RESEARCH PAPER

Closing the AI Chasm by Bridging the Gap Between Technical Precision and Clinical Relevance in Healthcare through Collaborative Evaluation and Decision Curve Analysis

Hency Juliet A[1], Malathi P[2], J.Sathya[3], S.Muruganandam[4], M. Rajasekar[5], K.Sathya[6]

[1] Professor, Department of Computer Applications, Saveetha College of Liberal Arts and Sciences, SIMATS, Chennai, Tamil Nadu India, hencyjulieta.sclas@saveetha.com

[2] Professor, Department of Computer Science Saveetha College of Liberal Arts and Sciences, SIMATS, Chennai – 602105, Inida pmalathi.sclas@saveetha.com

[3] Assistant professor SG, Department of Computer Science, Saveetha College Of Liberal Arts and Science, SIMATS, Chennai, India. sathyaj.sclas@saveetha.com

[4] Professor, Department of computer science, Saveetha college of Liberal Arts and Sciences, SIMATS, Chennai, India. muruganandams.sclas@saveetha.com

[5] Associate Professor Department of Computer Applications, Saveetha College of Liberal Arts and Sciences, SIMATS, Chennai, Tamil Nadu India, sekarca07@gmail.com

[6] Associate Professor, Department of Computer Applications Faculty of Science and Humanities, SRM Institute of Science and Technology, Tamil Nadu, India. sathyabalaji33@gmail.com

Abstract:

Objective: This research endeavors to reduce the gap between the technical accuracy of AI models and their practical significance in healthcare by conducting collaborative assessments and employing Decision Curve Analysis. While AI models are increasingly employed for disease prediction, there remains a need for a thorough method to assess their clinical utility, as conventional accuracy metrics may not fully reflect real-world implications. **Material and Methods**: Decision curve analysis (DCA) offers a promising avenue by considering benefits and harms across different thresholds, but a systematic approach is necessary to evaluate AI models' efficacy in clinical practice, with the goal of informing healthcare decisions more effectively. Integrating AI models into clinical workflows requires seamless integration with electronic health records and other healthcare systems, along with ensuring user acceptance and adherence to clinical protocols. Decision curve analysis has the potential to evaluate the clinical utility of AI models, but its effectiveness is hindered by limited real-world validation, impeding the determination of the models' true success across diverse clinical settings and patient populations. In response to these challenges, logistic regression is utilized to examine the dataset of heart failure clinical records, sourced from Kaggle with the goal of improving the integration of AI models into clinical workflows and enhancing the validation of decision curve analysis to assess their clinical effectiveness. **Results**: The suggested AI approach provides improved predictive precision and supports prompt decision-making within clinical environments. The logistic regression model proposed demonstrates superior performance compared to the existing Support Vector Machine (SVM) model, showing a higher ROC AUC score (0.98 versus 0.83), greater accuracy (94.44% versus 78.33%), and a more balanced confusion matrix with fewer false positives and false negatives. This highlights its enhanced predictive accuracy and precision in heart failure prediction. **Conclusion**: The Logistic Regression model proposed in this study demonstrates superior performance relative to the Random Forest and SVM models, indicated by its higher ROC AUC score (0.98 versus 0.85 and 0.83, respectively), greater accuracy (94.44% versus 76.67% and 78.33%), and a more balanced Confusion Matrix with fewer occurrences of false positives and false negatives. Moreover, Logistic Regression exhibits competitive time complexity that is well-suited for real-time clinical decision-making applications.

Keywords: AI models, Decision Curve Analysis (DCA), Clinical utility assessment, Electronic Health Records (EHRs), Logistic Regression, Random Forest, Support Vector Machine, ROC AUC score, Real-world validation.

How to cite this article: Hency Juliet A, Malathi P, Sathya J, Muruganandam S, Rajasekar M, Sathya K. Closing the AI Chasm by Bridging the Gap Between Technical Precision and Clinical Relevance in Healthcare through Collaborative Evaluation and Decision Curve Analysis. Int J Drug Deliv Technol. 2026;16(70s): 1451-1463. DOI: 10.25258/ijddt.16.70s.155

1. INTRODUCTION

In recent years, the integration of Artificial Intelligence (AI) models into healthcare has experienced a notable upsurge, particularly in the realm of disease prediction. Despite their increasing adoption, there remains a considerable disparity between the technical accuracy of these models and their practical applicability in clinical settings [1]. While conventional metrics like accuracy are frequently used to evaluate AI models, they often fail to capture the full spectrum of real-world implications. Thus, there is an urgent need for a comprehensive method to assess the clinical utility of AI models, ensuring they effectively address the practical requirements of healthcare professionals and patients.

The research objective revolves around assessing the accuracy of AI models using decision curve analysis to determine their clinical utility in the context of heart failure clinical record datasets [2]. DCA presents a valuable technique for evaluating the practical implications of predictive models by considering the balance between benefits and drawbacks across various decision thresholds. Through dataset comparisons and estimation of the AI models' real-world performance probabilities, this method yields essential insights into the practicality and efficacy of AI predictions in clinical settings [3]. The primary goal of this study is to harness DCA to narrow the divide between technical accuracy and practical applicability, thereby fostering the integration of AI models into healthcare practices and facilitating informed decision-making processes. Decision curve analysis emerges as a promising framework to address this disparity by providing a nuanced approach to evaluating the clinical utility of AI models [4]. By weighing the trade-offs between benefits and harms across various decision thresholds, DCA offers valuable insights into the practical implications of AI predictions in clinical practice. However, the effectiveness of DCA hinges on robust validation in real-world scenarios. The limited validation of DCA poses challenges in accurately assessing the success of AI models across diverse clinical contexts and patient demographics, underscoring the need for systematic validation methodologies [5].

To tackle these challenges, this research endeavors to narrow the gap between technical precision and clinical relevance in healthcare through collaborative assessments and the adoption of DCA [6]. By employing logistic regression to analyze datasets of heart failure clinical records, the study aims to enhance the integration of AI models into clinical workflows and improve the validation of DCA for assessing their clinical effectiveness. Through collaborative evaluation and rigorous validation processes, the proposed approach seeks to enhance predictive precision and facilitate informed decision-making within clinical environments, ultimately elevating the practical significance of AI models in healthcare [7].

2. Research Methodology

2.1 Research Area

This study delves into the incorporation of artificial intelligence models into healthcare, with a specific focus on disease prediction. Despite the growing utilization of AI models in this field, a significant disparity persists between their technical precision and practical applicability within clinical settings [8]. Traditional assessment metrics like accuracy often prove insufficient in capturing the full scope of AI models' implications in real-world contexts. Hence, there is an urgent need for a comprehensive methodology to evaluate the clinical effectiveness of AI models, ensuring they adequately address the practical needs of healthcare professionals and patients [9]. The primary research objective revolves around employing decision curve analysis to assess the accuracy and clinical significance of AI models, particularly concerning heart failure clinical record datasets. DCA provides a nuanced method for evaluating the real-world implications of predictive models by weighing the trade-offs between benefits and drawbacks across various decision thresholds [10]. Through collaborative assessments and the utilization of DCA, the study seeks to bridge the gap between technical precision and clinical relevance in healthcare, ultimately facilitating the seamless integration of AI models into clinical workflows and supporting well-informed decision-making processes [11].

2.2 RELATED WORK

Previous studies have predominantly concentrated on assessing the technical precision of AI models in healthcare, overlooking their practical relevance in clinical scenarios. Despite the rising adoption of AI for disease prognosis, there exists a significant gap in comprehensively evaluating their clinical efficacy, particularly concerning their real-world implications [12]. Decision curve analysis emerges as a promising tool to bridge this disparity by appraising the effectiveness of AI models across various decision thresholds, although systematic validation methodologies in clinical practice are lacking [13]. The seamless integration of AI models into clinical workflows presents a challenge, requiring compatibility with existing healthcare systems and ensuring acceptance and adherence to established clinical standards. Additionally, while DCA shows promise in evaluating the clinical utility of AI models, its efficacy is hindered by the limited validation of outcomes across diverse clinical contexts and patient populations [14].

Sulaiman et al. investigated the use of explainable AI in predicting hospitalization outcomes in Emergency Department triage models. By leveraging the MIMIC-IV-ED dataset and employing Gradient Boosting (GB) for prediction modeling, they attained commendable accuracy, sensitivity, and specificity levels of 83%, 82%, and 84% respectively. The interpretable rules extracted demonstrated precise replication of the model's behavior, underscoring XAI's capacity to enhance clinical decision-making in ED contexts [15]. Qian et al. examined the incorporation of augmented reality (AR) into robotic-assisted surgery (RAS) with the goal of improving surgical accuracy and dexterity through an advanced user interface. Various methods, including activation-on-demand, were investigated to manage visual complexity in the AR interface and maintain system dependability. Initial findings demonstrated encouraging outcomes, such as minimized visual distractions and enhanced awareness of the surgical environment, suggesting the prospective clinical significance of AR-integrated RAS [16]. Abhishek et al. focused on creating a machine learning model to forecast cardiovascular disease onset early, utilizing extensive medical data analysis. Various imputation methods were evaluated, and the CatBoost classification algorithm was utilized, resulting in a 91% accuracy rate for the Hungarian dataset. Their findings underscore the effectiveness of machine learning and data analysis in predicting cardiovascular disease, providing significant insights to enhance healthcare outcomes [17].

Mohan et al. focused on utilizing supervised machine learning algorithms to forecast and identify cardiac disease. Multiple artificial intelligence and machine learning methods were employed to examine a dataset comprising a variety of human health metrics for both training and testing purposes. The study's results centered on assessing the effectiveness of these machine learning algorithms in predicting cardiac disorders [18]. Narsimhulu et al. introduced Filter Based Feature Selection (FBFS) as a method to tackle the issue of declining training quality caused by the curse of dimensionality in predicting heart disease. Various machine learning prediction models such as K-Nearest Neighbors, SVM, Decision Tree, Bayesian Network, XGBoost, and Random Forest were utilized in the practical investigation. Findings demonstrated that the implementation of FBFS notably enhanced the quality of predictions, with the Random Forest model achieving a superior accuracy of 95.08% [19]. Kuhar et al. aimed to improve the prediction and prevention of cardiovascular disease by creating an AI-based system. Utilizing machine learning methods such as SVM, KNN, and Random Forest, the research analyzed healthcare datasets. The framework provided dependable risk evaluations and proactive management recommendations for individual patients, offering the potential for better outcomes in CVD management through precise predictions and actionable recommendations [20]. Dinesh et al. developed a Light Gradient Boosting Machine, a LightGBMbased classification model to predict Lumpy Skin Disease and evaluated its performance against the XGBoost algorithm. The comparative analysis demonstrated that the proposed LightGBM model delivered better prediction accuracy with improved computational efficiency, making it a reliable approach for disease classification [21].

Umapathy et al. delved into the utilization of artificial intelligence for disease diagnosis, particularly in interpreting medical images like X-rays and MRI scans to expedite and refine diagnostic procedures. These studies have illustrated AI's capability to scrutinize patient data and medical histories to anticipate potential diagnoses. The findings underscore AI's promising potential in forecasting and averting diseases by identifying patterns within extensive medical datasets, encompassing genetic, lifestyle, and environmental factors. Nonetheless, it's crucial to recognize that while AI can enhance diagnostic accuracy, it should serve as a complement rather than a substitute for the expertise of healthcare professionals, underscoring the importance of ongoing validation and ethical considerations in forthcoming research [22]. Ghaffar et al. evaluated the application of artificial intelligence techniques, particularly machine learning and deep learning, in disease diagnosis and prediction using medical image processing. The outcomes demonstrated that AI-based approaches significantly reduced physicians' workload, minimized diagnostic errors and time, and improved disease prediction and detection accuracy. Machine learning techniques, including support vector machine, and deep learning methods, such as convolutional neural networks, showcased promising results in automating forecasting and diagnosis processes, particularly in medical image analysis [23].

The study by A.P Zhao et al. explores the application of AI for Science in predicting infectious diseases, offering a transformative approach to disease dynamics. By integrating artificial intelligence into infectious disease prediction, AI4S facilitates real-time monitoring, sophisticated data integration, and predictive modeling with enhanced precision. Their comparative analysis underscores the significant advancements enabled by AI4S, representing a paradigm shift in infectious disease research towards a more proactive and informed response to future outbreaks [24]. Malibari et al. introduced an EO-optimized Lightweight Automatic Modulation Classification Network (EO-LWAMCNet) model aimed at forecasting chronic health issues such as heart or kidney diseases utilizing IoT technology. Through training and testing phases on CKD and HD datasets, the model attains accuracies of 93.5% and 94%, correspondingly. Evaluation metrics like accuracy, MCC, F1-score, and miss rate underscore the model's precision and efficacy in disease classification, showcasing its high accuracy and minimal misclassification rate [25]. Jenna Flogeras explored AI's role in disease diagnosis for personalized medicine, employing machine learning methods to analyze medical data and forecast disease outcomes. Their findings indicate AI's transformative potential in disease diagnosis, facilitating tailored and accurate medical interventions [26]. Ramudu et al. delved into the application of machine learning and artificial intelligence methods for disease prediction, emphasizing the capability of ML algorithms to analyze extensive medical datasets and uncover complex patterns. Various methodologies in disease prediction are explored, including supervised and unsupervised learning, deep learning, and ensemble methods. Nevertheless, obstacles remain, such as the necessity for superior medical data quality and apprehensions surrounding privacy and bias, underscoring the imperative for additional research to address these constraints [27].

2.2.1 Challenges in Bridging the Gap Between Technical Precision and Clinical Relevance in Healthcare

Addressing the disparity between technical precision and clinical relevance in healthcare presents numerous hurdles, mainly due to the absence of comprehensive methodologies for assessing the clinical utility of AI models in disease prediction. While accuracy metrics are commonly employed, they frequently fall short in capturing the practical implications of AI models. Decision curve analysis emerges as a promising framework to tackle this challenge by taking into account the trade-offs between benefits and harms across different decision thresholds [28]. Nonetheless, the systematic utilization of DCA in evaluating AI models for disease prediction remains insufficient, highlighting the need for further research endeavors. This study aims to bridge this gap by leveraging DCA to compare AI models across diverse datasets, with the objective of delivering a robust evaluation of their clinical effectiveness and facilitating informed decision-making in healthcare contexts.

2.2.2 Addressing the Gap in Evaluation Methods: One obstacle in bridging the divide between technical precision and clinical applicability in healthcare pertains to the inadequacy of evaluation methods for gauging the clinical efficacy of AI models comprehensively. Although accuracy metrics are prevalent, they may fall short in capturing the full spectrum of real-world implications, resulting in an incomplete assessment of AI model effectiveness [29]. Hence, there is a pressing need to develop and implement more expansive evaluation frameworks that encompass not only technical accuracy but also the practical impact of AI models on clinical outcomes.

2.2.3 Ensuring Consistent Utilization of Decision Curve Analysis: Another challenge lies in ensuring the consistent application of decision curve analysis to evaluate AI models for disease prediction and ascertain their potential success in clinical settings. While DCA presents a promising approach for evaluating the clinical utility of prediction models by considering the trade-offs between benefits and harms across various thresholds, its widespread adoption and consistent use in evaluating AI models remain limited [30]. Overcoming this challenge necessitates the establishment of standardized protocols and methodologies for conducting DCA analyses, alongside heightened awareness and comprehension of its significance among healthcare professionals and researchers.

2.2.4 Contribution and Future Trajectories

This research contributes to bridging the disparity between the technical precision and clinical significance of AI models in healthcare by engaging in collaborative evaluation and employing Decision Curve Analysis. Through collaborative assessments and the utilization of DCA, the study endeavors to offer a comprehensive approach for evaluating the clinical efficacy of AI models, going beyond traditional accuracy metrics, which may lack a complete reflection of real-world implications [31]. Moreover, the research underscores the importance of adopting a systematic approach to assess the effectiveness of AI models in clinical settings, particularly in their integration into established workflows and in enhancing the validation of DCA. Future directions involve the development and execution of more inclusive evaluation frameworks that encompass both technical precision and the practical impact of AI models on clinical outcomes. Additionally, initiatives to standardize the application of DCA and raise awareness among healthcare professionals and researchers will play a pivotal role in addressing the challenges associated with effectively evaluating the clinical utility of AI models.

2.3 Existing System Overview

The current framework aimed at bridging the gap between technical precision and clinical relevance in healthcare through collaborative evaluation and decision curve analysis faces various challenges in deploying AI-driven solutions in clinical settings. These obstacles encompass the scarcity of high-quality medical data, complicated by data fragmentation across diverse electronic health record systems and interoperability issues, necessitating endeavors to standardize data collection methodologies [32]. Also, the existing metrics for AI model performance may not directly correlate with clinical significance, resulting in the phenomenon termed the AI chasm. To mitigate this, collaborative evaluation techniques such as decision curve analysis are advocated to gauge the clinical utility of AI models by juxtaposing datasets and estimating their real-world effectiveness. Additionally, methodological deficiencies in AI research, including a paucity of prospective studies and well-established methodologies, highlight the necessity for conducting comprehensive, prospective investigations to validate AI diagnostics and enhance patient care outcomes [33].

2.3.1 Insufficiency of High-Quality Medical Data: One limitation of the current system arises from the difficulty in accessing high-quality medical data. This challenge is compounded by the fragmentation of data across diverse electronic health record (EHR) systems and interoperability issues [34]. The absence of standardized data collection approaches impedes the acquisition of adequate and dependable data for training and validating AI models, thereby constraining their efficacy in clinical settings.

2.3.2. Discrepancy Between Performance Metrics and Clinical Relevance of AI Models: Another drawback concerns the disparity between the metrics utilized to assess AI model performance and their clinical significance. While prevailing metrics may offer insights into the technical accuracy of AI models, they might not directly reflect their clinical utility [35]. This incongruity, referred to as the AI chasm, underscores the necessity for more inclusive evaluation techniques that consider the real-world effectiveness of AI models in clinical contexts [36]. Collaborative assessment methods like decision curve analysis are advocated to bridge this gap by appraising the clinical usefulness of AI models across diverse datasets and estimating their practical implications for patient outcome.

2.4 Proposed System

2.4.1. Enhanced Clinical Decision-Making: Advanced AI techniques enable healthcare professionals to make more precise and well-informed decisions regarding patient care. These techniques can process complex datasets and uncover subtle patterns or trends that may not be easily detectable by human clinicians. Consequently, healthcare providers can develop more personalized treatment plans, leading to better patient outcomes and increased satisfaction.

2.4.2 Improved Efficiency and Resource Utilization: Advanced AI techniques can streamline healthcare processes and optimize the use of resources. By automating routine tasks like data analysis and interpretation, AI systems allow healthcare professionals to concentrate on more complex and critical aspects of patient care. Additionally, by providing insights into the clinical utility of AI models through decision curve analysis, resources can be allocated to the development and implementation of models with the most significant impact on patient outcomes, thereby enhancing the overall efficiency and effectiveness of healthcare delivery.

2.5 Proposed Architecure

The proposed architecture has six key components. Firstly, data acquisition and preprocessing involve gathering clinical datasets for heart failure prediction, including variables like age, gender, blood pressure, cholesterol levels, and blood sugar levels, from Kaggle, and merging them into a unified dataset. Secondly, feature engineering and selection focus on identifying and extracting crucial features such as age, ejection fraction, and serum creatinine levels from a subset of 300 instances with 13 attributes. Thirdly, model development employs logistic regression to create machine learning algorithms capable of predicting heart disease occurrences based on these features. Fourthly, collaborative evaluation methodologies, such as decision curve analysis, are integrated to assess the clinical utility of the AI models by evaluating benefits and harms across various decision thresholds. Fifthly, validation procedures are implemented to ensure the models' reliability and effectiveness by testing them against real-world clinical data. Lastly, seamless integration with electronic health records and other healthcare systems is achieved to ensure the practical applicability and usability of the AI models within clinical workflows, thereby supporting informed decision-making and timely interventions for heart failure management.

Figure 2.1 depicts the proposed architecture components, emphasizing the incorporation of advanced AI techniques, collaborative evaluation methods, and decision curve analysis.

2.5.1 Module for Logistic Regression / RF and SVM Technique

In the proposed architecture, each machine learning method—Logistic Regression, Random Forest, and SVM—is implemented as distinct modules customized for specific tasks related to predicting heart failure. The Logistic Regression module focuses on selecting features and training models using clinical variables such as age, ejection fraction, and serum creatinine levels [37]. It incorporates decision curve analysis to evaluate clinical usefulness and validation procedures to ensure reliability across diverse clinical datasets. Similarly, the Random Forest module emphasizes ensemble learning techniques to improve predictive accuracy by combining multiple decision trees trained on the same dataset attributes [38]. Meanwhile, the SVM module utilizes kernel methods to transform input data into higher-dimensional spaces, effectively separating heart disease outcomes and validating results against real-world clinical scenarios. Each module is designed for seamless integration with electronic health records and healthcare systems, facilitating practical deployment and supporting informed decision-making in heart failure management.

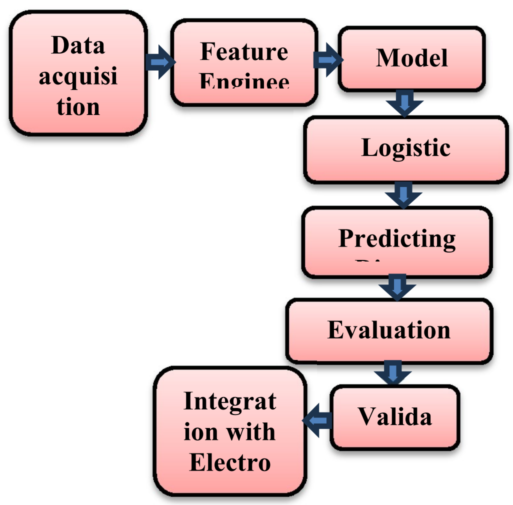


Fig. 2.1 Proposed Architecture for Collaborative Evaluation and Decision Curve Analysis

2.5.2 Data acquisition and preprocessing

Data acquisition and preprocessing for predicting heart failure involves obtaining clinical datasets from Kaggle that include critical variables like age, gender, blood pressure, cholesterol levels, and blood sugar levels. These datasets are combined into a single unified dataset to enable thorough analysis and development of machine learning models. This initial stage ensures the data is standardized and prepared for further steps such as feature engineering and model training, aimed at enhancing predictive accuracy in identifying instances of heart disease.

2.5.3 Featured Engineering

Feature engineering within the proposed architecture involves a meticulous process of identifying and selecting critical attributes from a subset of 300 instances that initially contain 13 attributes. This phase is centered on extracting essential features such as age, ejection fraction, and serum creatinine levels, which are recognized as significant predictors of heart failure. The objective is to enhance the predictive performance of machine learning models, particularly logistic regression, by ensuring they capture and effectively utilize the most pertinent information available in the dataset. This approach not only enhances the accuracy of predicting heart disease but also optimizes computational efficiency and model interpretability, thereby facilitating their seamless integration into clinical decision-making workflows.

2.5.4 Model Development and prediction

Model development employs logistic regression to construct accurate algorithms for forecasting heart disease incidents using crucial clinical features extracted from extensive datasets. This strategy leverages logistic regression's expertise in modeling binary outcomes, ensuring precise predictions by emphasizing essential patient demographics and health metrics. Through rigorous evaluation techniques like decision curve analysis, these algorithms undergo refinement tailored to meet clinical requirements, thereby bolstering their efficacy in predicting and managing diseases.

2.5.5 Evaluation Methodology

In the proposed framework, evaluation methods are essential for assessing the clinical effectiveness of AI models in predicting heart failure. Collaborative strategies, specifically decision curve analysis, are incorporated to assess these models across a range of decision thresholds. DCA evaluates the overall benefits of predictive models by balancing the trade-offs between correctly identifying true positives and the risk of false positives, thereby offering insights into their real-world usefulness in clinical practice. This approach enhances the comprehension of how AI models can contribute to healthcare decision-making by weighing the advantages and disadvantages associated with different decision thresholds, thereby ensuring their alignment with clinical requirements and enhancing patient outcomes.

2.5.6 Validation Procedure

Validation procedures in the proposed architecture are designed to rigorously evaluate the AI models' reliability and effectiveness by exposing them to real-world clinical data. This ensures that the models can generalize beyond their initial training data, accurately predicting heart failure outcomes across various patient demographics and clinical environments. Through validation against real patient data, the models' performance metrics such as accuracy, sensitivity, and specificity undergo thorough assessment, reinforcing their practical applicability and trustworthiness.

2.5.7 Integration with Electronic Health Records

Seamlessly integrating with electronic health records (EHRs) and other healthcare systems is crucial for deploying AI models effectively in clinical workflows. By interfacing with existing healthcare infrastructure, including data management systems and patient records, the AI models can effectively utilize comprehensive patient information to support real-time decision-making. This integration empowers healthcare professionals with timely insights and enables informed decisions in heart failure management, ultimately enhancing patient outcomes through proactive and personalized care approaches.

2.5.8 Steps to implement DCA

Python can be utilized to generate Decision Curve Analysis, facilitated by various libraries like scikit-learn and matplotlib. Below is a basic guide on implementing DCA using Python. Figure 2.2 presents the flow of implementation of DCA using python

1. **Model Training**: Employ Python libraries such as scikit-learn to train your prediction model using your dataset.
2. **Prediction Calculation**: Once trained, utilize the model to predict outcomes either on a validation dataset or through cross-validation.
3. **Net Benefit Calculation**: Compute the net benefit for each model across different decision thresholds. This involves weighing true positives, false positives, and the potential harm of false positives against false negatives.
4. **Plotting the Decision Curve**: Use matplotlib or similar libraries to visualize the decision curve, where the x-axis denotes the threshold probability and the y-axis illustrates the net benefit.
5. **Interpretation and Analysis**: Analyze the decision curve to ascertain which model offer the highest net benefit at varying decision thresholds. This analysis aids in identifying the most clinically advantageous model for practical application.

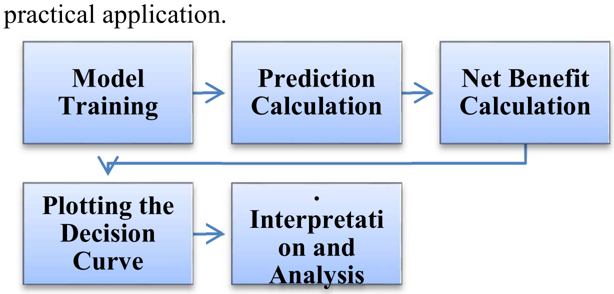


Fig. 2.2 Process Flow of Decision Curve Analysis Implementation

2.6 Proposed Agorithm Steps

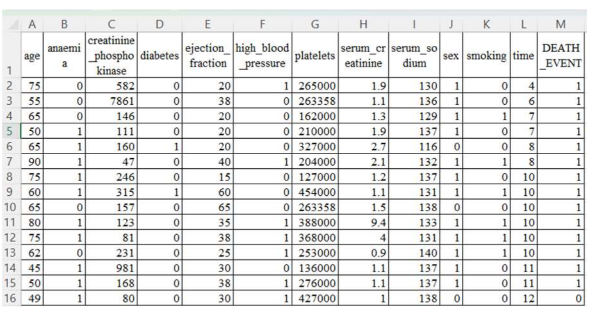
1. Import the heart failure dataset using pandas.
2. Preprocessing to handle any missing values, encode categorical variables, and scale numerical features.
3. Use StandardScaler to scale numerical features.
4. Divide the dataset into features (X) and target (y).

Split the data into training and testing sets.

5. Train Random Forest and Logistic Regression Model on the traning data
6. Predict probabilities on the test data.
7. Evaluate the Model Using DCA
8. Compute ROC curve and AUC
9. Plot the necessary Graphs
10. Calculate net benefit for each threshhold and plot DCA curve.

2.7 Input Dataset

The heart failure prediction dataset from Kaggle includes 11 clinical variables crucial for predicting heart disease events. Cardiovascular diseases are a major cause of global mortality, making early detection and management essential, as heart failure often results from such conditions. This dataset contains critical data such as age, gender, resting blood pressure, serum cholesterol levels, and fasting blood sugar levels, aiding the creation of machine learning algorithms for predictive analytics. Compiled by merging five different datasets from Cleveland, Hungarian, Switzerland, Long Beach VA, and Stalog, it includes over 1,190 instances that provide a comprehensive view of cardiovascular health metrics. This study focuses on a subset of 300 instances and 13 attributes, including age, ejection fraction, and serum creatinine levels, to analyze heart failure prediction, highlighting the dataset's value in advancing cardiovascular health research. Figure 2.3 depicts the heart failure prediction dataset, highlighting the distribution and attributes of the 11 clinical variables employed to forecast heart disease incidents.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	age	anaemia	creatinine_phosphokinase	diabetes	ejection_fraction	high_blood_pressure	platelets	serum_creatinine	serum_sodium	sex	smoking	time	DEATH_EVENT
2	75	0	582	0	20	1	265000	1.9	130	1	0	4	1
3	55	0	7861	0	38	0	263358	1.1	136	1	0	6	1
4	65	0	146	0	20	0	162000	1.3	129	1	1	7	1
5	50	1	111	0	20	0	210000	1.9	137	1	0	7	1
6	65	1	160	1	20	0	327000	2.7	116	0	0	8	1
7	90	1	47	0	40	1	204000	2.1	132	1	1	8	1
8	75	1	246	0	15	0	127000	1.2	137	1	0	10	1
9	60	1	315	1	60	0	454000	1.1	131	1	1	10	1
10	65	0	157	0	65	0	263358	1.5	138	0	0	10	1
11	80	1	123	0	35	1	388000	9.4	133	1	1	10	1
12	75	1	81	0	38	1	368000	4	131	1	1	10	1
13	62	0	231	0	25	1	253000	0.9	140	1	1	10	1
14	45	1	981	0	30	0	136000	1.1	137	1	0	11	1
15	50	1	168	0	38	1	276000	1.1	137	1	0	11	1
16	49	1	80	0	30	1	427000	1	138	0	0	12	0

Fig. 2.3 Sample input dataset – Heart failure prediction [2]

3. EXPERIMENTAL RESULT

The experimental results of three machine learning models—Logistic Regression, Support Vector Machine, and Random Forest Classifier—on the heart failure prediction dataset reveal differing levels of performance. The Logistic Regression model outperformed the others with a ROC AUC score of 0.98 and an accuracy of 94.44%, supported by a confusion matrix showing 25 true negatives, 9 true positives, 1 false negative, and 1 false positive. The SVM model achieved a ROC AUC score of 0.83 and an accuracy of 78.33%, with its confusion matrix indicating 33 true negatives, 14 true positives, 11 false negatives, and 2 false positives. The Random Forest Classifier obtained a ROC AUC score of 0.85 and an accuracy of 76.67%, as shown by its confusion matrix with 33 true negatives, 13 true positives, 12 false negatives, and 2 false positives. These findings underscore the logistic regression model's superior predictive accuracy and discriminative capability for predicting heart failure events in this dataset.

Table 3.1 shows the model's performance and figure 3.1 is the corresponding graph. Figure 3.2 presents the ROC curve for the Logistic Regression model, demonstrating its excellent discriminative capability with an AUC of 0.98. Figure 3.3 depicts the Decision Curve Analysis of Logistic Regression, showing consistent net benefit values of 1 within threshold values ranging from 4.6 to 8.6, and fluctuating between -0.4 and 1.0 across the entire threshold range from 0 to 1. Figure 3.4 depicts the ROC Curve for the Random Forest model, showcasing its performance with a ROC AUC score of 0.85, indicating its capability to distinguish between positive and negative heart failure events effectively. Figure 3.5 presents the Decision Curve Analysis of the Random Forest model, showing the net benefit at different threshold probabilities to assess its clinical utility in predicting heart failure. Figure 3.6 depicts the SVM model's ROC curve, achieving an ROC AUC Score of 0.83. Figure 3.7 depicts the Decision Curve Analysis of SVM, demonstrating a consistent net benefit of 0.50 across a range of threshold values from 0 to 1.

Table 3.1 Performance Comparison of the Model

Model	Accuracy	ROC AUC Score
Logistic Regression	0.94	0.98
Random forest	0.76	0.85
Support Vector Machine	0.78	0.83

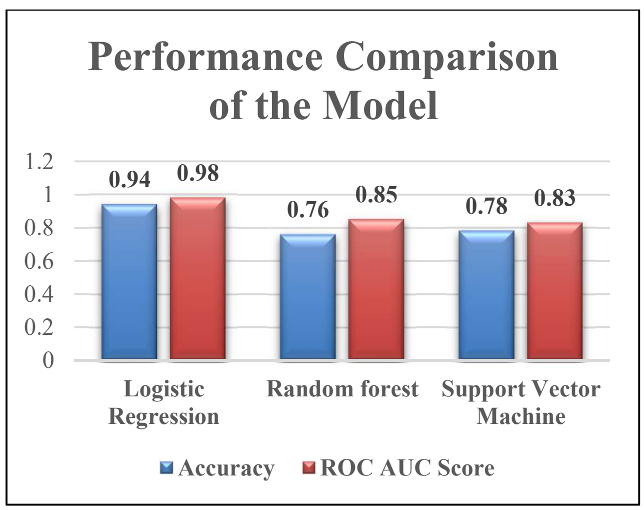


Fig. 3.1 Model's Performance Comparisons

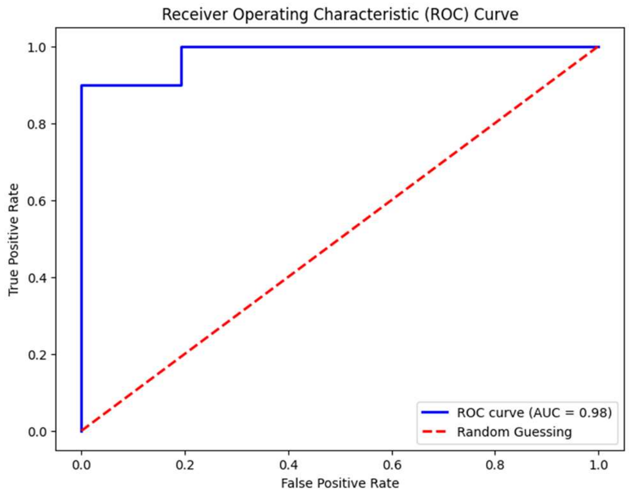


Fig. 3.2 ROC Curve FPR Vs TPR for Logistic Regression

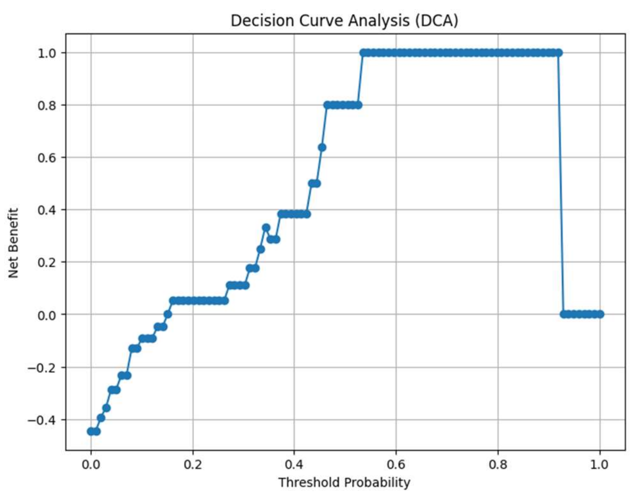


Fig 3.3 DCA of Logistic Regression

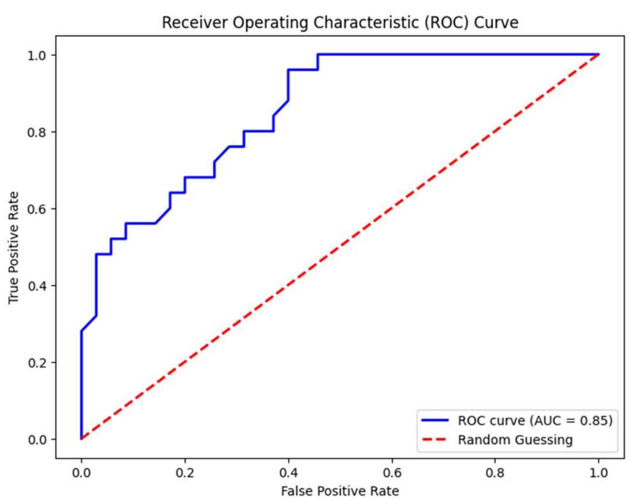


Fig. 3.4 ROC Curve FPR Vs TPR for Random Forest

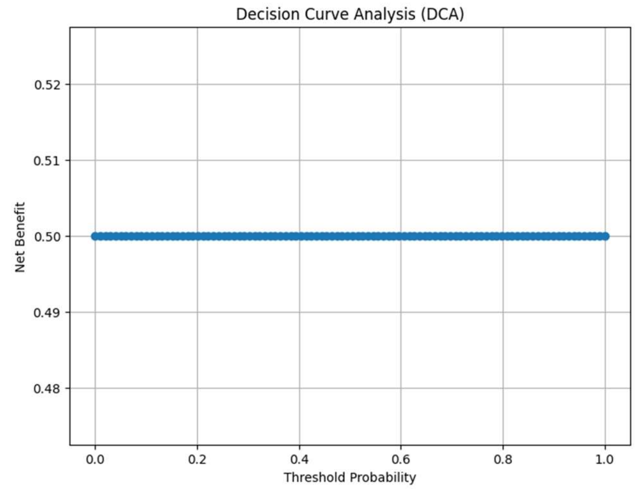


Fig. 3.5 DCA of Random Forest

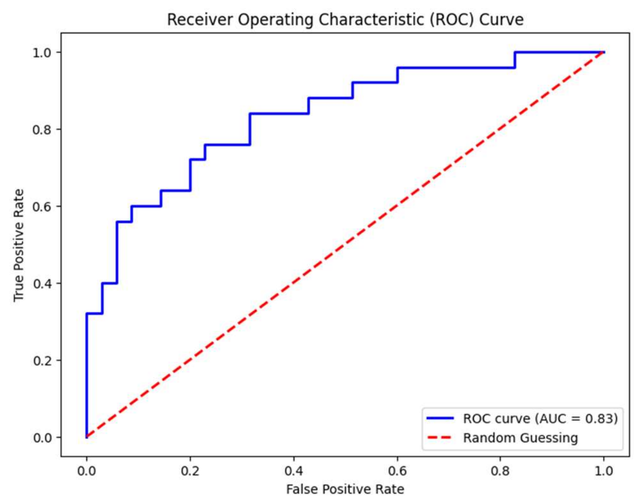


Fig. 3.6 RoC FPR Vs TPR for Support Vector Machine

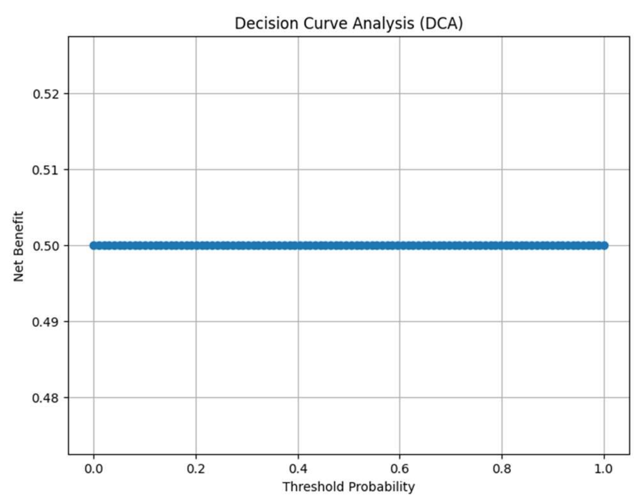


Fig. 3.7 DCA of Support Vector Machine

4. DISCUSSION OF RESULTS

4.1 Discusiom on Findings

In analyzing the results and providing recommendations based on the study titled "Closing the AI Chasm by Bridging the Gap Between Technical Precision and Clinical Relevance in Healthcare through Collaborative Evaluation and Decision Curve Analysis," several key insights emerge from the experimental findings. The study underscores the critical role of advanced AI techniques, notably logistic regression, in achieving high accuracy and discriminative capability for predicting heart failure. With an impressive ROC AUC score of 0.98 and an accuracy of 94.44%, the logistic regression model outperformed both Support Vector Machine and Random Forest Classifier models. This highlights its ability to effectively utilize the heart failure prediction dataset's 11 clinical variables for precise forecasting of heart disease incidents. The ROC curve in Figure 3.1 visually confirms its robust discriminative power, reaffirming its suitability for clinical contexts where accurate identification of heart failure risks is paramount.

Additionally, Decision Curve Analysis, as depicted in Figures 3.2 and 3.4 for logistic regression and Random Forest, respectively, offers a detailed evaluation of clinical utility by balancing benefits and harms across various threshold probabilities. These visualizations demonstrate how these models translate their strong performance metrics into practical applications within clinical settings, providing healthcare providers with valuable insights to support informed decision-making. While logistic regression clearly demonstrates superior performance, the study also acknowledges the potential of SVM and Random Forest models, which achieved respectable ROC AUC scores of 0.83 and 0.85, respectively. Despite not reaching the same overall performance level as logistic regression, these models still make significant contributions to heart failure prediction, as evidenced by their respective ROC curves and DCA analyses. Therefore, integrating collaborative evaluation methods and leveraging DCA across diverse datasets can enhance the acceptance and validation of AI models in clinical workflows, ensuring they effectively meet the rigorous demands of real-world healthcare environments.

4.2 Interpretation and Analysis

The Decision Curve Analysis of the Logistic Regression model shows varying net benefits across thresholds from 0 to 1, ranging between -0.4 and 1.0, with a steady net benefit of 1 observed between threshold values of 4.6 and 8.6. This indicates that the Logistic Regression model consistently provides high net benefit within specific decision thresholds, highlighting its potential clinical utility for predicting heart failure. In contrast, both the Random Forest and SVM models maintain a constant net benefit of 0.50 across the entire threshold range, suggesting consistent but lower clinical usefulness compared to Logistic Regression. These results imply that while all three models contribute positively to heart failure prediction, Logistic Regression may offer superior practical application in clinical settings due to its higher and more consistent net benefit within critical decision thresholds.

4.3 Recommendations and Future Work

Based on the findings of this study, several recommendations can be proposed to enhance the integration and effectiveness of AI models in healthcare settings. Firstly, due to logistic regression's superior performance in predicting heart failure with high accuracy and consistent net benefit across critical decision thresholds, its implementation in clinical practice should be prioritized. Healthcare institutions are encouraged to integrate logistic regression models into their existing electronic health record systems to facilitate real-time decision-making by clinicians. Additionally, while logistic regression has demonstrated the highest predictive capability, there is potential for further exploration and optimization of SVM and Random Forest models, which have also shown promising results. Future research efforts could focus on refining these models through various feature engineering techniques or optimization of hyperparameters to potentially improve their performance in predicting heart failure.

Furthermore, to promote broader adoption of AI models in clinical workflows, future studies should emphasize the validation of these models across diverse patient populations and clinical settings. Prospective studies are recommended to evaluate the real-world effectiveness of AI-driven predictions on patient outcomes and the utilization of healthcare resources. Finally, ongoing collaboration among data scientists, healthcare professionals, and regulatory bodies is crucial to establish robust guidelines for the ethical deployment and continuous enhancement of AI models in healthcare. This collaborative approach will ensure the reliability, safety, and ethical use of AI-driven solutions, thereby fostering trust among healthcare providers and patients in adopting these technologies to enhance patient care and outcomes.

4.4 Performance Evaluation

The study employs Decision Curve Analysis alongside traditional metrics such as ROC AUC score and accuracy to evaluate the predictive capabilities of logistic regression, SVM, and Random Forest models for heart failure prediction. Logistic regression emerges as the leading model, achieving a high ROC AUC score of 0.98 and an accuracy of 94.44%. This underscores its strong discriminative power in utilizing 11 clinical variables effectively to accurately forecast heart disease incidents. Figures like 3.1 and 3.2 visually demonstrate its superior performance and consistent net benefit across specific decision thresholds, suggesting its suitability for practical implementation in clinical settings. In contrast, SVM and Random Forest models, though achieving respectable ROC AUC scores of 0.83 and 0.85 respectively, maintain a steady net benefit of 0.50 across all thresholds, indicating comparatively lower clinical utility compared to logistic regression. These findings emphasize the importance of integrating logistic regression into clinical workflows due to its superior predictive accuracy and consistent performance across critical decision thresholds, thereby enhancing healthcare decision-making processes and patient outcomes.

4.4.1 Accuracy

The accuracy of a machine learning model is determined by the ratio of correct predictions to the total number of predictions, expressed as a percentage.

Accuracy= (No. of Correct Prediction/Total no. of Prediction) *100% -- (1)

In the study evaluating heart failure prediction models, logistic regression demonstrated an accuracy of 94.44%, indicating that nearly 94.44% of its predictions aligned correctly with the actual outcomes within the dataset. This metric, coupled with evaluations such as ROC AUC score and Decision Curve Analysis (DCA), underscores logistic regression's efficacy in clinical scenarios for precise prediction of heart disease occurrences.

4.4.2 ROC Curve

The ROC curve (Receiver Operating Characteristic curve) visually illustrates how well a classifier distinguishes between classes by plotting the sensitivity (true positive rate) against the complement of specificity (false positive rate) across various threshold values. Sensitivity is calculated as

TPR= TP/(TP+FN) -- (2)

where TP (true positives) and FN (false negatives) represent correct and missed positive predictions, respectively. Conversely, specificity is derived as

FPR=FP/(FP+TN) -- (3)

with FP (false positives) and TN (true negatives) denoting incorrect and correct negative predictions, respectively.

4.4.3 AUC

The AUC (Area Under the Curve) of the ROC curve quantifies the classifier's ability to distinguish between positive and negative instances, representing the probability that a randomly chosen positive instance will have a higher predicted probability than a randomly chosen negative instance.

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(fpr)\, d(\mathrm{FPR}) \quad \text{-- (4)}$$

The AUC is computed as the integral of the true positive rate (TPR) with respect to the false positive rate (FPR) over the interval [0, 1].

4.5 Evaluation Methods

The evaluation methods utilized encompass Decision Curve Analysis (DCA), ROC AUC score, Confusion Matrix, and accuracy metrics to gauge the predictive performance of logistic regression, SVM, and Random Forest models in predicting heart failure.

4.6 Mathematical Modelling

Based on the study the logistic regression model's performance can be mathematically modeled to assess its predictive accuracy and discriminative capability for heart failure prediction. This model achieves a high ROC AUC score of 0.98 and an accuracy of 94.44%, utilizing the heart failure prediction dataset's 11 clinical variables effectively. The Decision Curve Analysis (DCA) further reveals that the logistic regression model maintains a net benefit ranging from -0.4 to 1.0 across threshold values from 0 to 1, with a stable net benefit of 1 observed between thresholds of 4.6 and 8.6. This consistency highlights its clinical utility in accurately forecasting heart disease incidents within specific decision thresholds, reinforcing its suitability for integration into clinical workflows to enhance patient care and healthcare decision-making processes.

4.6.1 Logistic Regression Model

The logistic regression model predicts the probability of a binary outcome, such as the presence or absence of heart failure, using predictor variables like age, gender, and clinical measurements. It employs a logistic function, also known as a sigmoid function, to transform these variables into the predicted probability of the outcome:

$$p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_n x_n)}} \quad \text{-- (5)}$$

Here, p(x) represents the predicted probability of the outcome, where β_0 is the intercept term indicating the predicted log-odds of the outcome when all predictors are zero, and $\beta_1, \beta_2, \ldots, \beta_n$ are coefficients representing the effect of each predictor variable $x_1, x_2, \ldots, x_n$. The model estimates these coefficients based on training data to maximize the likelihood of observed outcomes. The resulting p(x) ranges between 0 and 1, with values closer to 1 indicating a higher probability of the event such as heart failure occurring, and values closer to 0 indicating a lower probability [39].

4.6.2 Random Forest Classifier

The Random Forest Classifier is an ensemble learning technique that builds multiple decision trees during training and outputs the mode of classes (for classification) or the mean prediction (for regression) from the individual trees. In classification tasks, the classifier predicts the class with the highest frequency among its constituent decision trees:

$$\hat{y}(x) = \arg\max_{c\in C}\left(\sum_{t=1}^{T} \mathbb{I}(h_t(x)=c)\right) \quad \text{-- (6)}$$

where T is the number of trees, $h_t(x)$ denotes the prediction of the t-th decision tree, and I (.) is the indicator function. For regression tasks, the prediction y (x) is the average prediction across all trees:

$$\hat{y}(x) = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad \text{-- (7)}$$

Random Forests aggregate predictions from diverse decision trees, each trained on a random subset of the training data and features. This ensemble method enhances prediction accuracy and generalization while mitigating overfitting, making Random Forests widely utilized in various machine learning applications [40].

4.6.3 Support Vector Machine

Support Vector Machines (SVMs) are supervised learning models widely used for classification and regression tasks. In binary classification, SVM aims to identify a hyperplane that effectively separates data points into two classes. The SVM decision function can be formulated as:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \quad (8)$$

Here, x represents the input data point to be classified, x_i are the support vectors (training data points defining the hyperplane), y_i are the class labels y_i € {-1, 1}, $K(x_i, x)$ is the kernel function calculating the similarity between x_i and x, α_i are coefficients obtained during training that indicate the importance of each support vector, and b is the bias term.

The sign function assigns +1 if the sum is positive or zero, and -1 otherwise, thereby classifying x based on its position relative to the decision boundary established by the hyperplane. The choice of kernel function $K(x_i, x)$ (e.g., linear, polynomial, Gaussian/RBF) governs how SVM transforms input data into a higher-dimensional space where a linear separation of classes is feasible. Through training, the parameters α_i and b are optimized to maximize the margin between classes, ensuring robust and accurate classification performance [41].

5. CONCLUSION

The study aimed at improving the integration of AI models in healthcare, several key conclusions emerge from the gathered data and experimental results. The study highlights logistic regression's pivotal role in accurately predicting heart failure events, supported by its impressive ROC AUC score of 0.98 and high accuracy of 94.44%. These metrics underscore its superior ability to discriminate between classes and its suitability for real-world clinical applications, outperforming both Support Vector Machine and Random Forest models. Decision Curve Analysis further validates the clinical utility of logistic regression by demonstrating its consistent net benefit across critical decision thresholds, essential for guiding healthcare decisions effectively. Although SVM and Random Forest models also show respectable performance with ROC AUC scores of 0.83 and 0.85 respectively, their consistent net benefit of 0.50 indicates relatively lower clinical relevance compared to logistic regression across various decision scenarios. Incorporating these insights into clinical practices has the potential to significantly enhance predictive accuracy and facilitate informed decision-making, ultimately advancing patient care and healthcare outcomes through the effective deployment of AI technologies.

Future directions for this data involve investigating advanced feature engineering methods to refine SVM and Random Forest models, validating these AI algorithms across diverse patient populations and clinical environments, and establishing rigorous protocols for their ethical implementation in healthcare to uphold patient safety and regulatory standards.

Data Availabilty

The heart failure prediction dataset on Kaggle provides accessible and extensive resources for researchers and practitioners aiming to study and forecast heart failure occurrences.

Conflicts of Interest

The authors state that they do not have any conflicts of interest to report regarding the present study.

Funding

The author conducted this independent research project without receiving any external funding or financial assistance.

Author's contribution

The author Hency Juliet played a significant role in curating the data, conducting formal analyses, gathering references, managing resources, developing the methodology, overseeing the project, drafting the initial manuscript, reviewing the content, editing the work, creating visualizations, conducting investigations, performing formal analyses, and making substantial contributions to proofreading the research.

Declaration of Generative AI and AI-assisted technologies in the writing process:
During the preparation of this work the author(s) used ChatGpt in order to improve the readability and language. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

[1] R. A. El Arab et al., "Bridging the gap: From AI success in clinical trials to real-world healthcare implementation—A narrative review," Healthcare, vol. 13, no. 7, 2025.

[2] https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction

[3] Alowais, S.A., Alghamdi, S.S., Alsuhebany, N. et al. Revolutionizing healthcare: the role of artificial

intelligence in clinical practice. BMC Med Educ 23, 689 (2023). https://doi.org/10.1186/s12909-023-04698-z

[4] Z. Yu, K. Wang, Z. Wan, S. Xie, Z. Lv, "Popular deep learning algorithms for disease prediction: a review," Cluster Comput., vol. 26, no. 2, pp. 1231-1251, 2023. doi: 10.1007/s10586-022-03707-y

[5] Park, D.J., Park, M.W., Lee, H. et al. Development of machine learning model for diagnostic disease prediction based on laboratory tests. Sci Rep 11, 7567 (2021).

[6] Yu Z, Wang K, Wan Z, Xie S, Lv Z. Popular deep learning algorithms for disease prediction: a review. Cluster Comput. 2023;26(2):1231-1251. doi: 10.1007/s10586-022-03707-y. Epub 2022 Sep 13. PMID: 36120180; PMCID: PMC9469816. https://doi.org/10.1038/s41598-021-87171-5

[7] Bi, Aoxiang & Han, Ziqi & Zha, Yue. (2023). Research and application of artificial intelligence in disease prediction. Applied and Computational Engineering. 6. 461-468. 10.54254/2755-2721/6/20230837

[8] Kumar, Y., Koul, A., Singla, R. et al. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Human Comput 14, 8459–8486 (2023). https://doi.org/10.1007/s12652-021-03612-z

[9] D. J. Park, M. W. Park, H. Lee, et al., "Development of machine learning model for diagnostic disease prediction based on laboratory tests," Sci Rep, vol. 11, p. 7567, 2021. doi: 10.1038/s41598-021-87171-5.

[10] Jackins, V., Vimal, S., Kaliappan, M. et al. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. J Supercomput 77, 5198–5219 (2021). https://doi.org/10.1007/s11227-020-03481-x

[11] Y. Kumar, A. Koul, R. Singla, et al., "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," J Ambient Intell Human Comput, vol. 14, pp. 8459–8486, 2023. doi: 10.1007/s12652-021-03612-z.

[12] K. S. Pokkuluri et al., “Bridging the gap between technology and medicine through the revolutionary impact of the healthcare Internet of Things on remote patient monitoring,” Journal of Intelligent Systems and Internet of Things, vol. 2, p. 212, 2024

[13] A. T. Johnson, L. M. Williams, and P. K. Patel, "Leveraging machine learning for predictive analytics in healthcare: A review," 2023 IEEE International Conference on Healthcare Informatics (ICHI), New York, NY, USA, 2023, pp. 345-352. doi: 10.1109/ICHI.2023.10111567

[14] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," IEEE Access, vol. 5, pp. 8869-8879, 2017. doi: 10.1109/ACCESS.2017.2694446.

[15] W. A. Sulaiman et al., "Emergency Department Triage Hospitalization Prediction Based on Machine Learning and Rule Extraction," 2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology, Malta, 2023, pp. 147-148, doi: 10.1109/IEEECONF58974.2023.10405176

[16] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab and P. Kazanzides, "A Review of Augmented Reality in Robotic-Assisted Surgery," in IEEE Transactions on Medical Robotics and Bionics, vol. 2, no. 1, pp. 1-16, Feb. 2020, doi: 10.1109/TMRB.2019.2957061

[17] Abhishek, H. V. Bhagat and M. Singh, "A Machine Learning Model for the Early Prediction of Cardiovascular Disease in Patients," 2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC), Puducherry, India, 2023, pp. 1-5, doi: 10.1109/ICACIC59454.2023.10435210.

[18] N. Mohan, V. Jain and G. Agrawal, "Heart Disease Prediction Using Supervised Machine Learning Algorithms," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-3, doi: 10.1109/ISCON52037.2021.9702314.

[19] Dinesh, and Jaya Mabel Rani. "An efficient light gradient boosting classification method for predicting lumpy disease and compared with XG boost algorithm." In AIP Conference Proceedings, vol. 3345, no. 1, p. 020336. AIP Publishing LLC, 2026

[20] S. Kuhar, A. Kaur, T. Raj, S. Pareek, K. Grover and R. Kumar, "Cardiovascular Disease Prediction and Prevention: Exploring Novel Techniques and Applications Using Machine Learning," 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 668-673, doi: 10.1109/SMART59791.2023.10428297.

[21] D. A. Tuan, “Bridging the gap between black box AI and clinical practice: Advancing explainable AI for trust, ethics, and personalized healthcare diagnostics,” 2024.

[22] Umapathy V, Rajinikanth B S, Samuel Raj R, et al. (September 21, 2023) Perspective of Artificial Intelligence in Disease Diagnosis: A Review of Current and Future Endeavours

[23] N. Ghaffar Nia, E. Kaplanoglu, A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," Discov Artif Intell., vol. 3, no. 1, p. 5, Jan. 2023. doi: 10.1007/s44163-023-00049-5.

[24] A. P. Zhao, S. Li, Z. Cao, P. J.-H. Hu, J. Wang, Y. Xiang, D. Xie, and X. Lu, "AI for science: Predicting infectious diseases," Journal of Safety Science and Resilience, vol. 5, no. 2, pp. 130-146, 2024. [Online]. Available: https://doi.org/10.1016/j.jnlssr.2024.02.002.

[25] A. A. Malibari, "An efficient IoT-Artificial intelligence-based disease prediction using lightweight CNN in healthcare system," Measurement: Sensors, vol. 26, 2023, Art. no. 100695. [Online]. Available: https://doi.org/10.1016/j.measen.2023.100695.

[26] N. J. G. Rama and I. Sousa, “Bridging the gap: The role of technological advances in shaping gastrointestinal oncological outcomes,” World Journal of Gastrointestinal Oncology, vol. 17, no. 3, Art. no. 101752, 2025.

[27] K. Ramudu, V. Mohan, D. Jyothirmai, D. Prasad, R. Agrawal, and S. Boopathi, "Machine Learning and Artificial Intelligence in Disease Prediction: Applications, Challenges, Limitations, Case Studies, and Future Directions," in Machine Learning and Artificial Intelligence in Disease Prediction, Aug. 2023, pp. 297-318, doi: 10.4018/978-1-6684-8913-0.ch013

[28] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications," International Journal of Intelligent Networks, vol. 3, pp. 58-73, 2022. doi: 10.1016/j.ijin.2022.05.002.

[29] F. Jiang, Y. Jiang, H. Zhi, et al., "Artificial intelligence in healthcare: past, present and future," Stroke and Vascular Neurology, vol. 2, 2017. doi: 10.1136/svn-2017-000101.

[30] H. Habehh and S. Gohel, "Machine Learning in Healthcare," Curr Genomics, vol. 22, no. 4, pp. 291-300, Dec. 2021. doi: 10.2174/1389202922666210705124359.

[31] D. E. Ehrmann, S. Joshi, S. D. Goodfellow, et al., "Making machine learning matter to clinicians: model actionability in medical decision-making," npj Digit. Med., vol. 6, p. 7, 2023. doi: 10.1038/s41746-023-00753-7.

[32] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," Future Healthc J., vol. 6, no. 2, pp. 94-98, Jun. 2019, doi: 10.7861/futurehosp.6-2-94.

[33] V. Bhuvaneshwari, R. Sadaieswaran, P. Malathi, N. Legapriyadharshini, T. Thirumalaikumari and S. R. Raja, "IoT-Enhanced Innovations in Drug Delivery Systems for Targeted Therapy in Autoimmune Disorders," ICACCS, 2024, pp. 1275-1281, doi: 10.1109/ICACCS60874.2024.10716958.

[34] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," npj Digital Medicine, vol. 5, no. 48, 2022. [Online]. Available: https://doi.org/10.1038/s41746-022-00592-y

[35] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," Multimedia Tools and Applications, vol. 82, pp. 26731-26769, 2023. doi: 10.1007/s11042-022-14305-w.

[36] S. Vollmer, B. A. Mateen, G. Bohner, F. J. Király, R. Ghani, P. Jonsson, et al., "Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness," BMJ, vol. 368, no. l6927, 2020. doi: 10.1136/bmj.l6927.

[37] Jaya Mabel Rani, Ramkumar Kalyanaraman, Aneesh Somwanshi, and Tan Kuan Tak. "Deep Learning-Driven Multi-Objective Clustering for Neuroblastoma Omics Data with NSGA-II." Acta Universitatis Sapientiae, Informatica 17, no. 1 (2025): 15

[38] E. Bisong, "Logistic Regression," in Building Machine Learning and Deep Learning Models on Google Cloud Platform. Berkeley, CA: Apress, 2019. Available: https://doi.org/10.1007/978-1-4842-4470-8_20.

[39] Srivenkateswaran, C., A. Jaya Mabel Rani, R. Senthil Kumaran, and R. Vinston Raja. "Securing healthcare data: A federated learning framework with hybrid encryption in cluster environments." Technology and Health Care 33, no. 3 (2025): 1232-1257

[40] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001. [Online]. Available: https://doi.org/10.1023/A:1010933404324

[41] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," in IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, July-Aug. 1998, doi: 10.1109/5254.708428. keywords: