IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DEEP - ENSEMBLE BLENDING BASED CARDIOVASCULAR DISEASE DETECTION SYSTEM

[1]Md. Sajida Begum, [2]M. Bindu, [3]T. Manoj Kumar, [4]P. Prem Vithin,[5]V.Rashmi

[1]B.Tech Student, [2]B.Tech Student, [3]B.Tech Student, [4]B.Tech Student,[5]Assistant Professor

[1,2,3,4,5]Information Technology,

[1,2,3,4,5]Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India

Abstract: Cardiovascular disease (CVD) is one of the major causes of mortality around the globe, necessitating the need for early detection that will enable better health care results. Existing diagnostic methods utilize significant clinical expertise along with extensive testing; hence, it becomes time-consuming in detecting diseases. This study proposes the "Deep-Ensemble Cardiovascular Disease Detection System" which uses the power of deep learning to detect CVDs at an early stage.

The proposed system incorporates a number of deep learning algorithms to develop an ensemble for better performance in predicting the results. In the data preprocessing process, the dataset is cleaned and balanced to improve model learning. To ensure effective prediction and detection of heart diseases, the ensemble uses the CNN and GRU architectures to capture the patterns from health data.

Primarily, rather than classifying patients in two categories, including No Disease and Disease, the categorical prediction is extended to four categories, including No Disease, Low Risk, Medium Risk and High Risk. Also, this categorical prediction is not limited to two categories, namely No disease and Disease but extended to four categories, namely No Disease, Low Risk, Medium Risk and High Risk.

First, instead of two categories of predictions, such as No Disease and Disease, the categorical prediction is expanded to four categories, including No Disease, Low Risk, Medium Risk, and High Risk. Second, it does not only involve the prediction of two categories, i.e., No disease and Disease, but the expansion to four categories, including No Disease, Low Risk, Medium Risk, and High Risk. Third, an online cardiovascular disease prediction and an internet-based application are realized. Lastly, a categorization of risk prediction is performed by the system. From experimental results, it can be observed that the proposed deep-ensemble model assures higher levels of accuracy, reliability, and usability compared to conventional machine learning methodologies.

The proposed system can be used by healthcare professionals for the purpose of early diagnosis, risk assessment, and preventive treatment planning, which helps improve patient care and reduce mortality.

Keywords: Cardiovascular Disease Detection, Deep Ensemble Learning, CNN-GRU Model, Multi-Class Classification, Risk Prediction, Web-Based Prediction System, Artificial Intelligence, Healthcare Analytics.

I.INTRODUCTION

Cardiovascular disease (CVD) is one of the leading causes of death globally and is a continuing public health issue. Every year, hundreds of thousands of people die from heart disease-related diseases including coronary artery disease (CAD), congestive heart failure (CHF), and cerebrovascular disease (stroke).The timely detection and diagnosis of CVDs can prevent the loss of many lives. Nonetheless, the conventional means of diagnosing cardiovascular diseases involve multiple medical procedures, thus rendering it extremely difficult to identify any signs at an earlier stage.

In light of the recent advancements made in the field of artificial intelligence (AI) and machine learning (ML), healthcare providers have been increasingly relying on intelligent algorithms for predicting and diagnosing ailments. It is possible to handle large volumes of patient information, recognize subtle patterns using the advanced technologies available today. Deep learning has proven to be rather effective when it comes to medical diagnosis because it can independently discover complex connections within the data.

Ensemble models based on deep learning have been utilized by a number of researches to overcome the challenges associated with the utilization of the above-mentioned traditional machine learning models like decision tree, random forest, logistic regression, and support vector machine. One challenge associated with all the traditional models listed above is that they often exhibit high levels of overfitting and poor generalization. Furthermore because of the unique characteristics of healthcare data, conventional machine learning algorithms have limitations on the amount and kinds of data they can process.

Recently, as AI and ML technologies have progressed, hospitals and other health systems have started using sophisticated models to assist in predicting and diagnosing diseases. This will assist in examining a large volume of patient data and finding correlations that would not otherwise be possible using conventional methods. Deep learning models have proven to be particularly helpful for diagnosing diseases because they can automatically identify complex patterns within given data, giving healthcare providers the ability to make much more accurate assessments of their patients' conditions than with traditional statistical approaches.

Many previous studies have used ML techniques such as logistic regression, decision trees, random forests and support vector machines to predict cardiovascular disease; however, these techniques have demonstrated several limitations due to issues such as overfitting, non-generalization and difficulty handling the vast variety of medical datasets that exist within healthcare today.Ensemble models have been suggested in order to minimize these issues by combining several models for increasing the predictive performance.

In this research, Deep-Ensemble Cardiovascular Disease Detection System will be proposed in order to increase prediction accuracy and provide a useful risk classification technique. In the proposed system, deep learning models such as CNN and Gated Recurrent Units(GRU) will be used to ensure feature relationship and sequential information processing.

Whereas current detection systems use binary classification models which only give the prediction whether the patient is affected by the disease or not, this research proposes multi-level risk prediction system that categorizes the affected people on the basis of risk level in four categories; namely, no disease, low risk, medium risk and high risk categories. This can allow healthcare professionals to take appropriate steps and enable individuals to avoid developing cardiovascular diseases in future.

Moreover, in order to facilitate usability, a Web-based application of the system has been proposed that enables individuals to enter their health parameters and receive cardiovascular disease predictions.

II. LITERATURE SURVEY

A number of authors have suggested the application of machine learning and deep learning approaches for predicting cardiovascular diseases. Algorithms like logistic regression, decision tree, random forest, and support vector machines have been used for detecting diseases. They yield reasonable accuracy but exhibit deficiencies when applied to complicated medical data.

Latest research employs deep learning techniques like CNN, RNN, and LSTM for achieving better results. There have been many ensemble learning techniques used to improve the accuracy of several different algorithms by combining their outputs into one useable prediction. Unfortunately, much of the current research in this area remains limited to binary classifications. This research improves existing work that has been conducted in this area by expanding the capabilities of these studies to include multi-level risk predictions and deploying them through a web-based interface.

III. PROPOSED SYSTEM

The above model incorporates a Deep-Ensemble Cardiovascular Disease Detection Model that aims to increase the precision rate and also give useful and clinically informative risk classification. In traditional models of prediction of heart diseases, only the binary approach was applied. That is the models used the two-class classifier to predict whether or not the patient has heart diseases.

However, there were limitations to this since this did not help in making informed decisions about prevention measures which can be adopted before the disease becomes severe. Therefore, this new system employs a multilevel categorical classification, whereby the patients will be classified under four categories: No Disease, Low Risk, Medium Risk and High Risk.

In this model, the convolutional neural network (CNN) and gated recurrent unit (GRU) will work together using ensemble learning techniques.The CNN model will be responsible for feature extraction, while the GRU model will capture dependency within the data set.

The proposed system's workflow starts with the collection of the dataset. The second step involves pre-processing of data, which is an important step before analysis of the data. After that, the process of feature extraction will be conducted, after which features can be used to train the deep learning model. Subsequently, the trained models will be combined using ensemble learning techniques in order to conduct multi-level risk classification using the outputs from the CNN and GRU models. The system will then be deployed via a web-based interface.

In general, the proposed solution not only boosts prediction performance but also facilitates usability.

IV. METHODOLOGY

This section covers the approach that we used to develop the deep-ensemble method that can classify

cardiovascular diseases. Our approach includes obtaining the data, preprocessing, developing a model and

testing its performance.

4.1 Dataset Description

The dataset for predicting cardiovascular diseases contains clinical variables such as demographics and vital

signs obtained from patients' medical history. In particular, the dataset provides the following information: age, gender, blood pressure, cholesterol levels, heart rate, type of chest pain, blood glucose test result, electrocardiogram result, exercise-induced angina, maximum heart rate achieved, slope of the peak exercise ST segment, major vessel numbers and thalassemia presence. Through a deep learning model, one can predict the probability of developing cardiovascular diseases from these features.

4.2 Data Preprocessing

Data preprocessing is an essential step not only for machine learning but also for deep learning. The initial dataset may have some problems such as the absence of values, duplicates, and inconsistency. These issues should be fixed via data preprocessing.
For data preprocessing, we removed all the absent values, duplicates, performed text to numeric transformation of categorical values, normalized the data, and scaled features. Scaling the features improves the performance of the model and speeds up the training. Afterward, the data set was split into two parts – one for training and another for model evaluation.

4.3 Deep Learning Model

Our system will use several deep learning algorithms.
Among them, there will be a Convolutional Neural Network (CNN). It is the stage of pattern recognition when the model extracts relevant patterns and associations between different input variables. Thus, conducting the procedure with CNN improves the overall performance of the model and reduces the amount of required data.
A similar technique used is Gated Recurrent Unit which is a recurrent neural network, used in learning from the information provided in sequential order. The unit can efficiently discover associations among the various features present in data, hence improving the accuracy of forecasts.

4.4 Ensemble Learning

This technique combines predictions made by the Convolution Neural Network and GRU model into one prediction. Essentially, using several models to make predictions and then combining them using ensemble learning enhances the performance of the whole system.
Using ensemble learning, we obtain more accurate results; it minimizes the risk of making the model biased towards the training set (hence increasing generalization), broadens the coverage of the model and finally we obtain more dependable results.

V. SYSTEM ARCHITECTURE

The new system will function according to a set procedure that will involve several steps to ensure accurate and efficient prediction of the risk of developing cardiovascular diseases. The first step in this process would be the provision of patient data to the proposed system as an input. The data will then be preprocessed to handle any missing values, standardize the input data, and prepare the data for further use. Feature scaling will then take place to optimize the prediction model performance and unify input variables. Preprocessed data will then be simultaneously passed through the CNN and GRU prediction models, and relevant patterns will be extracted during the process. Outputs from the two models will then be combined using the ensemble technique to enhance prediction efficiency and eliminate prediction errors. The final output of the classified risk level will then be generated and presented to users through the web application interface.

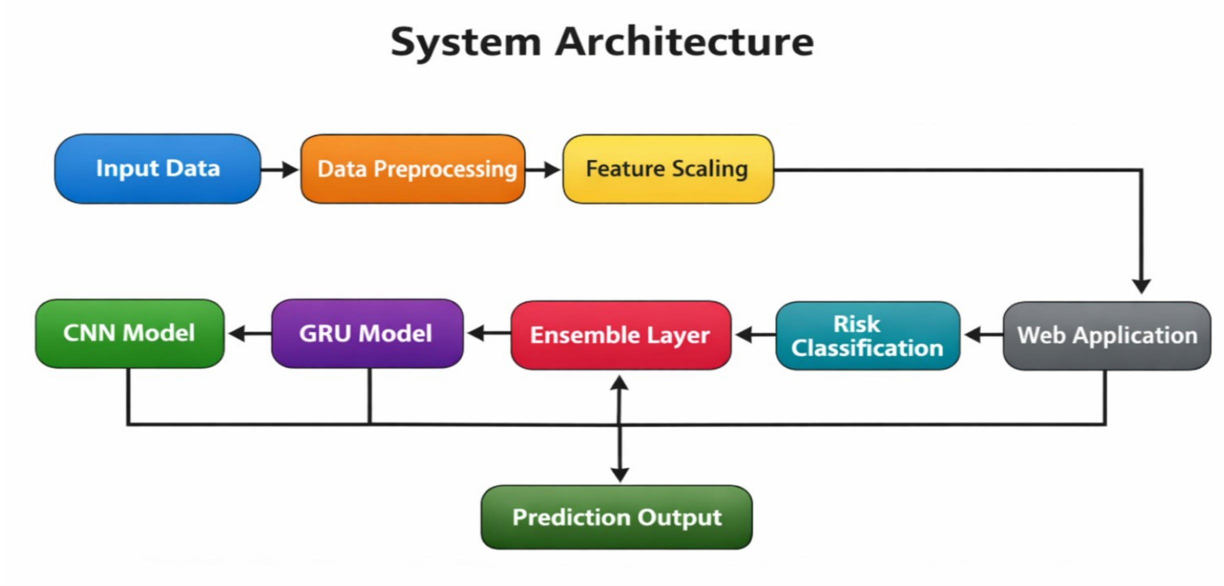

Fig 1: System architecture of Deep-Ensemble Cardiovascular Disease Detection System

VI. IMPLEMENTATION

The cardiovascular disease detection system proposed was built utilizing Python programming language and deep learning development framework. The decision to utilize Python as the primary programming language was based on several factors including that it is easy to use, flexible, and has many powerful tools available for developing machine learning/deep learning models.

Data processing, model training, evaluation, and deployment of the trained model to the web application are all steps in developing the system.

In the early development stages of the cardiovascular disease detection system, the dataset was imported and cleaned using various tools (such as Pandas and Numpy). Both Pandas and Numpy were part of the process of manipulating and cleaning data as well as transforming data into usable forms for training purposes.

The methods employed to prepare data for training purposes included filling in missing data, encoding categorical features, and scaling features.

The deep learning model was designed and implemented using TensorFlow and Keras software frameworks. In these software frameworks, deep ensemble was designed by incorporating the CNN and GRU architectures. While the CNN architecture was used to derive features from the data set, GRU was used to find relations between these features. Finally, the outputs from both models were combined using the idea of ensemble to achieve more precise results and minimize bias of the model.

Finally, after implementing the proposed model, performance assessment was done using various performance metrics such as accuracy, precision, recall, and F1 score. The purpose of the deployment of the trained model is to evaluate how the proposed system performs.

To enable real-world utilization of the model, it has been deployed through a web-based application constructed on Flask. Flask is a software framework that connects a website's backend (trained model) with its frontend (user-interface). The user-interface is built using HTML and CSS to provide a user-friendly experience. Users on the frontend will enter their health indicators into the website, and those indicators will be used to provide real-time predictions of risk for developing cardiovascular disease via the website.

The development of this project has resulted in effective data processing, accurate prediction results, and timely delivery of the prediction results. The combination of Python, Tensorflow, Keras, Pandas, NumPy, Flask, HTML, and CSS has created a reliable and extensible platform for detecting and classifying cardiovascular diseases.

VII. RESULTS AND DISCUSSION

The performance of the suggested deep-ensemble system for detecting cardiovascular diseases was tested based on certain evaluation criteria to test the efficiency and efficacy of the model. The evaluation criteria that will be employed in this case include accuracy, precision, recall, and F1 score. The use of such criteria ensures an objective assessment of the classifier performance and helps to make sure that the predictions generated by the model are reliable enough.

Accuracy will be used as a measure of the efficiency of the model in terms of its ability to classify patients based on the predicted probability of their having cardiovascular diseases. Precision will be employed in the evaluation to test the ratio of true positive predictions, whereas recall will show the efficiency of the model in detecting real positive cases.

The suggested method categorizes the risk of cardiovascular diseases into four classes namely no disease, low-risk disease, moderate-risk disease, and high-risk disease. The four-level categorization gives a deeper understanding of the medical condition of the patient as compared to other two-level categorization approaches. The multi-class risk prediction can help healthcare professionals diagnose the patients at an early stage and suggest precautionary steps before the disease develops further.

According to the experiment results, the suggested deep-ensemble model outperformed the traditional machine learning algorithms in terms of performance. The hybrid combination of CNN and GRU models improved feature extraction and learning abilities, which helped to increase the accuracy rate of predictions.
Additionally, the findings suggest that the designed model can significantly minimize the error rate and increase early identification of cardiovascular diseases. The multi-level risk evaluation and increased accuracy levels make it appropriate to use the designed system in real-time medical applications. As a conclusion, the designed deep-ensemble system of cardiovascular disease detection is a dependable and effective system.

VIII. WEB APPLICATION

A web-based application has been created in order to facilitate access and enable usability of the cardiovascular disease prediction system. Web application offers an interface where users can enter the health parameters of a patient and get the predicted cardiovascular disease risk levels. This will create a link between the trained deep learning model and the actual users who include doctors, patients, and other researchers in the field of medicine.

Web application includes an intuitive interface where users are able to fill the required information regarding the patient's health. These health parameters include age, gender, blood pressure, cholesterol level, chest pain types, heart rate, among others. After filling the form with the necessary data, the user submits the information that is then received by the backend of the application, which uses the trained deep ensemble model to generate results.

The outcome of the system is shown on the web page through categorical risk levels, namely No Disease, Low Risk, Medium Risk, and High Risk. This multilevel categorization would help users comprehend the current health condition related to their heart better and offer appropriate suggestions. Users can get the prediction results instantly.

The design of the web app was aimed at simplifying the process of using it for people who lack technical skills. The system makes the prediction platform more accessible since users have an opportunity to gain access to it from any device. Web-based system contributes to early prevention of cardiovascular diseases and provision of preventive health care due to its rapid predictions.

In general, the web app makes the prediction system usable due to several factors. First, users can obtain instant results. Second, the web app provides high accessibility and easy usability due to its interface.

CONCLUSION

To sum up, the proposed Deep-Ensemble Cardiovascular Disease Detection System offers a practical and dependable approach to cardiovascular disease prediction at initial stages. The use of CNN and GRU combined allow for the learning of complex patterns from health information of patients, thus improving the accuracy of prediction. Unlike the traditional binary classification of cardiovascular disease or no disease, this study predicted cardiovascular disease at four levels, namely No Disease, Low Risk, Medium Risk, and High Risk Diseases, helping healthcare professionals take necessary preventive action early.

In addition, the system's usability is enhanced by the integration of a web-based application through which users can provide health parameters for prediction swiftly. The experimental results show that the proposed model outperforms, has less error rate, and is more reliable than traditional Machine Learning techniques. Finally, the system can be used to assist doctors with early diagnosis, help patients determine their health status and reduce the mortality rate caused by cardiovascular diseases through timely intervention and preventive measures.

ACKNOWLEDGEMENT

Our special thanks go to **Mrs. V. Rashmi, M.Tech, PhD** and **Assistant Professor,Information Techonology, Prasad V Potluri Siddhartha Institute of Technology** for her encouragement and help to us during our study. She has always been very supportive of us throughout our study by her valuable suggestions and support. Her support to us and positive suggestions made us complete our study.

We wish to thank **Department of Information Technology, Prasad V Potluri Siddhartha Institute of Technology** for providing us all necessary infrastructure facilities to complete our study.

Authors would also like to thank the management of **Prasad V Potluri Siddhartha Institute of Technology** for their help and support.

REFERENCES

[1] "Heart Disease Prediction Using Novel Ensemble and Blending Based Cardiovascular Disease Detection Networks,"IEEE Access, Research Article ,1st July 2024

[2] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer,"SMOTE: Synthetic Minority Over-sampling Technique,"Journal of Artificial Intelligence Research, vol. 16, pp. 321–357, 2002.

[3] H. He, Y. Bai, E. A. Garcia, and S. Li,"ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning,"IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1322–1328, 2008.

[4] H. Han, W. Y. Wang and B. H. Mao,"Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning,"International Conference on Intelligent Computing, pp. 878–887, 2005.

[5] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner,"Gradient-Based Learning Applied to Document Recognition,"Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[6] K. Cho et al.,"Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation,"EMNLP, pp. 1724–1734, 2014.(Reference for GRU)

[7] T. G. Dietterich,"Ensemble Methods in Machine Learning,"
International Workshop on Multiple Classifier Systems, pp. 1–15, 2000.

[8] D. H. Wolpert,"Stacked Generalization,"Neural Networks, vol. 5, no. 2, pp. 241–259, 1992.(Reference for Blending/Stacking)

[9] UCI Machine Learning Repository,"Heart Disease Dataset,"Available: https://archive.ics.uci.edu

[10] J. Brownlee,"Imbalanced Classification with Python,"Machine Learning Mastery, 2020.

[11] World Health Organization (WHO),"Cardiovascular Diseases (CVDs) Fact Sheet,"Available: https://www.who.int

[12] M.F.Nazirkhan,"Heart Disease Dataset,"Kaggle,2021.
https://www.kaggle.com/datasets/mfarhaannazirkhan/heart-dataset