2024 IEEE Conference on Artificial Intelligence (CAI) | 979-8-3503-5409-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/CAI59869.2024.00163

Machine and Deep Learning Based Clinical Decision Making for Coronary Artery Disease and Chatbot Tool

1st Vincent Wei Jun Sim
Dept. of Cardiology
National Heart Centre Singapore
Singapore, Singapore
vincent.sim@mohh.com.sg

2nd Glades Tan
Dept. of Cardiology
National Heart Centre Singapore
Singapore, Singapore
gladeshmm@gmail.com

3rd Xinliu Zhong
Dept. of Biomedical Engineering
National University of Singapore
Singapore, Singapore
xinliuzhong@u.nus.edu

4th Terrance SJ Chua
Dept. of Cardiology
National Heart Centre Singapore
Singapore, Singapore
terrance.chua.s.j@singhealth.com.sg

5th Jie Jun Wong
Dept. of Cardiology
National Heart Centre Singapore
Singapore, Singapore
jiejun.wong@mohh.com.sg

6th Si Yong Yeo
Dept. of Digital Health
Lee Kong Chian School of Medicine
Singapore, Singapore
siyong.yeo@ntu.edu.sg

7th Angela S. Koh
Dept. of Cardiology
National Heart Centre Singapore
Singapore, Singapore
angela.koh.s.m@singhealth.com.sg

***Abstract*—Coronary Artery Disease (CAD) is a leading cause of cardiovascular disease. No dominating clinical feature determines how physicians make clinical decisions for patients with CAD. CAD clinical features are wide ranging, from symptoms, to risk factors, to kidney disease, to individualized factors such as family history. These diverse features are weighted differently by each physician in their evaluation of the patient, leading to variable clinical decision pathways. The aim of this study is to utilize a deep learning framework with a sequential attention mechanism that amassed CAD clinical features, to determine key features that dominated the final clinical decision.**

We identified a total of 10 variables inclusive of 7 cardiovascular risk factors, 2 symptoms as well as any previous cardiac testing done that would affect the clinical decision making for management of CAD, producing a total of 13,824 total scenarios. Subsequently, we focused on cardiovascular risk factors, selecting a total of 6 features, to produce a total of 384 scenarios for the parent dataset. A total of 3 decisions outcomes representing therapeutic decision making were chosen: 'Risk factor management alone', 'exercise ECG' as well as 'functional or anatomical testing'.

Features such as Diabetes Mellitus (approx. 0.083), Peripheral Vascular Disease/Cerebrovascular Accident/Family History (approx. 0.061) and Framingham Risk score Ten Year Risk (approx. 0.053) ranked high in importance in the SHAP analysis with the model attaining a good overall accuracy of 0.966. Feature optimization using only these top three features showed good model accuracy of 0.922.

On the other hand, the TabNet architecture reported high feature importance for Framingham Risk score Ten Year Risk (0.298), followed by chronic kidney disease (0.297) and Family history of premature death (0.235) which contributed to a best test accuracy of 0.940 out of 10 epochs. Optimizing the dataset with these three features retained the model accuracy of 0.940.

We further uploaded the algorithm to a web-based chatbot system as a future support tool for physicians.

Keywords – ***artificial intelligence; coronary artery disease; clinical decision-making; deep learning; neural networks***

I. Introduction

Coronary artery disease (CAD) is the leading type of cardiovascular disease across the globe. CAD management represents a complex clinical paradigm, spanning across disease diagnosis to disease therapeutics. Physicians often juggle with multiple CAD clinical features that may be unique to each patient, in order to arrive at physician-driven clinical decision pathways. CAD clinical features are wide ranging, from symptoms, to risk factors, to kidney disease, to individualized factors such as family history. These diverse features are weighted differently by each physician in their evaluation of the patient, leading to variable clinical decision pathways. Given the wide spectrum of clinical manifestations, we utilized deep learning frameworks to elucidate key features involved in clinical decision making in multiple CAD clinical scenarios.

II. Methods

A. Data Collection

We identified a total of 10 variables inclusive of 7 cardiovascular risk factors, 2 symptoms as well as any previous testing done that would affect the clinical decision making for management of CAD, producing a total of 13,824 total scenarios.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00163

Subsequently, we focused on cardiovascular risk factors, selecting a total of 6 features, to produce a total of 384 scenarios for the original dataset. A total of 3 decisions outcomes representing therapeutic decision making were chosen: ‘Risk factor management alone’, ‘exercise ECG’ as well as ‘functional or anatomical testing’.

Risk factor management includes control of lifestyle factors with dietary control, exercise, quitting smoking as well as pharmacological management of any underlying hypertension, hyperlipidemia, and diabetes mellitus. Exercise ECG involves recording the ECG of a patient’s heart while walking on a treadmill to look for any exercise-induced ischemia. Functional or anatomical testing involves echocardiogram-stress test, nuclear heart scans or a coronary angiogram.

We then conducted a survey with 20 scenarios disseminated across to 10 senior cardiologist to gather the decisions these cardiologists would make in these clinical scenarios focusing the choice of scenarios based on instances where clinical decisions are not straightforward and clearcut. For each scenario, responses were compiled in a row-by-row fashion to train the models via 2 methods: merged dataset vs aggregate decision column.

B. Model Development

The model was constructed using a supervised greedy layer-wise [1] multi-layered neural network method [2] where 70 percent of the dataset was utilized on pretraining and the remaining 30 percent on testing.

We used a combination of Shapley [3] (Local model-agnostic method) and TabNet architecture [4] (Global model-specific method) in our therapeutic decision making.

Local feature importance calculates the importance of each feature for each data point. In our case, we averaged the absolute Shapley values [5] per feature for every observation (local feature importance) across the entire dataset to get the global feature importance. This gives an overall view on which features are generally more influential in the model making process. The model performance is determined by several performance metrics including the Receiver Operator Curve Area Under Curve (ROC AUC), accuracy, recall, precision and F1 scores.

TabNet analysis was performed on parent and survey datasets using width values of 8 for both the decision prediction layer and attention embedding for each mask.

Upon successful training, we extracted feature importance from the trained TabNet model, which provided insights into which features were most influential in the classification decisions. These importance scores were visualized using a bar plot, highlighting the relative importance of each feature in the CAD decision-making process. This interpretability aspect is crucial for understanding the driving factors behind the model’s predictions and for gaining trust in its outputs.

C. Platform

A web-based chatbot system was then developed as a clinical decision support tool that can be used in clinical workflows in patients with significant cardiovascular risk factors to engage in secondary assessments of at-risk patients.

III. Results

A. Model accuracy

Using the supervised greedy layer-wise neural network parent model applied to the survey dataset, the parent model had a mean ROC AUC of 0.984 and a mean accuracy of 0.965 (Fig.1). The mean recall score, precision score, and F1 score were 0.848, 0.848, and 0.848 respectively.

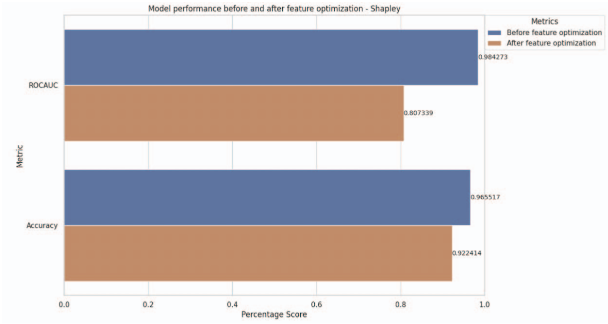

Fig. 1. Shapley Model Performance before and after feature optimization

B. Shapely feature importance

Shapley analysis after sequential features selection reported the highest global feature importance (Fig.2) for Risk Factor: Diabetes Mellitus (RF:DM) (approx. 0.083), followed by Risk Factor: Peripheral Vascular Disease/Cerebrovascular Accident/Family History (RF: PVD/CVA/FH) (approx. 0.061), Scoring: Framingham Risk score Ten Year Risk (Scoring: FHS TYR) (approx. 0.053), Risk Factor: Chronic Kidney Disease (RF: CKD) (approx. 0.032), Family history of premature death (approx. 0.027) and other risk factors (approx. 0.005) using mean SHAP values. Upon reducing the dataset to the top 3 features – RF: DM, RF: PVD/CVA/FH and Scoring: FHS TYR, the model performed fairly with an ROC AUC of 0.807 and a test accuracy of 0.922, albeit a decrease from before feature optimization. This reflects the significance of these 3 features in deciding on the optimal treatment for the patient.

C. TabNet feature importance

Model training utilizing a maximum of ten epochs produced a best epoch count of zero and a best test accuracy value of 0.940 (Fig.3). TabNet analysis reported the highest feature importance (Fig.4) for Framingham Risk score Ten Year Risk (0.298), followed by chronic kidney disease (0.297), Family history of premature death (0.235) and Diabetes Mellitus (0.106). Other risk factors and Peripheral Vascular Disease/Cerebrovascular Accident/Family History have the relatively lowest feature importance values at 0.057 and 0.0081, respectively. After extracting the top 3 features for dataset

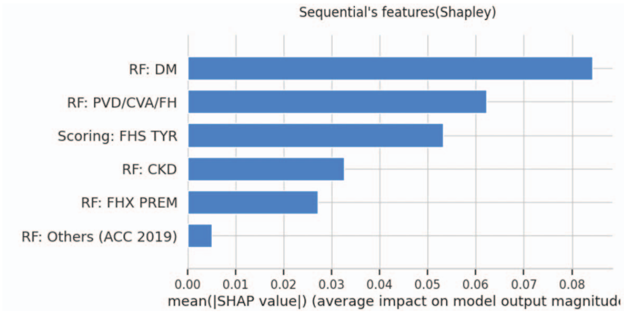

Fig. 2. Shapley feature importance of variables

optimization, the model attained the same accuracy score of 0.940 as before feature optimization.

In both SHAP and TabNet analysis, Scoring: FHS TYR was commonly identified as a highly important feature in both SHAP and TabNet analysis, indicating its predominant influence in the therapeutic decision making process.

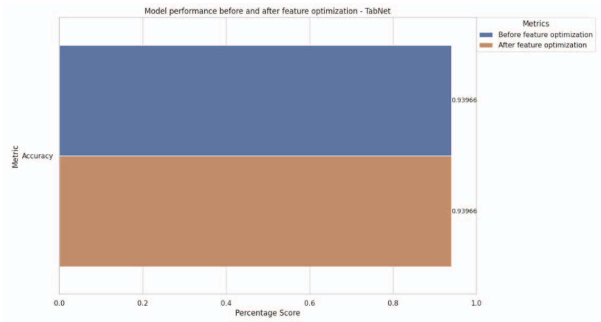

Fig. 3. TabNet Model Performance before and after feature optimization

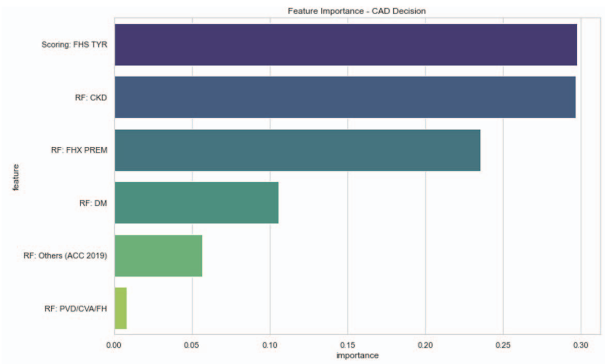

Fig. 4. TabNet feature importance of variables

D. Platform

We further uploaded the algorithm to a web-based chatbot system as a future support tool for physicians (Fig.5).

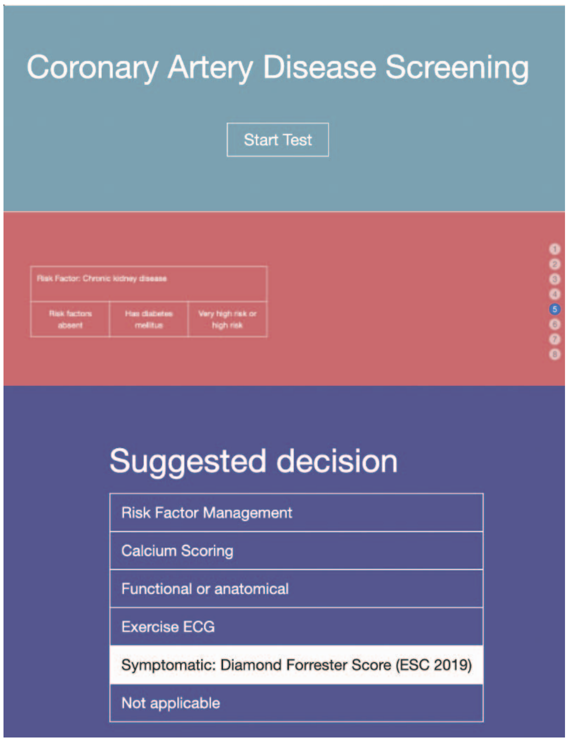

Fig. 5. Web based chatbot system

Discussion

We explore the use of an efficient pipeline method employing machine learning and deep learning techniques to model physicians' evaluation of patients for coronary artery disease. Traditional risk scores like FHS TYR are commonly used in daily clinical practice but fail to account for other important vascular factors such as peripheral vascular disease and kidney disease. Our results show the importance of such vascular factors in physicians' evaluation and decision-making process for patients.

While FHS TYR was a feature that showed up in the methods, our pipeline demonstrate relative importance of FHS TYR against other features. This is a novelty of this method otherwise not demonstrable by routine statistics.

There are several distinct Framingham risk models and one of the commonly used Framingham outcomes model is intended for use in non-diabetic patients aged 30-79 years with no prior history of coronary heart disease or intermittent claudication. However, based on the Shapley feature importance, we found that diabetes mellitus has the highest feature importance in influencing physicians' decision-making process for CAD treatment. This provides basis to incorporate diabetes mellitus into clinical tools for CAD decision making, using real-life physician derived data to build reliable chatbots.

One of the key limitations of this study is that the model's performance in this study is based on an original produced dataset, which may not fully represent the broader CAD patient population. There is a risk that the model's findings and the identified key features may not be applicable to different healthcare settings or populations that are not represented in the study's data.

Moving forward, data can be collected from a wider range of demographics, locations, and different clinical settings to clinically validate the findings seen in this study. Web-based chatbots can then be pushed forward in clinics once validated as a clinical decision support tool that can be used in clinical workflows in patients with significant cardiovascular risk factors to engage in secondary assessments of at-risk patients. Data can be collected from cardiologists to determine how they have benefitted from the chatbot system.

REFERENCES

[1] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H. (2006). Greedy layer-wise training of deep networks. Advances in neural information processing systems, 19.

[2] Haykin, S. (1998). Neural networks: a comprehensive foundation. Prentice Hall PTR.

[3] Covert, I., Lee, S. I. (2021, March). Improving kernelshap: Practical shapley value estimation using linear regression. In International Conference on Artificial Intelligence and Statistics (pp. 3457-3465). PMLR.

[4] Arik, S. Ö., Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 8, pp. 6679-6687).

[5] Sundararajan, M., Najmi, A. (2020, November). The many Shapley values for model explanation. In International conference on machine learning (pp. 9269-9278). PMLR.