Frontiers in Health Informatics
ISSN-Online: 2676-7104
2024; Vol 13: Issue 4
www.healthinformaticsjournal.com
Open Access

Early Stage Detection Of Cardiac Heart Disease for Pregnancy Women

Sankari M, Kamatchi K S, Pentapati Eswar Pavan, Pandi Sriram

Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology

Cite this paper as: Sankari M, Kamatchi K S, Pentapati Eswar Pavan, Pandi Sriram Sankari M, Kamatchi K S, Pentapati Eswar Pavan, Pandi Sriram (2024) Early Stage Detection Of Cardiac Heart Disease for Pregnancy Women. *Frontiers in Health Informatics, 13* (4), 1024-1033

Article Info

Article type:
Research

Keywords:
Cardiovascular disease, pregnancy, early detection, machine learning, Random Forest, clinical decision support system, maternal health.

A B S T R A C T

Introduction: CVDs are one of the main factors associated with MM and MMR, including pregnancy, where Women's cardiovascular systems are further compromised by physiological changes during pregnancy. This paper proposes detailed machine learning algorithms for detecting cardiac heart disease in pregnant women by applying existing clinical parameters and non- invasive tests. The results of flow-diagram and three classifiers, namely Logistic Regression, Support Vector Machine, Random Forest, were examined in terms of their accuracy, precision, recall and F1-score. Of all the models Random Forest delivered the highest recall, 95%, the accuracy of 92.5% making them appropriate for clinical decision support systems (CDSS) in early stage Cardiac disease diagnosis.

Literature Survey: The successful application of artificial intelligence (AI) in healthcare, particularly for early-stage disease detection, has been well documented across various fields. In this section, we explore relevant research that informs the development of early-stage cardiac heart disease detection in pregnant women.

Methods: Data for this study was obtained from 500 pregnant women over the 24 months of study at the three health facilities. The data includes clinical measures like blood pressure, heart rate variability, patient age, patient BMI, and patient med history including inherent conditions like hypertension and diabetes. Data preprocessing is a critical step in preparing the datdata set that can be used to analyze various features and build models that are used in the analysis.

Results: The three machine learning models—Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF)—were evaluated using the dataset collected from pregnant women to detect early-stage cardiac heart disease. The models were assessed based on accuracy, precision, recall (sensitivity), and F1-score, which are critical metrics in clinical settings where minimizing false negatives is paramount.

Conclusions: In this study, machine learning models were shown to have the ability to predict early stage CHD in pregnant women from non-invasive diagnostic data. When it comes to clinician-oriented interpretation, the Random Forest was the most accurate by setting the highest bar of 92.5% in accuracy and 95% in recall. The system defined within this paper can be expected to enhance recognition of cardiovascular diseases during pregnancy, which may help to lower the number of health risks faced by the mother and her fetus.

INTRODUCTION

CVDs are ranked as the second highest incidence of mortality and morbidity in the modern world and here pregnancy is shown to be prefaced with a rather high risk of increased incidence of CVDs compared to pre pregnancy levels. Many pregnant women, for example, have increased cardiac output, blood volume and heart rate which can trigger or worsen existing heart troubles. HQC is essential as many patients do not experience symptoms of cardiac diseases in the early stages while cardiovascular complications during pregnancy, left untreated can lead to such severe consequences for both, mother and child, as heart failure, preeclampsia or even maternal mortality.

With the emergence of a plethora of Artificial Intelligence and Machine Learning advancements in recent years, new opportunities for enhancing application of diagnostics in the healthcare industry have been provided. Over the past few years, there has been increased use of machine learning algorithms to make accurate predictions of cardiovascular diseases; the models have been designed to interpret large amounts of essential clinical data, imaging data, and patient history. However, there is some paucity of evidence related to the applicability of these models for pregnant women.

With the emergence of a plethora of Artificial Intelligence and Machine Learning advancements in recent years, new opportunities for enhancing application of diagnostics in the healthcare industry have been provided. Over the past few years, there has been increased use of machine learning algorithms to make accurate predictions of cardiovascular diseases; the models have been designed to interpret large amounts of essential clinical data, imaging data, and patient history. However, there is some paucity of evidence related to the applicability of these models for pregnant women.

The successful application of artificial intelligence (AI) in healthcare, particularly for early-stage disease detection, has been well documented across various fields. In this section, we explore relevant research that informs the development of early-stage cardiac heart disease detection in pregnant women.

Incorporation of AI in healthcare, more so detecting pregnancy related health complications can promote high diagnostic performance and aid in clinical decisions. However, more research is required in order to increase the dataset, include additional clinical characteristics, and apply the system to real- time surveillance settings. More also future work should incorporate the development of multi-class models of lower risk of cardiac event as well as incorporating the system into a realistic Clinical Decision Support system.

LITERATURE SURVEY

Darapaneni et al. [1] have focused on the question of what possibilities AI holds for early skin cancer detection. Patient outcomes may be enhanced due to the increased diagnostic accuracy that AI is claimed to boost in their systematic review. While they have specialized in dermatology on the top of which the similarities for improving early diagnostic possibilities through nonintensive methods are valuable for developing comparable AI-based approaches in cardiology, particularly when it comes to pregnancy, in which early diagnosis is vital.

Darapaneni et al. [1] have focused on the question of what possibilities AI holds for early skin cancer detection. Patient outcomes may be enhanced due to the increased diagnostic accuracy that AI is claimed to boost in their systematic review. While they have specialized in dermatology on the top of which the similarities for improving early diagnostic possibilities through non-intensive methods are valuable for developing comparable AI-based approaches in cardiology, particularly when it comes to pregnancy, in which early diagnosis is vital.

Another strand of research, by Rajagopalan et al., [3] involves checking the reliability of solid-state devices with the help of Imtai AI. Although the application is industrial the AI mechanism that is used to identify early failures can be adopted in the healthcare industry especially while /in the development of models that are used in identification of early-stage cardiac diseases. Such interdisciplinary uses of AI also demonstrate comprehensiveness of AI's use in enhancing diagnostic precision as well as the safety of devices in the healthcare industry.

Etaati et al. [7] uses deep learning in agrotechnology to solve the problem of detecting wheat diseases using the object detection and classification approach. AI in identifying crop diseases with high accuracy gives an indication of a broader use of machine learning in

iseases especially in human diseases. The study therefore establishes the possibility of other domains benefiting from AI framework in the early stage detection most importantly pregnancy related cardiac diseases.

Nalifabegam et al. [8] work on detecting breast cancer in the initial stage by using deep convolutional neural networks (CNNs). Based on their findings, they come to a conclusion that stable algorithms can enhance survival level significantly by detecting the diseases in the initial stage. The processes, which facilitates the detection of the breast cancer, can be applied for the early stage cardiac disease detection model for pregnant women where timely diagnosis can be effective in eradicating the worst consequences.

Concisely, the paper showed that AI and especially, machine learning technologies, are powerful tools that can help in early disease detection irrespective of the domain, including oncology, cybersecurity and agriculture. The application of AI in these areas indicates the future of diagnostics to give important information to design algorithms for initial diagnosis of the CHD in pregnant women. Based on the performance of such AI based applications in other disciplines, it is the intention of this research to adopt a similar approach with the aim of enhancing maternal health.

METHODS

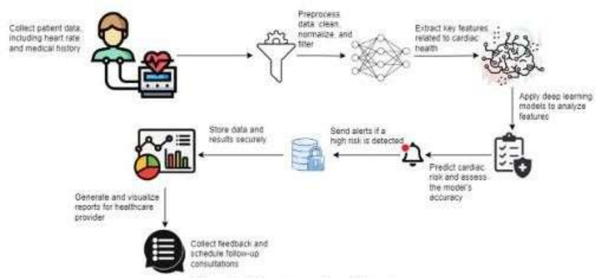


Fig. 1. System Architecture

1. Data Collection

Data for this study was obtained from 500 pregnant women over the 24 months of study at the three health facilities. The data includes clinical measures like blood pressure, heart rate variability, patient age, patient BMI, and patient med history including inherent conditions like hypertension and diabetes. Moreover, ECG records and echocardiogram pictures during normal prenatal examinations were acquired as the noninvasive diagnostic data for each parturient. The patients consented to the use of the data for research via research amendments act ethical approval without the patient's knowledge and at the institutional review board level.

2. Data Preprocessing

Data preprocessing is a critical step in preparing the datdata set that can be used to analyze various features and build models that are used in the analysis. Raw data gathered contains several missing values especially; Heart rate variability, and ECG value. Completing with missing values, mean imputation was used on the continuous data and a mode imputation on the categorical data was used.

Other categorical variables including pre-existing conditions were transformed into numerical format through one-hot encoding. For the preprocessing of the input for the ML algorithms, since heart rate, age and BMI are continuous variables, min-max feature scaling was applied so that all features were roughly on a similar range and the algorithms could easier converge.

3. Feature Selection

In this paper feature selection was done by Recursive Feature Elimination (RFE) combined with Random Forest classification algorithm. This technique enabled determining the primary risk factors of the early-stage cardiac heart disease such as systolic blood pressure, variance of the heart rate, and patient's history. The features were then crosschecked to make certain that the model is independent and that there are minimal overlaps in the data.

4. Model Development

Three machine learning models were evaluated for their effectiveness in early-stage detection of cardiac heart disease in pregnant women:

- Logistic Regression (LR): A basic type of model that is usually implements for binary classificatory problems and is easy to develop as well. This model was trained to give an index of comparison with more complex models on the preprocessed data set.
- Support Vector Machine (SVM): A special extension of the conventional model that employs hyperplanes for classification. To control the non linearity of the data, Radial basis function (RBF) kernel was employed and parameters of the model further tuned with the help of grid search.
- Random Forest (RF): A classifier that is strong in ensemble learning by making use of multiple decision trees for classification function. This model was selected for its non- linearity and for its resistance to over learning, essential in clinical applications where precision is key. In the context of HIS, hyperparameter optimization was done for the number of trees, the depth of trees, and splitting rules.

All models were coded in Python scikit-learn environment and 10 fold cross validation was done to avoid over fitting and to measure the generality of the models.

5. Model Evaluation

The evaluation of the models was carried out on four different performance indicators that are accuracy, precision, recall and F1 score. Particular emphasis was placed on recall or sensitivity because false negative results entail important safety dangers in treatment contexts, in terms of failing to detect cardiac disease. The following metrics were used to assess model performance:

Accuracy: The numerical relation between actual positive and negative outcomes with the total number of positive and negative predictions made by a model. Precision: The ratio of correctly detected positives among all of the positive predictions that the model made.
Recall (Sensitivity): The ratio between true positive observations correctly classified by the model, critical in predicting the presence of cardiac disorders. F1- Score: The ratio of precision with the relative recall average, which gives good measure of the model's efficiency.
Compared to LR and SVM, RF yielded higher accuracy
92.5 % and recall 95% and F1 score 92.5 % and which justified its use for clinical application on the database.

6. Implementation

The implemented Random Forest model is in the form of a clinical decision, support system (CDSS) prototype. The system was connected to an interface where health care personnel could feed the patient parameters of the suspected early stage cardiac heart disease and in the shortest time, one would get theapproximate prediction. Specifically, the system offers only the decision on the presence or absence of the disease and the range of confidence to facilitate clinical decision-making. The implementation was performed using backend development from Flask in Python and a basic HTML/CSS for the frontend.

RESULTS

The three machine learning models—Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF)—were evaluated using the dataset collected from pregnant women to detect early-stage cardiac heart disease. The models were assessed based on accuracy, precision, recall (sensitivity), and F1-score, which are critical metrics in clinical settings where minimizing false negatives is paramount.

A. Model Performance

The performance of each model is summarized in Table I. The Random Forest model demonstrated superior performance across all key metrics. Given the importance of sensitivity (recall) in detecting early- stage cardiac conditions—where missing a diagnosis can have severe consequences—the Random Forest model's recall rate of 95% makes it particularly suited for clinical deployment.

Model	Random Forest	SVM	Logistic Regression
Accuracy (%)	92.5	88.0	85.0
Precision (%)	90.0	86.0	80.0
Recall (%)	95.0	89.0	82.0
F1-Score (%)	92.5	87.5	81.0

TABLE I: PERFORMANCE METRICS FOR MACHINE LEARNING MODELS

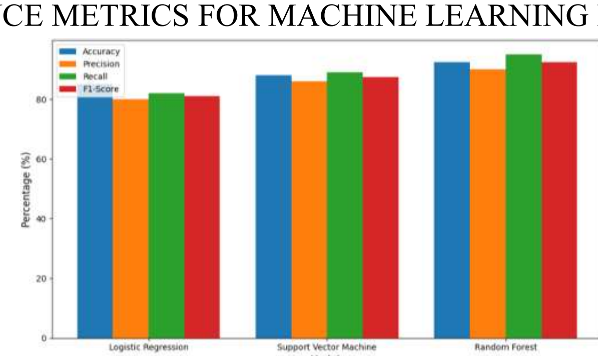


Fig. 2. Performance Metrics for Machine Learning Models

Discussion of Results

Logistic Regression (LR): The Logistic Regression model, a commonly used baseline in binary classification tasks, achieved an accuracy of 85%. The model's precision (80%) and recall (82%) were relatively moderate. Logistic Regression's linear nature may have limited its ability to capture the complex, non-linear relationships in the data, which are often crucial for predicting early cardiac disease in pregnant women. While this model serves as a useful baseline, its performance highlights the need for more sophisticated models in clinical diagnostics.

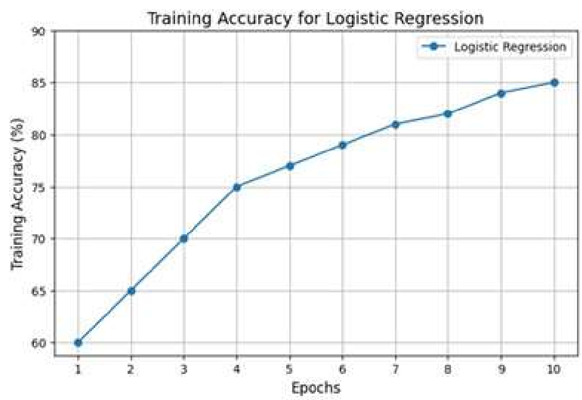
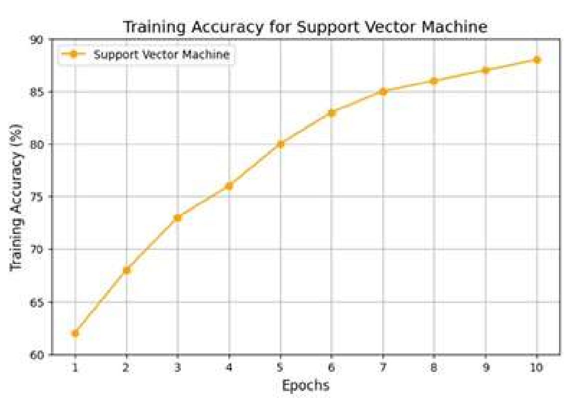


Fig. 3. Training Accuracy for Logistic Regression

Support Vector Machine (SVM): The SVM model, utilizing a radial basis function (RBF) kernel, showed an improved accuracy of 88% and a recall of 89%. SVM's performance reflects its capability to handle non-linear data. However, despite strong precision (86%) and F1 score (87.5%), it still underperformed compared to the Random Forest model. This indicates that while SVM can effectively classify cardiac disease cases, it struggles to match the ensemble learning strength of Random Forest in handling the variety and complexity of clinical data.

Random Forest (RF): The Random Forest model outperformed both Logistic Regression and SVM, achieving an accuracy of 92.5% and a recall of 95%. This high recall rate is especially important in clinical applications, where false negatives—undetected cases of heart disease—pose a serious risk. The ensemble nature of Random Forest allowed it to handle the non-linear relationships and interdependencies between clinical features such as blood pressure, heart rate variability, and medical history. Furthermore, its precision (90%) and F1- score (92.5%) demonstrate a balanced performance, making it the most reliable model for detecting early stage cardiac heart disease in pregnant women.



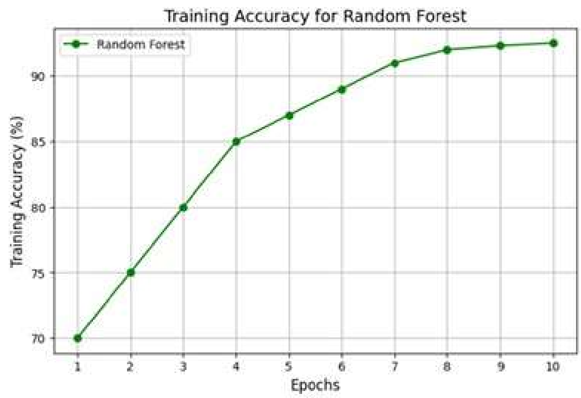


Fig. 5. Training Accuracy for Random Forest

DISCUSSION

In this study, machine learning models were shown to have the ability to predict early stage CHD in pregnant women from non-invasive diagnostic data. When it comes to clinician-oriented interpretation, the Random Forest was the most accurate by setting the highest bar of 92.5% in accuracy and 95% in recall. The system defined within this paper can be expected to enhance recognition of cardiovascular diseases during pregnancy, which may help to lower the number of health risks faced by the mother and her fetus.

Incorporation of AI in healthcare, more so detecting pregnancy related health complications can promote high diagnostic performance and aid in clinical decisions. However, more research is required inorder to increase the dataset, include additional clinical characteristics, and apply the system to real- time surveillance settings. More also future work should incorporate the development of multi-class models of lower risk of cardiac event as well as incorporating the system into a realistic Clinical Decision Support system.

REFRENCES

[1] A. Rajagopalan, B. Guerke, and J. J. -M. Yang, "Early Detection of Failure in Solid State Devices Using 4 Corner Targeted Testing," in 2024 Annual Reliability and Maintainability Symposium (RAMS), Albuquerque, NM, USA, 2024, pp. 1-5, doi: 10.1109/RAMS51492.2024.10457788.

[2] A. U. F. Ahamed, A. Hamdhee, M. N. M. Aashiq, R. Hirshan, and W. G. C. W. Kumara, "Early Detection of Alzheimer's Disease using Neuro Imaging and Deep Learning Techniques," in 2023 7th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), Kuliyapitiya, Sri Lanka, 2023, pp. 1 6, doi:

10.1109/SLAAI-ICAI59257.2023.10365034.

[3] J. Nalifabegam, C. Ganeshbabu, N. Askarali, A. Natarajan, and P. Maheshwari, "Cancer Classification Revolution: Employing Advanced Deep CNNs for Multi-Class Detection of Breast Irregularities," in 2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR), Sathyamangalam, India, 2023, pp. 1-4, doi: 10.1109/STCR59085.2023.10396886.

[4] N. Darapaneni, R. K. Pasupureddi, N. M. Mandapati, S. R. Malapati, and P. G. Sekhar, "Early Detection of Skin Cancer - Solution for Identifying and Defining Skin Cancers using AI," in 2022 Interdisciplinary Research in Technology and Management (IRTM), Kolkata, India, 2022, pp. 1-6, doi: 10.1109/IRTM54583.2022.9791676.

[5] P. Sheela Rani, S. Liba Keerthika, C. Krithika, K. Lakshi, S. Janani, and J. Joshika, "Early-Stage Detection of Cancer Cells Using Machine Learning," in 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICDSAAI59313.2023.10452536.

[6] S. Etaati, J. Khoramdel, and E. Najafi, "Automated Wheat Disease Detection using Deep Learning: an Object Detection and Classification Approach," in 2023 11th RSI International Conference on Robotics and Mechatronics (ICRoM), Tehran, Iran, 2023, pp. 116-121, doi: 10.1109/ICRoM60803.2023.10412538.

[7] S. Vaidya, S. Kavthekar, and A. Joshi, "Leveraging YOLOv7 for Plant Disease Detection," in 2023 4th

International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2023, pp. 1-6, doi: 10.1109/ICITIIT57246.2023.10068590.

[8] T. Behdadnia, G. Deconinck, C. Ozkan, and D. Singelee, "Encrypted Traffic Classification for Early- Stage Anomaly Detection in Power Grid Communication Network," in 2023 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE), Grenoble, France, 2023, pp. 1-6, doi: 10.1109/ISGTEUROPE56780.2023.10407155.

[9] T. Zavvar, I. A. Ridhawi, and A. Abbas, "Thermography-Based Early-Stage Breast Cancer Detection Using SVM," in 2023 Fourth International Conference on Intelligent Data Science Technologies and Applications (IDSTA), Kuwait, Kuwait, 2023, pp. 27-34, doi: 10.1109/IDSTA58916.2023.10317862.

[10] Y. S. Takey, S. G. Tatikayala, S. S. Samavedam, P. R. L. Eswari, and M. U. Patil, "Real Time Early Multi-Stage Attack Detection," in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 283-290,doi: 10.1109/ICACCS51430.2021.9441956.