strategy for catastrophic forgetting reduction in incremental learning," *Electronics*, vol. 12, no. 10, p. 2265, 2023. DOI: [10.3390/electronics12102265](https://doi.org/10.3390/electronics12102265)
- [19] S. Tian, L. Li, W. Li, H. Ran, X. Ning, and P. Tiwari, "A survey on few-shot class-incremental learning," *Neural Networks*, vol. 169, pp. 307–324, 2024. DOI: [10.1016/j.neunet.2023.10.039](https://doi.org/10.1016/j.neunet.2023.10.039)
- [20] A. Singh, M. B. Gurbuz, S. S. Gantha, and P. Jasti, *Class-incremental continual learning for general purpose health-care models*, 2023. arXiv: [2311.04301](https://arxiv.org/abs/2311.04301).
- [21] C. Aldrich, "Process variable importance analysis by use of random forests in a shapley regression framework," *Minerals*, vol. 10, no. 5, p. 420, 2020. DOI: [10.3390/min10050420](https://doi.org/10.3390/min10050420)
- [22] H. Meydan and M. Bal, "Children of the tree: Optimised rule extraction from machine learning models," *J. Data Anal. Artif. Intell. Appl.*, vol. 1, no. 1, pp. 14–35, 2025. DOI: [10.26650/d3ai.1606958](https://doi.org/10.26650/d3ai.1606958)
- [23] H. A. Alsaeedi and A. S. Alhegami, "An incremental interesting maximal frequent itemset mining based on fp-growth algorithm," *Complexity*, vol. 2022, no. 1, p. 1942517, 2022. DOI: [10.1155/2022/1942517](https://doi.org/10.1155/2022/1942517)
- [24] A. Al-Hegami and H. Alsaeedi, "A framework for incremental parallel mining of interesting association patterns for big data," *Int. J. Comput.*, vol. 19, no. 1, pp. 106–117, 2020. DOI: [10.47839/ijc.19.1.1699](https://doi.org/10.47839/ijc.19.1.1699)
- [25] T. Liu, W. Fan, and C. Wu, "A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset," *Artif. Intell. Med.*, vol. 101, p. 101723, 2019. DOI: [10.1016/j.artmed.2019.101723](https://doi.org/10.1016/j.artmed.2019.101723)
- [26] T. Liu, W. Fan, and C. Wu, *Data for: A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical-datasets*. DOI: [10.17632/x8ygrw87jw.1](https://doi.org/10.17632/x8ygrw87jw.1) [Online]. Available: <https://doi.org/10.17632/x8ygrw87jw.1>
- [27] V. Sam, *Cerebral Stroke Dataset*, Kaggle, Accessed: Jul. 2, 2026. [Online]. Available: <https://www.kaggle.com/datasets/viviansam/cerebral-stroke-dataset>
- [28] H. A. Marwai and G. H. AL-Gaphari, "A hybrid feature selection method based on binary pso algorithm for microarray data classification," *Sana'a Univ. J. Appl. Sci. Technol.*, vol. 2, no. 4, pp. 375–380, 2024. DOI: [10.59628/jast.v2i4.1023](https://doi.org/10.59628/jast.v2i4.1023)
- [29] V. Bhatnagar, A. S. Al-Hegami, and N. Kumar, "Novelty as a measure of interestingness in knowledge discovery," *Constraints*, vol. 9, p. 18, 2006. DOI: [10.5281/zenodo.1082389](https://doi.org/10.5281/zenodo.1082389)