AUIQ Humanities and Social Sciences

Manuscript 1012

Towards Sustainable Agentic Artificial Intelligence in Medical Diagnosis for Autonomous Heart Disease Prediction

Thura J. Mohammed

Dhiaa M. Abed

AUIQ HUMANITIES AND SOCIAL SCIENCES 2 (2026) 55–69

ORIGINAL STUDY

Towards Sustainable Agentic Artificial Intelligence in Medical Diagnosis for Autonomous Heart Disease Prediction

Thura J. Mohammed [a], **Dhiaa M. Abed** [b]

[a] School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia
[b] College of Biomedical Engineering, University of Technology, Baghdad, Iraq

ABSTRACT

Diagnosis of heart disease is a procedure that requires fast, accurate predictions and conventional fixed machine learning (ML) pipelines may not fully accommodate. Although models, such as Logistic Regression and Naïve Bayes are effective on structured clinical data however they do not have self-assessment or adaptive reasoning mechanisms. In this work, we present an agentic AI framework that turns the classical models into self-evaluating diagnostic agents managed by a performance conscious supervisory component. We trained and tested the system on the UCI Heart disease dataset over three different splits (70:30, 80:20, 85:15) to test its diagnostic stability with respect to data availability. The Logistic Regression model had the best overall (up to 85% accuracy) and F1-score performance across the experiments, while Naïve Bayes outperformed in recall for heart disease cases, indicating its beneficial use in sensitivity- oriented tasks. In each situation the supervisory agent automatically selected the superior model, demonstrating that agents can coordinate with one another even at relatively low levels of complexity. This work mediates interpretability with accountability by the construction classical models in a simple self-contained framework for a reliable transparent diagnosis without any additional computational effort. Conclusion: Agentic supervision may be the key to taking classical algorithms and developing them into scalable intelligent decision support tools for real-world cardiovascular care.

Keywords: Agentic artificial intelligence, Heart disease diagnosis, Clinical decision support, Logistic regression, Naïve Bayes

1. Introduction

Cardiovascular diseases are the leading cause of death worldwide; therefore, early and accurate heart disease diagnosis represents a significant challenge in contemporary healthcare [1, 2]. Diagnosis in clinic usually relies on the analysis of other physiological and diagnostic indexes, which is time-consuming and human subjective [3, 4]. As a result, data-driven ML methods are increasingly sought after as promising methodologies for assisting in implicit, automated and unbiased cardiovascular disease diagnosis [5]. Classical ML methods like Logistic Regression and Naïve Bayes are used in heart disease prediction depending on the interpretability, computational efficiency and clinical transparency [6]. Although these models achieve impressive results, most previous work has depended on static learning pipelines with isolated classifiers and lack self-evaluation or coordination mechanisms [7]. These methods compromise robustness and reduce the clinical utility of intelligent diagnostic systems in real-world clinical environments. The agentic artificial intelligence (AI) was introduced recently as the paradigm, which supports autonomy and situational awareness of performance at agents with ability to make decisions in an

Received 15 February 2026; revised 2 May 2026; accepted 22 May 2026.
Available online 30 May 2026

* Corresponding author.
E-mail address: thurajamal@gmail.com (T. J. Mohammed).

https://doi.org/10.70176/3106-7557.1012

interacting community of intelligent agents [8, 9]. In the medical diagnosis setting, agentic AI provides diagnostic systems with abilities to evaluate model behavior autonomously as well as compare performance results and provide decision support reasoning [10]. But what has been under-investgated is the fusion of agentic AI principles with such interpretable classical ML models for heart disease diagnosis [11]. In this study, agentic AI refers to an autonomous diagnostic framework in which machine learning models function as diagnostic agents, while a supervisory agent evaluates their performance and coordinates the final decision, rather than merely selecting a model through a static comparison process [12]. To this end, this study presents an agentic AI framework for heart disease classification that incorporates several autonomous diagnostic agents and a supervisory decision mechanism. This study employs Logistic Regression and Naïve Bayes models as independent diagnostic agents, and use a supervisory agent to assess their performance over several train–test settings for autonomous and reliable decision-making. This in turn provides greater diagnostic robustness and transparency without a corresponding increase in system complexity. The proposed approach is also tested on the UCI Heart Disease dataset with different levels of data split [8, 13]. The performance is consistent with good generalization, as we see from our experimental results by both the accuracy, loss, ROC analysis along with confusion matrix. It proposes an agentic diagnostic system based on classical models, devises a performance-oriented supervisor module, and elaborates results supporting intelligent heart disease diagnosis. However, existing heart disease prediction studies have largely focused on static machine learning classifiers, while the integration of interpretable classical models within an agentic AI framework with autonomous performance supervision remains insufficiently explored.

2. Related work

2.1. Agentic AI in healthcare

Agentic AI represents a new class of autonomous systems capable of goal-directed behavior, self-initiated action, and adaptive learning [14]. In healthcare, agentic AI enables AI systems to actively perceive data, reason through diagnostic tasks, and adapt decision-making based on outcomes [15]. Unlike traditional AI tools that work in fixed pipelines, agentic systems combine all components of the decision-making process (planning, memory, perception and action) in a holistic loop, so that they can react to clinical variations dynamically [16]. The latest developments indicate AI with agency is capable of improving the accuracy of diagnosis, mitigating human error and optimising healthcare process flow [17]. AI models have demonstrated the ability to triage patients in need of urgent care, interface with and interpret diverse kinds of patient data (e.g., images, genetics, text), as well as forecast a patient's decline or recovery using real-time telemetry [18]. These models are not limited to the static ones, since they learn through feedback and have adaptive thresholds handling interactions needed for certain tasks. Yet, even with these resources, agentic AI remains largely theoretical [19, 20]. Existing applications are constrained by absence of clinical validation, regulatory clarity and interpretability. Safety, bias, and transparency concerns make it challenging to deploy in high-stakes medical settings. As the veracity and value of data become clearer for many domains within the broader field, embedding agentic ability into explainable systems will be a key driver to clinical acceptance.

2.2. ML models for heart disease prediction

ML approaches perform quite well for heart disease prediction, especially in the structured clinical data. Logistic Regression and Naïve Bayes are popular due to their ease, interpretability and speed of calculations [21]. Many works have shown that this model can achieve reasonable high accuracy (e.g., > 80%), sometimes even comparable to, and more robust than, some other advanced algorithms on feature selection and preprocessing [22]. Logistic Regression demonstrates good performance when features are linearly separable and sample size is moderate to small; scores less than the threshold would be predicted as non-contraband facilities [23]. Though Naïve Bayes works under the assumption of feature independence, but it performs well with strong marginal distributions during training and has obtained high recall in detecting positive cardiac cases [24, 25]. Results from comparative studies also show that although sophisticated ensemble models might provide marginal gains of accuracy, traditional ones are still competitive and have the advantage of interpretation that plays a crucial role in clinical decision making [26, 27]. A dominant trend in the recent literature is hybrid approaches, primarily combining classical models with ensemble or multi-agent techniques aimed at balancing robustness and transparency [28, 29]. However, the vast majority of these models are trained on retrospective cohorts with limited generalizability to real-world clinical populations. Prospective validation and application within the live clinical environment are now essential components of future work. Prior studies on heart disease prediction have shown that classical machine learning models can achieve reliable diagnostic performance using structured clinical datasets. However, most existing approaches mainly compare classifiers in a static manner without incorporating autonomous performance monitoring or supervisory coordination. The proposed framework builds on these studies by treating Logistic Regression and Naïve Bayes as diagnostic agents and introducing a supervisory agent that evaluates their performance across different data-splitting scenarios. This connection allows the system to move beyond conventional model comparison toward an agentic supervision approach for transparent and adaptive clinical decision support.

2.3. Multi-agent and supervisory AI frameworks in diagnosis

The use of multi-agent AI systems is a known approach to organizing intelligent diagnostic systems [30, 31]. These frameworks use special-purpose agents each dedicated to a task, such as diagnosis prediction, monitoring or recommendation, via supervision [32]. Through the distributed properties of cognitive load across multiple agents, the system is able to handle complex workflows, cross validate decisions, and increase diagnostic reliability through inter-agent communication [33, 34]. Patterns of results in the new evidence base demonstrated substantial gains on diagnostic accuracy, differential diagnosis coverage, and distributed task performance from multi-agent systems [2, 35]. For instance, improved performance has been shown for lossless detection when plural diagnostic agents cooperate under a supervisory controller particularly when faced with indefinite or borderline findings [36, 37]. Such architectures mirror the holistic collaboration practiced in medical teams, where specialists bring complementary views on a patient's case [1, 7]. The role of the supervisory agent is to collect all agent responses, balance specific input/output requests and maintain decisional coherence [20]. However, above oversight has been known to enhance both accuracy and credibility. Nonetheless, the application of multi-agent systems in the healthcare domain presents issues such as communication delay, coordination complexity and development of efficient control logic [30]. Further, accountability gets multiplied in multi-agent scenarios and it is difficult to attribute mistakes to some specific components. To allay these concerns, human-in-the-loop supervision remains vital to keep the control in clinicians' hands and at the same time receive automated support.

3. Methodology

The proposed framework is organized into sequential phases, starting from data acquisition and feature selection, followed by preprocessing and data partitioning. Subsequently, two diagnostic agents, Naive Bayes and Logistic Regression, are employed to generate probabilistic and discriminative outputs. These outputs are then evaluated by a supervisory agent, which monitors performance and produces the final diagnostic decision. The overall methodology adopted in this study is illustrated in Fig. 1.

3.1. Dataset description and exploratory data analysis

This study is evaluated on the UCI dataset, but the focus of attention is given to the Cleveland Clinic dataset, widely used and clinically proven datasets in literature [38, 39]. There are 76 features in the original dataset, but only 14 clinically relevant features are mostly used in previous ML analysis [40, 41]. These are 13 predictive features, and one target value, whether disease is present or absent. The target is devised as binomial response of 0 (no heart disease) and 1 (some kind of heart disease). This binary characterization is consistent with clinical practice, and supports the supervised classification of cardiovascular risk.

3.1.1. Feature description

The selected attributes represent a comprehensive set of demographic, physiological, and diagnostic indicators relevant to cardiovascular health [42]. These include patient age and sex, chest pain type, resting blood pressure, serum cholesterol level, fasting blood sugar status, electrocardiographic results, maximum heart rate achieved during exercise, exercise-induced angina, ST-segment depression (oldpeak), slope of the ST segment, number of major vessels identified via fluoroscopy, and thalassemia status [43]. Together, these features capture both structural and functional aspects of cardiac performance, providing a rich input space for intelligent diagnostic modeling.

3.1.2. Statistical overview of the dataset

As an initial overview, we find many of the clinical traits exhibit high variability, even at baseline, which telegraphs realistic patient heterogeneity in the dataset. The continuous predictors as age, cholesterol, resting blood pressure (RBP), maximum heart rate (MHR) and the ST depression are in large numerical ranges and distribution is not symmetric [44, 45]. Binary and nominal variables such as sex, fasting blood sugar, and type of chest pain have limited range of values also clinically interpretable. This statistical

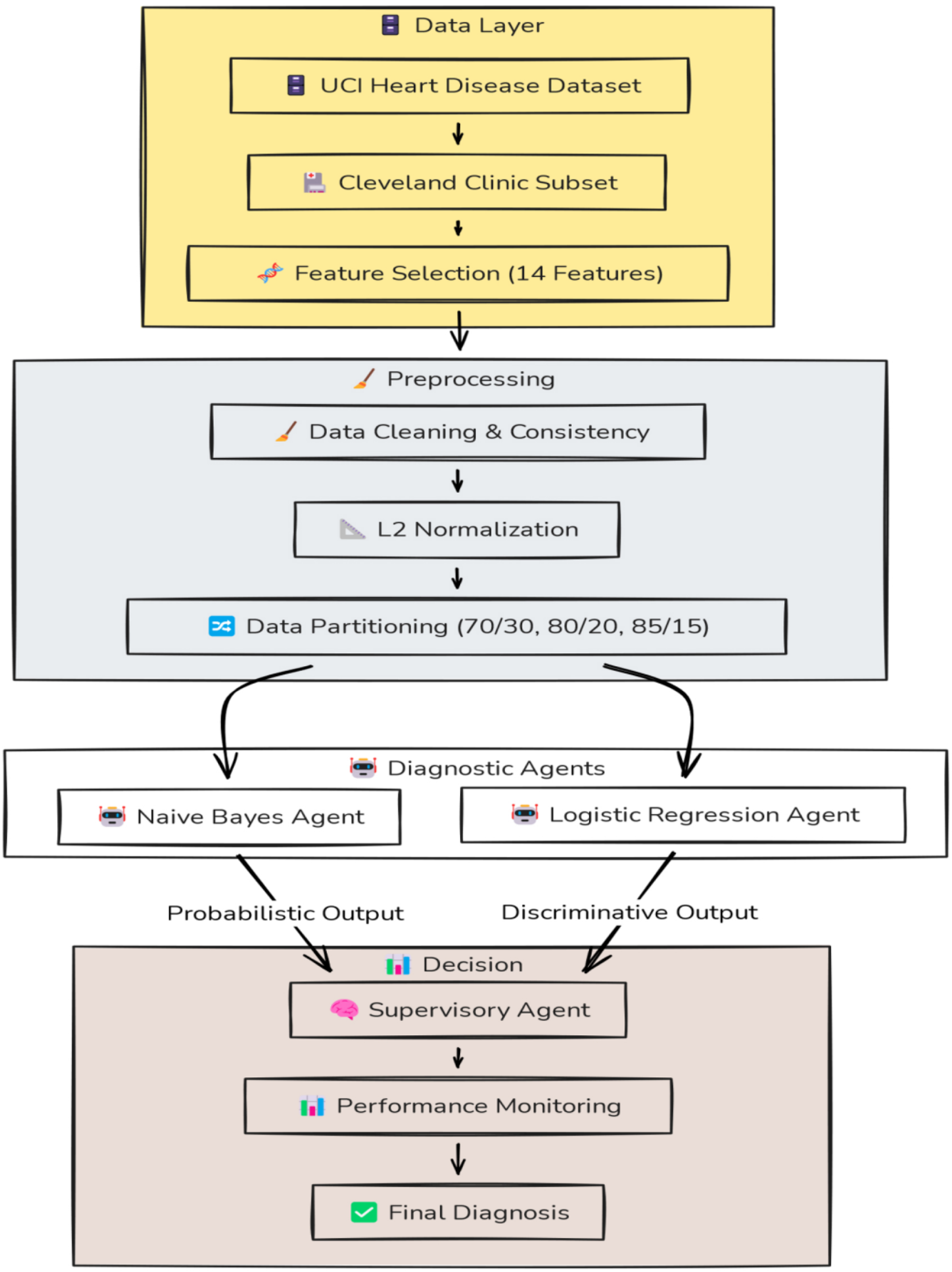

Fig. 1. Methodology phases diagram.

variation will encourage normalization (or training with respect to) and robust learning in the following steps of our approach [46, 47].

3.1.3. Exploratory data analysis (EDA)

To facilitate intuitive use of data and stimulate autonomous decisions within the agentic approach, exploratory data analysis (EDA) was performed through a variety of complementary visualization methods. Histograms for chosen clinical characters showing the distributions of their frequency were shown in Fig. 2. Various features such as cholesterol, ST depression suffer from asymmetry with long tails; other factors like age and maximum heart rate have wider spread in clinical ranges [48]. These insights give reasons why normalizing features is necessary and motivate

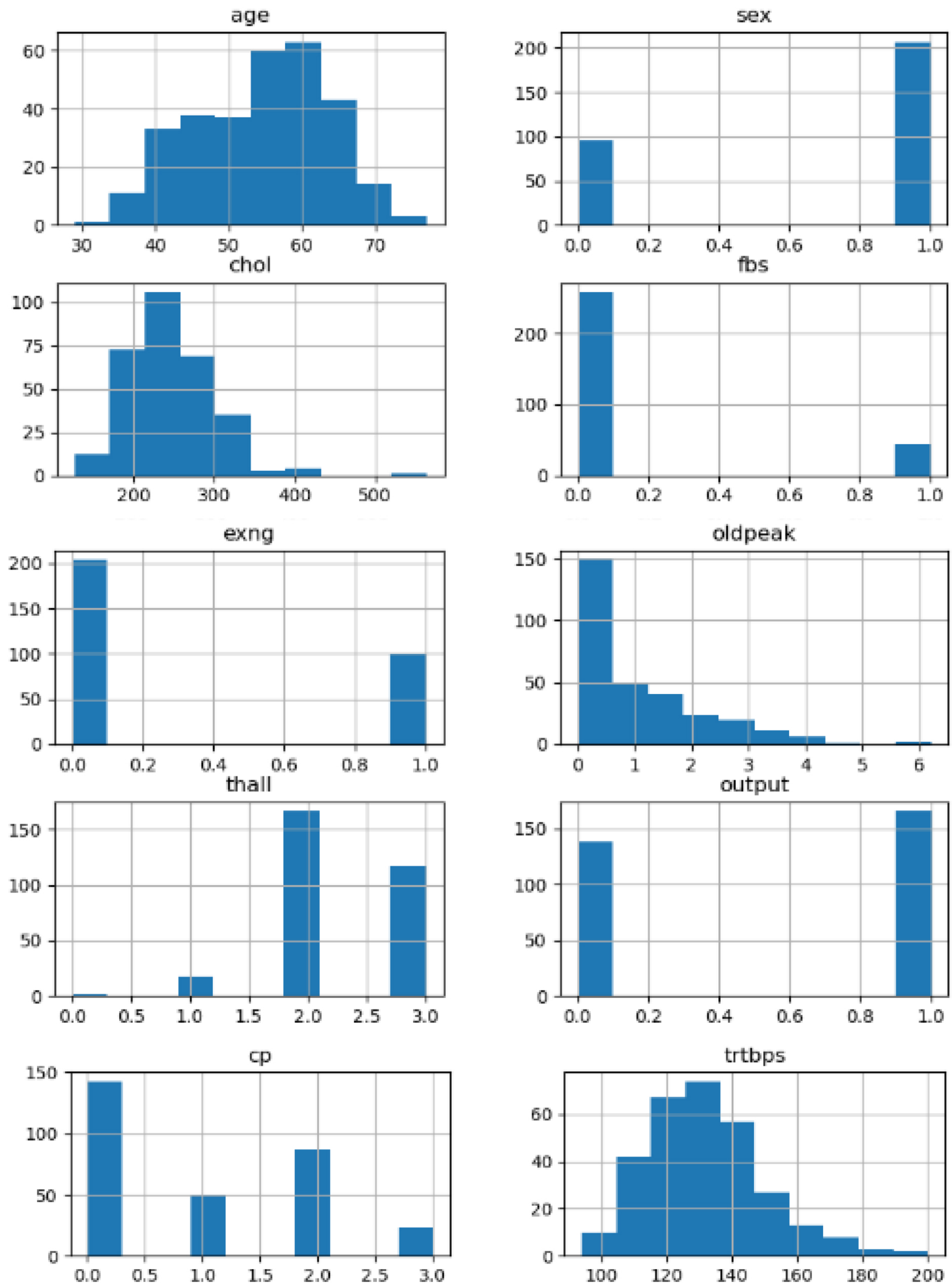

Fig. 2. Histograms of selected features.

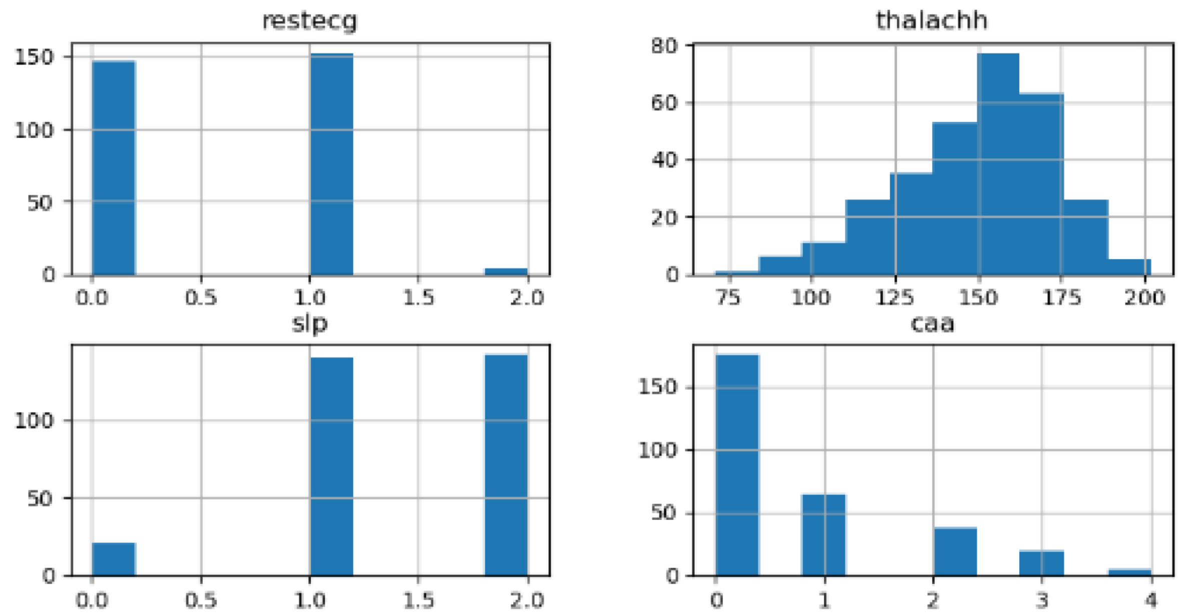

Fig. 2. Continued.

the importance of learning models that can accommodate non-Gaussian data distributions.

Fig. 3 shows density plots of the same features to have a smoothed estimate over their underlying probability distributions. In the density curves, multimodal distributions of some variables and overlapping regions between healthy and diseased patients. From an intelligent systems point of view, the overlap in these regions indicate that they are ambiguous with respect to diagnosis, revealing that it is crucial to have adaptive and probabilistic reasoning models for medical classification purposes [49].

Fig. 4 illustrates boxplots of selected features stratified by heart disease outcome. Clear shifts in median values and interquartile ranges are observed between the two classes for several attributes, particularly maximum heart rate and ST depression. These visual differences suggest strong discriminative potential and support the clinical relevance of the selected predictors for automated diagnosis.

3.1.4. Class-wise statistical comparison

To further quantify the observed differences between patient groups, a class-wise statistical analysis was performed on key continuous features, including age, cholesterol, maximum heart rate, and ST depression. Fig. 5 presents the mean $\pm$ standard deviation of these features for patients with and without heart disease.

The analysis indicates that patients diagnosed with heart disease tend to be older, exhibit higher levels of ST depression, and achieve lower maximum heart rates during exercise compared to healthy individuals [50, 51]. These statistically meaningful trends reinforce known clinical patterns and provide quantitative evidence supporting the discriminative capacity of the dataset.

3.2. Agentic data preprocessing module

3.2.1. Data cleaning and feature consistency

In the first step, an inspection of the dataset was performed in terms of inconsistent, missing or unreasonable values. Inconsistent or nonnumerical records for critical attributes were removed to make the data suitable for learning, and in order to avoid bringing noise into the learning process [52]. This self-cleaning step guarantees that the resulting feature representations provided to downstream diagnostic agents are consistent and valid, something of paramount importance in order to ensure a stable performance from any model operating on these data within medical decision-support systems.

3.2.2. Feature normalization strategy

Due to the heterogeneous nature of clinical features, ranging from demographic attributes to physiological measurements, the dataset exhibits significant variation in feature scales. To address this issue, L2 normalization was applied to the feature vectors prior to model training [53]. Formally, for a given feature vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$, L2 normalization is defined as:

$$\mathbf{x}' = \frac{\mathbf{x}}{\|\mathbf{x}\|_2} = \frac{\mathbf{x}}{\sqrt{\sum_{i=1}^{n} x_i^2}}$$

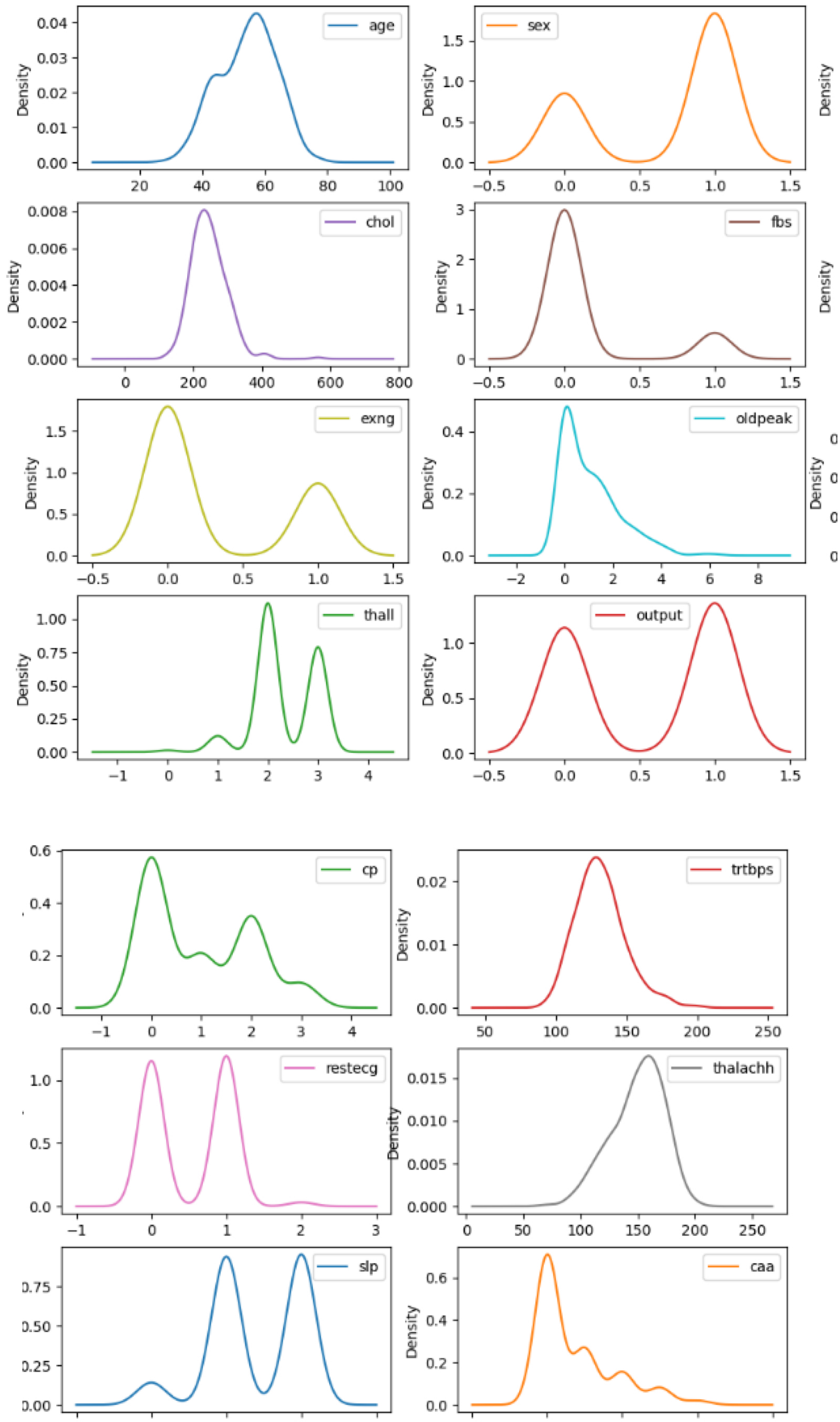

Fig. 3. Density plots of selected features.

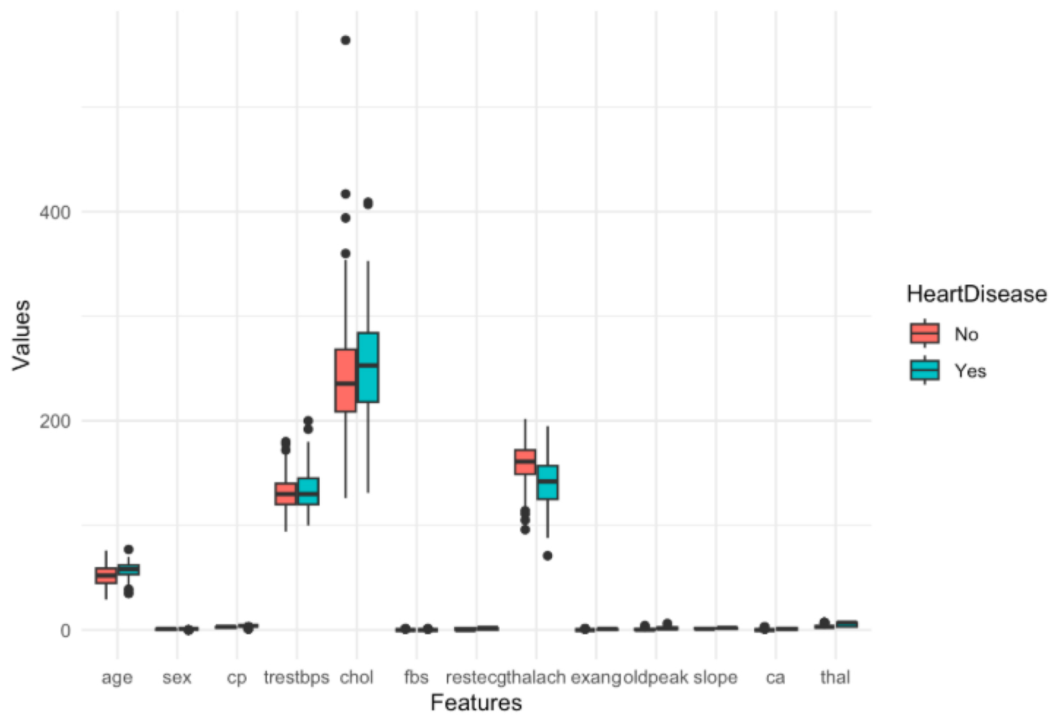

Fig. 4. Boxplots of Features by Heart Disease Outcome.

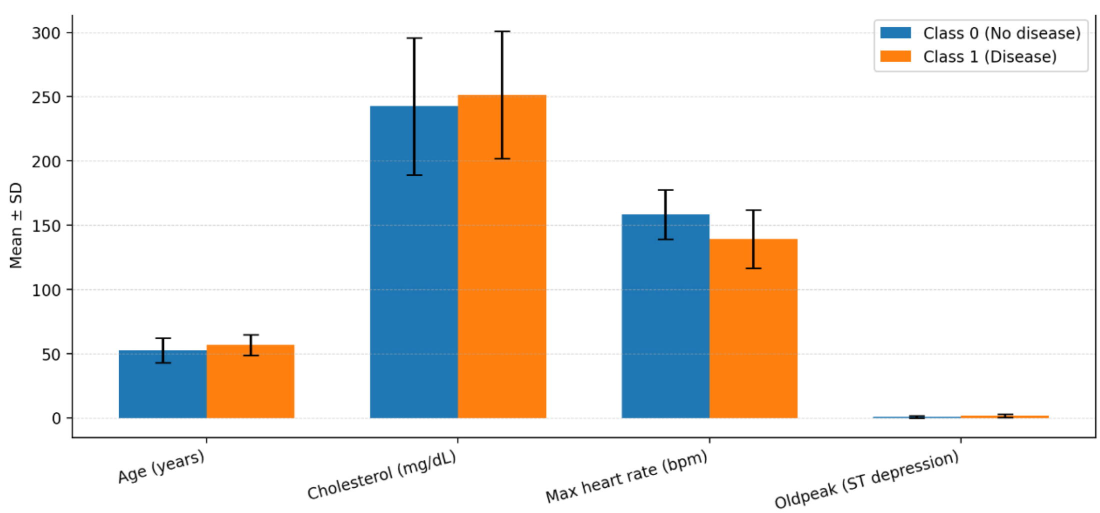

Fig. 5. Class-wise mean ± standard deviation of key clinical features from Heart Disease dataset.

L2 normalization was chosen because it scales each patient feature vector to a comparable magnitude, thereby reducing the dominance of high-value clinical measurements such as cholesterol and resting blood pressure while preserving the relative pattern of features within each patient record. This normalization strategy ensures that all features contribute proportionally to the learning process, preventing dominance of high-magnitude attributes such as cholesterol or resting blood pressure. From an agentic perspective, normalization enhances the perceptual uniformity of the environment, enabling diagnostic agents to make balanced and stable inferences.

3.2.3. Autonomous data partitioning

The preprocessing module automatically partitions the dataset into training and testing subsets under three experimental settings: 70:30, 80:20, and 85:15. Stratified sampling was applied during each partitioning process to preserve the original distribution of heart disease and non-heart disease cases in both the training and testing subsets data [54]. In addition, a fixed random seed was used to ensure reproducibility of the data splits and experimental results. These multiple partitioning settings enable the evaluation of the proposed agentic system under different levels of training data availability and provide a robustness assessment of diagnostic performance across varying experimental conditions [55].

3.3. Agentic AI architecture

The model presented in this work is based on the agentive AI category with the combination of different one agent to carry out heart disease classification. The architecture is made up of diagnostic agents which are directed by another higher-level executive agent, model necessarily subjected to variety and consistency [19]. It adopts proactive AI principles, decoupling the perception of data, diagnostic reasoning and evaluation for performance by means of synergistic autonomic steps. There are two private diagnostic agents, corresponding to different ML methods [56, 57]. The first agent is based on a naive model of Naïve Bayes classifier doing probability reasoning with assumption of conditional independence and the second uses Logistic Regression to train a discriminative classifier to distinguish diseases. Both agents take normalized feature input from the preprocessing modules and provide separate diagnostic outputs. A supervisory agent assesses the outputs of the diagnostic agents by observing their performance over experimental set-ups. Operating at a meta-decision level allows the supervisory agent autonomous assessment of diagnostic reliability and it facilitates to combine decision in congruent way. This hierarchical agentic architecture improves robustness and interpretability and is appropriate for autonomous medical decision-support systems.

3.4. Learning and adaptation strategy

This structure follows a supervised learning methodology where diagnostic agents learn from standardize patient data and are tested in multiple different training–test set-up. This may be seen as a way to measure how well the methods generalize with different levels of teacher input. Adaptation is done by performace based assesement instead of getting trained perpetually [58, 59]. A monitoring agent monitors diagnostic outcomes and the performance history thereof, so as to enable automatic detection of sustained and successful diagnostic operation by the system. This learning process provides both stability and interpretability, but it is also in line with the requirements of a real-world autonomous medical decision support system.

4. Results and discussion

4.1. Logistic regression results

The Logistic Regression diagnostic agent demonstrates consistently strong performance across all splitting scenarios, as summarized in Table 1.

With the model on the 70:30 split, testing accuracy reaches 0.83, while the testing loss is only 0.164. This shows that generalization has been effective. The precision–recall balance gives a great recall of cases for heart disease, which is especially important

Table 1. Performance comparison of Logistic Regression and Naïve Bayes models under different data splitting cases.

Model	Splitting Case	Class	Precision	Recall	F1-Score	Training Accuracy	Training Loss	Testing Accuracy	Testing Loss
Logistic Regression	70:30	Class 0	0.9143	0.7273	0.8101	0.86	0.136	0.83	0.164
		Class1	0.7857	0.9362	0.8544				
	80:20	Class 0	0.875	0.7778	0.8235	0.84	0.157	0.85	0.147
		Class1	0.8378	0.9118	0.8732				
	85:15	Class 0	0.8947	0.7391	0.8095	0.83	0.163	0.82	0.173
		Class1	0.7778	0.913	0.84				
Naïve Bayes	70:30	Class 0	0.8611	0.7045	0.775	0.83	0.16	0.8	0.197
		Class1	0.7636	0.8936	0.8235				
	80:20	Class 0	0.8333	0.7407	0.7843	0.83	0.169	0.81	0.18
		Class1	0.8108	0.8824	0.8451				
	85:15	Class 0	0.8571	0.7826	0.8182	0.83	0.167	0.82	0.173
		Class1	0.8	0.8696	0.8333				

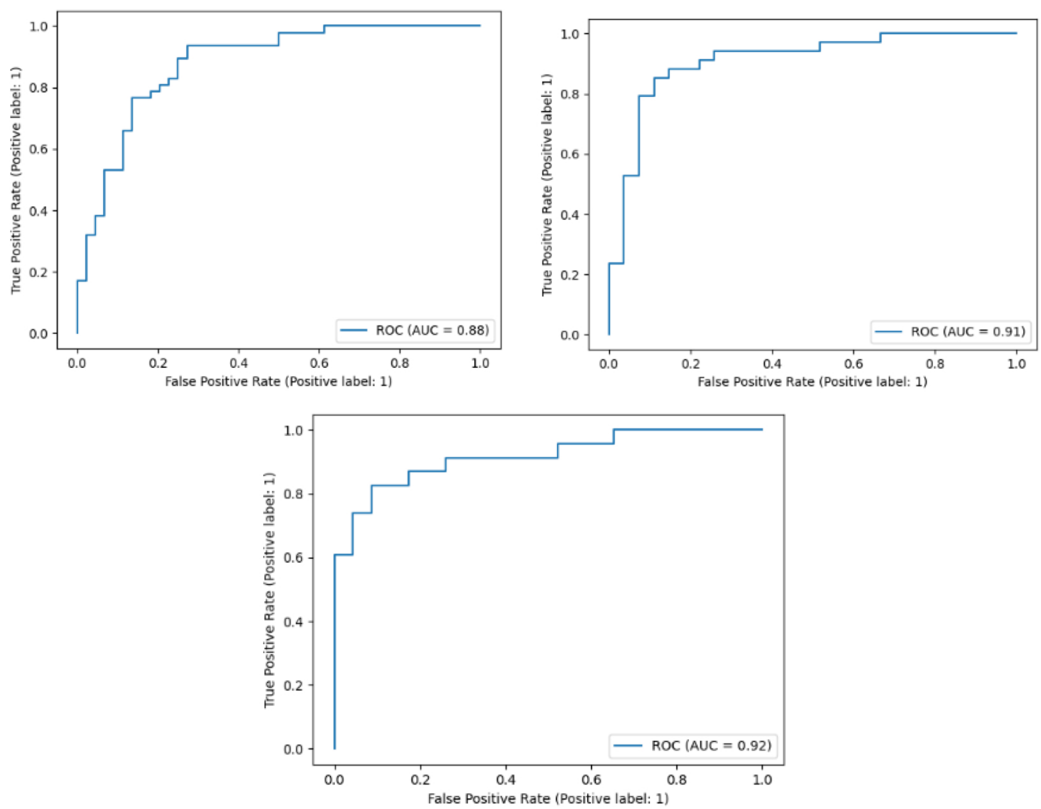

Fig. 6. ROC curves for Logistic Regression under Different Train–Test Split Ratios (a) 70:30 split, (b) 80:20 split, and (c) 85:15 split.

in clinical applications. Here, reduced negatives are vital. Whatevers is 80:20 split will thus have highest test accuracy (0.85) and least test loss (0.147), indicating that a good balance exists between the availability of training data in quantity with reliability at evaluation time. This arrangement is also good for both precision and recall results for every class, allowing the F1-scores to be further improved as a result.

Logistic Regression benefits from more training data whereas decision thresholds don't change. In the case of the 85:15 split, although training accuracy remains stable, a small reduction in testing accuracy is seen. This may be attributable to the small size of the test set leading to more uncertain performance estimates. Nevertheless, the model retains competitive precision and recall measures; suggesting that it is still sound. These results are further supported by the ROC curves shown in Fig. 6 which present good separation between classes for all partitions, with maximum curve characteristic obtained via an 80:20 layout. Confusion matrices in Fig. 7 show classifiers with balanced identifying behaviors and relatively few incorrectly classified samples. This is consistent performance characteristics for heart disease diagnosis using Logistic Regression.

4.2. Naïve Bayes results

The Naïve Bayes diagnostic agent also achieves satisfactory performance across all experimental configurations, as shown in Table 1. For the 70:30 split, the model attains a testing accuracy of 0.80, with higher recall for heart disease cases compared to non-diseased cases. This tendency reflects the probabilistic nature of Naïve Bayes, which favors sensitivity over specificity in overlapping feature spaces. In the 80:20 configuration, Naïve Bayes shows improved testing accuracy (0.81) and balanced F1-scores, indicating more stable classification behavior with increased training data. The 85:15 split further improves testing accuracy to 0.82, demonstrating that the model benefits from larger training sets despite its simplifying independence assumptions.

The 80:20 split provided the best overall balance because it offered sufficient training data for stable model learning while retaining an adequate testing subset for reliable evaluation. In contrast, the 70:30 split used less training data, whereas the 85:15 split had a smaller test set that may increase evaluation variability. This explains why the 80:20 split achieved the highest testing accuracy and lowest testing loss. ROC curves in Fig. 8 confirm reasonable discrimi-

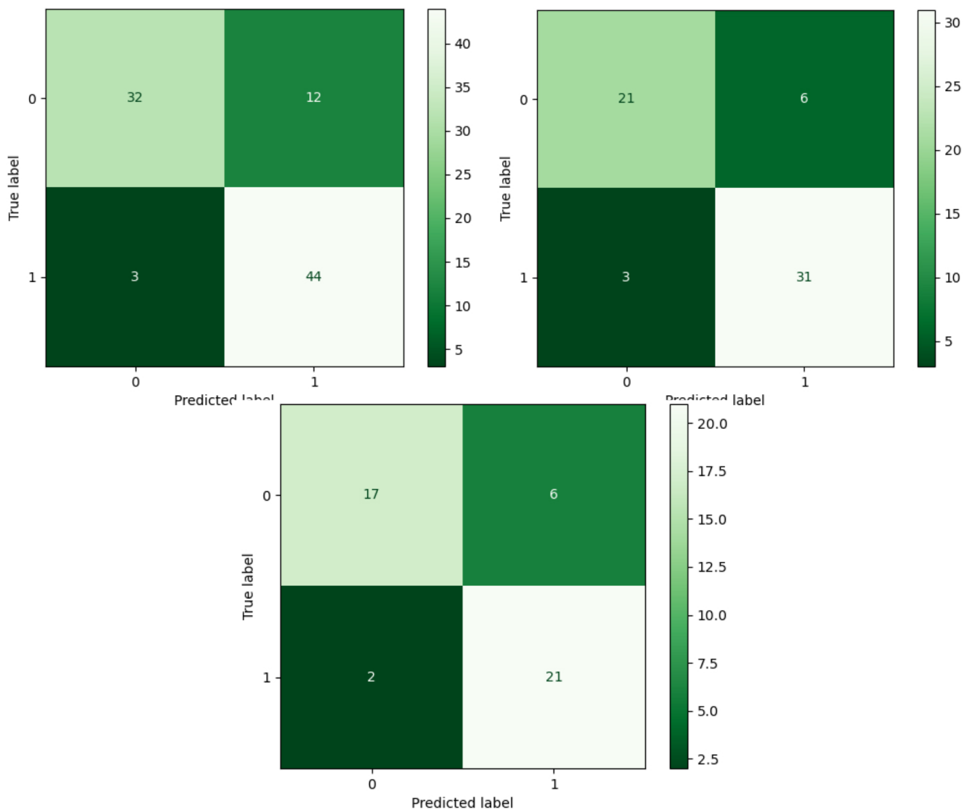

Fig. 7. Confusion Matrices for Logistic Regression under Different Train–Test Split Ratios (a) 70:30 split, (b) 80:20 split, and (c) 85:15 split.

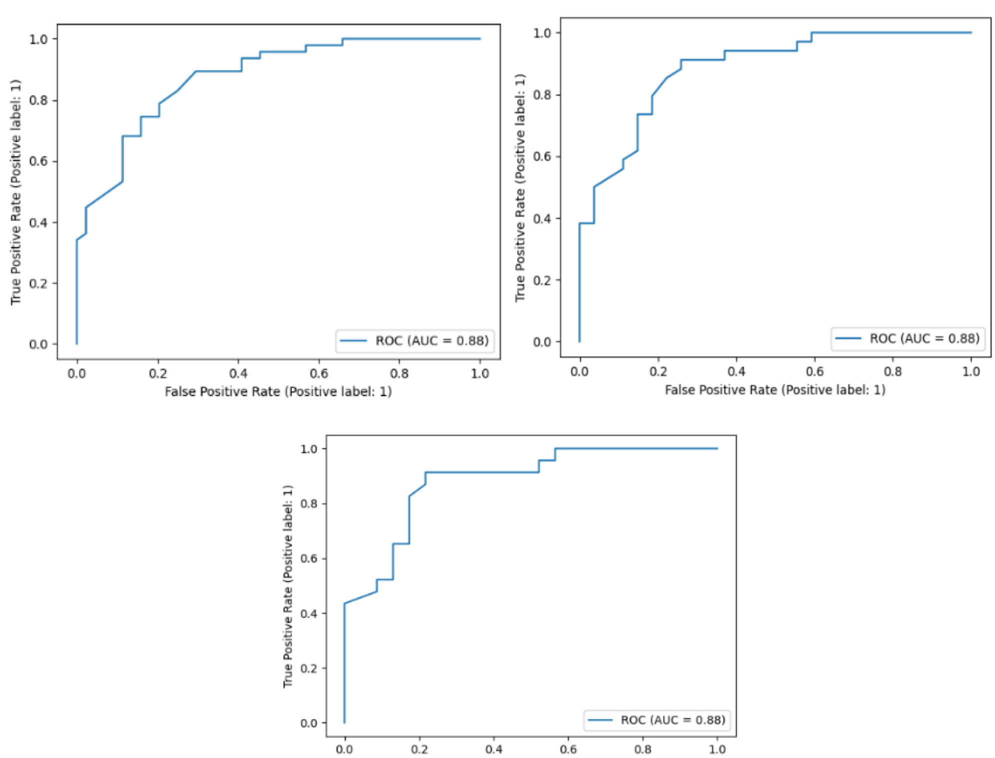

Fig. 8. ROC curves for Naïve Bayes under Different Train–Test Split Ratios (a) 70:30 split, (b) 80:20 split, and (c) 85:15 split.

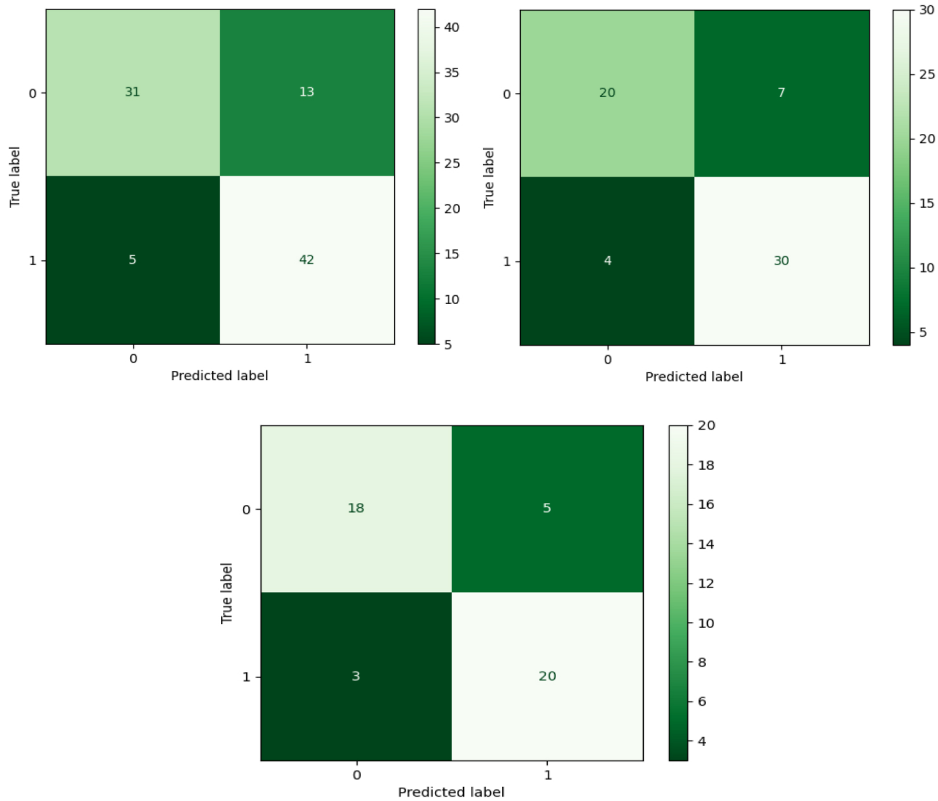

Fig. 9. Confusion Matrices for Naïve Bayes under Different Train–Test Split Ratios (a) 70:30 split, (b) 80:20 split, and (c) 85:15 split.

nation capability for Naïve Bayes across all splits, although the curves exhibit slightly lower separability compared to Logistic Regression. Confusion matrices in Fig. 9 reveal higher misclassification rates for borderline cases, particularly false positives, which can be attributed to correlations among clinical features that violate the independence assumption.

4.3. Comparative discussion

Here is a few direct comparison between Logistic Regression and Naïve Bayes. Logistic Regression always outperforms Naive Bayes in terms of testing accuracy and loss stability under the 80:20 split. This inferior performance is in line with the exploratory data analysis, which found overlapping and correlated feature distributions favoring discriminative learning methods. Naïve Bayes, on the other hand, exhibits high recall in all set-ups including those that involve heart disease cases thus making it useful to detect patients with positive predictions. Concerning clinical applications, these behaviors could be useful for screening situations in which sensitivity is more important than specificity. In the agentic AI setting, these performance deltas are actionable feedback to the supervising agent. By observing accuracy trends, loss magnitudes and error distributions under various different configurations, it can autonomously determine that Logistic Regression would be a more trustworthy diagnostic agent given the experimental conditions, yet continue to acknowledges merits in Naïve Bayes' strengths as well.

5. Conclusion & future work

In this research, an agentic AI model was presented for heart disease categorization using Logistic Regression and Naïve Bayes diagnostic agents. The proposed system includes self-sufficient prepossessing, agent-based diagnosis and performance-based supervision to facilitate trustworthy and understandable medical decision application. Experimental results on UCI Heart Disease dataset indicate that Logistic Regression is able to show better generalization performance over various data split cases, and Naïve Bayes estimates high sensitivity detection of heart disease cases. The results demonstrate the efficacy of the agentic framework in independently determining the choice of diagnostic agents, guided by quantitative performance feedback. The findings suggest that agentic supervision can enhance classical ML algorithms by transforming them into scalable, transparent, and reliable decision-support tools for cardiovascular care. However, the present study is limited by the relatively small size of the UCI Heart Disease dataset and the use of retrospective data, which should be addressed through validation on larger and more diverse clinical datasets in future work. Future work will involve scaling-up the method to a number of learning agents, including explainable AI techniques, and testing on larger clinical data to improve scalability and clinical application.

Conflict of interest

The authors declare no conflict of interest.

Author contribution

All authors contributed equally to the conception, design, data analysis, manuscript preparation, and revision of the study. All authors read and approved the final manuscript.

Data availability

The dataset analyzed during the current study is publicly available online.

References

1. Liu F *et al.* A foundational architecture for AI agents in healthcare. Oct. 21, 2025. doi: 10.1016/j.xcrm.2025.102374.
2. Liawrungrueang W. Artificial Intelligence (AI) Agents Versus Agentic AI: What's the Effect in Spine Surgery? Jun. 01, 2025, Korean Spinal Neurosurgery Society. doi: 10.14245/ns.2550308.154.
3. Rimal Y, Sharma N, Paudel S, Alsadoon A, Koirala MP, Gill S. Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. Sci. Rep. Dec. 2025;15(1). doi: 10.1038/s41598-025-93675-1.
4. Xi Y, Wang H, Sun N. Machine learning outperforms traditional logistic regression and offers new possibilities for cardiovascular risk prediction: A study involving 143,043 Chinese patients with hypertension. Front. Cardiovasc. Med. Nov. 2022;9. doi: 10.3389/fcvm.2022.1025705.
5. Narasimhan G, Victor A. Empirical analysis of predicting heart disease using diverse datasets and classification procedures of machine learning. Ain Shams Engineering Journal. Aug. 2025;16(8). doi: 10.1016/j.asej.2025.103470.
6. Chen X *et al.* Enhancing diagnostic capability with multi-agents conversational large language models. NPJ Digit. Med. Dec. 2025;8(1). doi: 10.1038/s41746-025-01550-0.
7. Collaco BG *et al.* The Role of Agentic Artificial Intelligence in Healthcare: A Systematic Review. Aug. 07, 2025. doi: 10.21203/rs.3.rs-7234499/v1.
8. Alsabhan W, Alfadhly A. Effectiveness of machine learning models in diagnosis of heart disease: a comparative study. Sci. Rep. Dec. 2025;15(1). doi: 10.1038/s41598-025-09423-y.
9. Marengo A. Navigating the nexus of AI and IoT: A comprehensive review of data analytics and privacy paradigms. Oct. 01, 2024. Elsevier B.V. doi: 10.1016/j.iot.2024.101318.
10. Ilyas AR, Javaid S, Kharisma IL. Heart Disease Prediction Using ML. MDPI AG. Oct. 2025;124. doi: 10.3390/engproc2025107124.
11. O'Hair DP *et al.* Enhanced Detection of Heart Valve Disease Using Integrated Artificial Intelligence at Scale. Annals of Thoracic Surgery. May 2022;113(5):1499–1504. doi: 10.1016/j.athoracsur.2021.04.106.
12. Ibrahim RM, Mohammed TJ, Alsaidi SAAA. Agentic AI for Drug Prescription: Machine Learning Models for Intelligent Healthcare Decision Support. in Beyond Intelligent Systems: Navigating the Power of the Agentic Artificial Intelligence for Driving Innovation, V. Tiberius, A. Malik Alnoor, M. Valeri, K. Dahleez, and M. S. Alazzawi, Eds., Cham: Springer Nature Switzerland. 2026:365–379.
13. Katarya R, Meena SK. Machine Learning Techniques for Heart Disease Prediction: A Comparative Study and Analysis. Health Technol. (Berl). 2021;11(1):87–97. doi: 10.1007/s12553-020-00505-7.
14. Sunge AS, Amali, Ahmad Turmudi ZY, Pramudito DK, Badruzzaman A, Purwanto. The model interpretability on SHAP and comparison classification selection feature for heart disease prediction. Procedia Comput. Sci. 2024;245(C):210–219. doi: 10.1016/j.procs.2024.10.245.
15. Santos MAG, Munoz R, Olivares R, Filho PPR, Del Ser J, de Albuquerque VHC. Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. Information Fusion. 2020;53:222–239. doi: 10.1016/j.inffus.2019.06.004.
16. Baseer KK *et al.* Healthcare diagnostics with an adaptive deep learning model integrated with the Internet of medical Things (IoMT) for predicting heart disease. Biomed. Signal Process. Control. 2024;92:105988. doi: 10.1016/j.bspc.2024.105988.
17. Hamid RA, Albahri AS, Albahri OS, Zaidan AA. Dempster–Shafer theory for classification and hybridised models of multi-criteria decision analysis for prioritisation: a telemedicine framework for patients with heart diseases. Springer Berlin Heidelberg. 2022;13(9). doi: 10.1007/s12652-021-03325-3.
18. Kalid N *et al.* Based on Real Time Remote Health Monitoring Systems: A New Approach for Prioritization 'Large Scales Data' Patients with Chronic Heart Diseases Using Body Sensors and Communication Technology. J. Med. Syst. 2018;42(4). doi: 10.1007/s10916-018-0916-7.
19. Karunanayake N. Next-generation agentic AI for transforming healthcare. Informatics and Health. Apr. 2025;2. doi: 10.1016/j.infoh.2025.03.001.
20. Angulo J, Yeste V. Agentic MCS: A Multilingual Clinical Summarization Framework. in CEUR Workshop Proceedings. 2025;136–147. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105019039482&partnerID=40&md5=5c8c9f181a7babc4d2d3e0e10ee0e5b1
21. Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. Int. J. Med. Inform. 2021;151:104484. doi: 10.1016/j.ijmedinf.2021.104484.
22. Salmi N, Rustam Z. Naïve Bayes Classifier Models for Predicting the Colon Cancer. IOP Conf. Ser. Mater. Sci. Eng. 2019;546(5). doi: 10.1088/1757-899X/546/5/052068.

23. Zaben SO. IDC-insight: boosting intrusion detection accuracy in IoT networks with Naïve Bayes and multiple classifiers. International Journal of Information Technology (Singapore). 2025;17(2):869–881. doi: 10.1007/s41870-024-02026-2.
24. Konstantinov AV, Utkin LV. Interpretable machine learning with an ensemble of gradient boosting machines. Knowl. Based. Syst. 2021;222:106993. doi: 10.1016/j.knosys.2021.106993.
25. Murphy MP, Brown NM. CORR Synthesis: When Should the Orthopaedic Surgeon Use Artificial Intelligence. Machine Learning, and Deep Learning? 2021. doi: 10.1097/CORR.0000000000001679.
26. Kang YJ, Il Yoo J, Cha YH, Park CH, Kim JT. Machine learning–based identification of hip arthroplasty designs. J. Orthop. Translat. 2020;21:13–17. doi: 10.1016/j.jot.2019.11.004.
27. Atik OŞ. Artificial intelligence, machine learning, and deep learning in orthopedic surgery. 2022. doi: 10.52312/jdrs.2022.57906.
28. Sniderman J, Stark RB, Schwartz CE, Imam H, Finkelstein JA, Nousiainen MT. Patient Factors That Matter in Predicting Hip Arthroplasty Outcomes: A Machine-Learning Approach. Journal of Arthroplasty. 2021;36(6):2024–2032. doi: 10.1016/j.arth.2020.12.038.
29. Khosravi B *et al.* Patient-specific Hip Arthroplasty Dislocation Risk Calculator: An Explainable Multimodal Machine Learning– based Approach. Radiol. Artif. Intell. Nov. 2022;4(6):e220067. doi: 10.1148/ryai.220067.
30. Thenuwara SS, Premachandra C, Kawanaka H. A multi-agent based enhancement for multimodal biometric system at border control. Array. 2022;14:100171. doi: 10.1016/j.array.2022.100171.
31. Yordanova Z, Hristozov Y. The Evolution Of Financial Analysis: From Manual Methods To AI and AI Agents. Economics - Innovative and Economics Research Journal. 2025;13(3):219–239. doi: 10.2478/eoik-2025-0063.
32. Jayachandran JJB, Ambigapathy S, Abirami P, Ishwaryalakshmi K. X - Ray Image Analysis in Identification of Bone Cancer Using Laws Features and Machine Learning Model. in 2022 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2022. 2022:1–5. doi: 10.1109/ICDSAAI55433.2022.10028844.
33. Calvaresi D, Marinoni M, Dragoni AF, Hilfiker R, Schumacher M. Real-time multi-agent systems for telerehabilitation scenarios. Artif. Intell. Med. 2019;96:217–231. doi: 10.1016/j.artmed.2019.02.001.
34. Dumitrescu C, Radu V, Gheorghe R, Tăbîrcă AI, Ştefan MC, Manea L. Crowd Panic Behavior Simulation Using Multi-Agent Modeling. Electronics (Switzerland). Sep. 2024;13(18). doi: 10.3390/electronics13183622.
35. Piccialli F, Chiaro D, Sarwar S, Cerciello D, Qi P, Mele V. AgentAI: A comprehensive survey on autonomous agents in distributed AI for industry 4.0. Expert Syst. Appl. 2025;291. doi: 10.1016/j.eswa.2025.128404.
36. Di Domenico J, Catalano M, Riccetti L. Scaling and forecasting in a data-driven agent-based model: Applications to the Italian macroeconomy. Econ. Model. 2025;147:107046. doi: https://doi.org/10.1016/j.econmod.2025.107046.
37. Bemthuis RH, Lazarova-Molnar S. Towards integrating process mining with agent-based modeling and simulation: State of the art and outlook. Expert Syst. Appl. 2025;281. doi: 10.1016/j.eswa.2025.127571.
38. Levy-Loboda T, Rav-Acha M, Katz A, Nissim N. Cardio-ML: Detection of malicious clinical programmings aimed at cardiac implantable electronic devices based on machine learning and a missing values resemblance framework. Artif. Intell. Med. 2021;122:102200. doi: 10.1016/j.artmed.2021.102200.
39. Chang AC, Limon A. Introduction to artificial intelligence for cardiovascular clinicians. in Intelligence-Based Cardiology and Cardiac Surgery: Artificial Intelligence and Human Cognition in Cardiovascular Medicine, A. C. Chang, A. B. T.-I.-B. C., and C. S. Limon, Eds., Academic Press. 2023:3–120. doi: 10.1016/B978-0-323-90534-3.00010-X.
40. Khosravi B *et al.* Machine Learning and Deep Learning in Cardiothoracic Imaging: A Scoping Review. Diagnostics. Oct. 2022;12(10). doi: 10.3390/diagnostics12102512.
41. Shemanski KA *et al.* Gender representation among leadership at national and regional cardiothoracic surgery organizational annual meetings. Journal of Thoracic and Cardiovascular Surgery. 2021;161(3):733–744. doi: 10.1016/j.jtcvs.2020.11.157.
42. Sargazi M, Taghian F. Royal Jelly bioactive compounds and exercise improved cardiomyopathy and redox status in patients under methadone maintenance therapy: Randomized clinical trial. J. Herb. Med. 2023;38:100634. doi: 10.1016/j.hermed.2023.100634.
43. Rajaganapathi R, Mahendran R, Sivaganesan D, Vadivel MrR, Joel MR, Kannan V. An IoT enabled computational model and application development for monitoring cardiovascular risks. e-Prime - Advances in Electrical Engineering, Electronics and Energy. 2024;8:100513. doi: https://doi.org/10.1016/j.prime.2024.100513.
44. Cosoli G, Spinsante S, Scardulla F, D'Acquisto L, Scalise L. Wireless ECG and cardiac monitoring systems: State of the art, available commercial devices and useful electronic components. Measurement (Lond). 2021;177:109243. doi: 10.1016/j.measurement.2021.109243.
45. Anuhya Ardeti V *et al.* Development of real time ECG monitoring and unsupervised learning classification framework for cardiovascular diagnosis. Biomed. Signal Process. Control. 2024;88:105553. doi: 10.1016/j.bspc.2023.105553.
46. Keller JA *et al.* Different cardiovascular risk factors are related to distinct white matter hyperintensity MRI phenotypes in older adults. Neuroimage Clin. 2022;35:103131. doi: 10.1016/j.nicl.2022.103131.
47. Chen X, Zheng C, Golestanirad L. Application of Machine learning to predict RF heating of cardiac leads during magnetic resonance imaging at 1.5 T and 3 T: A simulation study. Journal of Magnetic Resonance. 2023;349:107384. doi: 10.1016/j.jmr.2023.107384.
48. Peng F *et al.* Management and Treatment of COVID-19: The Chinese Experience. Canadian Journal of Cardiology. 2020;36(6):915–930. doi: 10.1016/j.cjca.2020.04.010.
49. Bartlett VL *et al.* Physical activity, patient-reported symptoms, and clinical events: Insights into postprocedural recovery from personal digital devices. Cardiovasc. Digit. Health J. 2021;2(4):212–221. doi: 10.1016/j.cvdhj.2021.06.002.
50. Bedzra EKS, Goh JL. The Residency Selection Process in the New World; Time for a Rethink. Semin. Thorac. Cardiovasc. Surg. 2021;33(3):825–829. doi: 10.1053/j.semtcvs.2020.12.011.
51. Jia J, Guo J, Liu X, Li G, Liu Y, Zhao X. Impact of serum phosphate on severity and functional outcomes after ischemic stroke in young adults. Nutrition, Metabolism and Cardiovascular Diseases. 2022;32(11):2553–2560. doi: 10.1016/j.numecd.2022.08.008.
52. Dapamede T, Li F, Khosravi B, Purkayastha S, Trivedi H, Gichoya J. DICOM LUT is a Key Step in Medical Image Pre-

processing Towards AI Generalizability. Journal of Imaging Informatics in Medicine. 2025;0123456789. doi: 10.1007/s10278-025-01418-5.
53. Trojani V, Bassi MC, Verzellesi L, Bertolini M. Impact of Preprocessing Parameters in Medical Imaging-Based Radiomic Studies: A Systematic Review. Cancers (Basel). 2024;16(15). doi: 10.3390/cancers16152668.
54. Hanlon M, Jackson B, Rice J, Walsh J, Riordan D. Audio Pre-Processing and Neural Network Models for Identification of Orthopedic Reamers in Use. 2020. doi: 10.1109/ISSC49989.2020.9180175.
55. Mohammed TJ, Hadi SM, Albahri AS. Improving the Prioritization Procedure of Patients With Covid-19 in Hospitals Based on Decision-Making Techniques: a Systematic Review. Iraqi Journal for Computers and Informatics. 2022;48(1):6–12. doi: 10.25195/ijci.v48i1.347.
56. Vandin A, Giachini D, Lamperti F, Chiaromonte F. Automated and distributed statistical analysis of economic agent-based models. J. Econ. Dyn. Control. 2022;143:104458. doi: https://doi.org/10.1016/j.jedc.2022.104458.
57. Chan A *et al.* Infrastructure for AI Agents. Transactions on Machine Learning Research. 2025; 2025-May, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105006738477&partnerID=40&md5=ecdf866bbe1063afa8de5235768e2d63
58. Song X, Tang H. Blood Cell Target Detection Based on Improved YOLOv5 Algorithm. Electronics (Switzerland). Dec. 2024;13(24). doi: 10.3390/electronics13244992.
59. Mohammed TJ, Xinying C, Albahri AS, Alnoor A, Khai Wah K. A Novel Trusted Framework for Orthopedic Disease Detection With Reliability Against Adversarial Attacks. IEEE Access. 2025;13:160193–160220. doi: 10.1109/ACCESS.2025.3609330.