AI-Based Platform For Predicting The Risk of Having NCDs

Ariane Shimirwa, Claudine Mahoro, Tyson Muvunyi
Information & Communication Technology
Carnegie Mellon University Africa
Kigali, Rwanda
ashimirw@andrew.cmu.edu claudinm@andrew.cmu.edu
tmuvunyi@andrew.cmu.edu

Abstract

Non-communicable diseases (NCDs) are chronic medical conditions that are not caused by infectious agents and are primarily associated with lifestyle factors. This paper highlights the significant impact of NCDs on global public health, particularly in low- and middle-income countries where they account for most premature deaths. A privacy-preserving AI platform is proposed as a solution to predict the risks of having NCDs, such as cardiovascular diseases and type 2 diabetes, using clinical data and advanced machine learning algorithms such as ANN, SVM, Naive Bayes, and Random Forest. ANN proved to be the best performing for Diabetes and SVM for Heart disease with an accuracy of 92.15% and 92.59 respectively. Utilizing the top-performing models for predictions, the mobile application alerts users who are at high risk and also offers them the opportunity for further medical consultation.

1 Introduction

Non-communicable diseases (NCDs), also known as chronic diseases, are long-lasting conditions that develop slowly and are not caused by infectious agents. They include cardiovascular diseases, cancer, chronic respiratory diseases, diabetes, and mental health disorders, among others [1]. NCDs require ongoing treatment and care, affecting long-term health [1]. Lifestyle factors such as unhealthy diets, physical inactivity, tobacco use, excessive alcohol consumption, and environmental factors contribute to the development of these diseases [2]. NCDs have become a major public health issue, particularly in low and middle-income countries. They are responsible for a significant number of deaths worldwide, with 17 million individuals dying prematurely each year before the age of 70. Around 86% of these deaths occur in low and middle-income countries because universal health coverage or access to healthcare services is often limited [3]. To address this growing health concern, it is crucial to prioritize early detection, prevention, and management strategies for NCDs.

This study aims at developing an AI-powered platform, considering the fact that mobile penetration in LMIC is increasing [4], we decided to use a mobile application to help people in the early detection of such diseases. That will be an act of raising awareness for people by being able to predict the risks of having heart and diabetes diseases, based on their health and lifestyle information. The platform utilizes supervised learning algorithms and a Machine Learning framework to uncover hidden patterns and relationships in the data. By employing techniques such as Random Forest (RF),

Support Vector Machine (SVM), Naïve Bayes (NB), and Artificial Neural Networks (ANN), it analyzes clinical data to predict the risks of diabetes or heart disease. The proposed platform not only enhances NCD risk prediction but also ensures privacy and security protection. It notifies individuals at high risk and offers the option to schedule appointments with medical professionals for further consultation.

1.1 Assumptions

This study presumes that the platform will be utilized by individuals previously undiagnosed with heart disease or type 2 diabetes, and who are interested in evaluating their disease risk. We also posit that users are informed about their health metrics and familiar with certain medical terms, allowing effective use of our platform. Moreover, we assume that users are knowledgeable about and provide accurate health information. We further assume that the developed platform can supplement clinical screening and diagnosis, with high-risk individuals identified by the platform seeking subsequent medical consultations.

2 Methodology

The methodology of this study includes several steps: data collection, pre-processing, exploratory analysis, model building, mobile application development, and security implementation. The data used were sourced from publicly available online datasets. The "Heart_Disease_Prediction" dataset from Data World was utilized for heart disease, containing 270 entries with 150 indicating the presence and 120 indicating the absence of the disease, across 14 features [5]. For diabetes, the "Early-Stage Diabetes Risk Prediction Dataset" from Kaggle was used, originally consisting of 510 entries [6]. After removing duplicates, this was reduced to 269 entries, with 173 records indicating the presence and 78 indicating the absence of the disease, across 17 features. Both datasets were independently analyzed and utilized for training and modeling.

The two datasets were pre-processed, and exploratory analysis was conducted to gain insights from the data. The visualizations from the data demonstrated how people around the age of 60, especially males, were more prone to heart disease risks, and more females had diabetes disease between the age of 40 and 50. This observation aligns with the research from Medicine in Novel Technology and Devices, which indicates that men have a higher risk of heart disease at an early age compared to women, whose risks tend to increase after menopause [7]. To build predictive models, forward stepwise regression was employed to select relevant features. Four supervised models, namely Random Forest, Support Vector Machine, Naïve Bayes, and Artificial Neural Network were used for each dataset. These models were evaluated using classification metrics such as accuracy, precision, recall, and F1 score to assess their performance.

Furthermore, a mobile application was created using Flutter for the app development, a Postgres database for storing patient records, and Fast API for backend services. To ensure security various measures were implemented. These include encryption, hashing of user credentials, authentication verification, and password protection. To ensure privacy, data anonymization, consent management, and audit trail measures were put into consideration. These measures were put in place to safeguard user's personal information and maintain data privacy. The mobile application enables users to input their health metrics, which are then passed into a machine-learning algorithm for predictions, and according to the prediction scores, the system outputs a result. The whole process can be seen in Fig 1 below.

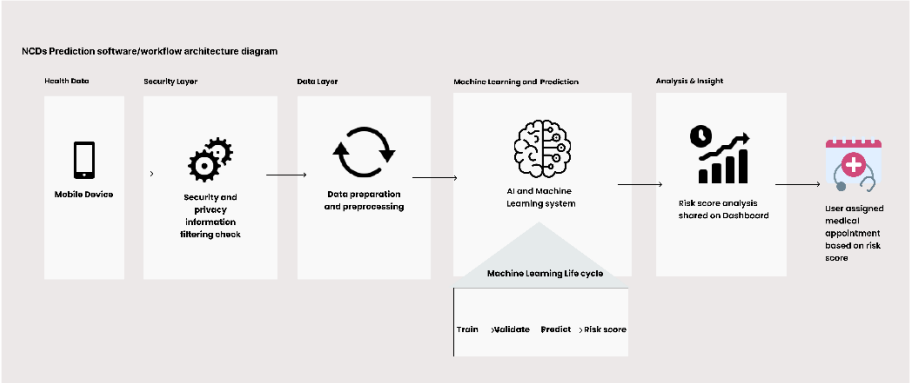

Figure 1: NCDs prediction architecture diagram

3 Results

In the modelling phase, the support vector machine model demonstrated superior performance in predicting cardiovascular disease with an accuracy of 92.59% and the Artificial Neural Network model outperformed other models in predicting type 2 diabetes, achieving an accuracy of 92.15%. As a result, these two models, SVM and ANN, were saved and integrated into the mobile application. The mobile application was successfully developed, allowing users to input their diagnostic data through the user interface and utilize the models for risk prediction (see Fig 2 below). The application returns the results to the user, evaluating their risk for the corresponding non-communicable disease.

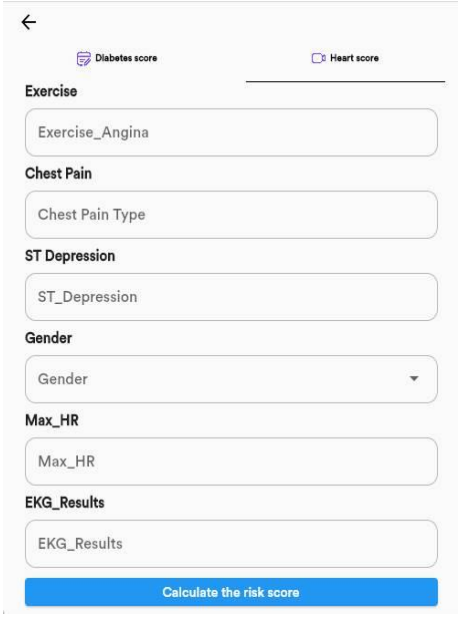

Figure 2: Interface for heart disease risks prediction

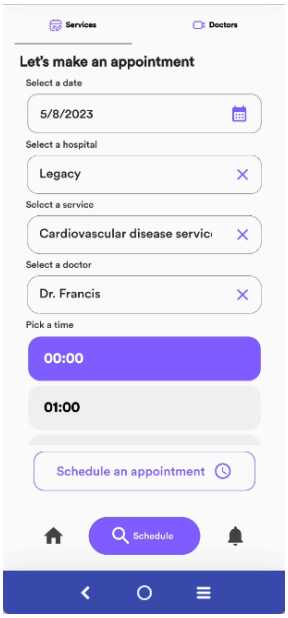

Figure 3: making an appointment

Additionally, the mobile application enables users to promptly schedule an appointment with a specialist for further diagnosis (see Fig 3 above). The booking system within the application facilitates the process of reserving a medical appointment at a hospital. If the NCD prediction indicates high risk, the user can choose to immediately make an appointment with their preferred hospital or reserve it for a later date. This functionality enhances the user experience by providing seamless access to specialized medical care based on their predicted NCD risk.

4 Conclusion

Detecting diabetes and cardiovascular diseases at an early stage poses a significant challenge in the healthcare industry [8]. In our research, we designed a smartphone application, which can predict cardiovascular disease and diabetes with high accuracy. We used clinical data from an online source, pre-processed it, explored it, used feature selection to select significant variables for prediction, and trained different machine learning algorithms, Random Forest, Support Vector Machine, Naive Bayes, and Artificial Neural Network with the selected features on each dataset.

The model's performance was evaluated using classification metrics; accuracy, precision, recall, and F1-score. All models show good results however Support Vector Machine and Artificial Neural Network outperformed other models in predicting heart disease and diabetes with an accuracy of 92.59% and 92.15% respectively. The best models were then deployed into the application built using Flutter. The user can use his mobile phone to log into the application and fill in the fields required to predict whether he has diabetes or heart disease. When the predicted results are positive, the system prompts the user to book an appointment at a desired hospital for further medical consultation and when it is negative, he gets a message that encourages him to keep exercising regularly, avoid smoking and eat healthy to prevent acquiring any of the NCDs.

5 Recommendation

Based on the findings presented in the paper, generally, it is recommended that AI-powered platforms that use clinical data and advanced machine learning algorithms should be considered by healthcare organizations for predicting the risk of non-communicable diseases such as cardiovascular diseases and diabetes, due to their potential for early disease detection. However, the study recognizes certain limitations like sampling bias, implying that the chosen dataset sample might not truly represent the actual population. Hence, future research is encouraged to use real clinical datasets collected from healthcare institutions to improve prediction accuracy and reliability, by providing a more representative sample of the target population.

Additionally, we suggest exploring the potential of wearable sensors for data collection, given the wide availability of specialized sensors in healthcare environments like hospitals and pharmacies. For example, Continuous Glucose Monitoring (CGM) devices are used to track interstitial glucose levels in individuals with type 2 diabetes [9]. For heart diseases, wearable sensors such as smartwatches can be employed to monitor heart activity, blood pressure, heart rate variability, and oxygen levels. Notable examples include the Apple Watch, Fitbit Sense, and Garmin Venu 2 [10]. These wearable devices offer an effective and convenient method for real-time data collection, potentially enhancing prediction accuracy and facilitating patient involvement in the screening process.

We also advocate for the exploration of additional machine learning algorithms and techniques to further enhance prediction accuracy. In light of future research, we emphasize the importance of using broader, more representative datasets. We also suggest expanding the application of AI-driven platforms to encompass other non-communicable diseases such as cancer and other respiratory disorders. By broadening the reach of these technologies to cover a diverse array of diseases, healthcare institutions can enhance disease detection and prevention, ultimately improving patient health outcomes.

References

[1] “Non-communicable diseases,” World Health Organization, September 2022.

[2] “Non-communicable diseases,” *UNICEF*. [Online]. Available: https://www.unicef.org/health/non-communicable-diseases.

[3] “Non-communicable diseases,” *World Health Organization*, September 2022.

[4] J. Sambira, “Africa’s Mobile Youth Drive Change | Africa Renewal,” United Nations, May 2013.

[5] “Heart disease prediction - dataset by Informatics-Edu,” data.world, https://data.world/informatics-edu/heart-disease-prediction

[6] I. Dutta, "Kaggle," [Online]. Available: https://www.kaggle.com/datasets/ishandutta/early-stage-diabetes-risk-prediction-dataset.

[7] Z. C. A. S. X. D Zujie Gao, "Gender differences in cardiovascular disease," Medicine in Novel Technology and Devices, vol. 4, no. ISSN 2590-0935, 2019.

[8] Health Care Quality Indicators - cardiovascular disease and diabetes - OECD, https://www.oecd.org/els/health-systems/hcqi-cardiovascular-disease-and-diabetes.html.

[9] A. J. Modaragamage, “Top wearable medical devices used in Healthcare | HealthNews,” *Health news*, Mar-2023.

[10] C. Vaughn, “10 different sensors you'll find in your smartwatch,” *MUO*, 08-Dec-2022.