2023 CMBEC45
Vancouver, B.C.
May 16-18, 2023
CMBES/SCGB

An Automated Online Recommender System for Stroke Risk Assessment

Shams Khan[1], Nour Dekhil[2], Ehsan Mamatjan[1], Safwat Hassan[3] and Yasin Mamatjan[1]

[1] Faculty of Science, Thompson Rivers University, Kamloops, Canada
[2] Electrical and Computer Engineering, Concordia University, Montreal, Canada
[3] Faculty of Information, University of Toronto, Toronto, Canada

Abstract— **A stroke is a complex emergency event that leads to major neurological impairments and patient disability. It is imperative to have an automated smart recommender system that can help with early stroke prediction and hence assist clinicians in stroke risk management. This study proposes the use of an automated online recommender system that can predict ischemic stroke risk levels based on the given patient-specific clinically identified stroke risk factors, such as systolic and diastolic blood pressure, age, gender, smoking habit, and cholesterol level. We integrate this model in an interactive Django-based web framework built upon software engineering best practices that can assist clinicians in monitoring patient-specific risk factors automatically. We use machine learning (ML) techniques to predict stroke risk levels and employ an Explainable Artificial Intelligence (XAI) strategy to explain the predictions to provide meaningful insights for stroke risk management. Additionally, we built a web framework for automated patient monitoring to monitor patient vital signals using a Smartwatch and transmit the data to the web application where the data is concurrently processed by the ML model and the output is displayed in an interactive dashboard. The results show that this automated online recommender system can predict ischemic stroke risk levels with an average Area Under the Curve of 0.98 for the independent test set. The automated recommender system can stratify stroke risk levels for optimal patient management and interpret the prediction associated risk factors for stroke in a trustworthy and transparent manner. Thus, our system can help clinicians to make better patient management in a busy clinic and guide ordinary users with potential stroke to improve their health habits to reduce the risk of stroke with greater accessibility.**

Keywords— **Telehealth, Explainable Artificial Intelligence (XAI), Online Recommender, Stroke Risk, Smartwatch**

I. Introduction

Acute ischemic stroke is an emergency neurovascular condition that affects the brain vessels carrying oxygen to the brain. Furthermore, the Center of Disease Control (CDC) states that stroke alone affects more than 795,000 people annually in North America, 87% of which are ischemic strokes, in which blood flow to the brain is blocked [1]. Also, individuals who suffered a transient ischemic attack (TIA) are at a recurring chance of suffering it again [2]. Thus, there is an urgent need for an automatic online recommender system that can not only aid in the early detection and classification of stroke risk but also help continuously monitor and track stroke risk factors. Previous studies employed machine learning (ML) techniques and created models to predict stroke risk levels based on clinically identified stroke risk factors [3, 10]. Although those studies employed similar methods, the resulting prediction is simply not explainable and hard to leverage, owing to the Black-box nature of these models. A study done by ElShawi et al. [4] tackled this by using interpretable ML techniques such that clinicians may understand the reasoning behind predictions and decisions made by the model. Previous work done by our lab applied these techniques and created an interpretable ML model using both Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive exPlanations (SHAP) explainer to provide meaningful insights into the predictions. Secondly, several studies employed an Internet of Things (IoT) network to monitor patient vital signs [5], but there are limitations with the feasibility and low-cost effectiveness of the sensors used in the system.

To address this challenge, we propose a novel online recommendation system to employ current Smartwatch technologies and automatically monitor 12 clinical stroke risk factors that are associated with stroke in individual patients. We apply a machine learning algorithm called CatBoost classifier [6] to predict stroke risk levels of patients and use SHAP to rank the risk factors for interpretation. The users enter their data from the web application, also having the option to send diastolic and systolic blood pressure data (dynamic risk factors) automatically through Smartwatches and generate results to guide clinicians and patients regarding stroke risk stratification and interpretation automatically.

II. Methodology

A. System Overview

In this study, we aim to build an automated online recommender system to guide clinicians to efficiently predict stroke risk levels based on the clinical risk factors of individual patients. We developed an online recommender system and employed commonly available Smartwatch devices that entail a

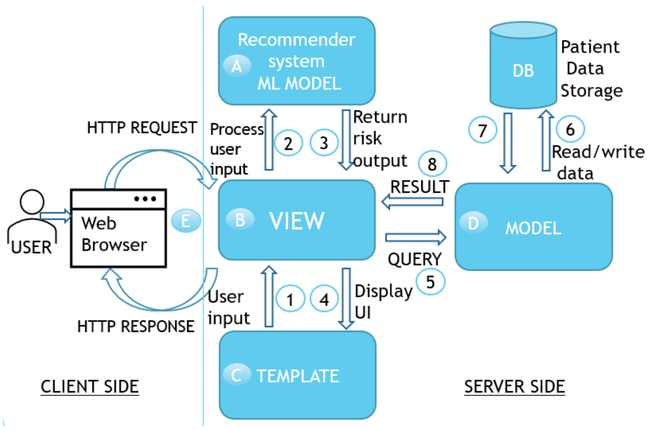

Fig. 1: The proposed recommender system framework for online stroke risk stratification.

complete framework for automated stroke risk prediction and guide users. First, our system consists of the ML component which can take the inputted patient data and process it using ML techniques to give a stroke risk level prediction. The SHAP explainer [7] is then processed on the data; this is the Explainable Artificial Intelligence (XAI) component that provides useful insights on the prediction. The application of this ML model is then extended to create a complete stroke risk management framework that consists of a web application built in Django framework, a Smartwatch device with clinically approved sensors, and a cloud server in its IoT network [5]. Specifically, the system follows a web application that is built in the Django framework, the application is adapted to integrate the ML model built by our lab. Furthermore, a Smartwatch device is also used that allows continuous patient stroke risk factor data acquisition dynamically. The recommender system can generate a patient-specific explanation report detailing a patient's stroke risk levels and its graphical representation. In Fig. 1, the graphical representation of the proposed software architecture demonstrates the working of the system which follows an 8-step process described in the following subsections.

B. Online recommender system framework

The primary unit of the stroke risk monitoring system is the web application based on the Django framework. This application has the ML model integrated into it and it also serves as a basis for the user to generate stroke risk results statically. The application is built on the MVT (Model, View, Template) software design architecture wherein the *"Template" (C)* refers to the front-end system that the user interacts with, the *"Model" (D)* refers to the database component in the application, and the *"View" (B)* is the controller that interlinks the various objects together and provides a channel of communication between them. The system also uses a Hypertext Transfer Protocol (HTTP) which is a client-server protocol used for fetching (HTTP request) or sending (HTTP Response) to the web server. The ML model for the recommender system is encapsulated in the *"ML Model component" (A)* wherein it contains the dataset and the SHAP explainer files, along with the driver code that generates the model after data processing.

C. Stroke risk factors

To build the classification model, we used the publicly available Stroke Analysis dataset [8, 11] composed of 4,798 patients. We split it into independent training (3,838 patients) and test sets (960 patients). This data consisted of stroke risk level output and 12 attributes associated with stroke namely gender, age, body mass index, smoking status, cholesterol level, systolic/diastolic blood pressure, glucose level, Thoracic Outlet Syndrome (TOS), and Modified Rankin Scale (MRS). These attributes were further classified into two parts: static and dynamic factors. Static factors refer to those attributes which remain constant for a patient over a brief periodic interval such as age, gender, smoking habit, cholesterol levels, TOS, and MRS. The dynamic factors are a group of factors that vary constantly, these are Systolic and Diastolic blood pressure and Glucose Levels. These factors are entered by the user in the application (1) and then sent to be processed (2) to the proposed ML component (*A*) of the recommender system through the View component *(B)* of the system.

D. Explainable AI model

XAI is a set of processes that enables people to understand and interpret the results generated by an ML model. Our classification model uses the CatBoost classifier for the four-stroke risk level classes categorized as no risk, low risk, moderate risk, and high risk for stroke. CatBoost [6] is a machine learning algorithm with the ability to use categorical features by transforming them into numbers without carrying out any data pre-processing. Furthermore, CatBoost employs Bayesian estimators to substantially decrease the chances of overfitting caused by generic gradient-boosting algorithms. Shapley Additive exPlanations (SHAP) are used for model interpretation [7]. It uses the concept of game theory to interpret predictions by evaluating the weightage of all the features in terms of their Shapley value. SHAP assists in understanding the contribution and results of the Black-box ML model. This in turn makes the model agnostic hence providing both global and local interpretability. SHAP has three different classes, from which TreeExplainer was used to interpret the results from the CatBoost classifier. The data is first processed and cleaned after which the CatBoost classifier classifies the risk level and the SHAP explainer ranks the risk factor (3). This data is then presented to the user in an interactive interface (4).

E. Secure database system

To store and manage patient records that entail various details such as patient stroke risk levels and the data points for the patient's stroke risk factors a database component is employed in the system. This allows the patient record to be stored in a database such that it can be Queried (retrieved) for future references (5). Sqlite3 database is used as a lightweight and robust database system that is compatible with nearly every operating system, thus optimizing the efficiency and portability of the system. The database is hashed using the PBKDF2 algorithm such that the data that is stored gets transformed into an encrypted value, hence all patient records are safe and can only be accessed by administrator users. Model (D) contains the application logic for the database system and is responsible for interacting with the Sqlite3 database. It can then either write data (6) or fetch data (7) from the Sqlite3 database and transfer it to Views (8) which would then in turn transfer it to the template (4) or send data to be processed by the recommender system ML component (2).

F. Continuous data acquisition model

The system also enables the user to employ a Smartwatch for sending the latest dynamic stroke risk data to the web application and get current prediction and explanation results. This is a novelty that is offered by this automated online recommender system. We have developed a platform-agnostic Smartwatch application able to run on any Android-based Smartwatch. This feature allows an autonomous approach to patient monitoring as the users have a choice wherein instead of statically entering data into the system, it can be collected dynamically through the Smartwatch device (E). For development purposes, this study used an Android-based Smartwatch - Galaxy Watch 5 developed by Samsung and operating on WearOS. The Smartwatch has sensors such as the FDA-approved ECG and blood pressure monitor, that will be used to collect blood pressure data from the patient user. These sensors would collect dynamic stroke risk factors and transfer them in the form of a JSON file to the cloud server with the help of Retrofit API which is an Inter Processing Communication (IPC) programming interface that allows bidirectional communication between client and server nodes (receiver and sender device) thus enabling data to be transmitted from the Smartwatch to the web application. The risk

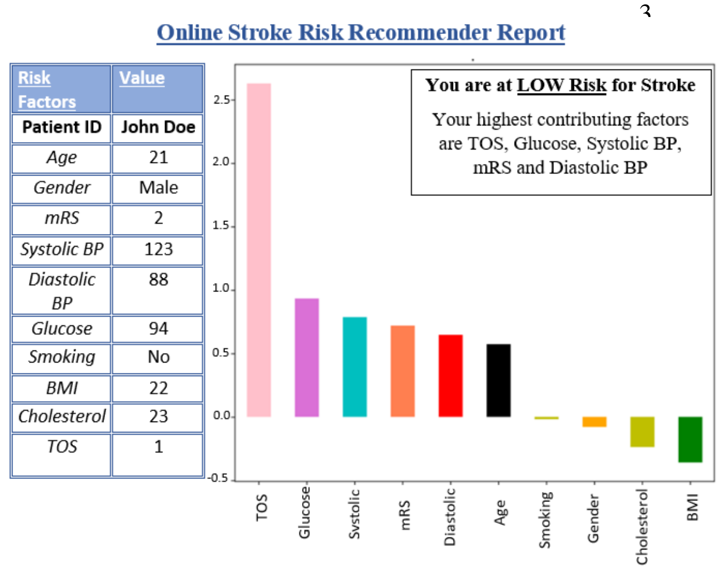

Fig. 2: Modified Patient-specific clinical report, generated by the automated online recommender system for John Doe (an anonymous user).

factor data is then processed by the recommender system ML component *(A)* in the web application (2).

III. Results

We demonstrated the proposed automated online recommender system for stroke risk assessment. Users can view the results of prediction and explanation in a patient-specific clinical pdf report generated by the system (5). The user data is first entered into the system, either manually in the application (2) or with the aid of the Smartwatch that can collect the data directly (1). This is then sent to the recommender system ML component present in the web application using WebSockets (3). After being processed, the results are saved by the database component and also available to the user as a clinical report (5) or to view in their dashboard (4). The classifier algorithm used in the system achieved an average area under the curve (AUC) of 0.98 using a one vs. all approach, a Cohen's Kappa score of 0.74, and a weighted average F1 score of 0.91 for the independent test set.

The proposed system is deployed on a web server and offers user registration and secure encryption of user data using the PBKDF2 algorithm. The system provides patient-specific stroke risk predictions, ranking of risk factors, and group explanations via an interactive user interface. A clinical report can also be downloaded as a pdf to guide patients. Patient data is stored in a secure database accessible to clinicians for diagnosis. The system supports continuous data acquisition through the use of Smartwatch devices to collect dynamic stroke risk factors such as systolic/diastolic blood pressure.

IV. Discussion and Conclusion

Our developed online automatic recommender system employs ML methods for effective stroke risk stratification and applies explainable AI strategies to make complex stroke information easy to understand predicted risk with associated risk factors for ordinary users. We have designed the automatic recommender system to have a simple web interface that is practical and publicly available as proof of principle.

The stroke risk prediction system employs the CatBoost classifier for accurate stroke risk prediction. The XAI component of the system utilizes SHAP to explain and interpret the results by ranking the risk factors based on their contribution towards stroke. To provide a user-friendly interface, we developed a web application using the Django framework in MVT software architecture. Users can upload their risk factor data and get results, which can also be downloaded as a clinical report (Fig 2). Our system also includes automatic data acquisition from a Smartwatch, which allows patients to send their heartbeat and blood pressure data dynamically using web sockets.

The stroke risk classification model in the web application achieves an AUC of 0.98 with the one vs all approach for the test set. SHAP values are used to rank the risk factors, and a graph visually represents the feature impact on a prediction. The XAI component is integrated and tested on the web application, which stores patient data for clinician use and generates a clinical report. The system assists clinicians in monitoring key risk factors and helps prevent stroke. The recommender system is publicly available at https://www.mamatjanlab.com/stroke/.

Previous studies have utilized IoT-based networks for patient health monitoring; however, the monitoring equipment used was experimental, expensive, and not publicly available. Furthermore, no existing system can predict stroke risk in a specific patient and subsequently manage and monitor it. Some studies have utilized machine learning techniques to predict stroke risk, but these lack interpretability and transparency, making it impractical for clinicians to rely on them for diagnosis [3, 10]. Dekhil et al. [9] applied interpretability techniques to stroke risk stratification, but their system did not integrate the Django web framework or current Smartwatch/Smartphone technologies.

We aim to provide a proof of concept for an online stroke risk assessment framework integrating current state-of-the-art technologies. The web application allows the system to be readily used while storing patient records securely. The intended purpose of this system is not for stroke diagnosis, but rather to serve as an ancillary pre-assessment tool to guide patients in taking preventative measures. The predictions can give the clinicians a second opinion to guide them in their diagnosis. Thus, we established the groundwork for such a smart recommender system by integrating an XAI module, a wearable Smartwatch device for patient monitoring, and an online system that could detect ischemic stroke risk early so that users can take necessary steps to reduce key risk factors.

V. Future Works

We acknowledge that wearable technology has limitations in terms of the accuracy of blood pressure measurement. This is not an integral part of our system, but an additional functionality that no one has implemented for stroke risk estimation or remote monitoring. Our recommender system offers real-time blood pressure and heart rate updates besides other risk factors used in this study. In the future, we intend to make it a practical tool for ordinary users to estimate stroke risk by integrating it with the current BP monitoring device with greater accuracy. Furthermore, our system is not intended to replace the NIHSS score which is a numeric range estimated by specialized clinicians. We excluded the NIHSS score from our ML/XAI models and categorized it into four risk levels. We also plan to evaluate our system with an additional external data set. While this study is still in progress, the proposed system can provide the reasoning behind stroke prediction to users and clinicians to make better patient-specific management to avoid high-risk factors.

References

1. “Stroke,” Centers for Disease Control and Prevention, 01-Sep-2022. [Online]. Available: https://www.cdc.gov/stroke/index.html.
2. S. Khare, “Risk factors of transient ischemic attack: An overview,” *J Mid-life Health*, vol. 7, no. 1, p. 2, 2016.
3. T. Tazin, et al. “Stroke disease detection and prediction using robust learning approaches,” *J. Healthc. Eng*, vol. 2021, pp. 1–12, 2021.
4. R. ElShawi, et al. ”Interpretability in Healthcare: A Comparative Study of Local Machine Learning Interpretability Techniques,” *Comput. Intell.*, 2020.
5. H.T. Yew, et al. ”IoT Based Real-Time Remote Patient Monitoring System,” *16th IEEE ICSPA*, pp. 176-179, Malaysia, 2020.
6. A.V. Dorogush, et al. “CatBoost: gradient boosting with categorical features support,” *CoRR*, vol. abs/1810.11363, 2018.
7. S.M. Lundberg and S. Lee, ”A Unified Approach to Interpreting Model Predictions”, *Advances in NeurIPS 30 (NIPS 2017)*, 2017.
8. V. Bandi, et al. “Stroke Analysis”, *Mendeley Data*, V1, 2020. https://data.mendeley.com/datasets/jpb5tds9f6
9. N. Dekhil, et al. ”A Novel Recommender System for Stroke Risk Stratification”, *Proc. 19th IEEE Conference on CIBCB,* 2022.
10. K. Harshitha, et al. “Stroke prediction using machine learning algorithms,” *IJIREM*, vol. 8, no. 4, 2021.
11. V. Bandi, et al. “Prediction of Brain Stroke Severity Using Machine Learning,” *Revue d’intelligence Artificielle*, Vol. 34, No. 6, 2020.