A Smart Recommender System to Stratify Heart Attack Risk

A. Bolatbekov[1], D. Salman[1], S. Khan[1], E. Mamatjan[1], M. Orhun[2] and Y. Mamatjan[1]

[1] Faculty of Science, Thompson Rivers University, British Columbia, Canada
[2] Department of Computer Engineering, Istanbul Bilgi University, Istanbul, Turkiye

Abstract— **Heart diseases are a major cause of mortality and morbidity. Faster detection of life-threatening emergency events and an earlier start of the therapy would save many lives and reduce successive disabilities. We developed an automated smart recommender system (Explainable Artificial Intelligence) for heart attack prediction and risk stratification. This study is focused on an improved recommender approach to identify key risk factors impacting risk stratification for a particular patient. To help patients and clinicians better understand specific risk factors associated with heart attack and the degree of association, we used CatBoost classifier to build the model and SHAP to interpret its results. The risk factors that can be measured using smart monitoring wearable devices are called dynamic, while other risk factors obtained from users directly are static. The system collects both factor types to predict a heart attack risk. We created a Django-based online application that uses patient data to update medical information about an individual's heart attack risk. Moreover, the individual can quickly locate nearby doctors and hospitals from the application. The smart recommender system achieved high accuracy in predicting a patient risk level with an average AUC of 0.85. Utilizing the SHAP interpretation technique, we provided insights into the reasoning behind the predictions, including group-based and patient-specific explanations. Additionally, we employed a smart monitoring wearable device, such as a Smartwatch to automatically gather dynamic risk factors from the patient. The recommender system is cost-effective, easy to use, and portable since the main component of the system is commonly available on smartwatches, smartphones, and the Django framework. Early detection can improve patient management and lower heart attack risk while timely therapy aids in avoiding subsequent disabilities.**

Keywords— **Artificial Intelligence (AI), Recommender System, Heart Attack, Risk Prediction.**

I. Introduction

Heart diseases are a significant cause of death and illness with 53,704 cases of heart disease in Canada [1]. Early detection and treatment of potentially life-threatening conditions can help save lives and prevent long-term disabilities. Recent innovations in the field of machine learning (ML) have paved the way for the development of models for predicting heart conditions. A recommendation (Explainable) algorithm has been developed previously for short-term cardiovascular disease risk that is based on the minimum and maximum normal values for each measurement. If any of the measurements do not meet the minimum or maximum, then the recommendation is determined according to that factor [2]. Other researchers have employed ML techniques to predict heart attack risk and assist clinicians in diagnosis [3], [4]. However, the common problem of researchers who used concepts of ML is the lack of transparency in black-box models. The phenomenon makes it difficult to understand exactly how the model made that prediction fully. The SHapley Additive exPlanations was introduced by Shawi et al. [5] as a technique that explains the predictions by calculating the influence of each risk factor on the classification decision [6], [7]. Moreover, Dekhil et al. [8] used SHAP to assess stroke risk factors and prioritize risk factors. We aim to enhance understanding of how the model can identify the risk of heart attacks by explaining each risk factor's influence. Furthermore, we want to develop a framework that will gather some input data from the individuals and their smartwatches to predict the heart attack risk, resulting in a novel approach. The risk analysis results will be displayed online, or on computers and smartphones to monitor patient-specific risk factors associated with the heart attack. Moreover, the framework will send out a warning message to the user if any of the risk factors look abnormal, and provide all nearby hospitals on the map integrated into the website. We will make a website publicly available for clinicians and users to test and generate patient-specific reports.

II. Methodology

Overview: In this study, we present a smart health recommendation system that utilizes a technique to identify and prioritize factors that are most significant in predicting the risk of a heart attack. We created methods that involved collecting dynamic data using a wearable device, using this derived information to predict the risk of heart attack, and identifying specific risk factors for individual patients. We introduced a method for predicting the risk of heart attack by collecting dynamic and static risk factors from the patient using a website and wearable device. In the first step, a CatBoost classifier [9] was used to train the model and make the prediction. Then, we utilized the SHAP method [6] to identify the key factors that significantly impacted the risk assessment. Finally, we presented the results to the patients on the website. The system would send out a warning notification if any of the risk factors were out of the normal range of values. The recommender system also can generate a patient-specific PDF report along with the closest hospitals on the map to provide comprehensive guidance to users. Fig. 1 shows the schematic flowchart of the steps for the proposed recommender system:

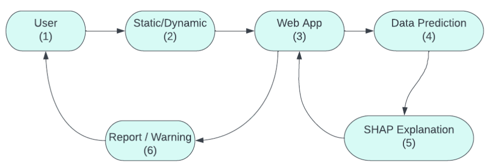

Fig. 1: A diagram showing the steps of the proposed method.

1. The user input from the web application and smartwatch;
2. Data processing step: static data is manually entered user data; dynamic data is retrieved from wearable devices;
3. The Django framework that saves patient data securely based on an individual user account;
4. Classification step creates a model from the training dataset and performs heart disease prediction using the test dataset;
5. The effect of each parameter on prediction is ranked and visualized using SHAP explanation.
6. The Django Web framework produces a patient-specific report and bar charts describing each parameter's effect on prediction and risk outcome. The map with the nearest hospitals is generated by taking the patient location.

Patient dataset: The heart attack dataset [10] used in this study contains 303 patient information and 13 risk factors with a heart attack risk outcome. The output response is the heart attack risk presented by 1 or 0 to show if there is a risk or no risk, respectively. The 13 risk factors used for predicting the risk of heart attack are as follows: age, sex, chest pain type (cp), resting blood pressure (trestbps), cholesterol (chol), fasting blood sugar (fbs), resting electrocardiographic result (restecg), maximum heart rate achieved (thalach), exercise-induced angina (exng), ST depression induced by exercise relative to rest (oldpeak), lope of peak exercise (slope), number of major vessels (caa) and thalassemia (thall).

Data preparation: Following data collection, the dataset was cleaned and divided into two independent subsets: the training set (80%), and the testing set (20%). Moreover, the Spearman correlation was applied to examine the relation between risk factors.

Heart attack risk classification model: To construct a classification model, we opted for the CatBoost classifier [9]. This ML algorithm possesses the unique capability to handle categorical features directly, eliminating the need for additional preprocessing procedures. Our selection of CatBoost was motivated by its user-friendly nature, efficiency, and exceptional suitability for handling categorical data. For model training, the number of iterations and learning rate were set to 1000 and 0.2 respectively.

SHapley Additive exPlanations (SHAP) explanation: We used the SHAP technique to explain the output of our ML model. It calculates the values for each feature, stating the contribution to the classification model's decision, which are called SHAP values [6]. The SHAP technique can produce the SHAP values for each feature output, without depending on the relation between the feature and the label.

Global-based explanation: The SHAP method offers the capability to generate group-based model explanations, enabling the assessment of feature importance for the whole dataset. Analysis of the mean and variance of the SHAP values of the features allows us to understand how the model predicts outcomes for that particular group.

Local-based explanation: In SHAP, local explanations are used to provide insights into the prediction of a specific instance, going beyond general global explanations of the model. The results give an understanding of risk factors including their ranking with SHAP values. Using this explanation we can identify what are the main contributed risk factors for the specific patient or one instance in the dataset.

Model evaluation methods: In this study, various ML evaluation methods were used. Accuracy, precision, F1 score, recall, and the Cohen Kappa score [11] were calculated to observe the results of the CatBoost classifier model. These evaluation techniques were observed to improve our model's performance by setting different iterations and learning rates.

Django-based web framework: For our heart attack prediction model, we utilized Django, an open-source framework that employs the Model Template View (MTV) architecture and has robust security features. The system, integrating Python with HTML, CSS, and JavaScript, serves as an interface for patients and clinicians, providing access to a machine-learning model for heart attack prediction and enhancing stroke risk management.

Map, warning and data privacy: Beside patient risk prediction, the system includes a practical feature to display nearby hospitals on a map. Another feature called real-time health alerts notifies users of deviations from normal metrics, urging immediate medical consultation. With a user-friendly interface, the system can encourage proactive health management, reducing heart-stroke risks. Moreover, data security is ensured through a PostgreSQL database, employing SHA256 encryption for patient data integrity and privacy. This setup allows for secure data storage and retrieval, essential for maintaining accurate patient records and analysis.

Data collection from Smartwatch-based wearable device: To enhance the data collection process, the web application incorporates a wearable Smartwatch (Galaxy Watch 5) allowing them to acquire their heartbeat data, measured live through the smartwatch. ECG and BPM measurements in the Samsung Galaxy Watch 5 were approved by Health Canada, this was the reason behind the selection of this model. Users can directly transmit their heartbeat and blood pressure data to the system via Smartwatch through the use of our native application that integrates the user's current heartbeat and measured user blood pressure from the smartwatch.

A Galaxy Watch 5-based Smartwatch (Health Canada approved) running on the Google Wear operating system was employed to integrate it into our Django framework. The development process took place within the Android Studio framework. The data collected from the smart wearable device user is converted into a JSON file, which is then transmitted to the web application using a socket API. Subsequently, the web application processes the received data.

III. Results

Model evaluation: In this research study, a CatBoost classifier model was developed and trained using a dataset of 242 patients (80% of the original dataset). To evaluate the performance of the model, a separate dataset of 61 patients was used for testing (20% of the original dataset). The evaluation results showed that the model achieved an average AUC (Area Under the Curve) of 0.85, the Cohen Kappa score, which measures inter-rater agreement, was found to be 0.704. Additionally, the weighted average f1 score, a measure of precision and recall, was calculated as 0.85.

Global-based explanation: Following the training of the model, we further analyzed the risk classification by employing SHAP (SHapley Additive exPlanations) methods. By utilizing the complete dataset, we assigned SHAP values to each risk factor and ranked them based on their global importance. The result (not shown) displayed the top 8 factors that exert the most significant influence on the final prediction indicating that the number of major vessels (caa), chest pain (cp) and thalassemia (thall) were the most influential features on the prediction outcome respectively.

Risk Factors	Value
Patient ID	**Rating**
Age	44
Sex	Male
cp	2
trestbps	140
chol	235
fbs	0
restecg	0
thalach	180
exng	0
oldpeak	0
slope	2
caa	0
thall	2

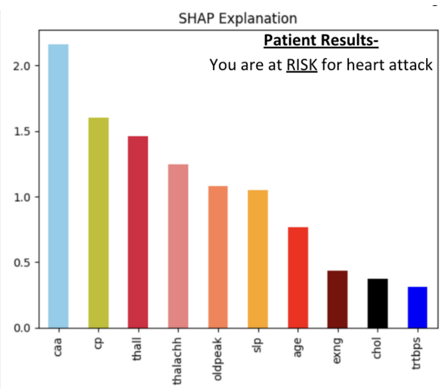

Fig. 2: User information and individualized recommendation output.

Patient-based explanation: Fig. 2 shows the risk factors that contribute to the prediction of individual patients. For example, we selected a 46-year-old man with the following characteristics: chest pain - 2, resting blood pressure - 140, cholesterol - 235, fasting blood sugar - 0, resting electrocardiographic result - 0, maximum heart rate achieved - 180, exercise-induced angina - 0, ST depression induced by exercise relative to rest - 0, number of major vessels - 0, thalassemia - 2. The model's prediction was accurate as it matched the actual class from the dataset, indicating that the individual has a risk of a heart attack.

User-centric role-based system access: The login features in the smart recommender system benefit clinicians and patients. Clinicians can securely access patient data, review predictions and risk factors, and make better decisions based on SHAP results. Whereas, patients can securely store their information, input risk factors, and receive personalized analysis. The login feature enables easy access to previous records, tracking of risk factors over time, and receiving updated recommendations based on the latest data. Overall, the login functionality enhances usability and security for both clinicians and patients. It provides seamless and personalized access, fosters collaboration, and improves the effectiveness of cardiovascular risk assessment and management.

Evaluation of the proposed recommender system: Our Django-based smart recommender system effectively enabled user operations such as account creation and data storage (Fig.1). The system collected user data, including static inputs from the website and dynamic inputs from a smartwatch, transmitted via a socket API and converted to JSON files (Steps 2 and 3). The prediction model then processed these risk factors to generate heart attack predictions (Step 4), with results displayed on the website, including model predictions and a local SHAP explanation (Fig. 2).

IV. Conclusion

In this study, we proposed an online recommender system to predict heart attack-associated risk and rank patient-specific risk factors that can be used by both clinicians and ordinary users. We aim to improve the model's ability to identify heart attack risk and offer explanations for the influence of each risk factor using SHAP technique. We integrated a framework that collects data from the user, and Smartwatch to display heart attack risk results with the user's data. Additionally, users can locate nearby hospitals for immediate action.

The heart attack analysis dataset was cleaned, processed, and divided into independent training and testing sets. The CatBoost classifier model accurately predicted heart attack risk, relying on the dataset of 303 patients' data. The model produced an average AUC of 0.85 and a Cohen Kappa score of 0.704 for the test dataset.

The risk factors that we used for our input from the dataset were: age, gender, chest pain type, resting blood pressure, cholesterol, fasting blood sugar, resting electrocardiographic result, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, number of major vessels and thalassemia. The dynamic factors which are heart rate and blood pressure were obtained in real time from wearable devices, while the static factors were collected from the user's input.

For dynamic data retrieval, we used Samsung Galaxy Watch 5, where we built a watch application with the Android Studio framework. With this application, individuals can measure their heart rate and blood pressure along with sending the collected data to the web application using the socket API to perform the prediction and explain the results.

Fig. 2 illustrates the findings of the explainable AI (XAI) analysis, showcasing the influence of various risk factors on the prediction. It highlights the key risk factors that have a significant impact on the final prediction, with taller bars representing higher SHAP values and indicating a greater influence. The model produced a high level of accuracy in predicting the risk of a heart attack for individual patients.

In prior studies [3] [4], ML algorithms were employed to predict heart attacks. However, these approaches lacked explanatory capabilities, leaving users and doctors uncertain about the reasoning behind the model predictions. Our approach uses XAI in the form of SHAP to address this limitation, predicting heart disease risk accurately as well as explaining the specific factors that contribute to each prediction for users and medical professionals.

Furthermore, we implemented a website where users can create an account by entering their username, email, and password. To log in, users need to write their username and password. The user data is secured and encrypted in the database. Once the user is logged in and inputted all risk factors, the system will produce a prediction and an explanation (Fig. 2). If any of the risk factors seem to be abnormal, the system sends out a warning notification. The map function that displays all of the nearest hospitals to the user to quickly locate nearby hospitals for immediate actions. The Django website can be accessed through the link below: https://www.mamatjanlab.com/heart/.

There is a limitation with the current smartwatches as they are not as precise as a medical-grade device. However, with this Health Canada approved smartwatch, users can dynamically measure their heart attack risk and receive early warning as a pre-assessment tool that guides users in taking preventative measures. The classification model also has a limitation due to the limited number of females in the dataset (100 out of 300), while 73/100 of them were found to have a risk of heart attack. Due to the imbalanced dataset, we could not perform gender-based risk stratification. In the future, we will test the model with additional datasets to further improve our recommender system.

The smart recommender system can be used by patients to predict heart attack risk and identify patient-specific risk factors. Clinicians can use this system to predict the heart attack risk of the patient along with viewing the most influential risk factors. Thus, it may guide them to understand underlying risks and further take the right step to prevent heart disease.

References

1. Statistics Canada. "Heart and Stroke Month." 2022. Available: https://www.statcan.gc.ca/o1/en/plus/353-heart-and-stroke-month.
2. R. Lafta, et al. "An intelligent recommender system based on short-term risk prediction for heart disease patients," *In2015 IEEE/WIC/ACM International Conference on WI-IAT,* Vol. 3, pp. 102-105, 2015.
3. Y. Yazici and Y. Mamatjan, "A Novel Approach for Efficient and Accurate Detection of Myocardial Infarction," *International Conference on Advanced Technology and Sciences* (ICAT 14), Turkey, 2014.
4. S. Abu-Alru, et al. "Smartwatch Electrodiagrams for Automated and Manual Diagnosis of Atrial Fibrillation: A Comparative Analysis of Three Models," *Front. Cardiovasc. Med.*, 2022.
5. R. ElShawi, et al. "Interpretability In Healthcare: A Comparative Study of Local Machine Learning Interpretability Techniques," *Comput. Intell.* 2020.
6. P. Gohel, et al. "Explainable AI: current status and future directions," *arXiv preprint arXiv* :2107.07045 (2021).
7. S.M. Lundberg and S.I. Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems* 30, NIPS, 2017.
8. E. Dekhil, et al. "A Novel Recommender System for Stroke Risk Stratification," *Proc. 19th IEEE Conference on CIBCB,* 2022.
9. A.V Dorogush, et al. "CatBoost: a gradient boosting with categorical features support," *Computing Research Repository (CoRR),* abs/1810.11363, 2018.
10. D. Lapp, "Heart Disease Dataset," *Kaggle*, 2019. Available: https://www.kaggle.com/datasets/johnsmith88/heart-disease-dataset.
11. M.L. Mchugh, "Interrater reliability: the kappa statistic Importance of measuring interrater reliability," *Biochem Med.* 22(3), 276–82, 2012.