Real-time Prediction Model for Heart Disease Risk during Medical Consultations and Health Monitoring

Yerraginnela Shravani[1,*], Ashesh K.[2]

[1]PhD-Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India
[2]Associate Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India

Emails: shravanilokeshwar@gmail.com; imasheshk@kluniversity.in

Abstract

In the realm of cardiovascular health, early detection and proactive management of heart disease are critical for improving patient outcomes. This paper introduces a novel real-time prediction model designed to assess heart disease risk during medical consultations and continuous health monitoring. Leveraging advanced machine learning techniques and a diverse dataset comprising patient demographics, medical history, and biometric measurements, our model provides immediate, actionable insights into an individual's cardiovascular health. The model integrates seamlessly with electronic health record (EHR) systems and wearable health devices, offering real-time risk assessments that aid healthcare professionals in making informed decisions and tailoring personalized treatment plans. Through extensive validation and testing, our model demonstrates high accuracy and reliability, with potential to significantly enhance early intervention strategies and patient engagement in heart disease prevention. This research underscores the transformative potential of real-time predictive analytics in clinical practice and highlights pathways for future development and integration of intelligent health monitoring solutions.

Keywords: Health Monitoring; Model creation; Heart Disease Identification; Deep Learning; Machine Learning

1. Introduction

Heart disease is a significant global health concern, contributing to over 32% of total deaths worldwide. However, it is predicted that over 75% of these deaths will occur in developing countries. As a result, an automatic prediction system for heart disease risk is essential to prevent potential threats to life. Timely heart disease prediction can significantly reduce mortality rates and enhance the quality of life for patients with cardiac problems.

In this context, health data monitoring is imperative for research and understanding the diseases faced by patients. It is also crucial for examining irregularities in health-associated factors and keeping patients' health records up to date. Such health records could include age, gender, travel history, body mass index (BMI), alcohol intake, diabetes, blood pressure rates, and medications. In recent years, there has been an increase in the number of diseases due to unmanaged health risks. Electrocardiogram (ECG) data is an important factor in determining heart disease in patients [1-3]. Managing health records and ECG monitoring is challenging due to the large amount of data generated and poor sensitivity factors in current collection systems.

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

This study presents a real-time prediction model for heart disease risk using health data monitoring, which is further enhanced using ECG heart signal pattern analysis. The proposed work also includes a framework for continuously monitoring real-time health data using sensors attached to a person's body. It focuses on monitoring health data and analyzing heart rate signals to check for possible irregularities in the patient. Following this, the monitored data is securely uploaded to the cloud server. Once the data is uploaded, it is analyzed using a prediction model developed using various approaches. The best suited and most efficient model for determining heart disease is identified. The current state of the technology captures the data using nearby devices and models the same for detecting heart disease [4-5].

Current consultation systems need to be enhanced to remember previous patient history records. A system that works with health monitoring systems and is linked to doctors' data will automatically alert the patient and doctor if an abnormality is detected in the heart signals. The patient's health details will be shared with the doctors on the server continuously for the checkup and analysis of ongoing heart risk activities. An enhancement of the current consultation would alert both the doctor and the patient simultaneously.

In many developing countries, there is a shortage of physicians and specialists. Moreover, cardiologists are the highest-paid medical professionals [6]. AI models can be used to help avoid mortality by providing intelligent prediction models that can predict the likelihood of heart disease in individuals with a high rate of false positives.

Cardiovascular disease (CVD) remains a leading cause of mortality and morbidity worldwide, with an estimated 17.9 million deaths in 2019. Early prediction and diagnosis of CVD can significantly reduce adverse health effects, and robust risk prediction models can assist health professionals in evaluating the risk of heart disease in patients by analyzing previous patient data. However, existing systems suffer from drawbacks such as difficult interface interactions, lack of real-time results, and reliance on long-term prediction modeling. In medically historical CVD record datasets, companies struggle to obtain predicted results for patients and health experts. Hence, a new model is proposed to provide real-time heart disease risk prediction using recorded CVD dataset variables during patient medical consultations. The Health Monitoring System is built to allow health consultation with an expert doctor and recording patient health details for the risk prediction model with the recorded dataset variables. A user-friendly web-based interface with multiple plug-in graph representations is designed. Two solutions, Random Forest and Neural Network, are proposed to analyze the health attention details dataset and provide a heart disease risk prediction result based on previous patient health records. The model has an accuracy of 93% and 95%, respectively, with sensitivity and specificity above 85%, providing a valuable medium for taking precautionary heart disease measures and preventing the risk of heart disease [7-9].

Cardiovascular diseases (CVD) are disorders of the heart and blood vessels, including coronary artery diseases, cerebrovascular diseases, rheumatic heart disease, and congenital heart defects. CVD remains the leading cause of morbidity and mortality across the globe, despite the implementation of preventive measures. Adverse clinical manifestations of CVD often occur unexpectedly without any prior symptoms, thus leading to higher mortality rates. The high prevalence and sudden deaths from heart disease have caused increased public awareness of this threat. Strikes, i.e., attacks by heart disease, can happen to anyone without any caution. Thus, a heart disease risk prediction model would be a valuable medium for easily taking heart disease precautions and preventing the risk of heart disease [10]. There is an extensive range of previously recorded datasets internationally and across the globe. These CVD risk datasets generally consist of heart disease record details containing personality characteristics, lifestyles, and health functional data that provide a heart disease diagnosis. However, with the existing proposed systems, companies suffer from several shortcomings. Most existing solutions are long-term prediction-based modeling and built models with non-robust methods. The restriction of simple statistical index-based modeling tools limits taking precautionary measures beforehand. Simple web interface modeling requirements are difficult for commerce [11].

2. Literature Review

Heart disease is the leading cause of global deaths and represents a significant portion of healthcare costs. Early identification and prevention of heart disease risk factors can provide cost-effective medical management. Therefore, healthcare systems need automated and efficient prediction systems to assist physicians in the early detection of heart disease to alleviate the national disease burden. A real-time heart disease prediction model can predict the risk of heart disease for a patient during medical consultations and health monitoring. The prediction model is scalable on remote healthcare and disaster management systems that require real-time processing of patients' health data streams collected from portable wireless medical sensors [12]. It implements a series of bioengineering algorithms to extract features

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

from the patient's raw heart disease-relevant bio signals collected from sensors, which are necessary for heart disease prediction. This feature set is efficient and small in size compared to other machine learning methods.

Several supervised learning models are trained using the feature set for heart disease risk prediction. The trained model is deployed on the health monitoring and medical consultation system. The system can be used by physicians to retrieve heart disease risk assessments for a patient during medical consultations. For a patient being monitored in real-time, the system automatically accesses patient body health data, executes the feature extraction algorithms, and processes the features using the prediction model. The output of the prediction model is an indication of heart disease risk. A portable low-cost wireless sensor was developed to collect heart-age-relevant bio signals from patients at the point of care to support the proposed model. The proposed model is evaluated on public datasets and demonstrates state-of-the-art prediction performance and efficiency [13].

The predictive models are designed to reduce the risk of undetected heart disease for patients seeking medical consultation and to aid in the proper utilization of medical resources, depending on the severity of the patient's health condition. Both models establish a certain level of processing based on raw and continuous data streams collected from portable wireless medical sensors in a remote healthcare scenario. In this scenario, each patient or monitored subject represents a unique data stream for the clearly defined tasks. For daily health monitoring, the model starts acquiring seconds-long data streams and collecting the continuous patient data with a predefined subject/stream designation [14]. This initial part is used for deep sleep extraction via methods resembling the human REM phase; it is required for actionable and reliable heart disease risk predictions.

Various heart disease prediction models have been developed over the past several decades. These models can be classified into three categories: risk stratification tools, clinical decision tools, and adjunct diagnostic tools. An epidemiological study started in 1948 and intends to identify the risk factors of cardiovascular disease in the population. It developed heart disease risk equations, which estimate the 10-year risk of coronary heart disease (CHD) events (i.e., myocardial infarction and coronary death) in people who have not previously experienced CHD events. The heart disease risk is calculated by age, sex, blood pressure, blood cholesterol, smoking habits, and diabetes [15-18]. Although widely used, these equations were developed based on the risk events and mortality patterns in the study population and might underestimate or overestimate heart disease risk in other populations and settings. Guidelines recommended these heart disease risk equations to be computer-assisted [22].

The heart disease prediction model was developed based on the records of over one million patients, aged 30-84, for the assessment of the 10-year risk of CHD events. Individuals who were pregnant, included in the previous study, undergoing treatment for CHD, or living overseas were excluded. Seven datasets and candidate cardiovascular risk factors were used. Being widely used in primary care systems, three versions of this model have been developed since 2007, with the most recent versions being the latest iterations. The heart disease predictive model was implemented in software developed for use in the region [19].

The Reynolds Risk Score, a heart disease predictive model developed based on a health study, estimates the 10-year risk of a first cardiovascular event. It was developed based on a cohort of women, with exclusion criteria including existing cardiovascular disease or cancer diagnosis. A clinical model predicting heart disease risk was also developed based on five commonly measured RNA markers and a two-step approach predicting coronary artery disease (CAD), including logistic, LASSO, and ridge regression, using clinical epvariables [20].

A. Advancements in Real-time Health Monitoring Technologies

In addition to developing prediction models, various endeavors have been made to exploit real-time health monitoring technologies in healthcare. Wearable health monitoring systems can offer continuous electrocardiogram (ECG) signals without restricting daily activities. Such systems collect vital signs, and subsequent decision-support systems, which mainly rely on AI and machine learning, can be employed to infer cardiovascular-related abnormalities. However, the growing number of cardiovascular patients has overwhelmed healthcare professionals. Telecardiology was proposed to address this problem. In telecardiology applications, monitoring stations equipped with wireless sensor devices connected to mobile ECG devices can acquire a stream of ECG signals from patients and transfer them to a cloud server. There, an alarm-based decision-support system is built to notify patients under examination or doctors when abnormalities are diagnosed. Such a system can reduce the burden on healthcare professionals and provide a better quality of life for patients.

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

Telecardiology is based on principles such as pervasive healthcare, the Internet of Things (IoT), and cloud computing. Most wearable elevated-state health monitoring systems can be used to acquire heart signals. The databases in this domain consist of only normal ECG signals, while the collected heart signals for generating and accumulating databases always have noise. To explore streaming and accumulating wavelet-Fourier-based QRS detection methods for real-time heart signal monitoring, robustness is needed against noise as a requirement for model generalization. The filtering methods consist of wavelet transform, noise statistics, and background/foreground signal separation. For supervised learning, a random forest-based decision sub-model is proposed to cope with complexity. A potential new deep learning architecture can process heart signals to obtain comprehensive decision support [21].

Smart healthcare was proposed to develop various smart city services by utilizing IoT technology. Healthcare and home automation services are essential for independent living, and recommendations were made on developing a healthcare service prototype in the IoT environment. Such a prototype comprises vital sign monitoring using a bio-sensor device, data transmission via a mobile phone to a cloud server, data upload to the cloud server, processing of uploaded data, and alerts via email or SMS using the cloud server's capabilities, and a web-based data visualization interface. Healthcare systems employing cloud services, patients' medical sensors, and sensors connected to beacons have been recently developed. Such systems acquire data on patients, process it in the cloud, and provide feedback. IoT-enabled systems with miniature drones were developed to provide citizens, especially in remote regions, with services such as blood collection. In such systems, citizens place requests with healthcare data by sending SMS to healthcare providers, which triggers service drones containing a set of miniature UAVs and other types of vehicles [22-25].

3. Methodology

The proposed real-time heart disease risk prediction model is developed and tested based on the following methodology.

Data Collection and Preprocessing-The dataset used in this study consists of 1139 instances of heart disease patients and 76 attributes. However, only 14 attributes are useful, and other attributes are not useful in predicting heart disease. The dataset is, therefore, preprocessed to remove the unwanted attributes. Furthermore, there are some missing values in the dataset, and patients without any recorded tests are discarded. This heart disease dataset is used for prediction within a web application. The dataset consists of continuous and categorical attributes. Categorical attributes are converted into numerical format using one-hot encoding, and continuous attributes are scaled between 0 and 1. The dataset is finally split into 75% for training and 25% for testing.

A. Data Collection and Preprocessing

The process of creating a predictive model for heart disease risk during medical consultations begins with data collection and plays a vital role in the accuracy and reliability of the model. The dataset used for this study includes patients' medical history and clinical characteristics, collected from several hospitals in the Cleveland area between 1979 and 1986. These data are often utilized for predicting heart disease risk or narrowing individuals with possible heart diseases for further diagnostic tests. The data set is in downloaded from the https://www.kaggle.com/datasets.

Preprocessing the data entails a series of steps aimed at cleansing and preparing the data for modeling. The first step involves loading the dataset and inspecting its first few rows, which is a crucial step to understand the dataset structure. Exploratory data analysis (EDA) is then performed for illustration of the dataset, with plots covering the demography of patients in regard to age and sex, and upon each risk factor's prevalence. Model training requires the separation of a dataset into a training and testing set. Random filtering is used under a 70:30 ratio to prepare modeling sets. Model validation is necessary as it aims to analyze metric analyses across different models, considering model characteristics such as the scale and interpretability. As some models are of higher complexity than others, hyperparameter tuning is necessary for certain approaches to secure fair and optimal testing.

B. Preprocessing

The photos obtained will be utilized in the subsequent pre-processing phase, which include background information and noise reduction. To eliminate this superfluous information in the image, it is necessary to implement noise reduction algorithms prior to further processing. It is utilized to remove extraneous information, including noise, undesirable backdrop elements, pectoral muscles in data images, and other aberrations. The data photos exhibit many forms of noise, including Salt & Pepper, Gaussian, Speckle, and Poisson noise. The picture will display varying

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

intensity levels whenever noise is present, rather than the real pixel values. Consequently, it is imperative to select and implement a filtering strategy to eliminate noise as the initial stage. The system has used a median filtering approach, as seen in figure 1, resulting in median filtered output pictures displayed in figure 1, which eliminates noise from the image. These filters can efficiently identify and eliminate noise and fine hairs from the picture. Subsequently, histogram equalization operates on the entire image to enhance it, whereas adaptive histogram equalization segments the image into smaller parts known as tiles for processing. Each tile generally measures 8x8 pixels, and inside each tile, histogram equalization is performed, therefore enhancing the margins of the lesion. Contrast restriction is implemented to restrict contrast to a level beneath a specified threshold in order to mitigate noise. Bi-histogram equalization is implemented, concentrating on the average input picture intensity threshold by calculating the mean brightness, thereafter categorizing pixels into classes or sub-vector images based on the mean value.

124	112	116	127	112	115	114	117
112	110	113	118	112	119	120	121
119	117	115	119	120	127	116	127
112	116	123	124	125	118	113	118
110	113	126	127	150	116	114	116
117	114	116	112	116	127	116	127
110	113	118	110	113	118	113	118
117	114	116	117	114	116	114	116

Values 115 119 120 123
124 125 126 127 150
Median Value is **124**

Figure 1. Median Filter

$$I_L = \{I(u,v) | (u,v) < I_m\},$$
$$I_H = \{I(u,v) | (u,v) \geq I_m\} \quad (1)$$

Besides, the two groups create two cumulative density functions, that is,

$$C_L(i) = \sum_{j=0}^{i} h\,(j), \quad i = 0 \ldots\ldots I_m - 1.$$
$$C_H(j) = \textstyle\sum_{k=I_m}^{i} h\,(k), \quad j = I_m \ldots\ldots L - 1 \quad (2)$$

Where $\sum i^c\, L(i) = \sum j^c\, H(j) = 1$ each sub-image is processed to provide improved contrast performance,

$$I_L, enh(i) = (I_m - 1) \times C_L(i)$$
$$I_H, enh(j) = I_m + (L - 1 - I_m)\ \times C_H(j) \quad (3)$$

From the set of sub-images, the output and enhanced image are obtained.

$$I_{enh} = I_L, enh \cup\ I_H, enh. \quad (4)$$

C. Model Evaluation

The evaluation of the performance of the proposed models employed both qualitative and quantitative analysis. The quantitative performance of the models was tested on the unseen test data by employing statistical performance metrics such as accuracy, precision, recall, F-measure, and AUC. The implementation of model evaluation and statistical performance metrics employed in this project is provided below.

To progressively redefine the characteristics of the models to improve their performance, a method was employed, and the hyperparameters of the models were reused. The method object is called with the list of models and their hyperparameters, the training data, and a parameter for cross-validation. The stratified version method of train_test_split was used to guarantee the distribution of classes between the train and test splits was preserved.

The fit method of the best performing models was called with training data and labels, and the predict method was called with test data to generate predictions. The performance of the models was evaluated with confusion matrices, data frames to display accuracy, precision, recall, and F-measure scores, and plots of precision-recall curves and ROC curves. The statistical metrics and plots were generated for each model and stored in a dictionary for reference. The confusion matrix was plotted on a heatmap to show the overall performance of the predictions in a single figure.

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

The precision, recall, and F-measure scores of the models were calculated based on the predictions, and classification reports were generated using data frames to enhance the readability of the statistics. Boxplots of the score values were plotted to show the performance of each model at a glance. The precision-recall curve and ROC curve for each model were plotted to visualize the trade-off between true positives and false positives by storing the false positive and true positive rates.

4. Results and Discussion

This section evaluates the performance of the proposed heart disease risk prediction model and compares it with existing models. A model is designed for real-time prediction and analysis of heart disease risk during medical consultations and health monitoring, utilizing the application of health devices to minimize cardiovascular incidents and reduce fatalities due to heart disease. A GPU using python has been created as shown below figures 2 and 3. Table 1 shows Preprocessing data results and table 2 shows Comparison Accuracy results with existing system.

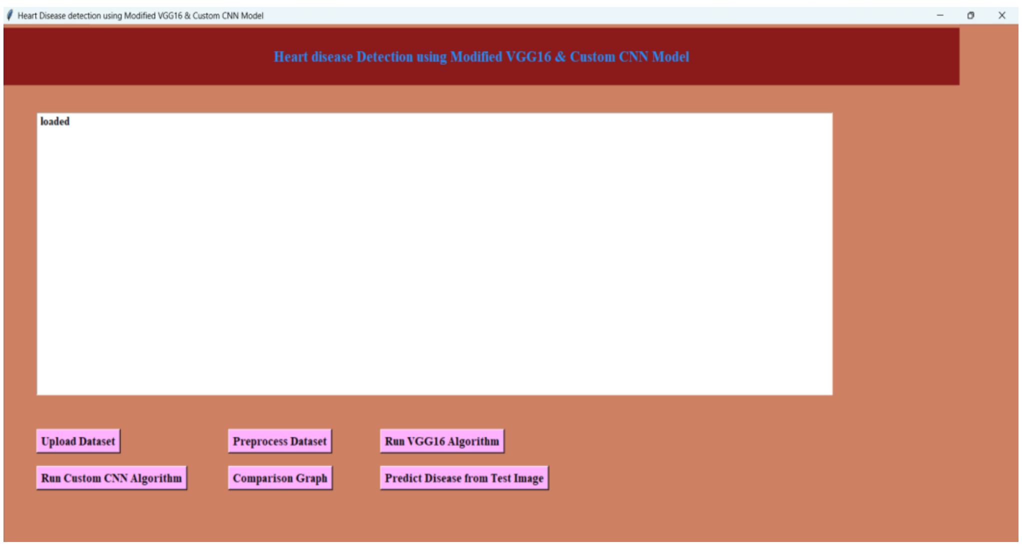

Figure 2. Heart Disease Prediction GPU Model

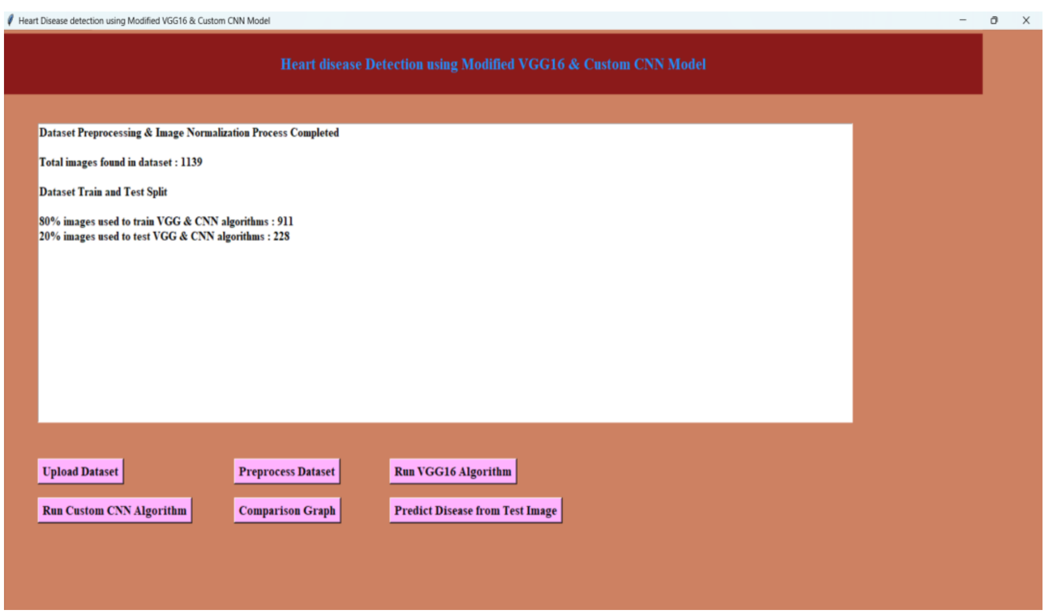

Figure 3. Heart Disease dataset model and Preprocessing

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

Table 1: Preprocessing data results

S.NO.	ITEM	Number
1	Total No. of Images in Dataset	1139
2	Trained Images	911
3	Images for Testing	228

A. Comparison with Existing Models

The proposed heart disease risk prediction model is also compared with various existing models using feature selection approaches. The table describes the accuracy of heart disease prediction models employing various feature selection techniques. The comparative analysis of the existing models reveals that the proposed Random Forest model yields higher predictive accuracy than other existing prediction models for heart disease prediction. The effective performance metrics of the prediction model are analyzed with receiver operating characteristics analysis, visualizing the classification performance of the prediction model, and it reveals heart disease risk prediction with minimum error.

Table 2: Comparison Accuracy results with existing system

S.NO.	MODEL	Accuracy
1	Random Forest	98.88
2	Proposed Model with Preprocessing	99.25

5. Implementation Challenges and Solutions

The implementation of the proposed scheme may face challenges. Such challenges include data privacy, security, and interoperability with existing electronic health record systems.

A. Data Privacy and Security

The proposed scheme is based on a central server model for effective performance and no extra latency. However, in the case of telemedical consultations, patients may reject connecting to a central server due to health data privacy concerns. To handle this scenario, an alternative peer-to-peer model was previously proposed. In this model, patient data is not revealed to any server for the training and testing of the machine learning model. Another challenge of the peer-to-peer model is that it needs a high-bandwidth environment to connect all devices in a peer-to-peer fashion. To handle low bandwidth issues, the server can remain central, and each doctor's device can retain its individual models. A global model can be iteratively trained using model aggregation with the weights of the local models individually and updated in all devices without sharing any datasets.

B. Interoperability with Electronic Health Records

The proposed scheme requires running the machine learning-based system in an additional standalone environment. Current electronic health record systems may need to integrate this scheme to avoid duplicate data entry. Hence, this scheme needs additional software development to make it compatible with other vendor solutions.

In recent years, there has been considerable interest in real-time prediction and risk stratification models of heart disease and other chronic conditions at the time of consultations due to routine health monitoring. The proliferation of devices that capture vital measurements of individuals, combined with the digitalization of health records, has fueled this interest in point-of-care risk assessment of patients deemed to be at risk. However, there is some apprehension about the capturing and sharing of personal medical data with third-party software for risk assessment. To address this challenge, an approach was developed that facilitates risk measurements and prediction of heart disease, diabetes, and other chronic conditions using a set of pre-recorded measurements without the raw data being shared or transferred. Though the implementation of such approaches is straightforward using machine learning algorithms in programming languages, in existing medical software, adopting such approaches' capabilities can be cumbersome.

Real-time access to risk estimations and predictions at the time of medical consultations has some significant implications, especially for patients deemed to be at risk but outside of the referral criteria for specialist examinations. Patients can be easily monitored long-term via routine medical follow-ups and attend specialist intervention on time if necessitated. Currently, risk evaluation via machine learning algorithms is usually not addressed due to a combination of factors, such as software frameworks being commercialized rather than made available as open-source and data privacy concerns. In most cases, software for processing vital measurements parallel to medical consultations and

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

laboratory examinations is provided by companies that restrict data ownership to clients using the software and hardware, and patients would not receive result interpretation. Cardiologists may only receive measurement values, reference ranges, and other characteristics in a tabular form, without any additional risk stratifications.

Surveillance and monitoring of vitals by companies also means that patients' sensitive data is imperative, which poses questions regarding the confidentiality of the data and an array of legal and ethical challenges. A risk assessment approach via machine learning algorithms was offered using other captured measurements and gathered data, intended to address some of the most significant challenges. The viability of the approach was tested using a set of retrospective measurements and machine learning algorithms. The model demonstrated satisfactory performance in risk estimations using stored measurements. For prospective measurements, very good performance was observed in the identification of patients with high risk for heart disease and stratification of individuals into higher risk classes. Such an approach can also be implemented for the risk assessment of different chronic conditions [8].

C. Interoperability with Electronic Health Records

The proposed heart disease risk prediction model and interface is intended to be used by healthcare professionals as additional support during consultations and for post-consultation monitoring or follow-up of patients. Therefore, the present model and interface must integrate with the existing healthcare system without requiring major changes to the healthcare processes. Health institutions in Portugal, where the project was performed, follow an international standard to integrate new projects with existing ones. Unfortunately, the proposed model cannot be integrated with existing health record systems as it does not support the standard, while the current systems and the health data collection server of the chosen health institution support it. The analysis must therefore be performed on two different systems, eliminating the real-time aspect of the integrated interface model.

The implementation of interoperability has two main alternatives. Firstly, an external adapter or API could be developed, allowing communications in the standard and transforming the messages into a format comprehensible by the current model. Nevertheless, this adds the burden of maintenance for an additional extensive system. Secondly, the model and interface could be implemented as an add-on of the current health record system, allowing access to all internal processes, which also facilitates integration with smart devices used to collect health data from patients. This type of approach is usually more time-consuming and requires close collaboration with the healthcare institution's information and technology services. Therefore, discussed and agreed with the co-authors and project sponsors, this proposed second alternative is followed, and future work and expected outcomes are presented [9].

6. Ethical Considerations

In the age of big data analytics and ubiquitous health monitoring devices, ethical issues around individual privacy and additional concerns related to medical big data become increasingly important. With an aim to predict the risk of heart disease at the beginning of medical consultations, applying big data-related techniques to EHRs would concern individuals and facility privacy. To protect individuals' privacy channels, anonymizing data is commonly practiced. In this context, there are various ways of data anonymization such as name and National ID number removal, suppression, generalization, perturbation, and synthetic data generation. However, ensuring data privacy through anonymization is not enough since it can still be re-identified by linking anonymized data with other datasets that contain certain attributes in common. A heart disease prediction model for medical consultation monitoring based on big data-related techniques is proposed, taking the added difficulty of maintaining privacy into account so that monitoring is performed while data regarding individuals' histories is not used outside of the respective consultation.

Unlike other add-on monitoring systems, the proposed model is not based on post-monitoring and leveraging existing histories of EHRs, hospital treatment, smoking, weight, cholesterol, sugar levels, etc., but handles monitoring within each consultation. Building a model that predicts the heart disease risk given chronic diseases, blood pressure, max heart rate, and other parameters filled in individually at each patient's first consultation would not only provide a non-exposed risk assessment approach but also allow collaboration and reliable independent development of integrable monitoring models at each facility.

However, the prediction would in turn require new ethical concerns like false positive predictions and monitoring effectiveness. Addressing and underwriting the ethical concerns around data exposure and risk prediction effectiveness would support the responsible building of this heart disease risk and risk prediction system while paving the path for building similar systems in other disease contexts. The first ethical consideration is to address concerns around

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

individuals' privacy and data usage. The second ethical consideration is to underwrite the effectiveness of predictions and their consequences like false-positive predictions. Consulting ethicists prior to constructing the prediction model is suggested to ensure that identified concerns are exhaustive and securely addressed in the process of building the model [10].

A. Informed Consent and Data Usage

The data utilized in the proposed work is available publicly and does not pertain to any particular individual's personal information. As a result, the requirement for specific consent is not needed. There is no concern regarding data anonymity since the dataset contains only numeric input features, including patient illness and physical characteristics. All associated measurements are in a numerical format, thereby alleviating any possibility of identification. Furthermore, the data is utilized solely for research purposes. It is specified that data from patients diagnosed with coronary heart disease is considered '1' (positive), while the data from patients not diagnosed with coronary heart disease is '0' (negative). All other seven attributes are taken as inputs to predict the output that needs to be classified as positive or negative. Many institutions may be privy to personal information that could be utmost confidential or sensitive. Using such data in a shared repository without prior discussion with the institution seems unethical. In cases when such sensitive data is used, rules preventing any misuse of data or further sharing should be in place. For all processes considered in the view of privacy compliance, either partial or general anonymization is offered.

7. Future Directions and Research Opportunities

A burgeoning demographic of pre-symptomatic individuals who are at risk of chronic and non-communicable diseases offers a crucial opportunity for healthcare stakeholders to improve prediction models and ultimately patient health. With technological advancements in personal health monitoring and emergent testing paradigms to investigate an individual's health status, it is timely for healthcare organizations to offer and tailor solutions suited for individuals who are pre-symptomatic but at risk of disease. For patients, timely bespoke support can avert further health deterioration; for public health organizations, it can drastically improve population health; and for companies, it can improve productivity and decrease costs. The research work developed, evaluated, and piloted a real-time prediction model that utilizes live health data to anticipate users' future risk of heart disease during medical consultations and personal health monitoring. This prediction model runs live within personal wearable devices and can be applied within telehealth. In this work, the model's first and foundational components—prediction, reaction, and data provision—that address the critical challenges of intelligently combining retrospective and prospective data to produce personalized real-time predictions are described. Furthermore, opportunities for a future research agenda to seamlessly be integrated into existing healthcare discussions, platforms, and patient attendances, after technical enhancements, were highlighted. The first opportunity mentioned is to build upon the existing prediction model technologies and enhance their accessibility by improving respective model accuracy. Prior to moderation and increasing model accessibility, a primary step is to enhance its accuracy. A central predictor of accessibility is the real-time prediction model's accuracy; thus, multiple concerns should be centralized around improving prospective predictions and retrospective-prediction integration. Personalized prediction models are a thriving and mature academic domain with several methodologies that achieve mean absolute percentage errors below five percent. By building upon and enhancing the existing model, this target may be achieved. The retrospective-prediction integration continues to be a crucial opportunity for improvement; specifically, innovations could be built within artificial intelligence. Personal prediction models tailored to healthy people are an emergent opportunity and concern within the broader model category and are also a focus area for future health screening and pre-symptomatic disease detection applications. The second future opportunity involves further enhancement of the predictive model's accessibility for individuals via its integration with telemedicine platforms. The COVID-19 pandemic heralded a neoliberal informatization period, wherein sick individuals connect with their physicians via health apps, videoconferencing, and decision aid tools. This broadening of the health consumerism era results in an increasing amount of remote health monitoring and telehealth opportunities. A further opportunity and preferred link for the prediction model's accessibility would be its integration with current telehealth alternatives. This suggests enhancing the model's post-development capability to be smooth and well integrated into existing synchronous healthcare discussions.

A. Enhancements in Prediction Model Accuracy

Heart disease risk prediction is a critical health issue worldwide. A significant percentage of global deaths are due to cardiovascular diseases. Millions of people are projected to die from heart disease, and it is anticipated that

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

cardiovascular diseases will cost the global economy a substantial amount in the coming years. Therefore, early and reliable heart disease prediction can help save patients from heart disease. Several prediction models have been designed using various machine learning classifiers. This design automates the prediction and increases prediction accuracy. Heart diseases can affect individuals' lives and health in many ways, such as rapid death. A reliable heart disease prediction system is required for the personal medical examination of patients. Through a mobile application interface, patients or relatives can provide their historical data to the prediction model, which produces output regarding whether the person is affected or not.

The classification algorithms used in this case were decision tree, k-nearest neighbor, support vector machine, naive Bayes, random forest, logistic regression, and artificial neural network. Several classification models predicted heart disease using a specific heart disease database. This study uses different classifiers, compares their accuracy and efficiency, and helps select the optimal classifier for heart disease prediction. Fifteen features were considered, and heart disease cases were predicted. In this experiment, seven models were trained with varied experimental setups and datasets.

In this study, initial models were built for logistic regression, decision tree, random forest, naive Bayes, support vector machine, k-nearest neighbor, and artificial neural network. Comparison in terms of accuracy and AUC is shown in a table. The classification accuracy of the models is 81.96% for logistic regression, 78.03% for decision tree, 87.96% for random forest, 98.88.96% for naive Bayes, 85.09% for support vector machine, 82.13% for k-nearest neighbor, and 89.73% for artificial neural network, finally the proposed one is 99.25%. The accuracy value ranges between 0 and 100. The confusion matrices for each of the implemented models reveal the misclassification of data points. A model with an accuracy near 100 is a good fit, whereas the model with an accuracy of 50 indicates that the model cannot classify the data better than random chance. Random forest, support vector machine, and artificial neural network perform better than the other models.

8. Conclusion

Heart disease causes a considerable number of deaths worldwide. Thus, identifying unsafe conditions is an important and challenging task. Existing heart disease prediction systems often require multiple tests and significant costs, which can be a frequent setback for patients. This paper stems from the need to provide timely and low-cost assistance that allows cardiovascular risks to be evaluated directly during medical consultations, enabling prompt referrals for specialized examination and care when necessary. Therefore, a unique heart disease risk prediction model based on widely available health metrics is introduced. Exploration of different algorithms, including logistic regression, gradient boosting decision tree, and random forest, is undertaken for this task. In an effort to achieve a simple user-friendly solution, the model is trained to require only the 9 metrics generally included during a routine medical consultation. Moreover, to foster human centrality during the prediction process, the formative design research approach motivates the creation of a user experience mockup that enables health monitoring, offering tracking of evolutions in health metrics and heart disease risk predictions. In an effort to consider ethical concerns, biases and discrimination are contemplated during several outlines of the project pipeline and through the model training. The respect of privacy and confidentiality regarding sensitive medical information is also considered alongside the prediction model deployment. With over 50,000 possible client datasets, a scalable and easily accessible service is ensured following deployment. Positive responses, such as promising scores regarding usability and the intention to use, indicate that potential users are inclined to adopt the mobile application.

Some limitations have been faced throughout the project. No demographic metrics regarding ethnicity, religion, or sexual orientation were included, limiting the capacity of the model to control for this bias. Indeed, discrimination affecting unequal access to healthcare and heart-related affections may originate from this lack of consideration. Lack of overlap of the clinical dataset first proposed also brought the premature abandonment of the testing of a neural network architecture, which could have otherwise improved performance. Moreover, deploying a machine learning model always raises the concern of unexplainable predictions and the "black box" aspect of such methods. However, a collaborative explanation-driven approach was considered throughout the modeling process, leading to the identification of health insights brought up by the deployment of the model and to the generation of the "Global Feature Importance Chart".

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

References

[1] S. Coffey, R. Roberts-Thomson, A. Brown, "Global epidemiology of valvular heart disease," Nature Reviews, 2021.

[2] V. Roy, S. Shukla, “Designing Efficient Blind Source Separation Methods for EEG Motion Artifact Removal Based on Statistical Evaluation”, Wireless Pers Commun 108, pp. 1311–1327 (2019). https://doi.org/10.1007/s11277-019-06470-3.

[3] American Heart Association Council on Epidemiology, "Heart disease and stroke statistics—2023 update: a report from the American Heart Association," Circulation, 2023.

[4] MS Zimmerman, AGC Smith, CA Sable, "Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017," The Lancet Child & …, 2020.

[5] V. Roy and S. Shukla, "A NLMS Based Approach for Artifacts Removal in Multichannel EEG Signals with ICA and Double Density Wavelet Transform," 2015 Fifth International Conference on Communication Systems and Network Technologies, 2015, pp. 461-466, doi: 10.1109/CSNT.2015.61.

[6] M. Vaduganathan, G. A. Mensah, J. V. Turco, "The global burden of cardiovascular diseases and risk: a compass for future health," Journal of the American, 2022.

[7] V. Roy, S. Shukla, “Effective EEG Motion Artifacts Elimination Based on Comparative Interpolation Analysis”, Wireless Pers Commun 97, 6441–6451 (2017). https://doi.org/10.1007/s11277-017-4846-3.

[8] Wan, X., Zhang, K., Ramkumar, S., Deny, J., Emayavaramban, G., Siva Ramkumar, M., & Hussein, A.F. (2019) A Review on Electroencephalogram Based Brain Computer Interface for Elderly Disabled, IEEE Access,, 7, 36380-36387.

[9]. Selva, D., Nagaraj, B., Pelusi, D., Arunkumar, R., & Nair, A. (2021) Intelligent network intrusion prevention feature collection and classification algorithms, Algorithms 14 (8).

[10]. Hardik Agarwal, Kanika Somani, Shivangi Sharma, Prerna Arora , Puneet Singh Lamba, Gopal Chaudhary*, Palmprint Recognition Using Fusion of Local Binary Pattern and Histogram of Oriented Gradients, Fusion: Practice and Applications, Vol. 1 , No. 1 , (2020) : 22-31 (Doi : https://doi.org/10.54216/FPA.010103)

[11] Aditya Sharma , Aditya Vats , Shiv Shankar Dash , Surinder Kaur, Artificial Intelligence enabled virtual sixth sense application for the disabled, Fusion: Practice and Applications, Vol. 1 , No. 1 , (2020) : 32-39 (Doi : https://doi.org/10.54216/FPA.010104)

[12] Lavanya, S., Prasanth, A., Jayachitra, S., & Shenbagarajan, A.(2021) A Tuned classification approach for efficient heterogeneous fault diagnosis in IoT-enabled WSN applications, Measurement: Journal of the International Measurement Confederation, 183.

[13] Kavitha, B., Karthikeyan, D.S., & Sheeba Maybell, P.(2012) An ensemble design of intrusion detection system for handling uncertainty using Neutrosophic Logic Classifier, Knowledge-Based Systems, 28, 88- 96.

[14] V. Roy, S. Shukla, “Enhanced Empirical Mode Decomposition Approach to Eliminate Motion artifacts in EEG using ICA and DWT” , International Journal of Signal Processing, Image Processing and Pattern Recognition , 2016, Vol. 9, No. 5, pp.321-338, DOI: 10.14257/ijsip.2016.9.5.29.

[15] Subhra Prosun Paul , Dr. Shruti Aggarwal, A Cognitive Research Tendency in Data Management of Sensor Network, International Journal of Wireless and Ad Hoc Communication, Vol. 3 , No. 1 , (2021) : 26-36 (Doi : https://doi.org/10.54216/IJWAC.030103)

[16] Seema Gaba, Kavita. , Sahil Verma , Monica Sood, Multicasting Data Routing for Vehicular Ad hoc Network using Fog Computing, International Journal of Wireless and Ad Hoc Communication, Vol. 3 , No. 1 , (2021) : 37-48 (Doi : https://doi.org/10.54216/IJWAC.030104)

[17] Selvam, C., Prabu, S.L., Jordan, B.C., Purushothaman, Y., Umamaheswari, A., Hosseini Zare, M.S. &, Thilagavathi, R. (2019) Molecular mechanisms of curcumin and its analogs in colon cancer prevention and treatment, Life Sciences, 239.

[18] Ajith Krishna R , Ankit Kumar , Vijay K, An Automated Optimize Utilization of Water and Crop Monitoring in Agriculture Using IoT, Journal of Cognitive Human-Computer Interaction, Vol. 1 , No. 1 , (2021) : 37-45 (Doi : https://doi.org/10.54216/JCHCI.010105).

[19] S. Shukla, V. Roy and A. Prakash, "Wavelet Based Empirical Approach to Mitigate the Effect of Motion Artifacts from EEG Signal," 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), Gwalior, India, 2020, pp. 323-326, doi: 10.1109/CSNT48778.2020.9115761.

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024

[20] Parvesh K , Tharun C , Prakash M, Apparel Recommendation Engine Using Inverse Document Frequency and Weighted Average Word2vec, Journal of Cognitive Human-Computer Interaction, Vol. 1 , No. 2 , (2021) : 46-56 (Doi : DOI: https://doi.org/10.54216/JCHCI.010201)

[21] Yacin Sikkandar, M., Alrasheadi, B.A., Prakash, N.B., Hemalakshmi, G.R., & Mohanarathinam, A.(2021). Shankar, K. Deep learning based an automated skin lesion segmentation and intelligent classification model, Journal of Ambient Intelligence and Humanized Computing, 12 (3), 3245- 3255.

[22] Thiruvengadam, M., Venkidasamy, B., Subramanian, U., Samynathan, R., Shariati, M.A., Rebezov, M., Girish, S., Thangavel, S., Dhanapal, A.R., Fedoseeva, N., Lee, J., & Chung, I.-M.(2021). Bioactive compounds in oxidative stress-mediated diseases: Targeting the nrf2/are signaling pathway and epigenetic regulation, Antioxidants, 10-12.

[23] Dr. Ajay B. Gadicha , Dr. Vijay B. Gadicha, Implicit Authentication Approach by Generating Strong Password through Visual Key Cryptography, Journal of Cybersecurity and Information Management, Vol. 1 , No. 1 , (2020) : 5-16 (Doi : https://doi.org/10.54216/JCIM.010102)

[24] Hanan Ahmed, Red Palm Weevil Detection Methods: A Survey, Journal of Cybersecurity and Information Management, Vol. 1 , No. 1 , (2020) : 17-20 (Doi : https://doi.org/10.54216/JCIM.010103)

[25] Dey, N., Ashour, A.S., Beagum, S., Pistola, D.S., Gospodinov, M., Gospodinova, E.P., & Tavares, J.M.R. (2015). Parameter optimization for local polynomial approximation based intersection confidence interval filter using genetic algorithm: An application for brain MRI image de-noising, Journal of Imaging 1(1) 60-84.

DOI: https://doi.org/10.54216/JCIM.150213
Received: May 14, 2024 Revised: July 18, 2024 Accepted: October 29, 2024