Heart Attacks and Brain Strokes in Early, Mid and Late Age: Study From Data Science Perspective

Ankur Singh Bist[1*], Untung Rahardja[2], Ninda Lutfiani[3], Rifqa Nabila Muti[4], Santiago Ramirez[5]
[1]Dept. of Computer Science, Graphic Era Hill University, India
[2]Indonesian Lecturers Association Banten, Indonesia
[3]Pandawan Group, Indonesia
[4]Faculty of Economics and Business, University of Raharja, Indonesia
[5]iLearning Incorporation, United States
[1]ankur1990bist@gmail.com, [2]untung@raharja.info, [3]ninda@pandawan.id, [4]rifqa@raharja.info, [5]tiago.ramirez@ilearning.co
***Corresponding Author**

Article Info

Article history:

Submission January 29, 2026
Revised March 13, 2026
Accepted April 3, 2026
Published May 28, 2026

Keywords:

Data Science
Heart Attack
Brain Stroke
Mid Age
Early Age

ABSTRACT

Heart attacks and brain strokes have become major global health concerns due to the increasing number of cases occurring among not only older adults but also middle-aged and younger populations. Unhealthy lifestyles, smoking habits, obesity, hypertension, and diabetes have significantly contributed to the rising risk of cardiovascular and neurological diseases. This study aimed to investigate heart attack and brain stroke risks across different age groups from a data science perspective while evaluating the effectiveness of machine learning techniques for disease prediction. The research employed quantitative experimental analysis using healthcare datasets containing clinical and lifestyle-related variables such as age, blood pressure, cholesterol level, glucose level, smoking behavior, body mass index, and diabetes condition. Two supervised machine learning algorithms, namely Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost), were implemented to classify disease risks. The experimental results showed that the SVM model achieved superior predictive performance with an accuracy of 99.5%, while XGBoost achieved 94.3% accuracy. The novelty of this study lies in integrating multi-age cardiovascular and neurological risk analysis with artificial intelligence-based predictive modeling to support the development of an intelligent mobile healthcare alert system. The study concluded that machine learning technologies have strong potential to improve early disease prediction, preventive healthcare, and proactive patient monitoring in modern healthcare environments.

DOI: https://doi.org/10.34306/heal.v1i2.105

1. INTRODUCTION

In earlier times, heart attacks and brain strokes did not cause many deaths due to the good lifestyle of people. As people have adopted a sedentary lifestyle and abstain from performing physical activity, they prefer to eat unhealthy food, due to which more and more people are becoming victims of heart attacks and brain strokes. Stroke is a neurological disorder characterized by blockage of blood vessels [1]. In particular, brain strokes occur due to blockage of blood vessels, which hampers the flow of blood to different parts of the body. According to a study, brain stroke is one of the leading causes of death globally [1]. Another common type of disease is a heart attack, which is prevalent among people today. Among other factors, age also plays a critical role n the deterioration of cardiovascular functionality, resulting in an increased risk of cardiovascular disease (CVD) in older adults [2, 3].

This study focuses on identifying the reasons for the transition of cardiovascular disease from older to middle and late ages from a data science perspective. Using data science, one can employ various machine learning algorithms, such as K-nearest neighbour, logistic regression, random forest (RF) classifier, and support vector classifier, to predict diseases in a timely manner so that preventive measures can be taken promptly [4].

The data usually required for predicting said disease are clinical data, ECGs, and lab findings, which must be sufficiently large to enable accurate predictions. Our study suggests that the required algorithms are complex and computationally demanding. From the existing literature, it is found that Convolutional Neural Networks (CNNs) typically identify the correct set from a group of about 10 sets and are ideally suited to medical image processing [5]. However, clinical symptoms are text-based, and these discrete sets can run into the 100s. Therefore, the diagnosis requires filters such as Bayesian filters, along with word2vec distance as priors.

This paper is structured as follows: Firstly, a general discussion on heart attacks and brain strokes is provided. Then we provide a statistic on how the disease transitions from old age to middle age and even to early age. Data is collected and analyzed to learn the pattern for this shift. Next, a list of symptoms is presented, based on data, to develop an AI-based alert system that enables a patient to learn about the disease before it becomes severe [6]. Finally, the conclusion is presented and discussed.

2. HEART RATE AND BRAIN STROKES ANALYSIS USING MACHINE LEARNING

This section reviews recent work on the application of machine learning algorithms to the prediction of heart attack and brain stroke [7, 8]. Analyze existing machine learning algorithms for their effectiveness and accuracy [9, 10], and address the gaps with our proposed methodology. Here summarize and analyze the most relevant.

In this work [11], the authors have proposed a methodology called retrospective study using a prospective cohort that enrolled patients with acute ischemic stroke. In the process, three models were developed, such as deep neural networks, random forest and logistic regression and their predictability was compared to find the best model for predicting stroke. To assess model accuracy, the results were compared with the standard score, namely the Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score. However, argue that to achieve favourable results, one must use a sufficient number of variables with machine learning algorithms. In this work, the authors used only six variables for the ASTRAL score, which is why the performance of machine learning models did not differ significantly from that of the ASTRAL score.

In yet another study [12], the authors used various ML algorithms, including Deep Learning (DL), Support Vector Machine (SVM), Random Forest (RF), XGBoost (XGB), and conventional LR models, to evaluate the reliability and clinical utility of machine learning in predicting stroke prognosis. Calibration was evaluated using a reliability diagram and Expected Calibration Error (ECE) to assess the reliability of estimates of the relationship between predicted and actual outcomes.

In this work [13], 10 different machine learning classification algorithms, namely LR, DT, NB, RF, ANN, KNN, GB, SVM, AB, and ET, are implemented to select the best model for early and accurate detection of heart disease [14, 15]. To assess the performance of classification algorithms, various evaluation metrics were used, including accuracy, sensitivity, specificity, AUC, F1-score, MCC, and ROC curve. However, further optimization can be explored to achieve more accurate results.

In this approach [16, 17], DWI and FLAIR images from consecutive patients with acute ischemic stroke were analyzed using an automated image-processing method. The approach uses three machine learning models named support vector machine, random forest and logistic regression for binary classification. To evaluate the performance of ML models, the sensitivity and specificity for identifying patients within 4.5 hours were compared with those of human readings of DWI-FLAIR mismatch. However, the proposed method has a few limitations, such as relatively small patient populations and the lack of an appropriate sample size.

According to medical professionals, abnormal bio signals are also considered to be a major indication of potential stroke. In [18], the authors proposed a system based on ML and deep learning algorithms that measures and accurately assesses biosignals in real time, helping patients receive appropriate treatment quickly.

From the existing literature, it was found that there is no ML-based alert system that covers all age groups and predicts heart attack and brain stroke using machine learning algorithms [19–21]. Our proposed AI-based alert system uses support vector machine and XGBoost models to alert a patient to a heart attack or a brain stroke in a timely manner [22], so that appropriate medication and steps can be taken.

3. METHODOLOGY

This study employed a quantitative data science approach to analyze the occurrence of heart attacks and brain strokes across early, middle, and late age groups. The research focused on identifying major health-related factors contributing to cardiovascular and neurological diseases while evaluating the effectiveness of machine learning algorithms in disease prediction [23]. Secondary datasets were utilized from publicly available medical repositories related to heart disease and brain stroke prediction [24]. The heart disease dataset consisted of several clinical and lifestyle-related variables, including sex, age, smoking habits, cigarettes per day, blood pressure, diabetes, hypertension, cholesterol level, body mass index (BMI), and glucose level. Meanwhile, the brain stroke dataset included variables such as age, NIHSS, mRS, systolic blood pressure, diastolic blood pressure, glucose level, paralysis condition, smoking status, BMI, and cholesterol level.

The collected datasets were analyzed using supervised machine learning techniques, namely Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) classifiers. Before model implementation, the datasets were processed through data cleaning and feature preparation stages to ensure consistency and improve predictive performance. SVM was selected due to its strong capability in handling classification problems with high-dimensional medical data, while XGBoost was utilized because of its efficiency and robustness in predictive analytics. The experimental process aimed to compare the predictive performance of both algorithms in detecting the risk of heart attack and brain stroke occurrences among different age groups.

The analysis results were further utilized to support the proposed development of an AI-based alert system and mobile healthcare application intended to assist early detection and preventive healthcare management for heart attacks and brain strokes.

4. EXPERIMENTAL ANALYSIS

The role of technology in health care can't be underestimated [25]. Machine learning-based techniques are widely used for data analysis. Using support vector machine and XGBoost classifier to study heart stroke and brain stroke. There are various datasets available for this analysis. An early heart disease detection dataset based on various features is available here: GitHub Heart-Disease-Risk-Prediction. There are following features used to analyze heart attack.

1. Sex
2. Age
3. Current Smoker
4. Cigarettes per Day
5. Blood pressure
6. Diabetes
7. Hypertension
8. Cholesterol Level
9. Body Mass Index
10. Glucose Level

Brain stroke is even more serious cause of deaths across the world. As heart failure prediction as prevention, our motive is to analyze brain stroke too. Data link for brain stroke prediction is as follows: Dataset Stroke Analysis. There are various features taken for brain stroke analysis.

1. Age

2. NIHSS
3. mRS
4. Systolic Blood pressure
5. Diastolic Blood pressure
6. Glucose
7. Paralysis
8. Smoking
9. BMI
10. Cholesterol

These variables were selected because they represent major cardiovascular risk factors commonly associated with heart disease occurrence. Meanwhile, the brain stroke dataset consisted of variables such as age, NIHSS score, mRS score, systolic blood pressure, diastolic blood pressure, glucose level, paralysis condition, smoking status, BMI, and cholesterol level. The inclusion of these variables enabled the models to learn complex relationships between patient health conditions and the probability of stroke occurrence.

Before model implementation, the datasets underwent preprocessing procedures to improve data quality and ensure reliable experimental outcomes. The preprocessing stage involved handling missing values, normalizing numerical attributes, organizing categorical data, and selecting relevant features to reduce data inconsistency and noise. Feature preparation is considered essential in healthcare machine learning because medical datasets often contain heterogeneous and highly sensitive clinical information. After preprocessing, the datasets were divided into training and testing sets to evaluate the predictive capability of the proposed algorithms objectively.

Support vector machine and XGBoost classifier are used to make the analysis. Figure 1 and Figure 2 shows the general architecture of Support vector machine and XGBoost classifier. We achieved 99.5% and 94.3% accuracy on above mentioned datasets using SVM and XGBoost respectively. SVM performs better than XGBoost.

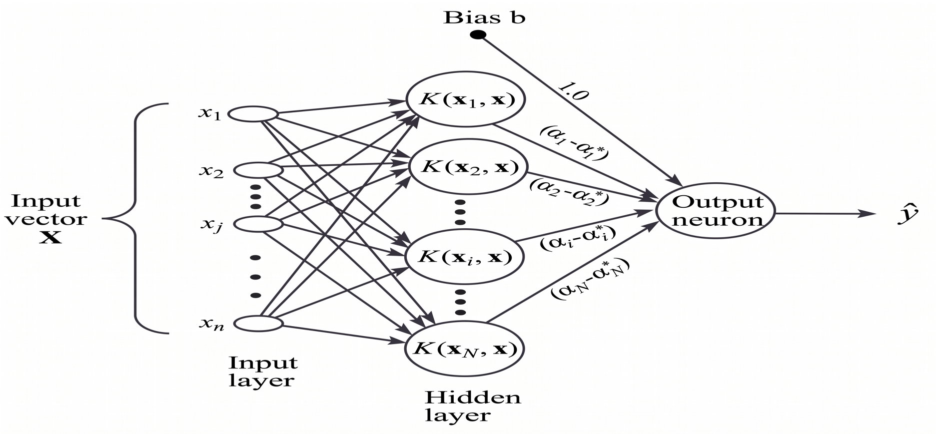

Figure 1. Support Vector Machine Architecture

The network topology illustrated in figure 1 represents a structural diagram of a Support Vector Machine (SVM) configuration tailored for non-linear regression, commonly referred to as Support Vector Regression (SVR), which shares an architectural paradigm with Radial Basis Function (RBF) networks. The network comprises three fundamental layers: the input layer, the hidden layer, and a singular output neuron. The input vector $\mathbf{X}$, consisting of elements ranging from x_1 through x_n, is fed directly into the input layer without any weight transformations. These inputs are subsequently mapped onto a high-dimensional feature space within the hidden layer via a series of kernel functions, denoted as $K(\mathbf{x}_i, \mathbf{x})$. This kernel mapping enables the network to effectively handle non-linear decision boundaries by computing the inner products or similarity measures between the input vector and the predetermined support vectors.

In the subsequent feedforward phase, the outputs generated by the hidden nodes are modulated by the network weights before converging at the output layer. Specifically, these connection weights are defined by the Lagrange multipliers $(\alpha_i - \alpha_i^*)$, which represent the dual parameters optimized during the training phase to establish the regression hyperplane. Concurrently, a structural bias b, scaled by a constant weight of 1.0, is directly integrated into the output neuron to shift the regression function appropriately. Ultimately, the output neuron aggregates the weighted linear combination of the kernelized hidden outputs along with the bias term, yielding the final predicted estimate $\widehat{y}$. This systematic integration of localized kernel transformations and global linear aggregation forms the mathematical backbone essential for robust generalization in complex predictive modeling tasks.

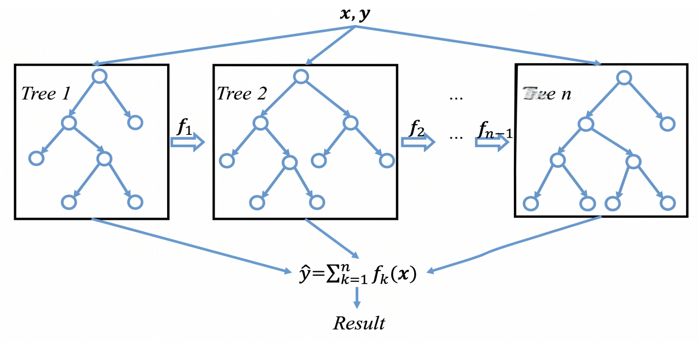

Figure 2. XGBoost Architecture

The provided diagram illustrates the architecture of a Gradient Boosting Decision Tree (GBDT) ensemble method, a powerful machine learning framework widely used for regression and classification tasks. The model processes the input feature vector and corresponding target variables, denoted as (x, y), by sequentially constructing a series of base learners represented as *Tree 1*, *Tree 2*, through *Tree n*. Unlike traditional bagging approaches that build trees independently, this sequential boosting mechanism fits each subsequent tree to the residuals or pseudo-residuals left unresolved by the preceding iterations.

The transition functions $f_1, f_2, \ldots, f_{n-1}$ represent the incremental optimization steps where each new tree aims to minimize the overall loss function relative to the prior ensemble's predictions.

Ultimately, the final predictive model aggregates the outputs from all individual decision trees to generate the definitive outcome. As mathematically formulated in the diagram, the final prediction $\hat{y}$ is obtained through the additive summation of the functions of the input data across all n trees, expressed as:

$$\hat{y} = \sum_{k=1}^{n} f_k(x) \tag{1}$$

This collective aggregation effectively combines multiple weak learners into a single, highly robust predictive model. The final *Result* yields a significantly optimized prediction that minimizes variance and bias, making this framework highly effective for handling complex, non-linear datasets in scientific applications.

5. MANAGERIAL IMPLICATIONS

The findings of this study provide important managerial implications for healthcare institutions, policymakers, and digital health technology developers. The high predictive performance achieved by the Support Vector Machine model demonstrates the potential of artificial intelligence-based systems in supporting early detection and preventive healthcare strategies for heart attacks and brain strokes across different age groups. Hospitals and healthcare providers can utilize AI-driven predictive systems to improve clinical decision-making, prioritize high-risk patients, and enhance preventive treatment planning. Furthermore, the proposed mobile-based alert system can support public health awareness by enabling continuous monitoring of critical health indicators such as blood pressure, glucose level, BMI, smoking behavior, and cholesterol level. From a strategic perspective, integrating machine learning into healthcare services may help reduce emergency cases, lower healthcare operational costs, and improve patient outcomes through earlier intervention and personalized healthcare recommendations.

6. CONCLUSION

Heart attacks and brain strokes have become major global health challenges affecting not only older adults but also middle-aged and younger populations. This study investigated the transition of cardiovascular and neurological disease risks across different age groups using a data science perspective. By analyzing multiple clinical and lifestyle-related variables, the research identified important health indicators associated with disease occurrence, including blood pressure, glucose level, cholesterol level, smoking behavior, BMI, and diabetes conditions. The study highlights the importance of utilizing machine learning technologies to support early disease prediction and preventive healthcare management.

The experimental analysis employed two machine learning algorithms, namely Support Vector Machine (SVM) and XGBoost classifiers, to evaluate predictive performance on heart attack and brain stroke datasets. The results demonstrated that SVM achieved superior performance with an accuracy of 99.5%, while XGBoost achieved 94.3% accuracy. These findings indicate that machine learning models, particularly SVM, can effectively support disease classification and risk prediction in healthcare environments. The study also confirms that AI-based approaches have significant potential to improve the accuracy and efficiency of early medical diagnosis systems.

Furthermore, this research contributes to the development of an AI-based alert system and mobile healthcare application designed to support timely preventive actions for patients at risk of heart attacks and brain strokes. The proposed system may assist healthcare professionals and individuals in monitoring health conditions continuously and identifying warning symptoms before diseases become severe. Future work will focus on expanding the dataset through collaboration with hospitals and healthcare institutions, incorporating additional clinical features, and evaluating other variants of SVM and advanced machine learning models to improve prediction robustness and real-world healthcare implementation.

7. DECLARATIONS

7.1. About Authors

Ankur Singh Bist (AB) https://orcid.org/0000-0002-7267-7366

Untung Rahardja (UR) https://orcid.org/0000-0002-2166-2412

Ninda Lutfiani (NL) https://orcid.org/0000-0001-7019-0020

Rifqa Nabila Muti (RM) https://orcid.org/0009-0008-2980-3823

Santiago Ramirez (SR) -

7.2. Author Contributions

Conceptualization: AB; Methodology: UR; Software: NL; Validation: AB and UR; Formal Analysis: RM and SR; Investigation: AB; Resources: UR; Data Curation: UR; Writing Original Draft Preparation: NL and RM; Writing Review and Editing: NL and RM; Visualization: UR; All authors, AB, UR, NL, RM, and SR, have read and agreed to the published version of the manuscript.

7.3. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.4. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.5. Declaration of Conflicting Interest

The authors declare that they have no conflicts of interest, known competing financial interests, or personal relationships that could have influenced the work reported in this paper.

REFERENCES

[1] B. Rawat, H. Pant, and A. Bist, "Interpretable machine learning for heart disease risk assessment: Leveraging shap values to identify clinically actionable predictors," *Biomedical and Pharmacology Journal*, vol. 19, no. 1, 2026.

[2] B. Rawat, A. S. Bist, P. Das, J. K. Samriya, S. C. Wariyal, and N. Pandey, "Alleviating the issues of recommendation system through deep learning techniques," in *Second International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2021*. Springer, 2021, pp. 1–9.

[3] N. Roozbeh, F. Montazeri, M. V. Farashah, V. Mehrnoush, and F. Darsareh, "Proposing a machine learning-based model for predicting nonreassuring fetal heart," *Scientific Reports*, vol. 15, no. 1, p. 7812, 2025.

[4] V. Krishna, J. S. Kiran, P. P. Rao, G. C. Babu, and G. J. Babu, "Early detection of brain stroke using machine learning techniques," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE, 2021, pp. 1489–1495.

[5] S. S. Kshatri and D. Singh, "Convolutional neural network in medical image analysis: a review," *Archives of Computational Methods in Engineering*, vol. 30, no. 4, pp. 2793–2810, 2023.

[6] P. P. Sianita, Y. Harwani, D. Permana, and E. S. Imaningsih, "Driving hospital revisit intentions through a technopreneurship approach to ewom care quality and patient experience," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 2, pp. 469–477, 2026.

[7] W. W. A. Effendi, M. Maswanto, M. Murod, and I. Maria, "Artificial intelligence enhancing financial regulatory compliance through regtech governance applications," *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, vol. 7, no. 2, pp. 216–225, 2026.

[8] E. V. Ivanovich, "The mechanism of brain strokes and heart attacks may be the same," *Applied Sciences Research Periodicals*, vol. 1, no. 2, pp. 01–04, 2023.

[9] B. Rawat, A. S. Bist, P. A. Sunarya, M. Hardini, N. A. Santoso, and R. Tarmizi, "Unveiling happiness disparities: A machine learning approach to city-village comparison," in *2023 11th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2023, pp. 1–5.

[10] M. Hardini, M. H. R. Chakim, L. Magdalena, H. Kenta, A. S. Rafika, and D. Julianingsih, "Image-based air quality prediction using convolutional neural networks and machine learning," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 5, no. 1Sp, pp. 109–123, 2023.

[11] T. Carval, C. Garret, B. Guillon, J.-B. Lascarrou, M. Martin, J. Lemarié, J. Dupeyrat, A. Seguin, O. Zambon, J. Reignier *et al.*, "Outcomes of patients admitted to the icu for acute stroke: a retrospective cohort," *BMC anesthesiology*, vol. 22, no. 1, p. 235, 2022.

[12] S. Demir and E. K. Şahin, "Liquefaction prediction with robust machine learning algorithms (svm, rf, and xgboost) supported by genetic algorithm-based feature selection and parameter optimization from the perspective of data processing," *Environmental Earth Sciences*, vol. 81, no. 18, p. 459, 2022.

[13] B. S. Ahamed, M. S. Arya, S. Sangeetha, and N. V. Auxilia Osvin, "Diabetes mellitus disease prediction and type classification involving predictive modeling using machine learning techniques and classifiers," *Applied Computational Intelligence and Soft Computing*, vol. 2022, no. 1, p. 7899364, 2022.

[14] B. Rawat, U. Rahardja, M. Hardini, and A. R. Dina, "Driver drowsiness detection using novel deep learning," *Health, Empathy, and AI Learning (HEAL)*, vol. 1, no. 1, pp. 1–6, 2025.

[15] A. Rozi, J. Junengsih, S. Alam, A. Pawar, W. Sumarjo, and D. Sunarsi, "Impact of hr management on ai implementation and data protection in indonesian manufacturing," *International Journal of Cyber and IT Service Management (IJCITSM)*, vol. 6, no. 1, pp. 51–64, 2026.

[16] E. M. Z. Akay, J. Rieger, R. Schöttler, J. Behland, R. Schymczyk, A. A. Khalil, I. Galinovic, J. Sobesky, J. B. Fiebach, V. I. Madai *et al.*, "A deep learning analysis of stroke onset time prediction and comparison to dwi-flair mismatch," *NeuroImage: Clinical*, vol. 40, p. 103544, 2023.

[17] I. B. Alaya, H. Limam, and T. Kraiem, "Automatic triaging of acute ischemic stroke patients for reperfusion therapies using artificial intelligence methods and multiple mri features: a review," *Clinical Imaging*, vol. 104, p. 109992, 2023.

[18] Y.-A. Choi, S.-J. Park, J.-A. Jun, C.-S. Pyo, K.-H. Cho, H.-S. Lee, and J.-H. Yu, "Deep learning-based stroke disease prediction system using real-time bio signals," *Sensors*, vol. 21, no. 13, p. 4269, 2021.

[19] U. Rahardja, S. Wijono, T. Wahyono, I. Sembiring, I. R. Widiasari *et al.*, "Effective ddos detection through innovative algorithmic approaches in machine learning," in *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*. IEEE, 2024, pp. 1–7.

[20] S. Anggoro and A. Nuche, "Leadership configurations supporting togaf-based information system architecture at jenderal achmad yani university," *International Journal of Cyber and IT Service Management (IJCITSM)*, vol. 5, no. 2, pp. 134–143, 2025.

[21] J. F. Scheitz, L. A. Sposato, J. Schulz-Menger, C. H. Nolte, J. Backs, and M. Endres, "Stroke–heart syndrome: recent advances and challenges," *Journal of the American Heart Association*, vol. 11, no. 17, p. e026528, 2022.

[22] X. Chen, J. Gu, and X. Zhang, "Brain-heart axis and the inflammatory response: connecting stroke and cardiac dysfunction," *Cardiology*, vol. 149, no. 4, pp. 369–382, 2024.

[23] J. Siswanto, U. Rahardja, I. Sembiring, E. A. Lisangan, M. I. N. Hakim, F. Wibowo *et al.*, "Hybrid deep learning model of lstm and bilstm for transjakarta passenger prediction," in *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)*. IEEE, 2024, pp. 1–7.

[24] A. Ruangkanjanases, N. Ivanov, M. Muhtarom, M. Salmi, and N. Lutfiani, "Advancing preventive community medicine for enhancing public health awareness," *Health, Empathy, and AI Learning (HEAL)*, vol. 1, no. 1, pp. 67–77, 2025.

[25] S. D. Retnandari, K. Khaeroman, A. T. Winarni, and W. Busse, "Digital innovation and technology driven evaluation of maritime safety systems performance in indonesia," *Aptisi Transactions on Technopreneurship (ATT)*, vol. 8, no. 2, pp. 551–564, 2026.