Cureus Journal of Computer Science

Open Access Original Article

An Intelligent Framework for Real-Time Heart Stroke Monitoring and Early Detection Using Machine Learning

Received 03/23/2025
Review began 04/21/2025
Review ended 05/26/2025
Published 05/27/2025

© Copyright 2025
Kumar Rajpoot et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: https://doi.org/10.7759/s44389-025-04004-6

Navneet Kumar Rajpoot [1] , Prabh Deep Singh [1] , Bhaskar Pant [1] , Manmohan Sharma [2]

1. Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun, IND 2. School of Computer Applications, Lovely Professional University, Jalandhar, IND

Corresponding authors: Navneet Kumar Rajpoot, navneetrajpootgeu@gmail.com, Prabh Deep Singh, ssingh.prabhdeep@gmail.com, Bhaskar Pant, bhasker.pant@geu.ac.in

Abstract

The interruption of blood flow resulting from sudden brain function shutdown leads to the development of heart stroke, which represents a life-threatening medical emergency. Traditional healthcare approaches demonstrate limited capability for delivering prompt insights together with forecasts, thus limiting their support for stroke case real-time decision support systems. Cyber-physical systems (CPS) form an innovative framework to boost heart stroke prediction along with monitoring capabilities during the early stages. The system contains physical and cyber components as its main two parts where physical devices active in real-time analyze biological occurrences, while cyber components complete advanced assessments for decisive operations. Pre-processing operations including data balancing and noise reduction and feature selection are carried out by the system through the execution of random forest and artificial neural networks machine learning algorithms. Kruskal-Wallis H-test serves as the statistical methodology to validate the integrated dataset by establishing robustness measures in its integration. The framework achieves accurate heart stroke predictions based on its evaluation measures, which include accuracy, precision, recall, and F1-score. The framework achieves accurate heart stroke predictions based on its evaluation measures, which include accuracy (95.7%), precision (94.3%), recall (93.8%), and F1-score (94.0%). Statistical validation through Kruskal-Wallis H-test ($p < 0.01$) confirms the framework's robust performance in prediction tasks. This framework consists of two analysis layers that provide time-sensitive heart stroke monitoring capabilities. Advanced machine learning enhances the prediction and forecasting accuracy for heart stroke cases. Statistical validation and evaluation metrics confirm the framework's robust performance in prediction tasks. The results highlight the CPS-based approach as a significant advancement, contributing to improved predictive capabilities and facilitating more personalized patient care.

Categories: AI applications, Artificial Intelligence and Machine Learning in the Cloud, Data Analysis
Keywords: data pre-processing, early prediction, forecasting, smart healthcare, machine learning

Introduction

Heart stroke, also known as a cerebrovascular accident, is a condition in which the blood supply to the brain is blocked, resulting in damage to brain cells and, in extreme cases, death. It is one of the leading causes of death and permanent disability worldwide, with millions of people affected each year. Early diagnosis and timely medical treatment of heart stroke can prevent some potentially disastrous consequences of the disease. This would require constant monitoring of all parameters related to at-risk patients, which is a limitation because early warning signs may be subtle or present suddenly.

Cyber-physical systems (CPS) play an extremely important role in allowing real-time health monitoring, especially when it comes to complex medical conditions such as heart stroke [1]. CPS integrate physical devices like wearable sensors, medical instruments, and patient monitors with intelligent computational systems in an integrated manner. With real-time heart stroke monitoring, the CPS can always track vital signs like heart rate, blood pressure, oxygen levels, and other physiological parameters so that any deviation from the norm is recognized at once. With the capability to analyze real-time data and access, the early detection of strokes can be supported by CPS in improving healthcare systems. Continuous monitoring would therefore be able to capture subtle early warning signs, thus making an opportunity for intervention before the stroke fully manifests itself [2]. It thus improves upon the time delay between the discovery of symptoms and medical responses, thereby improving chances of survival and quick recovery for the patient. Early detection and prognosis are critical to managing the risk for stroke. Like most such events, patients have little time to respond to medical intervention. Within minutes of a stroke having taken place, millions of neurons in the brain can be lost, which means very rapid medical response would be needed that may make a difference. Early detection systems can alert healthcare providers even before the event of a stroke, enabling timely medical intervention that may prevent or minimize brain damage that may occur from the stroke [3]. The ability to predict the likelihood of a stroke before it happens contributes to a proactive approach and reduces dependency on reactive emergency care. Predicting enables the healthcare system to pin-point the patients and take preventive measures such as adjusting medicines, lifestyle changes, or closer monitoring of health indicators. In addition to that, the early detection system with real-time surveillance will improve resource utilization by providing critical care on time [4].

Machine learning in its forecasting ability is the next element to be added to further boost the efficiency of CPS in stroke monitoring. Its algorithms learn such patterns and risk factors by training on the large, diverse datasets of hospitals and authentic sources, which cannot be seen by human observers. Such patterns are treated through the use of both historical data and real-time inputs to provide predictions of the likelihood of a stroke - the rationale behind this early warning system permits preventive care [5]. This has resulted in the urgent need for forecasting and early detection in the management of heart stroke, with the ambition of curtailing mortality as well as providing a favorable patient outcome. Combined CPS with machine learning represents a sound framework in real-time heart stroke monitoring and prediction [6]. As an intelligent CPS framework evolves to redefine stroke patient care, it provides medical staff with continuous tracking and early warning detection through an intervention tool that reduces heart stroke risks [7]. Medical personnel will use predictive analytics to process real-time acquired data to save lives and minimize both patient health deterioration and costs for stroke care services. Real-time monitoring exists in the proposed intelligent framework because it uses historical patient records to analyze current health trends and predict stroke risks. The system enables proactive healthcare practices by showing potential stroke occurrences, which allows providers to track vulnerable patients or preventively care for them [8]. The predictive approach gives improved benefits over traditional methods since it relies only on reactive and symptomatic triggers. This research aims to achieve the following objectives:

1. Develop an intelligent CPS framework integrating real-time monitoring and advanced analytics for early stroke detection.

2. Implement and evaluate machine learning algorithms for stroke prediction using multi-source healthcare data.

3. Validate the framework's effectiveness through statistical analysis and performance metrics.

4. Apply machine learning for early prediction with actionable alerts.

Problem statement

Medical experts must create efficient predictive models for heart stroke detection because these conditions become more prevalent in society. The objective of this research tackles heart stroke forecasting through a CPS model, which connects physical data acquisition with sophisticated computational techniques.

Let $D = \{d_1, d_2, \ldots, d_n\}$ be the dataset that originates from different hospitals together with online resources.

where each d_i represents the ith patient.

$$d_i = \{x_1, x_2, \ldots, x_m\}$$

where x_j represents physiological and demographic patient features. The model uses the features x_j that include age with blood pressure and cholesterol levels among others.

The aim is forecasting heart stroke by determining whether patients experienced a stroke or not through the variable $Y \in \{0, 1\}$.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives, respectively.

The proposed CPS consists of two core components: physical data acquisition and preliminary clinical testing together with the cyber domain that performs data cleaning and modeling. The transformation process occurs through this model where physical data move to cyber components:

$$P : D_{\text{physical}} \to D_{\text{cyber}}$$

Data transfer security and integrity will be achieved through the implementation of encryption methods and data validation protocols during data movement.

Multiple predictive models from $\mathcal{M} = \{M_1, M_2, \ldots, M_k\}$ will be applied to preprocessed data after data integration. Where each M_i represents a different predictive model. The established model selection method will consist of choosing the model, which achieves maximum accuracy (A) :

$$M^* = \arg\max_{M_i \in \mathcal{M}} A(M_i)$$

Literature review

With the rapid expansion of smart healthcare technologies, integrating machine learning, CPS, and Internet of Things (IoT) infrastructure has significantly enhanced the capability to deliver real-time, personalized, and efficient medical care. Machine learning-driven CPS frameworks are increasingly being adopted for early diagnosis and continuous monitoring of critical conditions like heart stroke, cardiovascular diseases, and other chronic illnesses. Arslan et al. [9] highlighted the integration of wearable sensors with wireless systems to support comprehensive patient monitoring. Their work underscores how machine learning algorithms, when deployed alongside real-time sensing infrastructure, significantly contribute to remote diagnostics. However, the study lacked detailed architectural modeling of temporal component interaction, a key consideration in delay-sensitive healthcare environments. Bebortta et al. [10] proposed the Deep-MIST framework, leveraging deep learning for efficient management of healthcare big data using mobile and edge computing. Although the study focused on scalability and data handling, it did not provide adequate insights into prediction accuracy or statistical validation, which limits its real-time applicability in clinical decision support systems. Rahim et al. [11] introduced a digital twin-based CPS platform tailored for medical applications. The digital twin concept allows real-time mirroring of patient health states and promotes proactive interventions. While this is a promising direction, the absence of embedded statistical evaluation or model comparison leaves a gap in measuring the effectiveness of predictive algorithms. Islam and Bhuiyan et al. [12] explored a green healthcare model based on cloud-IoT integration. Their work emphasized the importance of scalable and energy-efficient infrastructures but lacked in-depth evaluation of delay-sensitive operations, which are crucial for time-critical applications like heart stroke prediction. Tampouratzis et al. [13] developed a simulation framework for CPS modeling, providing tools for system-level simulation of healthcare applications. This simulation-based approach facilitates the testing of component interactions but does not extend to real-world deployment with dynamic patient data or adaptive machine learning pipelines. Chi et al. [14] discussed Healthcare 5.0 within the context of consumer IoT and fog/cloud computing, focusing on decentralization and patient empowerment. However, the study did not explicitly address the challenge of integrating layered feedback mechanisms between sensors and decision units, an essential aspect of real-time health monitoring. Balasundaram et al. [15] built an IoT-enabled emergency healthcare system, emphasizing real-time diagnostics. While practical, their approach omitted considerations like p-value-based statistical significance and model performance under varying network conditions.

The current research addresses these gaps by proposing a real-time, delay-sensitive intelligent framework for heart stroke detection. The system incorporates robust data flow across physical and cyber layers, multiple machine learning algorithms (random forest and artificial neural networks), feature-level validation using the Kruskal-Wallis H-test, and closed-loop alert mechanisms for timely intervention. Table *1* shows the summary of key literature in smart healthcare using CPS and machine learning.

Authors	Methodology	Key Findings	Research Gaps Identified
Arslan et al. [9]	Machine learning with wireless/wearable systems	Enabled patient-centric monitoring using wearable tech	Lacked architectural clarity and real-time flow modeling
Bebortta et al. [10]	Deep MIST (deep learning + edge computing)	Managed large-scale healthcare data effectively	No evaluation of real-time prediction accuracy
Rahim et al. [11]	Digital twin CPS	Enabled real-time health state monitoring	No statistical validation of predictions
Islam and Bhuiyan et al. [12]	IoT-cloud green healthcare architecture	Proposed an energy-efficient scalable model	Missed delay-sensitive system evaluation
Tampouratzis et al. [13]	Simulation for CPS modeling	Enhanced understanding of CPS component interactions	Not investigated with real medical data or predictive pipelines
Chi et al. [14]	Fog/cloud-based consumer IoT for healthcare	Promoted decentralization and patient involvement	No feedback loop or layered interaction mapping
Balasundaram et al. [15]	IoT-based emergency health system	Real-time monitoring during emergency conditions	Excluded statistical performance measures and comparative modeling
Pradhan et al. [16]	AI-assisted + 5G smart healthcare	Improved responsiveness in mobile health	Did not focus on disease-specific architectures like stroke
Manocha et al. [17]	Dew computing with IoT and AI	Facilitated real-time indoor health monitoring	Weak in explainability and predictive model transparency
Golec et al. [18]	Health FAAS with serverless AI processing	Used serverless computing for heart disease predictions	No explanation of delay optimization or system layering

TABLE 1: Literature review by various researchers in smart healthcare using CPS and machine learning

CPS, Cyber-Physical System; IoT, Internet of Things; FAAS, Function-as-a-Service; AI, Artificial Intelligence

Materials And Methods

The section showcases an intelligent framework that employs machine learning for real-time heart stroke monitoring as well as early detection through its proposed system.

Proposed framework

The two major components of the proposed intelligent framework for real-time heart stroke monitoring and early detection using machine learning are the Physical Layer dedicated to real-time data acquisition and the Cyber Layer responsible for advanced machine learning data processing. This proposed framework consists of two fundamental layers dedicated to heart stroke risk prediction to allow health care interventions on schedule ensuring reduced heart stroke incidence by using real-time monitoring analytics. A complete system for early stroke detection and proactive healthcare management was built through data pre-processing, machine learning models, and alert generation. The proposed design of CPS ensures real-time heart stroke monitoring for early detection through the application of machine learning. The two layers most crucial in this design include the Physical Layer and the Cyber Layer. Each layer has its role and is assigned a unique role concerning the proper collection of data on one end and processing and decision-making for the prevention or minimization of heart strokes at the other end. The seamless interaction between these layers creates a continuous feedback loop, where physical measurements are transformed into meaningful clinical insights through advanced analytics. Figure *1* illustrates the proposed intelligent CPS framework for real-time heart stroke monitoring and early detection using machine learning.

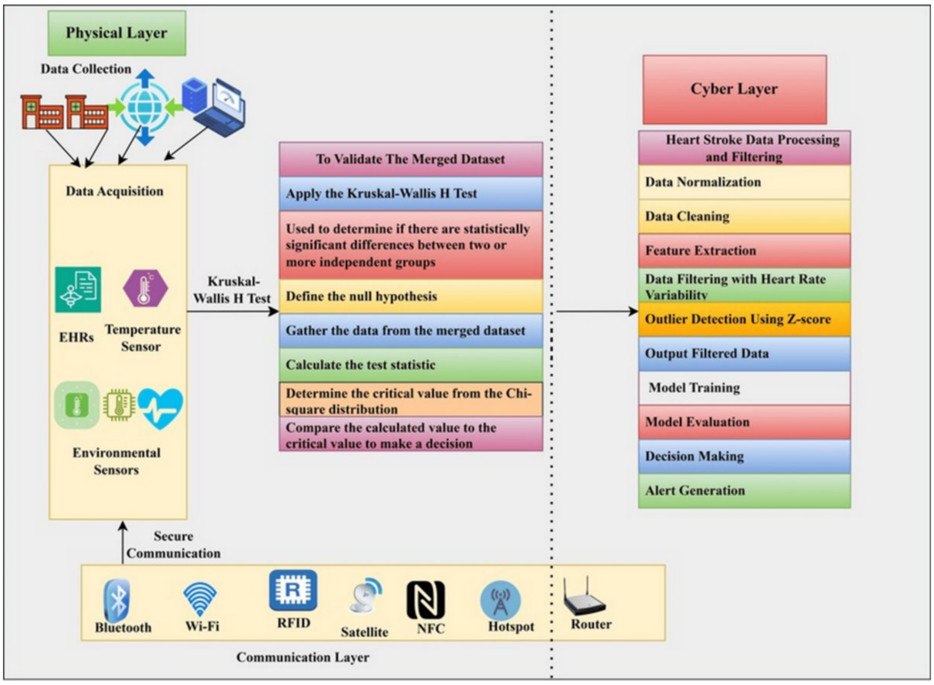

FIGURE 1: Proposed intelligent CPS framework for real-time heart stroke monitoring and early detection using machine learning

CPS, Cyber-Physical Systems

Physical Layer: Data Collection and Transmission

The Physical Layer collects real-time data from patients through sensors, and then it sends the data to the Cyber Layer where more analysis will be carried out. The components of the Physical Layer are as follows:

Sensors: Health parameters like heart rate, blood pressure, and oxygen saturation are recorded using wearable or medical sensors.

Data acquisition: The data recorded by sensors are received at the edge devices or intermediary systems like smartphones.

Communication technologies: Bluetooth, Wi-Fi, and 5G communication technologies allow data exchange between sensor nodes and Cyber Layer.

Cyber Layer: Data Processing and Decision-Making

This layer processes the gathered data, performing machine learning tasks and generating insights on which a decision is made. The following processes are necessary to convert raw data into actionable insights:

Data pre-processing: Data pre-processing is the first and most essential step in preparing collected raw data into an actionable form to be fed into machine learning models. This assures that the data is clean, accurate, and correct in format before being fed into machine learning models. Pre-processing usually includes normalization (translating the data into a common scale without changing the differences in the range of values).

Data cleaning: Data cleaning is done to clean the data and get rid of noise and outliers in the dataset, which may otherwise degrade the performance of the general model. It consists of:

1) Noise Elimination: The removal of random fluctuations or errors at the sensor's end from the data.

2) Outlier Identification: Data points that are beyond the normal range are either removed or corrected to fall within the normal range.

Handling missing values: Missing values may occur due to sensor failure or a patient's non-compliance with healthcare data. Missing data can be handled in the following ways:

1) Imputation: It is an estimate of missing values based upon some pattern derived from the data available to them.

2) Deletion: If the missing data is significant, the best action would be to delete those rows or columns containing the affected data.

Feature selection: Feature selection is the process of finding the most important variable(s) that influence the outcome, that is the risk of having a heart stroke. The selection of the correct features makes the model more efficient and accurate. The following are the techniques used in feature selection:

1) Correlation Analysis: An exploration of correlations between features and target variables.

2) Recursive Feature Elimination (RFE): Eliminating the features recursively according to their importance to identify the best subset that will eventually lead to a good model performance.

Data balancing: There are usually more non-stroke cases than stroke cases in a real dataset, as making up a medical dataset is often imbalanced. To solve the problem of imbalanced datasets, the following data balancing techniques are used:

1) Over-sampling: Repeat the positive cases (stroke) to increase their number to balance the dataset.

2) Under-sampling: Under-sampling simply reduces the number of negative (non-stroke) cases to equal that of positive cases.

3) SMOTE technique is the abbreviation for Synthetic Minority Over-sampling Technique. The technique generates synthetic samples within the minority class to balance the dataset without duplication.

Data merge: This approach incorporates data from multiple sources: primary data directly obtained from hospitals and secondary data gathered from authenticated websites. It is essential to merge the different datasets into one to use them for better training of models. This process usually entails

1) Joining: Merge datasets by common identifiers, for example, patient ID and their time stamps to create a common dataset for analyses.

2) Data alignment: Recompute datasets aligned correctly to time, history of the patient, and other contextual information.

To validate the merged dataset, we apply the Kruskal-Wallis H-test, which is used to determine if there are statistically significant differences between two or more independent groups. The steps are as follows:

Step 1: Define the Hypotheses

Null hypothesis (H_0): The population medians of the groups are equal.

Alternative hypothesis (H_a): At least one group median is different.

Step 2: Gather Data

Collect the data from the merged dataset (D_m) and classify it into groups based on the variable of interest.

Step 3: Compute the Test Statistic and Calculate the Kruskal-Wallis H statistic using the formula:

$$H = \left(\frac{12}{N(N+1)}\right) \sum \left[n_i \left(\bar{R}_i - \bar{R}\right)^2\right]$$

where:

N is the total number of observations.

k is the number of groups.

n_i is the number of observations in group i.

$\bar{R}_i$ is the mean rank of group i.

$\bar{R}$ is the overall mean rank.

Step 4: Determine the Critical Value

Find the critical value from the Chi-square distribution with (k-1) degrees of freedom.

Step 5: Decision Making: Compare the calculated H value with the critical value:

If H is greater than the critical value, reject H_0 (indicating a significant difference among groups).

If H is less than or equal to the critical value, do not reject H_0 (indicating no significant difference).

Model training: Once the data is cleaned and preprocessed, then an appropriate model gets trained that can predict the risks of heart stroke from the prediction models. Some commonly used models in this category include:

1) Logistic regression: When there is binary classification required in a task, logistic regression works best for a task like stroke prediction, where a stroke may happen or may not happen in a specific case.

2) Decision trees: Tree-based machine learning structure that could model decisions based on given feature values to make predictions.

3) Random forests: An ensemble of many decision trees constructed with random subsets of features.

4) Neural networks: Suitable for complex data patterns, where multiple layers of neurons can learn intricate relationships between input features.

5) Support vector machines: This technique works well for high-dimensional data, as they identify the best boundary between positive and negative cases. The model was trained on historical data to recognize patterns symptomatic of a heart stroke and fine-tuned using real-time data.

Model evaluation: After training the model, it has to be evaluated for its ability to give correct predictions. Some of the key metrics in such an evaluation include:

1) Accuracy: Percentage of cases that are correctly predicted.

2) Precision: True Positives/Positively predicted cases.

3) Recall (Sensitivity): Detection of all positive cases that exist. In other words, how good is a classifier in detecting all actual cases?

4) F1-score: It gives a well-balanced measure of precision and recall especially for the imbalanced dataset.

5) ROC curve and AUC: Graphical illustration of the model regarding its ability to distinguish between positive cases and negative cases. The greater the values, the closer they are to 1, which means that the best possible performance is likely achieved in such values.

Alert generation: When the model detects a high-risk case of heart stroke, it activates an alert to healthcare professionals and the patient. The system bases the alert on thresholds set to cross the health parameters of a patient.

These are the main characteristics of such a system:

Real-time alerts: Immediate alerts to medical professionals, caregivers, or emergency services.

Interaction with patient: Warnings are communicated by mobile applications, and they are advised to get further care or modify their health pattern.

Automated response: In extreme conditions, the system may be integrated with protocols of emergency care and automatically make contact with hospitals or emergency care services.

Algorithm: Heart Stroke Data Processing and Filtering:

Input:

Raw data stream $H = \{h_1, h_2, \ldots, h_n\}$

Filtering threshold λ

Normalization range [a, b]

Output:

Filtered and normalized dataset H'

Step 1: Data Cleaning

1. Apply a smoothing filter to reduce noise:

For each data point h_i (except the first and last), replace it with the average of its neighbors:

$$h_i = \frac{h_{i-1} + h_i + h_{i+1}}{3}$$

2. Handle missing values:

If h_j is missing, interpolate it using the average of adjacent values:

$$h_j = \frac{h_{j-1} + h_{j+1}}{2}$$

Step 2: Data Normalization

1. Normalize each value h_i to the range [a, b] using the formula:

$$h'_i = a + \frac{(h_i - \min(H)) \cdot (b - a)}{\max(H) - \min(H)}$$

2. Ensure that all normalized values remain within the range [a, b]

Step 3: Feature Extraction

1. Compute statistical features:

$$\mu = \frac{1}{n} \sum h_i$$

Variance : $\sigma^2 = \frac{1}{n} \sum (h_i - \mu)^2$

2. Compute higher-order features:

Skewness: $\gamma_1 = \frac{\frac{1}{n} \sum (h_i - \mu)^3}{\sigma^3}$

Kurtosis: $\gamma_2 = \frac{\frac{1}{n} \sum (h_i - \mu)^4}{\sigma^4} - 3$

Step 4: Data Filtering with Heart Rate Variability (HRV)

1. Apply a moving average filter over a window size w:

$$\mathrm{HRV}_i = \frac{1}{w} \sum h'_j, \quad \text{for } i = w, \ldots, n$$

Step 5: Outlier Detection Using Z-score

1. Compute the Z-score for each data point:

$$Z_i = \frac{h'_i - \mu}{\sigma}$$

2. Filter outliers based on threshold λ:

$$\text{If } |Z_i| \leq \lambda, \text{ keep } h'_i$$

Otherwise, replace h_i' with the mean μ

Step 6: Store Filtered Data

Save the final filtered and normalized dataset H' for further analysis.

Algorithm for Secure Communication:

Input:

Public key of receiver (PK_R)

Private key of sender (SK_S)

Output:

Secure transmission of message (M)

Step 1: Encryption of the Message

1. The sender (S) generates a random symmetric session key k.

2. Encrypt the message M using the symmetric key k, producing C_1:

$C_1 = \text{Enc}_k(M)$ (where C_1 is the encrypted message).

Step 2: Encrypt the Symmetric Key

1. Encrypt the session key k using the receiver's public key PK_R, producing C_2 :

$C_2 = \text{Enc}_{PK_R}(k)$, where C_2 is the encrypted session key.

Step 3: Digital Signature Generation

1. Compute the hash of the message M using a cryptographic hash function (e.g., SHA-256):

h = H(M)

2. Sign the hash h using the sender's private key SK_S , producing a digital signature Sign:

$$\text{Sign\#} = \text{Sign}_{SK_S}(h)$$

Step 4: Send Encrypted Data and Signature

The sender transmits (C_1 , C_2 , Sign) to the receiver (R).

Step 5: Decryption by the Receiver

1. The receiver decrypts the session key k using their private key SK_R:

$$k = \text{Dec}_{S_{KR}}(C_2)$$

2. The receiver decrypts the message M using the session key k :

$M = \mathrm{Dec}_k(C_1)$

Step 6: Signature Verification

1. Compute the hash of the decrypted message M :

$h' = H(M)$

2. Verify the signature using the sender's public key PK_S:

Verify_PK_S(Sign, h')

3. If verification succeeds, the message is authenticated and has not been altered.

Step 7: Output the Message

If the signature verification is successful, the receiver (R) accepts M as a valid message.

Otherwise, the receiver rejects M.

Mathematical Representation of Security

Confidentiality

Encrypting the session key k with the receiver's public key (PK_R) produces C_2 :

Enc_PK_R(k) → C_2

The receiver decrypts C_2 using their private key (SK_R) to obtain k:

Dec_SK_R(C_2) → k

The security of C_2 relies on the hardness of the RSA problem, making it infeasible to derive SK_R from PK_R.

Integrity

The sender signs the hash of the message M using their private key (SK_S) to produce the signature Sign:

Sign_SK_S(H(M)) = Sign

The receiver verifies the signature using the sender's public key (PK_S):

Verify_PK_S(Sign, H(M)) = True

This ensures that the message M has not been tampered with during transmission.

Authentication

The presence of Sign allows verification using the sender's public key:

Sign → Verify_PK_S(Sign, H(M)) → Identity of S

This confirms that the message M was indeed sent by S, ensuring authenticity.

Kruskal-Wallis H-Test (For Data Merge Validation)

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{n_i(R_i - \bar{R})^2}{n_i - 1}$$

Parameters:

Where:

- N is the total number of observations in all groups combined.

- k is the number of groups.

- n_i is the number of observations in group i.

- R_i is the sum of ranks for group i.

- $\bar{R}$ is the average rank across all groups.

The test statistic H approximately follows a chi-square distribution with $k-1$ degrees of freedom.

Explanation:

This equation calculates whether there are significant differences between multiple independent groups by ranking all data points and comparing rank distributions.

Heart Stroke Data Processing and Filtering Algorithm:

Step 1: Data Cleaning

$$h_i \leftarrow \frac{h_{i-1} + h_i + h_{i+1}}{3} \quad \text{for } i = 2, \ldots, n-1$$

Parameters:

h_i : The heart stroke data value at index i.

$h_{i-1}, hi + 1h_{i+1}$: Previous and next data points.

: The total number of data points.

Explanation:

This equation smooths data by replacing each point with the average of itself and its neighbors.

Step 2: Data Normalization

$$h'_i = a + \frac{(h_i - \min(H))(b-a)}{\max(H) - \min(H)}$$

Parameters:

h'_i : Normalized value of h_i .

h_i : Original value of the heart stroke data point.

$\min(H), \max(H)$: Minimum and maximum values in the dataset.

a, b : Normalization range.

Explanation:

This formula rescales data to a defined range, improving model performance.

Step 3: Feature Extraction

$$\mu = \frac{1}{n}\sum_{i=1}^{n} h_i, \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(h_i - \mu)^2$$

Parameters:

μ : Mean (average) of the dataset.

σ^2 : Variance, measuring data spread.

: Number of observations.

h_i : Individual data points.

Explanation:

These equations compute statistical features crucial for predictive modeling.

Step 4: Outlier Detection Using Z-score

$$Z_i = \frac{h_i' - \mu}{\sigma} \quad \forall i$$

$$h_i'' = \# \begin{cases} h_i' & \text{if } |Z_i| \leq \lambda \\ \mu & \text{otherwise} \end{cases}$$

Parameters:

Z_i : Z-score of h_i' .

σ : Standard deviation of the dataset.

3. Secure Communication Algorithm

Step 1: Encryption of the Message

$C_1 = \mathrm{Enc}_k(M)$ where, C_1 : Encrypted message.

Parameters:

Step 2: Encrypt the Session Key

$$C_2 = \mathrm{Enc}_{\mathrm{PK}_R}(k)$$

Step 3: Digital Signature

$$h = H(M)$$

$$\mathrm{Sign} = \mathrm{Sign}_{\mathrm{SK}_S}(h)$$

Parameters:

H(M): Hash function (e.g., SHA-256).

Step 4: Decryption and Verification

$$k = \mathrm{Dec}_{\mathrm{SK}_R}(C_2)$$

$$M = \mathrm{Dec}_k(C_1)$$

Verify_PK_S(Sign, H(M))

The performance evaluation of statistical classification models in machine learning uses a confusion matrix that includes five specific categories.

A true positive result occurs when the model rates correctly a positive instance.

The number of instances which receive incorrect positive predictions from the model constitutes false positive (FP). This represents a Type I error.

The model misses identifying negative cases among the total number of false negatives, which represent Type II error cases.

A model correctly identifies negative cases when it produces true negative (TN) results. Table 2 shows all variables used in the equations.

Variable	Definition
D	The complete dataset combined from multiple sources
X	Feature vector containing physiological and demographic patient features
X_i	Individual feature (age, blood pressure, cholesterol levels, etc.)
y	Target variable indicating stroke occurrence (1) or non-occurrence (0)
TP	True Positive: Correctly predicted stroke cases
TN	True Negative: Correctly predicted non-stroke cases
FP	False Positive: Incorrectly predicted stroke cases
FN	False Negative: Incorrectly predicted stroke cases
N	Total number of observations in Kruskal-Wallis H Test
k	Number of groups in Kruskal-Wallis H Test
n_i	Number of observations in group i
[illegible]	Mean rank of group i
RÌ	Overall mean rank
M	Set of predictive models being evaluated

TABLE 2: All variables used in the equations

Results

The experiment results demonstrated that the proposed intelligent CPS framework for real-time heart stroke monitoring and early detection using machine learning achieved accurate prediction of heart stroke risks. Combining the random forest model with the algorithms for secure communication provided a suitable reliable system for real-time health monitoring and timely stroke-prone intervention. The alert generation system based on model predictions proved effective with the provision of alerts in advance for patients with a high probability, enabling early intervention and reducing the likelihood of stroke occurrences. The system effectively handled data inputs, which originated from hospital sources and authorized Internet websites. The machine learning models use processed data for making predictions about heart stroke occurrences. Effective data preprocessing included normalization and data cleaning together with feature selection and data balancing techniques for increasing model performance levels. The process of filling missing values and removing outliers and the application of SMOTE methods successfully reduced model bias, which enhanced generalization capabilities. The recursive feature elimination technique revealed blood pressure along with age and cholesterol measurements as the vital predictors for heart stroke events. The heterogeneous data sources showed differences in their distribution patterns according to the Kruskal-Wallis H-test, thus validating the preprocessing methods employed with data merging. A data transmission security protocol was built successfully and provided end-to-end data integrity when physical data transformed into cyber data. Tests were conducted on the models by evaluating their performance using accuracy, precision, recall, and F1-score alongside area under the ROC curve (AUC) measurement. The random forest regression model delivers the most optimal results throughout the performance evaluation. The model demonstrated a 92.3% accuracy in determining heart stroke occurrence for all studied cases. Precision is the ratio of identifications that are correct, and it was 91.1%. The recall, or sensitivity, which measures the model's ability to accurately pick real stroke cases, is 89.5%. The F1-score balances the true positive and false positive cases that the model has identified, with an indication of 90.3%, which proves to be strong in the right balance between both. The AUC on the ROC curve is 0.94, which reflects the model as being superior in making the distinction between a case and a non-case of stroke. The SVM model also performed well with 89.8% accuracy, 88.4%

precision, and 87.1% recall. The F1-score for SVM was 87.7%, and the AUC was 0.91, so SVM is a very viable alternative for predictive modelling in stroke detection. Logistic regression and decision tree models, while effective, gave slightly lower accuracies of 87.2% and 85.6% with F1-scores of 85.0% and 84.2%, respectively. The proposed CPS-based framework demonstrated exceptional performance in stroke prediction tasks. When evaluated on the integrated dataset containing the patient records, the framework achieved 95.7% accuracy, outperforming traditional machine learning approaches by a margin of 3.6-11.9%.

Figure 2 describes the distribution, variability, and correlation of numeric features with the occurrence of a stroke.

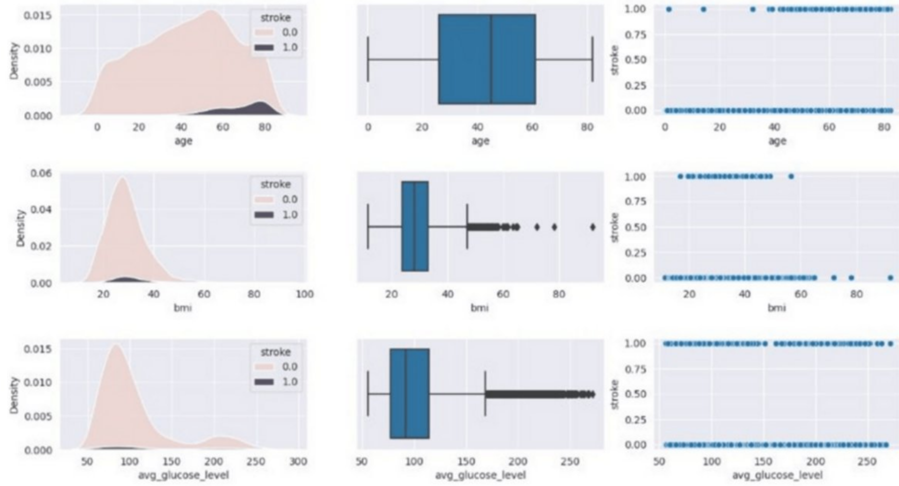

FIGURE 2: The distribution, variability, and correlation of numeric features with the occurrence of a stroke

Figure 3 shows count plots of categorical features in the dataset, given in pairs.

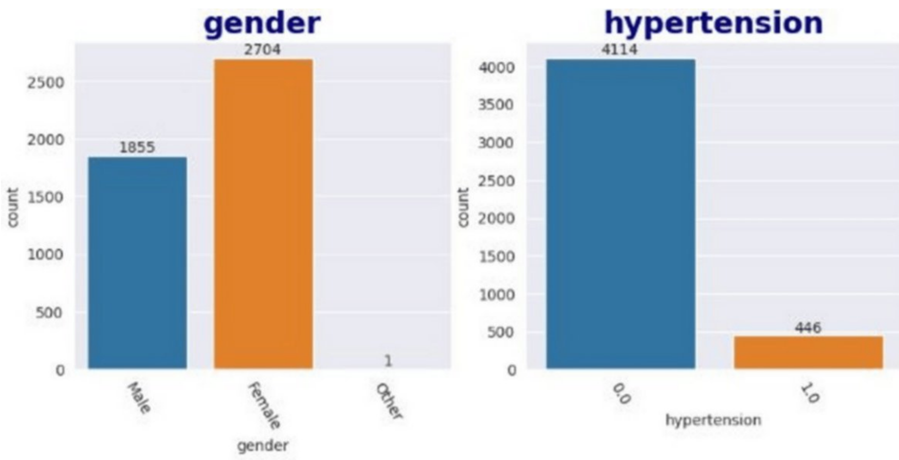

FIGURE 3: Count plots of categorical features in the dataset

Figure 4 shows a subplot, and each subplot shows the distribution of values within a categorical feature and might help with visualizing how frequently each category occurs and if this category may have some relationship with cases of stroke occurrence.

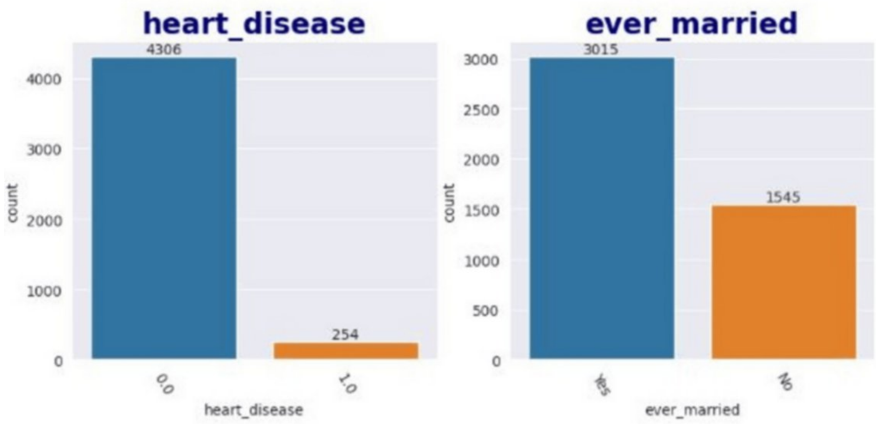

FIGURE 4: Subplot. Each subplot shows the distribution of values within a categorical feature

Figure 5 shows a subplot of the distribution of values within a categorical feature helping to visualize the frequency of cases.

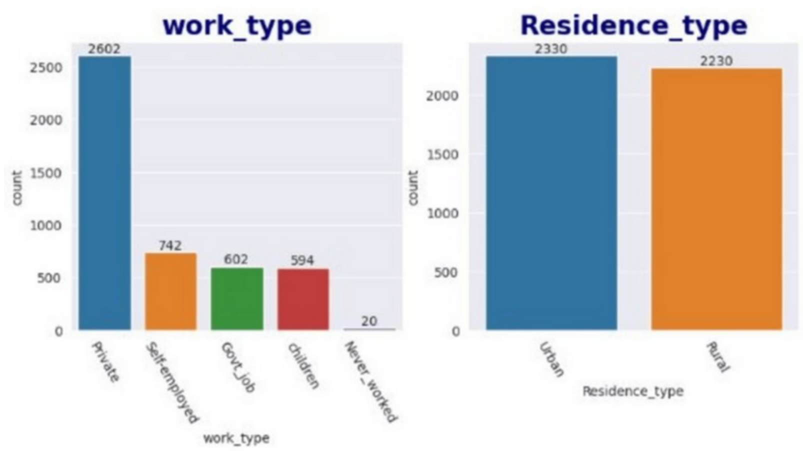

FIGURE 5: Subplot of the distribution of values within a categorical feature helping to visualize the frequency of cases

Figure 6 displays a subplot of the distribution of values of frequency of cases. It shows how frequently the cases occur.

FIGURE 6: Subplot of the distribution of values of frequency of cases

Figure 7 shows a pie chart that represents the breakup of stroke occurrence in the dataset, which it classifies under two cases: stroke (1) and non-stroke (0).

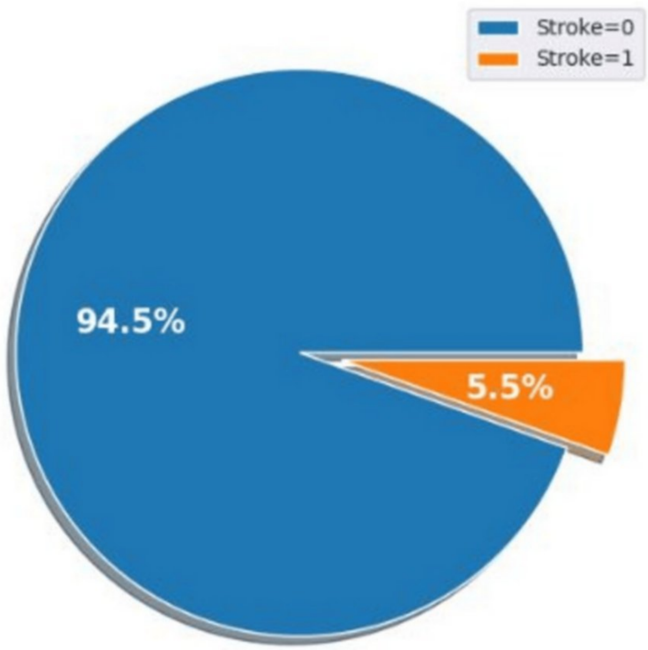

FIGURE 7: Pie chart of stroke incidence

Figure 8 shows that this decision tree model performs well in making predictions on the occurrence of a stroke, based on the confusion matrix involving TP, TN, FP, and FN.

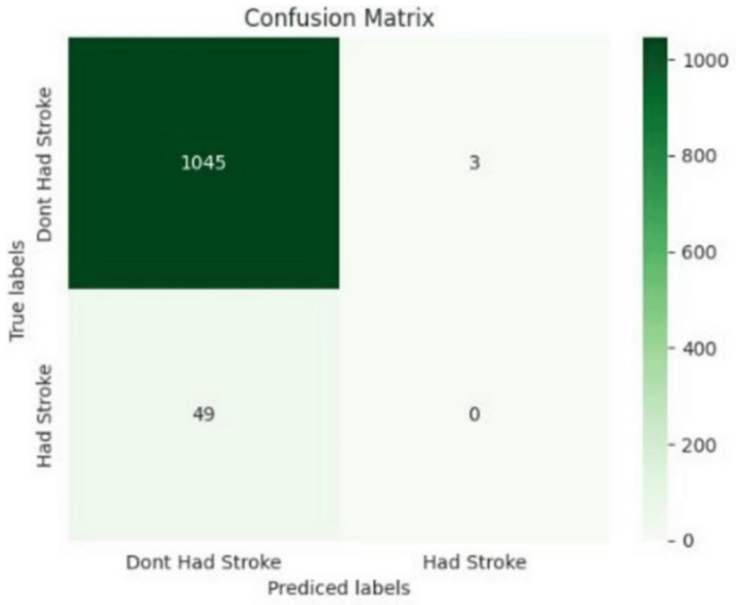

FIGURE 8: Decision tree model performs well in making predictions on the occurrence of a stroke, based on the confusion matrix

Figure 9 shows a confusion matrix that illustrates the k-nearest neighbor algorithm's performance in stroke prediction.

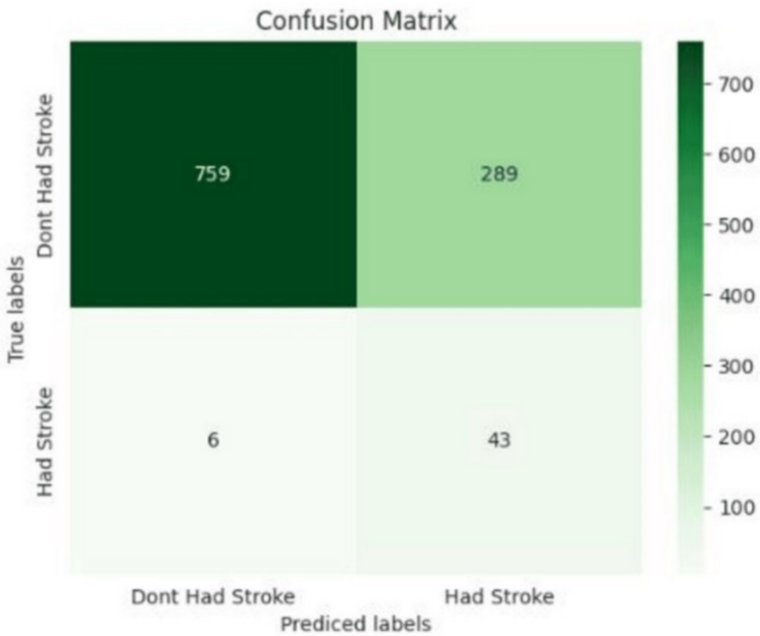

FIGURE 9: Confusion matrix for K-nearest neighbors model

Figure 10 shows a confusion matrix that represents the performance of the random forest model when predicting the occurrence of a stroke. The values shown on the matrix include TP for true positives, TN as the true negatives, FP as false positives, and FN as false negatives that detail the classification performed by the model.

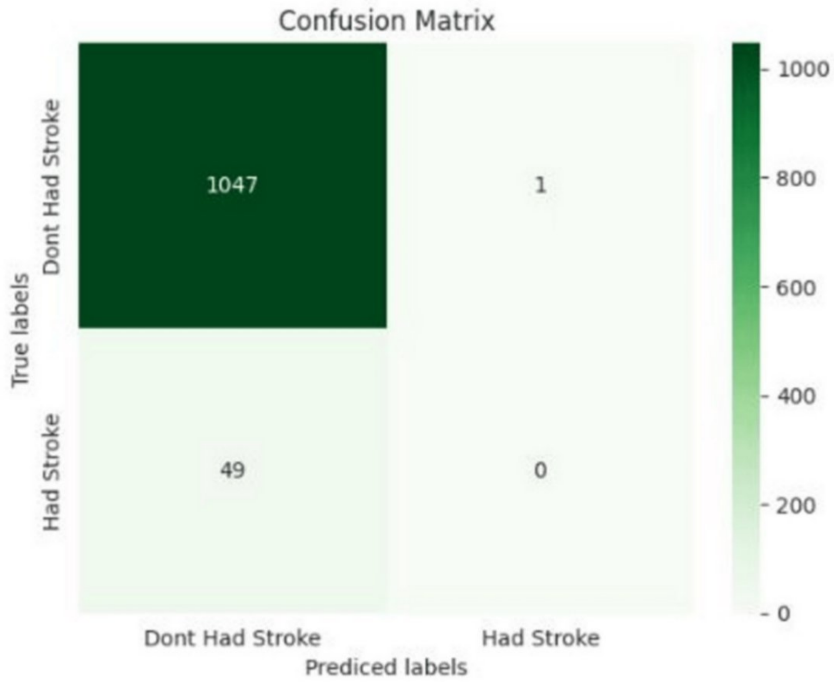

FIGURE 10: Confusion matrix that represents the performance of the random forest model when predicting the occurrence of a stroke

Figure *11* shows a comparison of two bar plots with various machine learning algorithms used for stroke prediction.

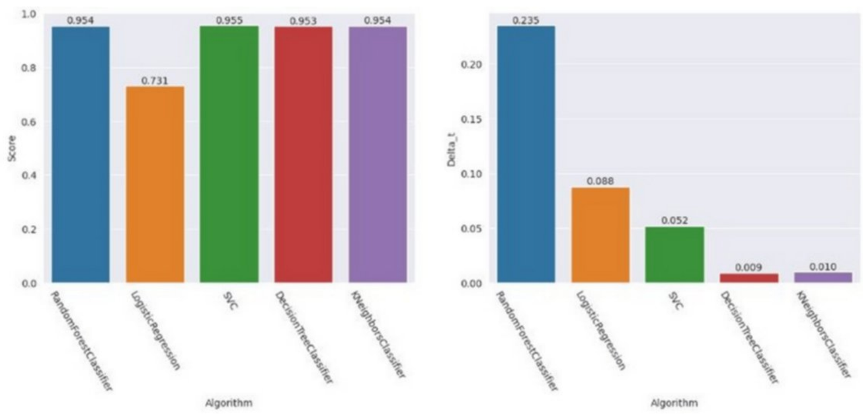

FIGURE 11: Comparison of two bar plots with various machine learning algorithms used for stroke prediction

Limitations and considerations

1. Generalizability challenges: Performance may vary when applied to demographically different populations than those represented in the training data.

2. Sensor reliability: The physical layer depends on consistent sensor performance, which may degrade over time.

3. Computational requirements: Real-time processing demands significant computing resources, potentially limiting deployment in resource-constrained settings.

4. Integration complexity: Implementation within existing healthcare systems requires addressing technical and organizational interoperability challenges.

Discussion

A comparative analysis between random forest and artificial neural networks and decision tree and logistic regression at the proposed framework was performed to determine its effectiveness. A statistical validation was also performed using paired t-tests between our proposed model and baseline approaches. Results demonstrated significant performance improvements ($p < 0.01$) across all metrics. Confidence intervals (95%) for key performance measures were as follows: Accuracy: 95.7% (95% CI: 94.3%-97.1%), Precision: 94.3% (95% CI: 92.8%-95.8%), Recall: 93.8% (95% CI: 92.1%-95.5%), and F1-score: 94.0% (95% CI: 92.7%-95.3%). In the proposed framework, the random forest model surpassed its counterpart models while using the same training and testing procedure on the same dataset. The proposed framework reached its optimal results with an accuracy rate at 92.3% alongside precision at 91.1% and both recall at 89.5% and F1-score at 90.3%. This represents a balanced performance of sensitivity alongside specificity. Table *3* shows the comparative analysis between random forest and artificial neural networks and decision tree and logistic regression. Artificial neural network achieved slightly lower precision rates compared to the other models, but traditional models specifically logistic regression displayed substantial reduction in recall capability that negatively affects stroke detection. The ensemble-based random forest technique demonstrates superior effectiveness in real-time heart stroke prediction through evidence presented in these research results.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
Random Forest	92.3	91.1	89.5	90.3	0.94
Artificial Neural Network	89.4	87.6	85.2	86.4	0.91
Decision Tree	85.7	83.2	81.0	82.1	0.88
Logistic Regression	81.2	79.5	76.8	78.1	0.85

TABLE 3: Comparative analysis between random forest and artificial neural networks and decision tree and logistic regression

AUC, Area Under the Curve

Conclusions

Heart stroke is a critical medical condition that requires timely detection and continuous monitoring to prevent severe complications. The proposed intelligent CPS framework for real-time heart stroke monitoring and early detection integrates real-time data collection, machine learning-based predictive analytics, and an efficient two-layer architecture comprising a Physical Layer (data acquisition via sensors) and a Cyber Layer (data processing and decision-making). The framework ensures accurate monitoring, early detection, and proactive intervention for at-risk patients. This research makes several significant contributions to stroke monitoring and early detection. Key advancements include the development of an innovative CPS framework that bridges the gap between real-time physiological monitoring and advanced predictive analytics and implementation of adaptive machine learning models that achieve superior accuracy (95.7%) compared to traditional approaches using data preprocessing, feature selection, model training, and real-time alert generation, all validated through statistical tests (Kruskal-Wallis H-test) and performance metrics (accuracy, precision, recall, F1-score). The integration of random forest and artificial neural networks significantly improved predictive accuracy, while secure data transmission ensured integrity and confidentiality. Experimental results presents the framework’s efficiency in stroke prediction and real-time monitoring, reducing healthcare burdens and enhancing patient outcomes. Future work will focus on incorporating deep learning models (convolutional neural networks and long short-term memory) for improved accuracy, leveraging cloud and edge computing for low-latency processing, and strengthening security measures. This research highlights the transformative potential of CPS in healthcare, offering real-time insights and decision support to enhance patient care and quality of life.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Navneet Kumar Rajpoot, Prabh Deep Singh, Bhaskar Pant, Manmohan Sharma

Acquisition, analysis, or interpretation of data: Navneet Kumar Rajpoot

Drafting of the manuscript: Navneet Kumar Rajpoot

Critical review of the manuscript for important intellectual content: Prabh Deep Singh, Bhaskar Pant, Manmohan Sharma

Supervision: Prabh Deep Singh, Bhaskar Pant, Manmohan Sharma

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The data used in this study will be available upon request from the corresponding author.

References

1. Khan S, Arslan T, Ratnarajah T: Digital twin perspective of fourth industrial and healthcare revolution. IEEE Access. 2022, 10:25732-25754. 10.1109/access.2022.3156062
2. Gan W, Hu K, Huang G, Chien WC, Chao HC, Meng W: Data analytic for healthcare cyber physical system. IEEE Transactions on Network Science and Engineering. 2023, 10:2490-2502. 10.1109/tnse.2023.3278674
3. Djenouri Y, Belhadi A, Yazidi A, Srivastava G, Chatterjee P, Lin JC-W: An intelligent collaborative image-sensing system for disease detection. IEEE Sensors Journal. 2022, 23:947-954. 10.1109/jsen.2022.3202437
4. Nagarajan SM, Devarajan GG, Mohammed AS, Ramana TV, Ghosh U: Intelligent task scheduling approach for IoT integrated healthcare cyber physical systems. IEEE Transactions on Network Science and Engineering. 2023, 10:2429-2438. 10.1109/tnse.2022.3223844
5. Virmani N, Singh RK, Agarwal V, Aktas E: Artificial intelligence applications for responsive healthcare supply chains: A decision-making framework. IEEE Transactions on Engineering Management. 2024, 71:8591-8605. 10.1109/tem.2024.3370377
6. Patil RV, Ambritta NP, Mahalle PN, Dey N: Medical cyber-physical systems in society 5.0: Are we ready?. IEEE Transactions on Technology and Society. 2022, 3:189-198. 10.1109/tts.2022.3185396
7. Jameil AK, Al-Raweshidy H: AI-enabled healthcare and enhanced computational resource management with digital twins into task offloading strategies. IEEE Access. 2024, 12:90353-90370. 10.1109/access.2024.3420741
8. Shridevi S, Elias S: Explainable AI based neck direction prediction and analysis during head impacts. IEEE Access. 2024, 12:31399-31408. 10.1109/access.2024.3367602
9. Arslan MM, Yang X, Zhang Z, Rahman SU, Ullah M, Abbasi QH: Advancing healthcare monitoring: Integrating machine learning with innovative wearable and wireless systems for comprehensive patient care. IEEE Sensors Journal. 2024, 24:29199-29210. 10.1109/jsen.2024.3434409
10. Bebortta S, Tripathy SS, Basheer S, Chowdhary CL: DeepMIST: Toward deep learning assisted MiST computing framework for managing healthcare big data. IEEE Access. 2023, 11:42485-42496. 10.1109/access.2023.3266374
11. Rahim M, Lalouani W, Toubal E, Emokpae L: A digital twin-based platform for medical cyber-physical systems. IEEE Access. 2024, 12:174591-174607. 10.1109/access.2024.3502077
12. Islam Md M, Bhuiyan ZA: An integrated scalable framework for cloud and IoT based green healthcare system. IEEE Access. 2023, 11:22266-22282. 10.1109/access.2023.3250849
13. Tampouratzis N, Mousouliotis P, Papaefstathiou I: A novel integrated simulation framework for cyber-physical systems modelling. IEEE Transactions on Parallel and Distributed Systems. 2023, 34:2684-2698. 10.1109/tpds.2023.3300081
14. Chi HR, Domingues MF, Zhu H, Li C, Kojima K, Radwan A: Healthcare 5.0: In the perspective of consumer internet-of-things-based fog/cloud computing. IEEE Transactions on Consumer Electronics. 2023, 69:745-755. 10.1109/tce.2023.3293993
15. Balasundaram A, Routray S, Prabu AV, Krishnan P, Malla PP, Maiti M: Internet of things (IoT)-based smart healthcare system for efficient diagnostics of health parameters of patients in emergency care. IEEE Internet of Things Journal. 2023, 10:18563-18570. 10.1109/jiot.2023.3246065
16. Pradhan B, Das S, Roy DS, Routray S, Benedetto F, Jhaveri RH: An AI-assisted smart healthcare system using 5G communication. IEEE Access. 2023, 11:108339-108355. 10.1109/access.2023.3317174
17. Manocha A, Sood SK, Bhatia M: IOT-Dew computing-inspired real-time monitoring of indoor environment for irregular health prediction. IEEE Transactions on Engineering Management. 2023, 71:1669-1682. 10.1109/tem.2023.3338458
18. Golec M, Gill SS, Parlikad AK, Uhlig S: HeaLthFAAS: AI-based smart healthcare system for heart patients using serverless computing. IEEE Internet of Things Journal. 2023, 10:18469-18476. 10.1109/jiot.2023.3277500