AL Fahim

ALFahim054015_B11-Heart Disease Chat gpt

My Files

My Files

University

Document Details

Submission ID

trn:oid:::3618:103551741

Submission Date

Jul 5, 2025, 6:10 PM GMT+5:30

Download Date

Jul 5, 2025, 6:11 PM GMT+5:30

File Name

ALFahim054015_B11-Heart Disease Chat gpt.docx

File Size

123.4 KB

5 Pages

1,931 Words

11,941 Characters

21% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

33 Not Cited or Quoted 18%
Matches with neither in-text citation nor quotation marks

4 Missing Quotations 2%
Matches that are still very similar to source material

1 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

14%	Internet sources
11%	Publications
13%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **33** Not Cited or Quoted 18%
  Matches with neither in-text citation nor quotation marks
- **4** Missing Quotations 2%
  Matches that are still very similar to source material
- **1** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

14%	Internet sources
11%	Publications
13%	Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Submitted works	University of Arizona on 2025-05-10	2%
2	Internet	pmc.ncbi.nlm.nih.gov	1%
3	Submitted works	King's Own Institute on 2024-08-13	1%
4	Submitted works	University of Birmingham on 2024-12-13	1%
5	Internet	www.medrxiv.org	1%
6	Submitted works	Sir George Monoux College on 2025-03-16	<1%
7	Internet	jcsdf.nfsu.ac.in	<1%
8	Publication	Dinesh Goyal, Bhanu Pratap, Sandeep Gupta, Saurabh Raj, Rekha Rani Agrawal, I...	<1%
9	Publication	HuaZhong Yang, Zhongju Chen, Huajian Yang, Maojin Tian. "Predicting coronary ...	<1%
10	Internet	www.frontiersin.org	<1%

11	Internet	acikerisim.aydin.edu.tr	<1%
12	Submitted works	University of Essex on 2024-03-20	<1%
13	Internet	ojs.bonviewpress.com	<1%
14	Internet	patents.google.com	<1%
15	Internet	www.mdpi.com	<1%
16	Internet	pubs2.ascee.org	<1%
17	Publication	Jiayu Shi, Zexiao Wang, Jiandong Zhou, Chengyu Liu, Poly Z.H. Sun, Erying Zhao, L...	<1%
18	Submitted works	UCL on 2025-04-25	<1%
19	Submitted works	University of Wolverhampton on 2025-05-23	<1%
20	Internet	digitalcommons.uconn.edu	<1%
21	Internet	listens.online	<1%
22	Submitted works	Vrije Universiteit Amsterdam on 2025-07-03	<1%
23	Internet	www.sciencepublishinggroup.com	<1%
24	Internet	www.informatica.si	<1%

25	Publication	
	Houssein Dhayne, Rafiqul Haque, Rima Kilany, Yehia Taher. "In Search of Big Med...	<1%
26	Publication	
	Islam Daoud Suliman, D. Vasumathi. "Prediction of Heart Disease Using Machine ...	<1%
27	Publication	
	Widdowson, Denise. "Computational Optimization of Material Properties Via a M...	<1%
28	Publication	
	Sushree Chinmayee Patra, B. Uma Maheswari, Peeta Basa Pati. "Forecasting Coro...	<1%
29	Submitted works	
	University of Birmingham on 2024-12-13	<1%
30	Submitted works	
	Hartlepool College of Further Education on 2025-03-08	<1%

HEART DISEASE PREDICTION

ARIFUR RAHMAN [1], AL FAHIM [2], SABBIR AHMMED SHUVO [3]

Supervised by
Mohaimen-Bin-Noor

22-47900-2@student.aiub.edu
22-46402-1@student.aiub.edu
22-47181-1@student.aiub.edu

***Abstract*- Cardiovascular diseases remain the leading cause of death worldwide, with an estimated 17.5 million fatalities annually, primarily in low- and middle-income countries. Early detection plays a vital role in reducing these numbers, but traditional diagnostic tools often demand significant resources, limiting their availability in under-resourced areas. This study explores the application of machine learning techniques for forecasting heart disease. Using a private dataset containing 1,025 instances and 13 features, several algorithms were tested, such as Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression. The Decision Tree model achieved the highest accuracy at 99.2%, followed closely by KNN at 98.20%. These outcomes emphasize the potential of machine learning models to improve early diagnosis. Integrating these tools into mobile platforms could enable timely, accurate predictions, especially in environments with limited healthcare infrastructure. Future research will focus on enlarging the dataset, refining algorithms, and addressing limitations to broaden real-world use.**

***Keywords*_**Heart disease prediction, machine learning, early detection, classification algorithms.

I. INTRODUCTION

The heart is a vital muscular organ responsible for pumping blood throughout the body as part of the circulatory system. It works in coordination with arteries, veins, and capillaries to maintain proper blood flow. Disruptions in this process can lead to various heart conditions, collectively referred to as cardiovascular diseases (CVDs). Globally, CVDs are the leading cause of death [1]. According to the World Health Organization (WHO), heart disease and stroke cause approximately 17.5 million deaths annually, with over 75% occurring in low- and middle-income regions. Moreover, strokes and heart attacks account for about 80% of these fatalities [2].

This study aligns with Sustainable Development Goal (SDG) 3, which emphasizes ensuring good health and well-being for all. Diagnosing heart-related conditions typically involves evaluating symptoms, patient history, and risk factors such as smoking, high cholesterol, sedentary behavior, aging, stress, and pre-existing conditions like diabetes and hypertension [3]. Lifestyle changes, including exercise, diet, and stress management, are often recommended. Diagnostic procedures may include electrocardiograms (ECGs), echocardiograms, cardiac MRIs, and blood tests. Treatments range from medication and lifestyle changes to surgical procedures like angioplasty or the installation of devices such as pacemakers [4].

With the growth of Big Data in Electronic Health Record (EHR) systems, healthcare professionals now have access to vast patient data, paving the way for predictive models. Machine learning (ML) plays a crucial role in analyzing these large datasets and generating actionable insights [5].

The primary aim of this research is to develop a machine learning model capable of predicting heart disease accurately. Using a private dataset of 1,025 samples with 13 attributes, the study evaluates multiple ML algorithms and selects the most effective one for integration into an Android application. This application aims to provide real-time, accurate predictions using high-quality data. The key contribution lies in utilizing a robust, privately sourced dataset for model development.

II. LITERATURE REVIEW

Machine learning classification techniques have been extensively applied in the prediction of cardiovascular diseases (CVD) across numerous datasets. This section reviews recent studies and highlights the effectiveness of various algorithms in enhancing prediction accuracy. The focus is placed on models that extract meaningful features and improve the classification of CVD risk.

Logistic Regression (LR): LR is a well-established model frequently used for binary classification tasks in healthcare. A study referenced in [6] used LR on a dataset of 735 individuals, achieving an accuracy rate of 87.63% for predicting cardiovascular issues. Another investigation applying LR on 3,980 records reported a 70.44% prediction accuracy [7]. Additionally, research targeting coronary artery disease (CAD) in women demonstrated a 70% sensitivity rate using LR [8].

K-Nearest Neighbor (KNN): KNN is widely regarded for its straightforward approach and reliability in detecting patterns based on feature proximity. In one study by Souza and Lima, the KNN algorithm was applied to a clinical heart failure dataset and achieved an accuracy of 89.49%, with both precision and sensitivity around 0.89 [9]. Another study conducted by Al-Adhaileh and colleagues showed that KNN achieved 92% accuracy on a smaller dataset, outperforming more complex models like deep learning and random forest [10].

Decision Tree (DT): Decision Trees are known for their interpretability and use in medical classification problems. One study demonstrated that a DT-based model reached an 84% accuracy in predicting chronic diseases, reinforcing its viability in CVD detection tasks [11][12].

Overall, various ML classification techniques have shown substantial promise in identifying CVD risks. Algorithms such as LR and KNN consistently deliver high accuracy and precision. Moreover, ensemble learning techniques like bagging, AdaBoost, and voting mechanisms have further improved model performance over individual classifiers. Continued exploration and comparison of these methods are essential for enhancing diagnostic capabilities in cardiovascular health.

III. METHODOLOGY

This research utilized a heart disease dataset sourced from a publicly available repository. Prior to model development, several preprocessing steps were carried out to ensure the dataset was suitable for machine learning analysis.

To address missing entries, imputation techniques were applied, preserving data consistency. Continuous variables underwent Z-score normalization to achieve uniform scaling, while categorical data was transformed into numerical format using encoding methods such as one-hot or label encoding. The relevance of input features was determined using correlation matrices and feature importance evaluations.

A. Data Partitioning

The dataset was divided into training and test sets using an 80:20 ratio. This ensured that model validation could be conducted using a separate dataset not seen during training, enabling a fair evaluation of performance.

B. Algorithm Selection

The study focused on three machine learning models known for binary classification:

- Logistic Regression (LR)
- Decision Tree (DT)
- K-Nearest Neighbor (KNN)

These algorithms were chosen based on their effectiveness and interpretability in health prediction tasks.

C. Training Phase

Each model was initially trained using the default settings provided by their respective implementations within the ML framework. The training was conducted on the designated training subset of the data.

D. Evaluation Metrics

The effectiveness of each model was gauged using several metrics:

- **Accuracy** to measure overall correctness
- **Precision** to assess the proportion of positive identifications that were actually correct
- **Recall** to determine how many actual positives were correctly detected
- **F1-Score**, a harmonic mean of precision and recall, offering a balanced evaluation

E. Model Comparison

The models were compared using the evaluation metrics mentioned above. Performance trade-offs among them were considered to identify the most reliable model for heart disease prediction.

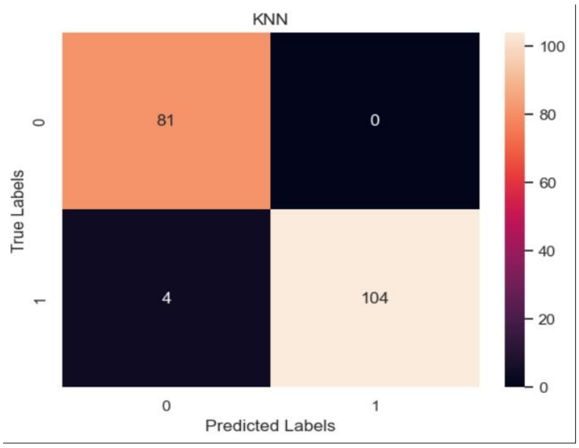
F. Development Environment

All experiments were performed in Python, leveraging popular libraries such as Pandas and NumPy for data handling, scikit-learn for implementing algorithms, and Matplotlib for visual representation.

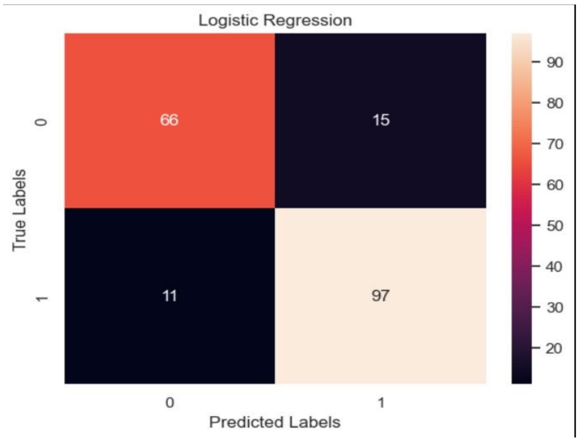
IV. ANALYSIS

This section presents the performance evaluation of three selected machine learning models applied to a private dataset containing 1,025 entries and 13 features. Each algorithm was assessed based on accuracy, precision, and F1-score.

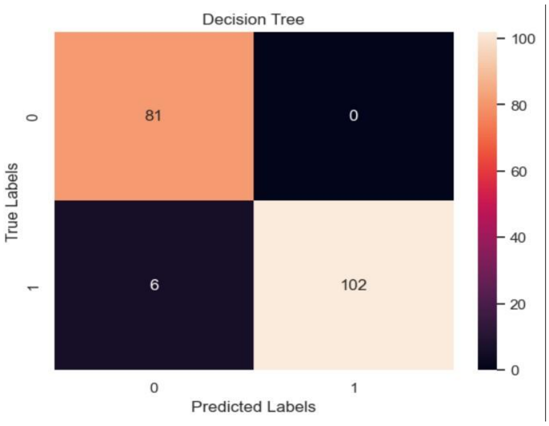
K-Nearest Neighbor (KNN):
The KNN algorithm achieved an accuracy of 98.20%. It also delivered a precision rate of 96.00% and an F1-score of 98.00%, indicating strong classification ability and reliability in detecting heart disease.



Logistic Regression (LR):
Logistic Regression yielded varied results. On one dataset, it reached 87.63% accuracy, while on another, it managed only 70.44%. This variation points to moderate but dataset-sensitive performance.



Decision Tree (DT):
The Decision Tree classifier outperformed the others with a top accuracy of 99.2%, confirming its suitability for high-precision classification tasks in this context.



Interpretation and Model Comparison

Preprocessing techniques like normalization and data balancing significantly enhanced model outcomes. Among the tested models, the Decision Tree emerged as the most accurate and robust, demonstrating excellent performance across metrics. KNN also produced consistently high results, validating its strength in medical prediction scenarios.

Significance of Findings:

- Accuracy and Early Detection: The high accuracy of the Decision Tree suggests it could serve as an effective tool for early diagnosis.
- Application Readiness: The simplicity and efficiency of these models make them ideal for embedding into mobile platforms, supporting on-the-go predictions for users in underserved areas.
- Public Health Relevance**:** By contributing to early diagnosis and intervention, this research supports international efforts such as SDG 3, aimed at improving global health outcomes.

V. CONCLUSION

This study highlights the significant role that machine learning can play in the early detection of heart disease—a condition that continues to be a leading cause of death worldwide. By examining a private dataset of 1,025 records with 13 selected features, the research evaluated the effectiveness of multiple machine learning models in predicting heart conditions.

Among the tested models, the Decision Tree classifier stood out with an exceptional accuracy of 99.2%, followed closely by the K-Nearest Neighbor algorithm at 98.20%, demonstrating their high reliability in classification tasks. These outcomes validate the strength of ML-based approaches for medical predictions when appropriate data preparation and algorithm selection are applied.

The integration of such predictive models into mobile applications could significantly enhance healthcare delivery by enabling early, accessible, and accurate diagnosis—particularly in low-resource environments. These advancements have the potential to reduce delays in treatment and ultimately save lives, aligning with global health initiatives aimed at improving early detection and medical responsiveness.

A.Limitations

While the results of this study are encouraging, there are several limitations to consider. The dataset used includes only 1,025 samples, which may not capture the full range of variability present in real-world populations. As a result, the models may perform differently when applied to larger or more diverse groups. Furthermore, the model depends on a fixed set of features; if similar data is unavailable or incomplete in other settings, the performance could decline. Since the dataset is privately sourced, there may also be inherent biases in data collection or labeling that were not accounted for.

B.Future Direction

To improve the utility and reliability of the models, future studies should aim to:

- Incorporate larger and more diverse datasets to increase generalizability.
- Develop robust techniques for managing missing or noisy data in real-world health applications.
- Pilot test the models in practical environments, such as mobile or cloud-based healthcare tools, to evaluate real-time performance.
- Investigate the effectiveness of ensemble techniques and the inclusion of new, potentially predictive features to boost overall model accuracy.

By overcoming these limitations and building upon the current findings, machine learning can continue to offer powerful tools for improving heart disease diagnosis and global health outcomes.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases (CVDs)," Jun. 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Z. Alom, M. A. Azim, Z. Aung, M. Khushi, J. Car, and M. A. Moni, "Machine learning approaches for early detection of heart failure," in *Proc. Int. Conf. Big Data, IoT and Machine Learning*, 2021, pp. 94–107.

[3] S. Gour, P. Panwar, D. Dwivedi, and C. A. Mali, "Heart attack prediction using machine learning techniques," in *Intelligent Sustainable Systems*, Singapore: Springer, 2022, pp. 741–747.

[4] C. Gupta, A. Saha, N. S. Reddy, and U. D. Acharya, "Predictive modeling of cardiac diseases using supervised ML algorithms," *J. Phys.: Conf. Ser.*, vol. 2161, no. 1, p. 012027, 2022. doi: 10.1088/1742-6596/2161/1/012027

[5] M.-L. Tsai, K.-F. Chen, and P.-C. Chen, "AI-driven cardiovascular risk assessment using EHR: A comprehensive review," *J. Am. Heart Assoc.*, vol. 14, no. 6, p. e036946, Mar. 2025. doi: 10.1161/JAHA.124.036946

[6] A. Rajkomar et al., "Scalable deep learning applications on electronic health records," *NPJ Digit. Med.*, vol. 1, p. 18, 2018. doi: 10.1038/s41746-018-0029-1

[7] Z. Du et al., "Machine learning for coronary heart disease prediction in hypertensive patients using EHR," *JMIR Med. Inform.*, vol. 8, no. 7, p. e17257, Jul. 2020. doi: 10.2196/17257

[8] T. Liu, A. Krentz, L. Lu, and V. Curcin, "EHR-based CVD risk prediction using ML: A systematic review," *Eur. Heart J.: Digit. Health*, vol. 6, no. 1, pp. 7–22, 2025. doi: 10.1093/ehjdh/ztae080

13 [9] V. S. Souza and D. A. Lima, "Diagnosing heart conditions with KNN using clinical heart failure data," *Artif. Intell. Appl.*, vol. 3, no. 1, pp. 56–71, Apr. 2024. doi: 10.47852/bonviewAIA42022045

[10] M. H. Al-Adhaileh, M. I. A. Al-Mashhadani, E. M. Alzahrani, and T. H. H. Aldhyani, "Enhanced heart attack prediction using ML and DL methods," *Iraqi J. Comput. Sci. Math.*, vol. 6, no. 2, Article 3, 2025. doi: 10.52866/2788-7421.1239

11 [11] C. Puelz et al., "Computational modeling of Fontan physiology with different circulatory modifications," *Comput. Biol. Med.*, vol. 89, pp. 405–418, 2017. doi: 10.1016/j.compbiomed.2017.08.017

[12] N. R. Rusyana, F. Renaldi, and D. Destiani, "Disease risk prediction using C4.5 decision tree," presented at the Conf. Dept. Inform. and Info. Syst., Univ. Jenderal Achmad Yani, Cimahi, Indonesia, n.d.