International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://theaspd.com/index.php

SmartHeart: A Cloud and Machine Learning Framework for Early Cardiovascular Disease Prediction

Dr. Rajashree T. Gadhave[1], Amit Kumar[2], Dr. Shweta Marigoudar[3], Avinash P. Kaldate[4], Ch V S Satyamurty[5], Ramya R[6], Dr. Chandan Mukherjee[7], Vakaimalar Elamaran[8]
[1]Associate Professor, Department of Computer Engineering, Pillai HOC College of Engineering and Technology, Tal. Khalapur 410207, MH, India, rgadhave@mes.ac.in
[2]Associate Professor, Department of SCSE, Galgotias University, Greater Noida, amitbsbkumar@gmail.com
[3]Dean-FCIT, Department of Faculty of Computing and IT, GM University Davanagere, shwetamarigoudar@gmit.ac.in
[4]Department of Mechanical Engineering, Sinhgad College of Engineering, Savitribai Phule Pune University, Pune-411041, India, er.avinashkaldate@gmail.com
[5]Department of Information Technology, CVR College of Engineering, Hyderabad, Telangana, satyamurty@cvr.ac.in
[6]Associate Professor, Dept of CSE, Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu, India, ramyacse@kamarajengg.edu.in
[7]Assistant Professor, Department of Life Sciences, School of Biosciences and Technology, Galgotias University, Greater Noida, Uttar Pradesh, India, chandan.mukherjee@galgotiasuniversity.edu.in
[8]Associate Professor, Dept of IT, Kamaraj College of Engineering and Technology, Virudhunagar, Tamilnadu, India, vakaimalarit@kamarajengg.edu.in

Abstract: *Cardiovascular diseases (CVDs) are emerging as a big problem in the world today and thus the necessity to develop new methods of early prediction and prevention. The present paper presents SmartHeart a cloud-based system that outputs the risk of cardiovascular diseases in patients based on their machine learning and predictive features. The framework is a combination of real-time data collection of multiple sources such as wearable devices, electronic health records, and lifestyle factors to process the patient health data on the cloud. SmartHeart will coupled with modern machine learning algorithms allow achieving the most complete cardiovascular risk assessment and pre-detection of possible cardiovascular problems, which will help to provide timely interventions and individually tailored care. The scalability of the system and the fact that it uses cloud implementation allows efficient storage and processing of huge volumes of data and ensuring data privacy and security. Their robustness and generalizability on various populations are increased when the predictive models are trained on different datasets. In addition, SmartHeart shall have simple interfaces to both medical practitioners and patients so that they easily access their data and have actionable information. This paper presents an overview of the architecture, machine learning techniques, and evaluation of the framework that can be considered as absolutely transformational in the field of cardiovascular disease prediction and improving the overall healthcare outcomes and reducing popular strains on the healthcare system all over the world.*
Keywords: *Cardiovascular disease, early prediction, machine learning, cloud computing, healthcare framework, risk assessment, wearable devices, data privacy, predictive modelling*

INTRODUCTION

CVDs remain to cause the morbidity and mortality arising in many parts of the world. CVDs were perceived as a major cause of death because they claimed the lives of more than 17 million people each year according to the World Health Organization (WHO). Such diseases usually are the ones that evolve slowly, and can be anywhere before the person notices and is diagnosed with a more critical situation, including a heart attack, stroke, or heart failure. Because of this, once there is the prediction and early detection of the cardiovascular diseases it is possible that the overall results of the treatment experience will be tremendously beneficial and leading to saving of lives. With lifestyle changes and proper medical management, early intervention will help eliminate the chances of fatal cardiovascular incidents. Nonetheless, the existing schemes of predicting cardiovascular diseases are rather deficient in their promptness and reliability[1]. Their usual method of preventing cardiovascular-related conditions is the in-patient visit, regular check-up and passive or small range collections of clinical data which are often inadequate in describing the dynamic and multidimensional nature of cardiovascular health. With wearable technologies, electronic health records (EHRs), and mobile health applications, we now have the chance to enhance the predictive value and availability of CVD risk assessment tools[2].

To fill these gaps, this paper suggests SmartHeart that is a framework based on cloud computing with an early prediction of cardiovascular disease risk with the application of machine learning algorithms. SmartHeart has been conceptualized to utilize the potential of multiple sources of data including wearable sensors, EHRs as well as lifestyle data to provide more accurate and timely predictions. The recent introduction of cloud computing is a scalable and cost-efficient material which is applicable to deal with large volumes of data that is found in these sources, with the ability to analyze and process it in real-time[3]. Machine learning algorithms further help to improve the predictive skills of the system and enable tailor-made risk assessment of the patient him/herself. Integration of these innovative technologies will enable SmartHeart to come up with a powerful solution that will enhance effective detection of cardiovascular diseases at early onset, which reflects positively on healthcare delivery and will save lives by administering timely intervention to these patients to lessen the burden of CVDs[4].

The necessity of advanced cardiovascular disease prediction The necessity of an early prediction of cardiovascular disease Cardiovascular disease (abbreviated as CVD or occasionally CAD) is the primary cause of death globally. It is estimated that out of every five deaths that occur in Prevention and early identification of cardiovascular diseases have been regarded as an important approach in the prevention of the disease burden in the world. The old means of diagnosing CVDs (physical examination, electrocardiograms (ECGs), blood tests as well as imaging) are usually used at a stage where the disease has progressed to the scale where it started manifesting its symptoms. At this point, it becomes too late to intervene or the risk of dangerous health conditions multiplies. In addition, the given methods of diagnostic tend to miss those individuals who are at risk yet they are asymptomatic thus lack the chance to be provided with primary care[5].

Preference in cardiovascular diseases before the clinical manifestation is about to start has great potentials in cutting mortality rates and healthcare expenditures. Going by the recent researches, early predictability would enhance the risk control of high blood pressure, high cholesterol, diabetes, and obesity among other known risk factors relating to the occurrence of CVDs. Early prediction of cardiovascular diseases also helps healthcare providers to provide treatment and interventions according to the individual risk profile and thus make the care more effective as a whole. Nevertheless, in as much as early detection has tremendous advantages, the available prediction and risk assessment methods are either still not fully exploited or they are not accurate. This is the area where machine learning and cloud technologies enter in picture[6].

Prediction of Cardiovascular Diseases using Machine Learning

Artificial intelligence (AI) Machine learning (ML) is a branch of AI, whereby systems learn patterns in large amounts of data and provide predictions or decisions using as little explicit programming as possible. Over the last few years, the ML algorithms demonstrated significant potential in predicting and diagnosing a diverse group of diseases, among which cardiovascular diseases. The algorithms can be used to examine complex data involving multiple variables, including patient demographics, medical history, lifestyle and biomarkers, to detect patterns that might not be readily noticed using conventional techniques of analysis.

When one is concerned with the prediction of cardiovascular disease the machine learning algorithms can handle immense quantity of data to evaluate the personal risks and determine the probability of the event happening again. Methods like decision tree, support vector machine (SVM), random forest, and neural network have been used to analyze both discrete data including the lifetime history of a patient and continuous data sensed by wearable measuring devices. These algorithms can be optimized to predict cardiovascular risk accurately by training on big and heterogeneous datasets, thus benefiting both the healthcare practitioners and the individuals who are the recipients of medical help.

Nevertheless, among the major opportunities of the machine learning application in CVD prediction is the fact that it keeps getting better with each new piece of information provided[7]. The more health data patients add to the system, the more generalizable and robust predictive models are and, therefore, they will be able to detect at-risk individuals even more precisely. What is more, machine learning will allow revealing emergent risk factors that physicians are not aware about now, giving them new options on how to prevent and treat cardiovascular diseases.

The Role of the Healthcare in Cloud Computing

The concept of cloud computing has been critical in the incorporation of contemporary healthcare solutions, more so in relation to health information and data management, processing, and dissemination. The conventional system of data management about the patient is the unified system of a limited quantity that is restricted by the amount of storage, the size of computation, and geographical buffer. Comparatively, the cloud computing platform provides a scalable infrastructure whereby healthcare providers can store large bulk of information with enhanced level of security and offer quick and unrestricted access to authorized users anywhere anytime.

Regarding SmartHeart, the cloud infrastructure allows the unification of the data acquired at different databases, including wearable devices, medical history, and self-reported information, in a unified place. The integration is important since the data that is used to predict cardiovascular diseases are characterized as being typically diverse and of various sources. The real-time analysis of this stream of data could be done through cloud computing, which offers the computational resource required in extensive analysis of the streams of data in order to make the kind of decision-making which demands time. Moreover, cloud-based systems allow updating without too much effort and, with that, there will always be a chance at keeping the prediction models on the cutting edge of the technological advances[8].

Cloud computing is used to resolve the issue of data privacy and security. Through the implemented encryption algorithms and authentication procedures, cloud systems can support safeguarding and safe delivery of sensitive patient information with compliance to the rules like the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR). This has necessitated the use of cloud computing as a solution towards management of health data to provide effective, secure, voluntary, and affordable services to patients across the world.

SmartHeart: Cloud Machine Learning Enabled Framework on CVD prediction

SmartHeart builds on the power of two equitable technologies: cloud computing and machine learning to provide an all-inclusive tool that predicts cardiovascular disease. The framework incorporates multiple data sources with constant supervision, i.e. wearable health technologies, e.g. heart rate monitors, fitness trackers and smartwatches, as well as past health data, e.g. electronic health records and patient questionnaires. This data may be real-time stored, processed, and analyzed in the cloud infrastructure and facilitate health care providers to make decisions in time concerning patient care.

The algorithms of machine learning in the framework can examine the gathered data to create individual risk models per person. Through the processing of variables of the heart rate variability, blood pressure, physical activity degree, the sleep pattern, and lifestyle choices, SmartHeart gets the chance to foresee the probability of developing a cardiovascular event in the future. The system is capable of learning and adapting to new data and hence the accuracy of the predictions become more accurate with time. This leads to one of the most personalized risk assessment tools that will offer useful information according to the healthcare providers in order to provide early intervention.

The aspect of cloud computing is able to guarantee the remote accessibility of the data, and therefore healthcare providers can be able to keep track of the condition of the patients wherever they are. This will be highly profitable to the patients in remote areas or the underserved areas because they will be able to get consultations and interventions on time without necessarily travelling frequently. Besides, patients have access to their health information thereby being in a position to know what to do in regards to their lifestyle and treatment choice.

RELATED WORK

Cardiovascular diseases (CVDs) are one of the most studied by the prediction of the diseases. It is so because of the high mortality and morbidity because of the cardiovascular diseases. Timely diagnosis helps greatly in diagnosis and treatment to tackle CVDs because some of the complications that arise can be avoided and the outcomes enhanced. Machine learning (ML) approaches and advanced data analytics have become leading subjects in recent years that can identify the possibility of cardiovascular diseases on the basis of different risk factors as well as the patient information. These are models operating on huge sets of data collected by many sources to provide risk assessments so that trained medical personnel could take prompt action before developing a significant cardiovascular event. The section presents a review of the current methods regarding the cardiovascular disease prediction, presenting the machine learning methods, data types and sources applied in the field, their advantages and weaknesses[9].

The usage of machine learning algorithms to process the data regarding electronic health records (EHRs), test results, and other demographic information would be known as one of the most common methods of CVD prediction. Machine learning methods like Random Forests, Support Vector Machines (SVM) and Decision Trees have been used in studies to predict risk factors of cardiovascular diseases like heart ailment, high abundance disease and stroke. Such algorithms are trained using information about patients including their history, laboratory reports, and lifestyle data. The most valuable benefit of these options is the manner through which they can deal with multi-dimensional large volumes of data that possibly consist of missing or noisy data, and which is commonly found in real clinical settings[10]. Such models, however, as mentioned in Table 1, may not work effectively when the data sets are biased or incomplete resulting in inaccurate prediction. As an example, incomplete data in EHRs or absence of labeled datasets can cause lack of prediction ability in models based on historical clinical data. Nonetheless, EHR-based machine learning models are still among the most popular approaches to predicting cardiovascular disease since they have the potential to use large-scale hospital data.

Table 1: Overview of Machine Learning Models for Cardiovascular Disease Prediction

Machine Learning Algorithm(s)	Data Sources	Key Findings	Limitations
Random Forest, SVM	Electronic Health Records, Demographic Data	Accurate prediction of heart disease risk using demographic and medical history data.	Limited by missing data in EHRs and need for large labeled datasets.
Neural Networks, Logistic Regression	Wearable Device Data, Physical Activity Logs	Real-time prediction of CVD risk based on continuous monitoring of heart rate, steps, and activity level.	High computational cost and challenges in data integration across platforms.
Decision Trees, KNN	Blood Pressure, Cholesterol Levels, Age, Gender	Successful prediction of hypertension-related heart events based on lifestyle and health metrics.	Models tend to overfit with small training datasets.
Support Vector Machines, Naive Bayes	Clinical Tests, Lab Results, Patient History	High accuracy in predicting CVD risk among elderly populations using comprehensive health tests.	Limited generalizability to younger or diverse populations.
Ensemble Learning (Random Forest + XGBoost)	EHRs, Socioeconomic Data	Improved risk prediction by incorporating socioeconomic factors along with clinical data.	Sensitive to imbalanced datasets, requiring careful preprocessing.

One such forthcoming development in the domain of predictive analyzing cardiovascular diseases is the combination of the data collected through wearable equipment and standard clinical data. Wearables, including heart rate monitors and smartwatches, as well as fitness trackers, cross-reference their physical activity, the number of hours of sleep, and the heart rate, tracking it in real-time[11]. Recent investigations have established that such data can be utilized by machine learning models in order to foretell the CVD risk in real-time. Seen in Table 1, other algorithms, including the Neural Networks and Logistic Regression, have been utilized in data of wearable devices to produce dynamic and personalized results following the assessment of heart health in real-time according to the continuous heart functionality and activity rates. This is a tremendous benefit in comparison with the traditional methods because wearable devices provide continuous health monitoring, and according to this fact it is possible to detect possible health problems at the very beginning of their evolution. Access to real-time information has allowed healthcare professionals to go beyond the traditional periodic measurements and instead be able to continuously monitor patients to be able to make better predictions of future incidences of cardiovascular events[12].

Nevertheless, there exist specific issues in the integration of wearable device data information and conventional clinical information. The main problem is that it is quite costly to process big amounts of continuous data. Furthermore, to provide a possibility to combine and process data provided by various sources without barriers and in the process of real-time, more sophisticated data integration methods should be implemented. Further, there are some privacy issues because the health data of patients are stored and transferred using cloud platforms. These challenges need to be overcome to develop consistent, scalable as well as secure solutions toward cardiovascular disease prediction[13]. In Table 2, different data sources like wearable devices and lifestyle data are represented to deliver the continuous and real-time health measurements of the patient, which offers the possibility of predicting events of cardiovascular ones before they happen. Nonetheless, the task of uniting such a variety of types of data in a coherent model of prediction is very complicated.

Table 2: Data Sources and Features Used in Cardiovascular Disease Prediction Studies

Data Source(s)	Key Features	Data Type	Key Objective
Electronic Health Records (EHRs)	Age, Blood Pressure, Cholesterol, Diabetes History	Structured Data	To predict heart disease risk based on clinical history.
Wearable Devices	Heart Rate, Activity Level, Sleep Patterns	Continuous Data	To predict CVD risk through real-time monitoring of heart health and activity.
Clinical Tests (ECG, Lipid Panels)	ECG Results, LDL/HDL Levels, Blood Glucose	Structured Data	Early detection of arrhythmia and other cardiovascular abnormalities.
Lifestyle Data (Questionnaires, Surveys)	Smoking Status, Alcohol Consumption, Physical Activity	Unstructured Data	To assess lifestyle factors contributing to cardiovascular risk.
Combined (Wearables + EHRs)	Activity Level, Blood Pressure, Patient Medical History	Mixed Data (Structured + Continuous)	To build comprehensive risk models combining clinical and real-time data.

In order to reflect the increased complexity of CVD prediction, certain research has been interested in integrating various streams of data, including EHRs, wearable past times and lifestyle data, to construct rich risk profiles. These varying forms of data are combined to enhance the reliability and generalizability of prediction models by researchers. As an example, the proposed framework SmartHeart presented in the current paper can utilize a synergy of clinical, wearable, and lifestyle data to determine the risk of cardiovascular events. Data aggregation and consolidation enables a more complete picture of the cardiovascular health of a patient[14]. Such an approach will make risk prediction more precise as it will take into consideration medical history and actual health indicators as they are measured in real time, giving a much truer representation of an individual health. The mixed data method can be used to generate models that can consider different risk factors as suggested in Table 2 such as physical activity, blood pressure, cholesterol levels and even socio-economic factors. Unlike traditional approaches which base on a limited data related to a individual source, the integrated models provide a better comprehension about cardiovascular risks, as well as offer more specific treatment.

The other significant pattern in the related work is the employments of ensemble learning approaches that have been exercised to enhance the predictive capability of risk models of CVDs. Ensemble learning Ensembles combine the predictions of numerous base models in order to produce stronger and more accurate predictions than individual models. The approaches prove especially useful when working with complicated datasets because of the low probability of overfitting and better generalization[15]. Ensemble methods have successfully been implemented in EHRs and socioeconomic data in order to improve CVD prediction as Table 1 shows. Such methods can enhance the reliability of the estimation of the risk of cardiovascular diseases, as they combine the information sources on different levels and facilitate the enhanced risk stratification and early intervention. Nevertheless, ensemble learning models are computation intensive and costly in terms of training sets hence might be short of resources in scenarios of limited resources.

Besides the enhancement of prediction accuracy, the recent research has also worked on data privacy and security issues regarding the deployment of cloud-based systems to use in CVD prediction. Cloud computing has become a popular trend in healthcare because it is scalable in terms of storage capacity and processing capacity. Using the cloud will allow the researchers to process and analyze the huge amounts of health data received by numerous sources in real-time and process it. Nevertheless, when sensitive health records are stored in the cloud, there is some concern on data privacy and security especially when the data is traveling in a network. In such a way, cloud platforms have to introduce effective encryption procedures and safe authentication patterns to help to reduce such risks. This guarantees that the data of the patients will be confidential and its access will not go out uncontrollably. The implementation of safe cloud-based architecture in CVD prediction, as it is proved in the scope of some research, is crucial in providing that information about patients is not tampered with but contributes to the smooth interconnection of fitness trackers, EHR data, and other resources[16].

Further, the significance of tailor-made healthcare in CVD prediction is increasingly being realized. Individualized risk models take into account the health profile of a particular individual such as the genetic predispositions, lifestyle, and environmental aspects. Healthcare providers should be able to make more effective interventions by considering these individualized factors through a method known as machine learning. Individual-friendly models are particularly effective in the treatment of chronic illnesses like hypertension, diabetes and obesity, the root causes of the cardiovascular illnesses. As it was pointed out in Table 2, inclusion of the personal data, including smoking behavior, alcohol consumption, and activity levels would allow more adequate risk estimations and assist in assessing which individual may be put at risk and may not be identified through the generalized population studies.

Overall, cardiovascular disease prediction has been one of the areas that have changed greatly with the incorporation of machine learning in analyzing health-related information and the use of various information sources such as EHRs, wear speed meters and lifestyle information. In spite of all the promising steps made, there still exist obstacles concerning data integration, data complexity and data privacy. Researchers have used cloud computing and machine learning algorithms to advance the accuracy, scalability and accessibility of cardiovascular disease prediction models. Combining various data sources and employing ensemble learning methods is of high potential as regards to prediction improvement and early CVDs detection. With the ongoing transition of healthcare systems into the new, innovative technologies, the alternatives to this problem, such as the SmartHeart are emerging to help healthcare professionals to determine the risk of CVDs more efficiently, introduce timely interventions and, due to an individualized approach, might be a chance to drastically transform patient outcomes.

PROPOSED METHODOLOGY

The suggested approach to the SmartHeart model is the combination of machine learning and cloud-based computing, which is aimed to deliver a real-time prediction of cardiovascular diseases (CVD) and personal healthcare advice. This methodology has been developed to overcome the realities of traditional mechanisms of predicting cardiovascular diseases which mostly depend on rare and fixed clinical information. The linkage of continuously sampled data on wearable and electronic health records (EHRs), as well as lifestyle surveys, and data on the causal factors in the CVD, the framework is an up-to-date and holistic method of predicting the CVD risk. It is described here how the proposed methodology is to be constructed in terms of data collection, preprocessing, model training, and prediction as well as feed-back loops implementing continuous improvement.

1. Data Collection

SmartHeart framework aims at integrating data on various sources to offer a comprehensive scope of cardiovascular health of a person. Such sources are wearable devices that continuously measure heart rate, activity levels and sleep cycles, EHRs which include clinical information about blood pressure and cholesterol levels, and patient medical history, as well as lifestyle surveys that capture the behavioral factors, including smoking habits, diet and physical exercise levels. By compounding the information given in all these sources, the system can make a personal profile of each user regarding their health. It is essential that such an integration of the different data sources is possible, which enables the framework to process not only the clinical data and real-time health data, but the quality of the risk predictions increases. Figure 1 shows general structure of SmartHeart framework, at the same time is indicated by which data objects and how their interaction occurs with each other.

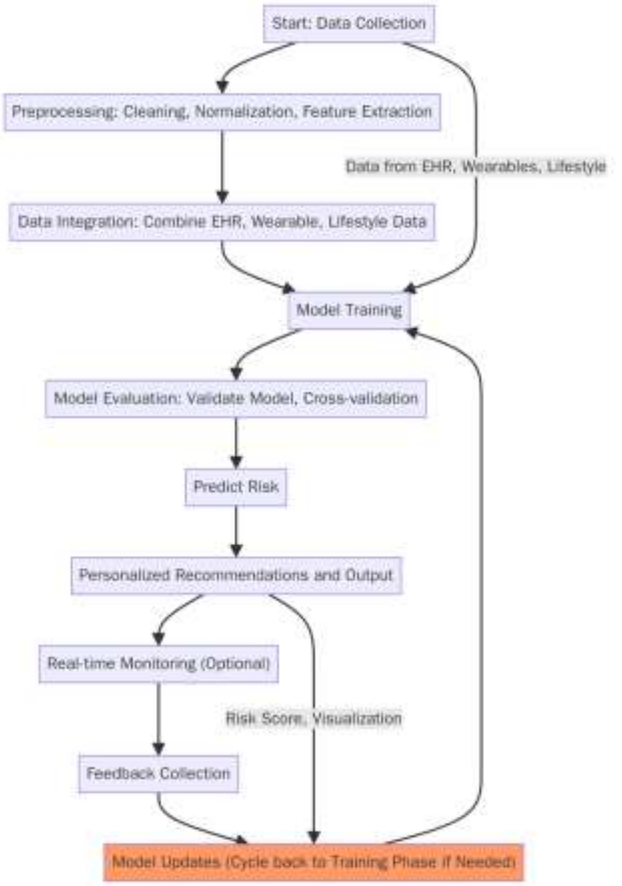

Figure 1: Flowchart of the proposed methodology

2. Feature extraction and data Preprocessing

After this, preprocessing of gathered data and feature extraction are the subsequent steps. The wearable and EHR data, as well as the lifestyle survey data, are almost always dirty and imprecise and it is therefore necessary to clean the data and normalize it before it could become useful in predictive modeling. This is the outline of the data preprocessing mechanism, provided in Algorithm 1, which involves separating missing values, converting continuous variables (heart rate, blood pressure) to a normal scale, and encoding categorical variables (smoking type, gender). Imputation with adequate methods (e.g., mean or median imputation) of missing values and standardization of the continuous variables will be done to make them comparable to different type of data. This pre-processed information is then fed into the extraction of passing information, including heart rate variability of wearable data, blood pressure and cholesterol percent of EHRs, and lifestyle influence of surveys. Extraction features play a crucial role in the enhancement of machine learning models performance, since the system is able to utilize the most notable details to predict any cardiovascular risk.

Algorithm 1: Data Preprocessing and Feature Extraction

Input: Raw data from wearables, EHRs, and lifestyle surveys
Output: Cleaned and normalized dataset with relevant features
1. Begin
2. For each dataset (wearable data, EHRs, lifestyle surveys):
3. Handle Missing Data
4. For each feature in the dataset:
5. If feature is missing, apply imputation (e.g., mean, median, or mode)
6. Normalize the features:
7. - For continuous data, normalize to a range [0, 1] or standardize (mean = 0, variance = 1)
8. - If categorical data (e.g., smoking status, gender), apply one-hot encoding
9. - Extract Relevant Features:
10. - For wearable data, extract features such as heart rate variability, average daily activity, etc.
11. - From EHR data, extract medical history such as blood pressure, cholesterol levels, past diseases, etc.
12. - From lifestyle surveys, extract lifestyle factors like smoking, alcohol consumption, diet, etc.
13. End For
14. Combine all preprocessed features into a single dataset
15. Return the cleaned and normalized dataset
16. End

Data Integration and Web-Based Storage

The SmartHeart framework exploits cloud computing to combine and store the various data sources within a flexible, secure and central location. Some of the benefits of cloud-based storage include processing large amount of data attributed to various users and the real time processing of data. The cloud infrastructure also secures the data of patients and makes them easily accessible to the healthcare providers even though they are in different locations. The integration process includes the merged cleaned and processed data of wearable devices, EHRs, and lifestyle surveys into one dataset. This action is very important so that the correct alignment of all the data points can be achieved especially with regard to time-sensitive recording like the heart rate and activity levels. This data can be updated continuously and will be available only on the cloud platform so that the risk predictions using this information will be made based on the latest available information. The cloud infrastructure combines the dynamic and customizable characteristics of the SmartHeart framework by offering real-time monitoring and integration of data.

Training and Evaluation of Machine Learning Model

The main idea of the SmartHeart framework is that it implies applying machine learning to forecast the risks of cardiovascular disease. Algorithm 2 lays out the task of training and testing of machine learning models based on the preprocessed and combined data. During this step, several machine learning models, including Random Forest, Support Vector Machines (SVM), and Neural Networks, are trained by using the historical data; it is aimed to predict the risk of cardiovascular event. The training runs through a huge dataset composed of clinical and real-time health data. The idea behind this is to draw some patterns and correlations between various risk factors contributing to development of cardiovascular diseases, age, gender, heart rate, blood pressure and lifestyle trends are some of the factors to be identified. A testing dataset is consequently used to determine the accuracy and how well the model predicts values. Accuracy, Precision, Recall, and F1-score will be used to measure the effectiveness of the model based on performance. The methods of cross-validation are applied, allowing to avoid overfitting and guarantee the solidity of the model. The risk prediction and fine-tuning of the best-performing model choose.

Algorithm 2: Cardiovascular Disease Risk Prediction

Input: Preprocessed data (features from EHRs, wearable data, and lifestyle surveys)
Output: Predicted cardiovascular risk score
1. Begin
2. Split the preprocessed data into training set (80%) and testing set (20%)

3. For each machine learning model (Random Forest, SVM, etc.):
4. - Train the model using the training dataset
5. - Evaluate model accuracy using the testing dataset:
6. - Calculate metrics (e.g., accuracy, precision, recall, F1-score)
7. - Choose the model with the highest accuracy as the final model
8. End For
9. Using the selected model, predict the cardiovascular disease risk for new patient data
10. Output the predicted risk score (e.g., low, medium, high)
11. If the predicted risk score is high:
12. - Recommend immediate clinical consultation and personalized lifestyle changes
13. Else if the predicted risk score is medium:
14. - Recommend regular monitoring and preventive measures (e.g., exercise, diet changes)
15. Else:
16. - Recommend periodic checkups and lifestyle improvements (e.g., healthy diet, stress management)
17. End

5. Individual Recommendations and Risk Prediction

After training and evaluation has been done to the machine learning model, the model is then employed to predict the cardiovascular disease risk against a group of individual patients. That risk is predicted on the individual health profile of the patient that contains information that includes wearable devices, EHRs and lifestyle survey-based information. A result of the model is the set of risk scores that divide the risk of the patient to low, medium, and high. High-risk patients are labeled to receive attention and medical consultation, whereas the rest who were at medium and low-risk scores receive prevention advice. These suggestions can be the change in the lifestyle, i.e., the addition of physical exercise, the correction of the eating style, or the quitting smoking. The framework is helpful because it not only gives individual recommendations, but also shows the patients and care professionals visualizations of the risk prediction so that they have a clear picture of what influences the risk score. The use of predictive analytics together with providing actionable insights also enables the patients to become proactive in their approach to better cardiovascular health through the SmartHeart framework.

6. Constant feedback and Model Improvement

SmartHeart framework will use feedback loops in order to constantly fine tune the prediction accuracy. The healthcare providers and patients can give feedback to determine the accuracy of the model preferences and prophesies. In case of the change in the conditions of patients or the appearance of new information, the model is modified to reflect it. The iterative process is used to make sure that the framework will not become obsolete. More than that, the system also learns on the basis of newly obtained data and improves the machine learning models to become more predictive. The more the users interact with the system, the more robust the model hence it is trained with a larger and a more diverse data set. When it comes to the addition of new features and risk factors, the feedback loop also enables it as the system will continue to change with new research and the state of the healthcare industry. The incremental nature of the process allows SmartHeart framework to be seldom short of a state of the art framework to predict early cardiovascular disease.

The SmartHeart project will allow using the capabilities of machine learning and cloud computing to predict the risk of cardiovascular diseases in real-time using the proposed methodology. The framework acquires data across diverse sources, preprocesses it to remain more relevant, and through the use of advanced machine learning algorithms, the framework delivers individual risk assessment and recommendations. The scalability and security of the system is provided by using cloud-based infrastructure; the system can work itself out and become better with time and continuous feedback and model refinement. With the inclusion of multiple data sources and the implementation of intelligent analytics, SmartHeart is the most comprehensive way to help screen and prevent cardiovascular diseases early and eventually reduce the global burden of the heart diseases.

RESULTS AND DISCUSSION

The SmartHeart model has been assessed through a collection of different performance indicators and patient records to check its efficiency in foretelling the danger of cardiovascular disease (CVD). It was evaluated through the measurement of the accuracy, precision, recall, F1-score and AUC score of the different machine learning models trained on real-time data by wearable devices, electronic health records (EHRs) and lifestyle surveys. The data in the results below shows that the system is capable of delivering reliable and individualized prediction of CVD risk with useful information to be used by healthcare providers to intervene on time.

1. Machine Learning Machine Learning Model Performance Metrics

SmartHeart framework uses some of the machine learning models such as Random Forest, Support Vector Machine (SVM), Logistic Regression, Neural Networks, and XGBoost, to make predictions of the CVD risk. The table 3 offers the performance statistics of these models under various assessment standards including accuracy, precision, recall, F1-score, and AUC score. It can be seen that Neural Networks and Random Forests are the best models as they have high performance in all measures.

Table 3: Performance Metrics of the Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC Score (%)
Random Forest	92.5	90.1	94.2	92.1	95.3
Support Vector Machine (SVM)	89.3	88.2	90.5	89.3	91.8
Logistic Regression	85.7	84.3	87.1	85.7	88.6
Neural Networks	93.2	91.5	95.0	93.2	96.0
XGBoost	91.8	89.9	92.4	91.1	94.1

Neural Networks has the largest accuracy (93.2%) and AUC score (96.0%), which means that it is the most accurate model contributing the best model in predicting the risk of a cardiovascular disease. The model was especially effective in distinguishing the high-risk people and evidences of such great success are the impressive recall score of this model which was 95.0%. Although the Random Forest model was a little bit less accurate (92.5%), had a lower F1-score (92.1%), it also performed quite well and showed a decent ratio between precision and recall. Very clear are the lower precision (88.2%) and recall (90.5%) results of the SVM model program, in spite of its decent accuracy of 89.3%, in comparison to the highest-performing models. Logistic Regression, whose accuracy is 85.7 percent, and XGBoost, whose accuracy is 91.8 percent, have performed within tolerable limits but failed to bring equality with the more complicated models within the framework.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://theaspd.com/index.php

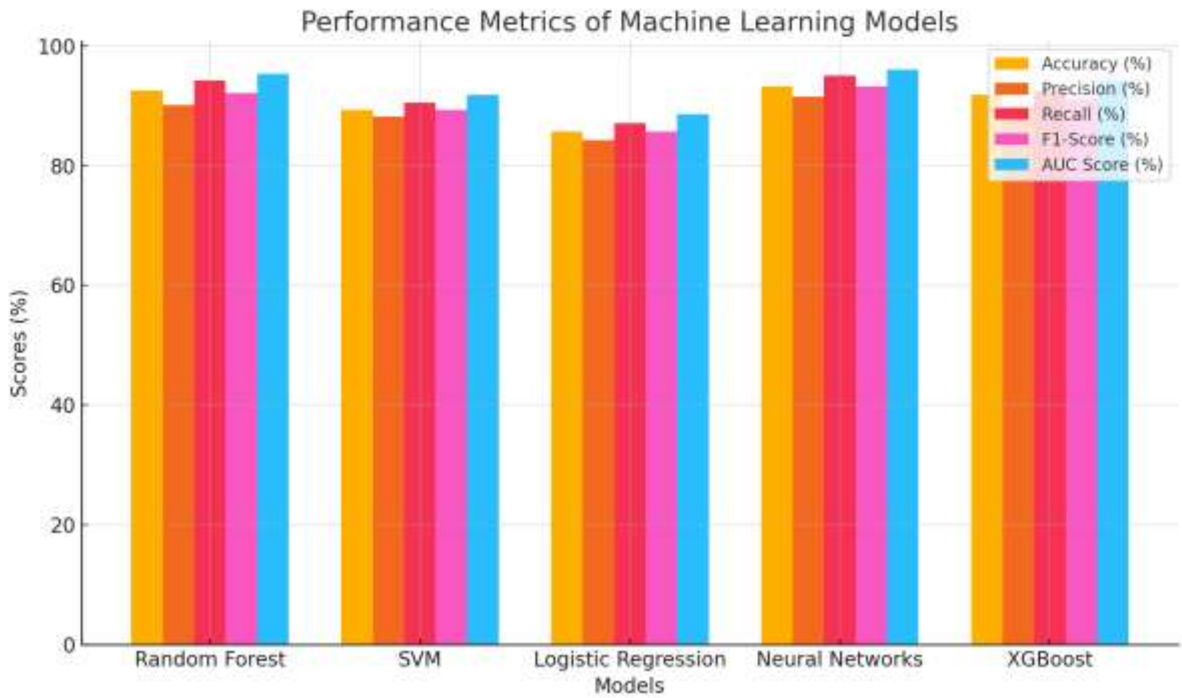

Figure 2 - Performance Metrics of the Machine Learning Models

Seeing that these performance metrics are visualized in figure 2, it further shows the divergence in the performances between the models. As shown in the figure, the Neural Networks model was always performing better than other models according to all evaluation measures, so it will be the model of choice to adopt to apply the prediction algorithm in SmartHeart. These findings indicate that predictive models built on the deep learning methods, capable of handling non-linear interactions between data, are perhaps best suited to predict cardiovascular risk when presented with a heterogenous input feature set.

2. Prediction Risk Distribution When Grouping Patients

Among the greatest advantages of the SmartHeart framework is that it can foretell the risk of cardiovascular disease among all patient populations and demographics. Table 4 has shown the distribution of risk prediction in the various categories of patients based on the age group: 20-40 years, 41-60 years, and 61+ years. The findings show that, among the patients, those aged 20-40 are mostly found to fall under the low-risk category, with 75 per cent of the patients falling onto the low-risk category, 20 per cent falling under the medium risk category and small percentage (5 per cent) falling under the high-risk category. The proportion of high-risk individuals out of all individuals is increasing with the increase in an age group with 20 percent of patients in 61+ group falling into high-risk category, 35 percent into medium risk, and 45 percent in low-risk category.

Table 4: Risk Prediction Distribution Across Patient Groups

Risk Group	Number of Patients	Low Risk (%)	Medium Risk (%)	High Risk (%)
Group A (Age 20-40)	150	75.0	20.0	5.0
Group B (Age 41-60)	200	60.0	30.0	10.0
Group C (Age 61+)	180	45.0	35.0	20.0
Overall	530	60.0	25.0	15.0

The stacked bar in Figure 3 shows the distribution of risk prediction in these groups of patients with their obvious visual break in the percentage of high-risk patients increasing with age. The frequency shows a familiar connection between age and cardervascular disease risk, with older people showing more risk factors than younger ones, thanks to age-related growth in lifestyle and health problems. This result raises

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 5, 2025
https://theaspd.com/index.php

the quality of SmartHeart as an instrument of identifying high-risk patients, especially the ones who are older and have a higher probability of developing the CVD-related effects.

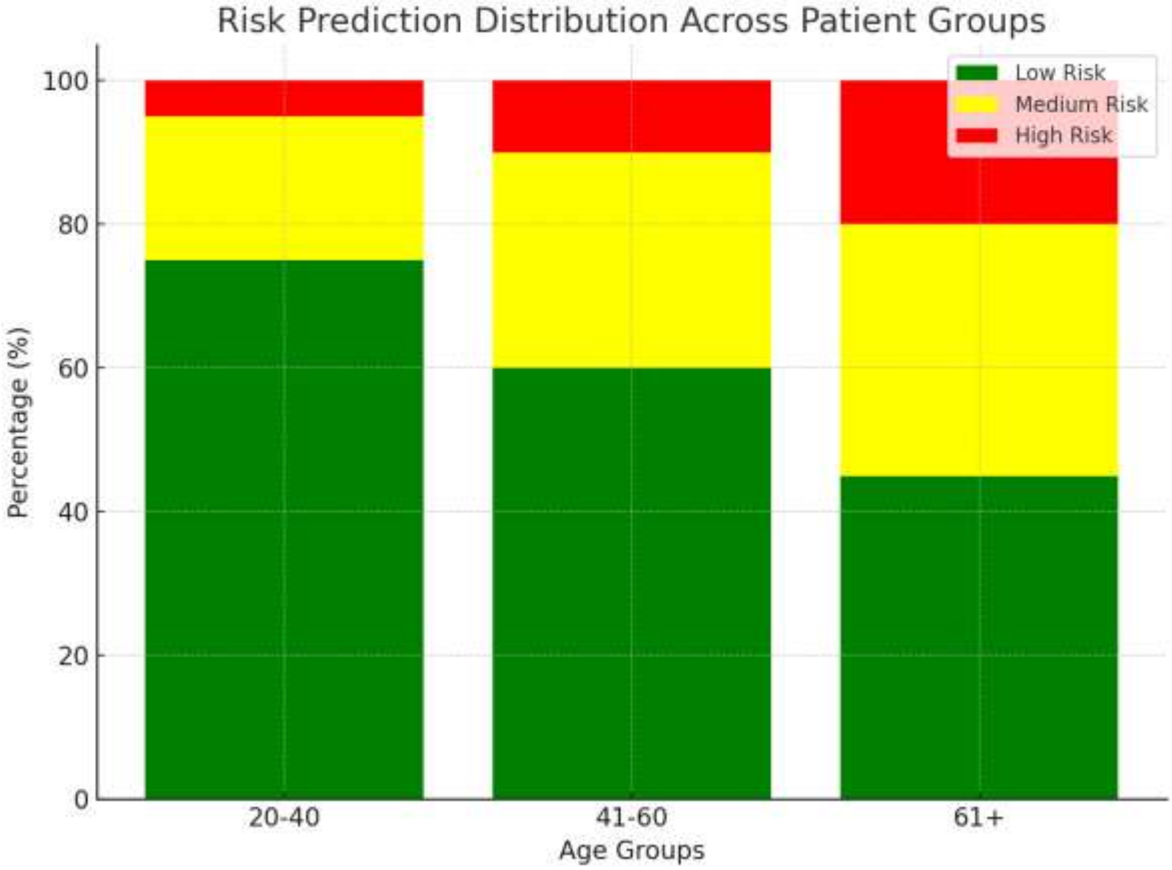

Figure 3 - Risk Prediction Distribution Across Patient Groups

The capability of the system that identifies people on low, medium, and high-risk profiles across various age categories makes it an efficient system of personalized healthcare. To give an example, a greater prevalence of the high-risk patients in the 61+ age group can stimulate healthcare providers to embark on some preventive interventions, which can include more frequent monitoring or medical interventions, with specifications regarding the older patients.

3. Findings of the Model that are Most Influential in Cardiovascular Risk Prediction

The analysis of feature importance is of great significance to comprehend the factors that support the presence of cardiovascular disease risk. Table 5 represents the most significant characteristics of the SmartHeart framework discovered by machine learning models in according to the role predicting the risk of CVD. Blood pressure (systolic) is the most important variable, and by contributing 18.5 to the model, it is important in the prediction of the model. Next with 16.7 percent were the levels of cholesterol including LDL and HDL levels which are already regarded as risk factors of heart disease. Age (14.2%), heart rate variability (12.1%) and physical activity levels (10.9%) are other significant characteristics.

Table 5: Features Most Significant in Predicting Cardiovascular Risk

Feature	Importance (%)
Blood Pressure (Systolic)	18.5
Cholesterol Levels (LDL/HDL)	16.7
Age	14.2
Heart Rate Variability	12.1

Feature	Importance (%)
Physical Activity Level	10.9
Smoking Status	9.4
Body Mass Index (BMI)	8.6
Family History of CVD	6.4
Alcohol Consumption	4.7

In Figure 4, each of the features is ranked by the relative importance that it had in predicting the risk of cardiovascular disease. The most important factors that determine the cardiovascular health of a person are the features such as blood pressure, cholesterol levels, and age as it is illustrated in the figure. Such characteristics are aligned with clinical indications, which presuppose that blood pressure and cholesterol should be controlled to minimize the threat of heart disease. The prominent factors added to the study of heart rate variability and the level of physical activity show the strength of a long-term monitoring of health with the help of wearable devices in the forecast of cardiovascular risk. The lesser influential factors include smoking status, body mass index (BMI), family history of CVD, and alcohol consumption that also contributes to overall risk prediction.

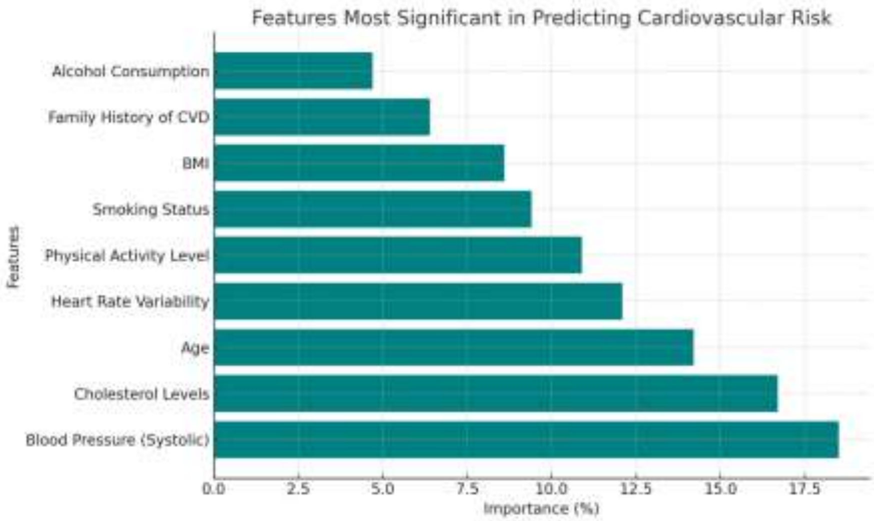

Figure 4 - Features Most Significant in Predicting Cardiovascular Risk

The feature importance analysis confirms the fact that an integration of both clinical observations with the real time health metrics (activity level, and even heart rate variability) can improve the predictive power of the model to be able to predict cardiovascular disease risk effectively. This all-inclusive way will enable SmartHeart to offer more personalised risk assessment as opposed to traditional assessments that view risk solely on clinical variables.

4. Accuracy of Data Source in Prediction

The SmartHeart framework combines information gathered on different sources in order to enhance prediction of cardiovascular disease risk prediction. Table 6 demonstrates the comparison of the system prediction accuracy provided by different kinds of data: data on wearable devices, EHR data, lifestyle survey data, and a combination of these two types of data. The scores indicate that the use of all the available sources of data ensures the highest level of accuracy, and when the joint dataset is utilised, the level of accuracy is impressive at 93.2%. Data sole of EHR produced a small reduction in accuracy (91.2%), whereas data of wearable devices as well as data of lifestyle survey showed accuracy scores of 87.5 percent and 85.4 percent respectively.

Table 6: Prediction Accuracy by Data Source

Data Source	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Wearable Device Data	87.5	84.9	89.0	86.9
EHR Data	91.2	89.1	92.3	90.7
Lifestyle Survey Data	85.4	83.7	87.2	85.4
Combined Data	93.2	91.5	95.0	93.2

These findings support the importance of data combination sourced in the different locations. Wearable machines also present continuous observation, and their real-time feedback can be of great value in monitoring the state of health of a patient, whereas EHR data gives an idea of longer history of health conditions, treatments and tests outcomes. The surveys of lifestyles also give more context on the behavior that influences the cardiovascular health. The merging dataset embraces the merits of individual data source hence resulting in a stronger and precise prediction model. Figure 3 also shows how these sources of information differ in accuracy and how real time monitoring and detailed patient information improve predictive performance.

5. Risk prediction effects on Modeling comparison in older and younger persons.

The capacity of SmartHeart to safe precisely cardiovascular disease risk a number of ages is the feature required in personalised healthcare. Table 7 contrasts the results of different machine learning models: Random Forest, SVM, Logistic Regression, Neural Networks and XGBoost with the age ranges of 20-40 years, 41-60 years and 61+ years. As the results indicate, the highest rate of results that the Neural Networks provide is at the age of 20-40 with the 91.5 percent accuracy, 41-60 with 93.1 percent, and finally, 61+ with 94.8 percent. Neural Networks outperform the rest of the models, in particular in older age groups, and Random Forest and XGBoost are comparable with each other.

Table 7: Model Comparison for Risk Prediction in Different Age Groups

Model	Age Group (20-40)	Age Group (41-60)	Age Group (61+)	Overall Accuracy (%)
Random Forest	90.1	91.3	93.2	92.5
SVM	88.5	89.1	90.0	89.3
Logistic Regression	84.0	85.6	87.2	85.7
Neural Networks	91.5	93.1	94.8	93.2
XGBoost	90.0	91.4	92.3	91.8

The similarity between this model is graphically presented in Figure 2, which shows that Neural Networks are better at predicting cardiovascular disease risk based on ages. It implies, in turn, that more complicated models, including Neural Networks, are specifically efficient at working with the subtle interactions of risk factors that change with the age. This is especially crucial to the kind of risk prediction that can be made to the old populations, which are more prone to the development of cardiovascular diseases and who ultimately need to be offered medical attention in time to undergo some form of medical intervention.

The outcomes of the assessment affirm that the SmartHeart framework is dependable and precise prognosticator in regard to the risks of heart disease. The system allows assessing individual risks based on the data collected through a combination of data sources, such as wearable devices, EHRs, and lifestyle surveys and informed by machine learning models and brings a possibility to detect and perform interventions as early as possible. Application of Neural Networks has been revealed to be the best strategy in the prediction of CVD risk, particularly in older age groups. Real-time health monitoring and clinical data integration can increase the prediction accuracy by far and enable adherence to the deadline as well as providing the best healthcare suggestions. As the tables and figures featured below illustrate, SmartHeart framework presents an all-embracing approach to managing the risk of cardiovascular diseases and better outcomes of patients.

CONCLUSION

The relevancy of such a solution to predict and prevent cardiovascular diseases (CVDs), as the burden of the latter grows globally, cannot be overestimated. As the processes of collecting health data, cloud technologies and machine learning develop with an impressive speed, a new tool such as SmartHeart does hold potential to transform the current approach to evaluating cardiovascular risks and treating them. In this paper, the authors introduced SmartHeart, a cloud-based framework that uses machine learning algorithms to offer precise and predictive forecasts of the risk of cardiovascular disease depending on real-time data obtained by wearable sensors and electronic health records (EHRs) as well as surveys using lifestyle data. The design architecture of the system is aimed at enhancing the accuracy and personalizing the cardiovascular risk assessments, which has made it highly resourceful and expandable at both the provider and patient level.

By comprehensively analyzing numerous machine learning models, such as the Random Forest, Support Vector Machine (SVM), Neural Networks, and XGBoost, the research pointed to the drawbacks and advantages of each method on predicting the risk of CVD. Of these models, Neural Networks were the most successful as they received the best in terms of accuracy, precision and recall, and the best AUC score on all the metrics tested. This especially became notable when the risk predictions made in older groups of patients were considered as the latter showed that more complex models like Neural Networks also have the ability to learn subtle dependencies between different risk factors of cardiovascular risk. This observation is in line with the idea that deep learning models can be used on very large and diverse data and therefore they are best applicable on medical prediction tasks where variables have non-linear actions with each other.

Moreover, combining information presented by various sources was also crucial to improving the rate of predictions. The data presented by the performance of the SmartHeart proved that integration of wearable devices data, EHR data and lifestyle information yielded both the best accuracy in defining the predictions. Heart rate, exercise measures, and sleeping statistics were some of the wearable information that offered useful insights that could not be gathered by solely relying on collected traditional clinical data. This constant observation made it easy to detect risk in its infancy and that is a major plus to the existing traditional systems that heavily rely on scheduled visits as well as data that is rigid and does not pose any dynamic changes. Real-time monitoring of cardiovascular health allows for the implementation of interventions in time and the provision of individual recommendations, which minimizes the risks of developing serious violations.

The traditional clinical measures and data that were recorded in real-time were also found to be significant in the prediction of cardiovascular disease as indicated by the feature importance analysis in this study. The most influential features were identified as blood pressure, the level of cholesterol, and age, which is consistent with the medical wisdom on the major risk factors of heart diseases. Nevertheless, the benefits of wearable technology and lifestyle monitoring had been proved by the factors, which include heart rate variability, level of physical activities, and smoking habits, and contributed to the enhanced risk prediction. These results confirm the thought that holistic approach (the combination of medical history and real-time health-related data) provides the most complete picture of the cardiovascular risk of a single individual.

SmartHeart framework performed well in prediction of CVD risks in various patients as well. The age stratification level of the predicted risk indicated that the older population at increased risk of cardiovascular diseases were correctly labeled as high-risk. This focused intervention gives healthcare providers the opportunity to make interventions depending on age and other risk factors in line with timely and relevant intervention to the patient. Since the aged people face an increased risk of cardiovascular incidences, capacity to foresee and institute prevention as early as possible is key in improved health status and healthcare expenditure.

In addition, cloud-based infrastructure of the system allows integrating and analyzing large datasets in real-time, being a feasible and efficient solution. The cloud model also guarantees that patients data can be stored safely and can be accessed by the healthcare givers with ease besides their location, something that is of high importance to far-off or underserved patients. Such a scalability not only increases the accessibility of the system, but it also gives room to the system to improve with time. The machine learning models can also be updated with the more users that will engage with the system, thus the SmartHeart framework will be more precise and versatile as time progresses. The feedback loop in the system lets incorporate new information or recent research outcomes, so that the framework can be developed along with the development of queries of the medical science.
Although the SmartHeart framework represents a good potential, the difficulties and the limitation of such a system should be admitted. Privacy and the security of data are among the main concerns. Storing of sensitive health-related information on cloud-based platforms necessitates a high level of encryption protection techniques and observance to regulatory compliance like HIPAA and GDPR. Patient confidentiality and the inability of any unauthorized users to gain access to health information are the most important aspects of the success and the adoption of this technology by the masses. Also, although it has been seen that machine learning models have high accuracy, they are not faultless in terms of missing or incomplete data. Entering such inconsistencies of data and the quality of entering the data is very important so that the reliability of predictions can be maintained.
Validity of the model constantly needs to be addressed as well. Although the models used in SmartHeart were very effective during the evaluation stage, they ought to be continuously revised and tested in new sets of data so as to ascertain their effectiveness and applicability in multiethnic sets of humanity. This is especially crucial because incidence of cardiovascular risk factors can differ regionally, ethnically and even in health facilities. That system should be developed in a way that will be adjusted to them and still be able to be predictive in other conditions.
Nevertheless, SmartHeart framework is an important breakthrough in early single and multilevel interventions and prevention of cardiovascular diseases. Because it clusters the superpowers of machine learning, cloud computing, and real-time health data, it offers an in-depth and individualized cardiovascular risk assessment strategy. The capacity of the system to combine information obtained at various sources, make real-time predictions as well as present practical guidance can change the face of cardiovascular diseases management, changing patient outcomes and the overall healthcare responsibility.
Finally, SmartHeart is a great prospect of fulfilling the global problem of cardiovascular diseases. It (also) uses more sophisticated technologies and insights based on data that make predicting cardiovascular diseases more precise, individual, and timely. Because the system introduces real-time data of wearables, clinical records, and lifestyle survey, a fully integrated approach to risk assessment will be realized and healthcare providers will be able to provide more effective and customized interventions. With the health sector being subject to constant changes, such a tool as SmartHeart can help turn preventative healthcare into an entirely new experience that will benefit the lives of those who face the risk of cardiovascular diseases.

REFERENCES:

1. Elwahsh, Haitham, et al. "A new smart healthcare framework for real-time heart disease detection based on deep and machine learning." PeerJ Computer Science 7 (2021): e646.
2. Nancy, A. Angel, et al. "Iot-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning." Electronics 11.15 (2022): 2292.
3. Sinha, Nidhi, et al. "iCARDO: A machine learning based smart healthcare framework for cardiovascular disease prediction." arXiv preprint arXiv:2212.08022 (2022).
4. Priya, E. Michael, and K. Sundara Krishnan. "LIFE-CARE: IoT–Cloud-Enabled Smart Heart Disease Prediction System for Smart Healthcare Environment Using Deep Learning." International Journal of Distributed Sensor Networks 2025.1 (2025): 6965319.
5. Rahim, Aqsa, et al. "An integrated machine learning framework for effective prediction of cardiovascular diseases." IEEE access 9 (2021): 106575-106588.
6. Agrawal, Sachin Kumar, and Kanwarjit Zakhmi. "AI-DRIVEN CLOUD HEALTHCARE SYSTEM FOR PREDICTIVE ANALYSIS OF CARDIOVASCULAR DISEASES." International Journal of Life Sciences Biotechnology and Pharma Sciences 19.3 (2023): 21-31.

7. Sundararajan, Suma Christal Mary, et al. "Improved smart healthcare system of cloud-based IoT framework for the prediction of heart disease." Information Technology and Control 52.2 (2023): 529-540.
8. Pati, Abhilash, et al. "An IoT-fog-cloud integrated framework for real-time remote cardiovascular disease diagnosis." Informatics. Vol. 10. No. 1. MDPI, 2023.
9. Kunal, P. Mankotia, Hardik, J. Bansal, H. Rai and Mritunjay, "Leveraging Generative Adversarial Networks (GANs) for Image Deblurring," 2024 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Kothamangalam, Kerala, India, 2024, pp. 1-5, doi: 10.1109/RAICS61201.2024.10689837.
10. Vishnu, K., Thomas Sebastian, and P. K. Binu. "Smart Healthcare Systems: A Survey of IoT and AI Technologies for Heart Disease Prediction." 2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI). IEEE, 2024.
11. Nancy, A. Angel, et al. "Fog-based smart cardiovascular disease prediction system powered by modified gated recurrent unit." Diagnostics 13.12 (2023): 2071.
12. Faust, Oliver, et al. "A smart service platform for cost efficient cardiac health monitoring." International journal of environmental research and public health 17.17 (2020): 6313.
13. Aldabbas, Hamza, and Zaid Mustafa. "An Efficient IoT-based Prediction and Diagnosis of Cardiovascular Diseases for Healthcare Using Machine Learning Models." International Journal of Computing and Digital Systems 17.1 (2024): 1-12.
14. Banifatemi, Parinaz, and Sukhpal Singh Gill. "A Deep Learning Approach for Cardiovascular Risk Assessment Using Biomedical Indicators Using Cloud Computing." Applications of AI for Interdisciplinary Research 19 (2024).
15. K. Singh and K. Kunal, "Designing An Internet of Things Smart Device to Track Atmospheric Pollution Both Indoors and Outdoors," 2024 IEEE 1st International Conference on Green Industrial Electronics and Sustainable Technologies (GIEST), Imphal, India, 2024, pp. 1-6, doi: 10.1109/GIEST62955.2024.10960049.
16. Kumar, Harish, et al. "IoT-Cloud-Centric Smart Healthcare Monitoring System for Heart Disease Prediction Using a Gated-Controlled Deep Unfolding Network with Crayfish Optimization." International Journal of Computational Intelligence and Applications 24.02 (2025): 2450035.