MILESTONE TRANSACTIONS ON
MEDICAL TECHNOMETRICS
VOL. 1, ISSUE. 1, 2023

ORIGINAL RESEARCH

Obtaining Better Accuracy Using Fusion of Two Machine Learning Algorithms for Prediction of Heart Attack

Soumyashree M[1] . A R S Bhaargav[2] . Prashanth B K[2] . S Ajith Kumar Reddy[2]

[1]Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, India
[2]School of Computer Science and Engineering, REVA University, Bangalore, India

Received: 29 June 2023 / Revised: 27 August 2023 / Accepted: 14 September 2023

DOI: 10.5281/zenodo.8344749

Abstract – There are numerous terms and labels for heart illness. The effects and processes of a cardiac condition can be very complicated, especially for those who are experiencing them themselves or through a loved one of the victim. Therefore, the goal of this paper is to understand why heart disease plays a factor in nearly half of all reported fatalities in society. This paper will start by outlining the fundamental concepts of what a cardiac illness is and how people typically view it. Additionally, this paper will explore the most prevalent risk factors for developing heart disease, and we'll do it by employing ML algorithms using a hybrid model.

Index Terms – Decision Tree, Random Forest, Hybrid Model, Model Ensemble, SDLC model like spiral V and V, Graphical User Interface, Django.

I. INTRODUCTION

Given that heart attacks cause 32% of all fatalities worldwide, they have emerged as one of the major threats to human existence. All age groups, starting with those under 21, are impacted by this issue. We created an ML model to estimate the likelihood of a heart attack based on symptoms like blood pressure, cholesterol, smoking status, and other parameters to stop this issue from getting out of hand. Based on training data, an ML model is developed to forecast the outcome. We constructed this machine learning (ML) model by combining two models using the model ensembling technique. We called it the hybrid model, and its accuracy is more than 96%, making its predictions more accurate than those made using the other models. Due to the dataset's focus on ML rather than deep learning, only numerical data is present. This effort uses a dataset based on certain symptoms like high blood pressure and tiredness to predict the likelihood of a myocardial infarction.

Motivation : The different challenges faced in the prediction is to find a cure immediately as discussed above. Many researchers have performed many classification techniques to overcome the challenges faced. Heart Attack prediction is a very fatal disease hence we have built this model to prevent it to a certain extent.

Contributions: We use the HD.CSV dataset from Kaggle and we pre process the dataset. The next step involved is we use dimensionality reduction algorithms for reducing the size of the dataset for proper analysis and then build our hybrid model using model ensembling.

Organisation: The paper is organised as follows: Section II includes the authors' prior related work in the field of ML. Section IV is about our outcomes and the findings acquired by our suggested model, followed by conclusion and references done in our article. Section III includes the model proposed, the algorithm utilised, and the description of the dataset used.

II. RELATED WORK

Other researchers have employed a variety of techniques to identify heart attacks at an early stage. We have mentioned some of these techniques, including those used by [1] Y. Rongfeng and X. Minhui, who found that nearly 9 individuals had a stroke before having a heart attack. R. A. Bauder, T. M. Khoshgoftaar, and T. Hasanin conducted study on the imbalances present in big data labs, which they report in [3]. The dataset was subjected to several pre-processing approaches in [4] by Durairaj, M. & Ramasamy, N. We used these techniques in our study.In [5] and [6], Das, R., Turkoglu, I. & Sengur and Yang, H. & Garibaldi, respectively, used a model ensemble approach and built a hybrid model, which we used as the foundation for our project. We combined the Decision Tree and Random Forest algorithms to increase the success rate for heart attack prediction, and we also used the model ensemble approach to increase the model's overall accuracy.

The authors of [7] Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., and Yarifard, A. A. used the neural networks idea to forecast the occurrence of heart attacks. We learned about the various implementation strategies for techniques like Principal Component Analysis (PCA), Decision tree, and Random Forest using [9]. The hybrid model is included in the form. With the assistance of [12] Adrian Holovaty and Jacob K. Moss, we have implemented the hybrid model in the form of a website where the user can enter heart attack symptoms like blood pressure and cholesterol levels, among others.

III. METHODOLOGY

This research work aims to predict the possibility of a person to suffer from a heart attack by giving certain symptoms like BP, cholesterol, walking style etc. as the input and the hybrid model built by us would predict if the person is prone to a heart attack. The framework used to create the graphical user interface(GUI) is Django which uses python as the backend for implementation .

We have tried to solve the following,

1. To remove the invalid, redundant and other inaccurate data from the dataset.
2. To make computing easier by reducing the dataset's number of dimensions.
3. The current algorithms, such as Decision Tree and Random Forest, that we used in our research

provide an accuracy of roughly 93–94%.

4. To increase the model's accuracy, we created a hybrid model that combines two methods utilising the model ensembling methodology to achieve an accuracy of roughly 96%.
5. We have adopted the hybrid approach using a website to enhance User Interface and User Experience.

Proposed System

This section explains the proposed system for solving the problem which is machine learning and the algorithms used are Random Forest Algorithm, which is a supervised ML algorithm and Decision Tree algorithm which is based on the method of ensemble learning to improve the performance of the model.

Dataset

This dataset came from the University of California Irvine data repository used to predict heart attack. The dataset consists of 22 columns and 2,53,681 rows and the main symptoms considered for prediction are blood pressure, High Cholesterol, Different walking style, smoker, BMI(Body Mass Index).

Theoretical Background

The algorithms used for computation are:

1. **Random Forest Classifier**

```
rn = RandomForestClassifier()
rn.fit(x_train,y_train)
rn_pred = rn.predict(x_test)
rn_ac = accuracy_score(y_test,rn_pred)
msg="Accuracy of Random Forest is: "+str(rn_ac)
return render(request,Module,{'msg':msg})
```

2. **Decision Tree Algorithm**

```
de = DecisionTreeClassifier()
de.fit(x_train,y_train)
de_pred = de.predict(x_test)
de_ac=accuracy_score(y_test,de_pred)
msg="Accuracy of Decision Tree is: "+str(de_ac)
return render(request, Module,{'msg':msg})
```

3. **Hybrid Model :**

```
hybrid = list()
hybrid.append(('RF', RandomForestClassifier()))
hybrid.append(('de', DecisionTreeClassifier()))
```

```
model_h=StackingClassifier(estimators=hybrid)
model_h.fit(x_train, y_train)
y = model_h.predict(x_test)
acc = accuracy_score(y_test, y)
print(acc)
msg="Accuracy of Hybrid Module is: "+str(acc)
return render(request,Module,{'msg':msg})
```

The above three algorithms are used to predict the output and of the three models the hybrid model is the most accurate algorithm as it is the fusion of the two algorithms. It provides an accuracy of about 96%. The figure. 1 below shows the accuracy of our hybrid model.

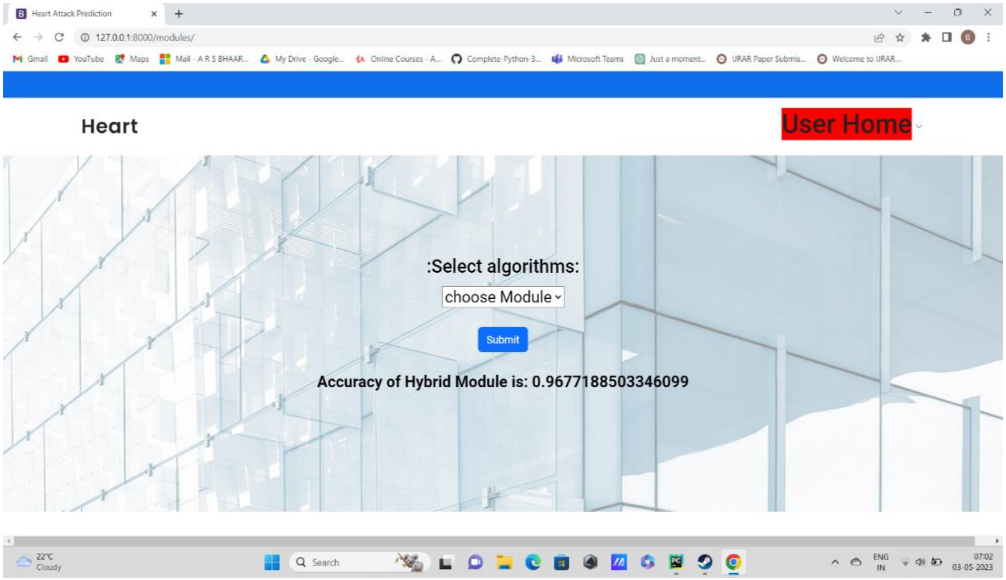

Fig. 1: Accuracy of the hybrid model

Processing

We obtained the dataset from Kaggle, and it was processed using the following steps: It is a benchmark dataset used at a well known data repository for heart attack prediction. To simplify the analysis, the dataset is pre-processed using dimensionality methods. After that, the dataset is separated into a training set and a testing set, with a 70% to 30%, as shown in (Fig. 2). The dataset will then go through feature scaling, a procedure that converts whole number values into decimal values in order to optimise the algorithms. The findings are analysed in the form of a matrix called the confusion matrix, whose dimensions are represented in the form of n components, the values of which can range from 2,5,7 depending on our requirement. Next, we employ the Decision tree and Random forest algorithms for our study. Following analysis of the aforementioned two algorithms, the accuracy rates of the two models were determined to be 93% and 95%, respectively.

We used the model ensembling strategy to create a hybrid model, which combines both algorithms, to increase the accuracy of our model, as shown in (Fig. 1). This strategy increases model accuracy since it combines two algorithms, giving it an accuracy of 96%, which is 1% more than the Random Forest algorithm. As a result, we take the hybrid model as the foundation for our prediction of heart attack and implement it using the Django framework to create a graphical user interface. The GUI contains a login page where users must first register before logging in. Once logged in, users can upload a dataset and evaluate the accuracy of all three models on the model page. The user can then enter the essential information, such as blood pressure, cholesterol, heart rate, whether they smoke or not, and other factors, as shown in figure 2, to determine whether they are at risk of having a myocardial infraction or heart attack.

Data Flow diagram

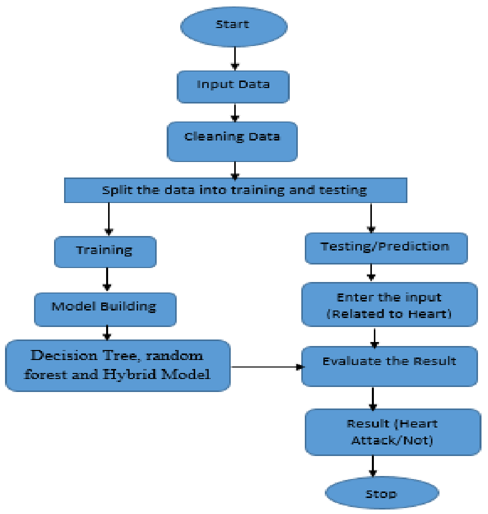

Fig 2: Outline Diagram

Procedure flow diagram:

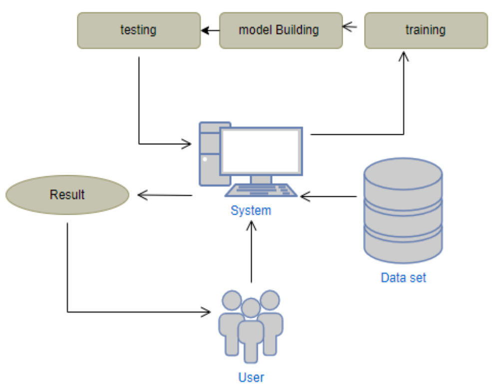

Fig. 3: Architectural Diagram

Experimental setup: The figure below shows the overall machine learning model setup and the terminologies used in our model.

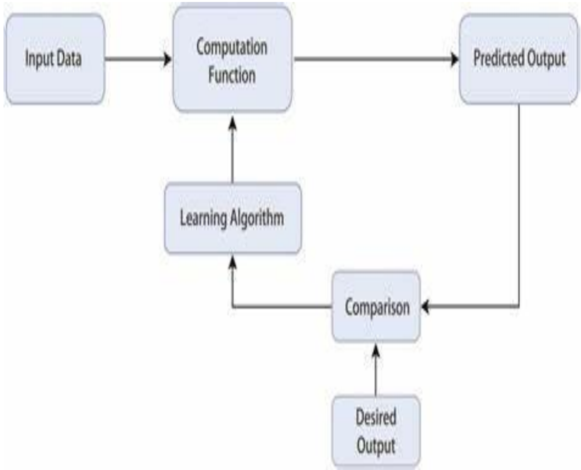

Fig. 4: Overall Experimental Setup

IV. RESULTS

The proposed model i.e. Hybrid Model used by us provides an accuracy of 96% for predicting heart attack in patients and hence it can be used to avoid heart attack at an earlier stage. The figure below shows the webpage where user has to give the inputs to know if he/she is prone to a heart attack.

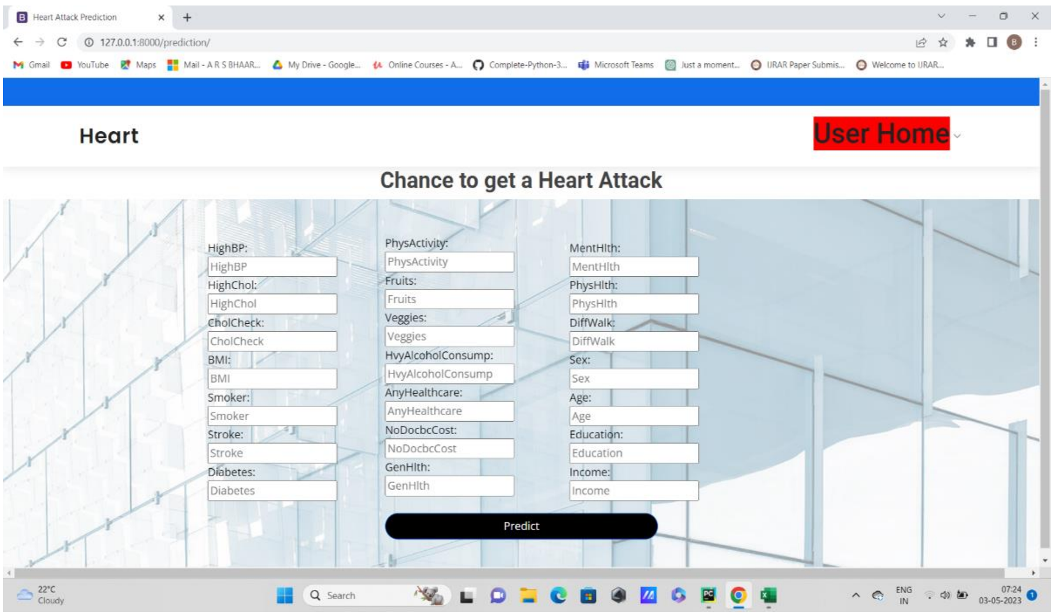

Fig. 5: Prone to heart attack

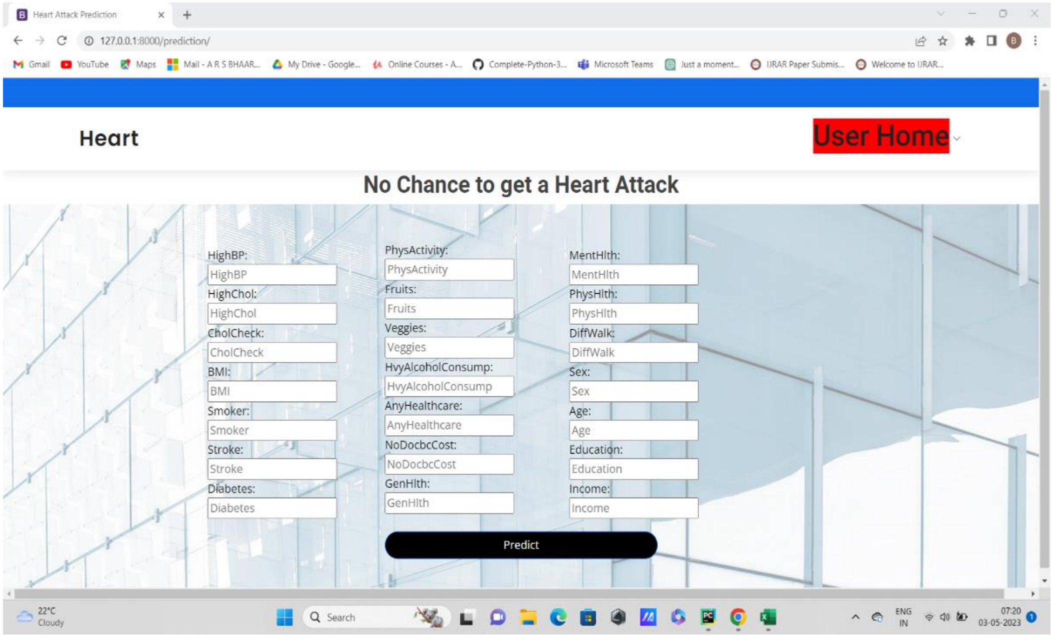

Fig.6: Not Prone to heart attack image

V. CONCLUSION

Thus, we draw the conclusion that the accuracy of the first two models—the Decision Tree and Random Forest models—was 93% and 95%, respectively. By improvising a model and calling it a hybrid model, we were able to boost the model's accuracy score and get a 96% accuracy rate. Using cholesterol and blood pressure as the root nodes, this gave a concrete response to the inputs supplied by the users and forecasted the outcomes.

REFERENCES

1. Wang, M., Yao, X., & Chen, Y. (2021). An imbalanced-data processing algorithm for the prediction of heart attack in stroke patients. *IEEE Access*, *9*, 25394-25404.
2. Bauder, R. A., Khoshgoftaar, T. M., & Hasanin, T. (2018, November). Data sampling approaches with severely imbalanced big data for medicare fraud detection. In *2018 IEEE 30th international conference on tools with artificial intelligence (ICTAI)* (pp. 137-142). IEEE.
3. Durairaj, M., & Ramasamy, N. (2016). A comparison of the perceptive approaches for preprocessing the data set for predicting fertility success rate. *Int. J. Control Theory Appl*, *9*(27), 255-260.
4. Das, R., Turkoglu, I., & Sengur, A. (2009). Effective diagnosis of heart disease through neural networks ensembles. *Expert systems with applications*, *36*(4), 7675-7680.
5. Yang, H., & Garibaldi, J. M. (2015). A hybrid model for automatic identification of risk factors for heart disease. *Journal of biomedical informatics*, *58*, S171-S182.
6. Ahmed, S. T., Sandhya, M., & Sankar, S. (2019). A dynamic MooM dataset processing under TelMED protocol design for QoS improvisation of telemedicine environment. *Journal of medical systems*, *43*(8), 257.
7. Kumar, S. S., Ahmed, S. T., Xin, Q., Sandeep, S., Madheswaran, M., & Basha, S. M. (2022). Unstructured Oncological Image Cluster Identification Using Improved Unsupervised Clustering Techniques. *Computers, Materials & Continua*, *72*(1).
8. Arabasadi, Z., Alizadehsani, R., Roshanzamir, M., Moosaei, H., & Yarifard, A. A. (2017). Computer aided decision making for heart disease detection using hybrid neural network-Genetic algorithm. *Computer methods and programs in biomedicine*, *141*, 19-26.
9. Samuel, O. W., Asogbon, G. M., Sangaiah, A. K., Fang, P., & Li, G. (2017). An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert systems with applications*, *68*, 163-172.

10. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J. J., Sandhu, S., ... & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *The American journal of cardiology*, *64*(5), 304-310.
11. Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile information systems*, *2018*, 1-21.
12. Al-Shammari, N. K., Alzamil, A. A., Albadarn, M., Ahmed, S. A., Syed, M. B., Alshammari, A. S., & Gabr, A. M. (2021). Cardiac stroke prediction framework using hybrid optimization algorithm under DNN. *Engineering, Technology & Applied Science Research*, *11*(4), 7436-7441.
13. Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. *IEEE access*, *7*, 81542-81554.
14. Holovaty, A., & Kaplan-Moss, J. (2009). *The definitive guide to Django: Web development done right*. Apress.
15. LK, S. S., Ahmed, S. T., Anitha, K., & Pushpa, M. K. (2021, November). COVID-19 outbreak based coronary heart diseases (CHD) prediction using SVM and risk factor validation. In *2021 Innovations in Power and Advanced Computing Technologies (i-PACT)* (pp. 1-5). IEEE.