PREDICTION OF CARDIAC DISEASE USING MACHINE LEARNING ALGORITHMS

Neelu Kumari K.S[1].,Anushka D.V[2].,Kalaiarasi S[3]., Umavathi.N[4].,
[1]*Associate professor,Department of ECE,*
E-Mail: neeluvijay@gmail.com
[2,3,4]*UG Scholar, Department of ECE,*
E-Mail: anuvijay3777@gmail.com, kalaiarasi2708@gmail.com,
umavathi8900@gmail.com
P.A College of Engineering and Technology, Pollachi, Coimbatore, India.

Abstract

Healthcare provision is amongst humanity's most formidable challenges. Diseases of the heart and blood vessels are major killers worldwide. Researchers will need to create a mechanism to detect this cardiac irregularity and stop the spread of heart problems, which are on the rise. In this study, we improved the accuracy of cardiovascular disease prediction by using a Cardiac MRI picture and other advanced image processing techniques. It aids in the diagnosis of cardiovascular illness and enhances the assessment of heart conditions in patients. To back up our assertions, we analysed the performance parameters of many machine learning algorithms, including their accuracy, precision, f1-score, and recall.

Keywords: Cardiac Magnetic Resonance (CMR) Imaging, segmentation, deep learning, U-net, Convolutional Neural Networks(CNN),Random Forest.

1. INTRODUCTION

The primary focus of this research is on the application of machine learning techniques to the problem of predicting cardiovascular disease. There is no more vital a part of a human body than the heart. Basically, it regulates the circulation of blood throughout our body. Any kind of cardiac problem may lead to discomfort elsewhere in the body. Any disturbance in the heart's normally rhythmic pumping action is considered cardiac disease. Heart disease has emerged as a major global killer in recent decades. Tobacco use, excessive alcohol use, and a diet heavy in saturated fat have all been linked to an increased risk of cardiovascular disease [1]. More than 10 million people worldwide die annually from heart disease, according to the World Health Organization. Providing high-quality services and correct diagnoses is the major problem in modern healthcare [2]. Cardiovascular disease may be the biggest cause of mortality globally, yet it is also a condition that can be treated and controlled. The effectiveness of medical treatment relies entirely on early diagnosis. Machine learning (ML) is a kind of data mining that makes quick work of very big, properly organised information. Many medical conditions may be diagnosed, detected, and predicted with the use of machine learning. This effort is motivated by a desire to provide medical professionals with a method for early diagnosis of heart illness [3]. Therefore, patients will get better care with fewer negative side effects.

2. LITERATURE REVIEW

Avinash Golande et al. look at several ML algorithms that might be used to categorise cardiac diseases. K-Nearest Neighbor (KNN) and K-Means (K-Means) classification accuracy was studied [1]. This research found that the decision tree was the most accurate tool, and that it could be improved upon by incorporating insights from other approaches and playing with the settings. Common cardiac image segmentation tasks, including examples using MRI, CT, and ultrasound. Model-based approaches (e.g., active shape and appearance models) and atlas-based methods were demonstrated to perform well in cardiac image segmentation before the advent of deep learning (5–7). Lord Krishnan S. J. Geetha S [8] created a system to predict the consequences of cardiovascular disease. The results of this system raise the risk of cardiovascular disease. Similarity in data structure to treatment parameters was taken into account. The information mining plan methodology is used to analyse these criteria within their framework. Both the explicit Decision Tree Algorithm and the Naive Bayes Algorithm, which are considered to be the two gold standards of machine learning, were used to build up their datasets in Python, and their results showed that the former provided the more accurate prediction of heart disease.

3. PROPOSED WORK

The series of steps involved in our proposed model are as follows: Data preprocessing,splitting datasets, Feature Extraction and Selection, Disease Prediction using Machine Learning algorithms as shown in fig.1.

3.1 DATA PREPROCESSING

Data preprocessing is a required task for cleaning the data and making it suitable for a machine learning model, which improves the model's precision and efficiency. Here's what you need to do: 1.Read picture, 2. In order to do step 2 (Image Resizing), Third, silence the background chatter (Denoise), 4 Segmentation 5 Morphology (smoothing edges).

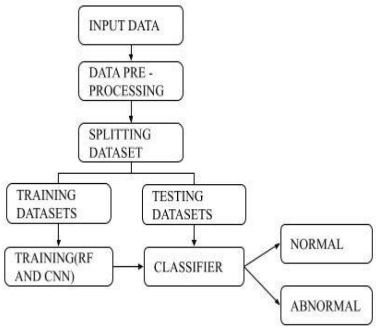

Fig 1 Flowchart for Proposed Work

3.2 CONVOLUTIONAL NEURAL NETWORK (CNN) ALGORITHM:

Convolutional neural networks (ConvNets) are a kind of deep learning network that can take an image as input, prioritise (using learnable weights and biases) certain qualities or objects in the picture, and then extract meaningful results. When compared to ConvNet, other classification methods need extensive preprocessing. ConvNets, given sufficient training, can learn these filters and properties, but essential approaches still need human intervention.

ARCHITECTURE OF CNN:

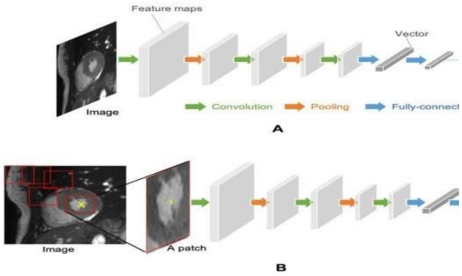

Fig.2 Architecture Of CNN

CNN receives an MR picture of the heart and uses a series of convolutions and pooling operations to learn the image's structural properties. These maps of spatial features are transformed to a vector form by flattening them using fully linked layers. Depending on the situation, this vector might take on a variety of forms. As can be seen in fig.2, these outputs may be probabilities for a collection of classes (image classification), bounding box coordinates (object localization), a projected label for the input's centre pixel (patch-based segmentation), or an actual value (regression issues).

3.3 RANDOM FOREST ALGORITHM:

The Random Forest machine learning technique is widely used for supervised learning. Both classification and regression issues may be tackled with the help of machine learning. The model's superior performance may be attributed to its foundation in ensemble learning, a technique for integrating several classifiers to address difficult problems. The accuracy of a dataset may be improved by using a classifier called a Random Forest, which takes an average of the decisions made by many decision trees applied to various subsets of the dataset. The random forest takes the predictions from many different trees and uses the ones with the most votes to determine the final result.

ARCHITECTURE OF RANDOM FOREST

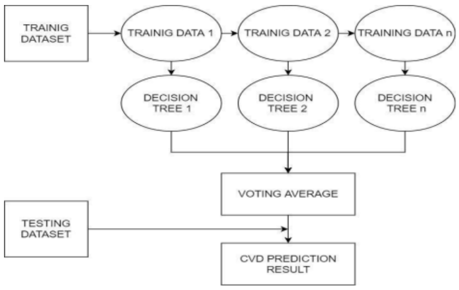

Fig 3 Block Diagram for Random Forest Classifier

Many different decision tree models compose the random forest. Each tree in fig.3 was constructed using bootstrap samples from the training dataset, and each node was partitioned using a randomly selected optimum subset of explanatory variables or features. The model prediction was chosen using a majority vote based on the class predictions produced by each tree in the forest. By randomly assigning samples to nonoverlapping training samples for building the prediction model and testing samples for determining sensitivity, specificity, accuracy, positive predictive value, and negative predictive value of prediction models, we were able to evaluate the performance of the random forest prediction model.

3.4 PERFORMANCE EVALUATION

Confusion matrix, accuracy score, precision, recall, sensitivity, and the F1 score are all employed in the assessment procedure. Confusion matrices are tabular representations of data that include both actual and expected (or true positive and true negative) outcomes. True positive (TP) is the first of four components of the definition, and refers to cases when the values have been recognised as true and are, in fact, true. The second kind is known as a false positive (FP), and it occurs when erroneous results are incorrectly labelled as true. The third kind is known as a false negative (FN), in which a valid result was incorrectly labelled as negative. In the fourth case, called true negative (TN), the value was really negative. Figure 4 displays the data table.

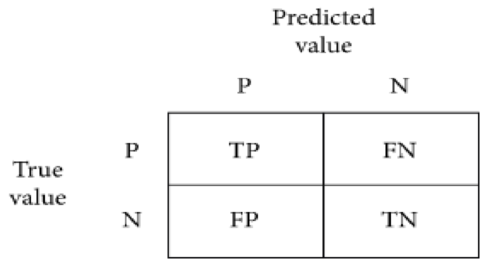

Fig 4 Confusion Matrix

In Figure 4, P = positive, N = negative, TP = true positive, FN = false negative, FP = false positive, TN = true negative.

Then, an accuracy score is utilised to evaluate a model's performance. It is calculated by dividing the sum of all true positive and true negative values by the sum of all true positive and all false negative values. The equation is

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Then there is sensitivity in which the proportion of actual positive cases got predicted as positive (or true positive). Sensitivity is also termed as recall. In other words, an unhealthy person got predicted as unhealthy. The formula is

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

4. RESULTS AND DISCUSSION

The classification's accuracy was measured by contrasting the anticipated with the actual type. The suggested approach analysed the dataset using the four classifiers shown in fig.5 to achieve its categorization goal. As can be seen in fig.5, the highest accuracy (99.98%) is achieved by the Random Forest tree classifier. Table II displays the agreement rates between training and testing. The Random Forest tree classifier was employed because it achieved the best levels of accuracy. We compare the accuracy of predictions made by Convolutional Neural Networks with the Random Forest approach for cardiovascular disease.

METHOD	DATARATIO	ACCURACY	SENSITIVITY	PRECISION
SVM	50:50	57%	38%	43%
KNN	50:50	75%	54%	63%
CNN	50:50	87%	90%	72%
RF	50:50	99.98%	92%	85%

Fig. Table for Performance Evaluation

5. CONCLUSION

Many of today's algorithms are not well suited to making reliable CVD disease predictions. Even medical professionals have trouble making accurate prognoses of the situation. So, the suggested strategy helps doctors with their predictions. In this script, we used convolutional neural networks and Random Forest to create a model for predicting the onset of disease. The research also compared the suggested method to the state-of-the-art algorithms. The suggested model consists of two convolutional layers, two dropout layers, and one output layer. The proposed network will have to process a large amount of data. Unlike the prior technique, which required separate solutions for preprocessing, feature extraction, and prediction, the present one does all three in a unified fashion.

6. FUTURE SCOPE

Findings from evaluations may be enhanced via the use of machine learning and other optimization techniques. It is possible to use several techniques for data normalisation and then compare the outcomes. For the benefit of both patients and medical professionals, further methods might be investigated for integrating ML and DL models trained on heart disease with specialised multimedia.

7. REFERENCES

[1] T.Nagamani, S.Logeswari, B.Gomathy, ‘Heart Disease Prediction using Data Mining with Mapreduce Algorithm’, International Journal of Innovative Technology and Exploring Engineering (IJITEE) , Volume-8 Issue-3, January 2019.

[2] Avinash Golande, Pavan Kumar T, ‘Heart Disease Prediction Using Effective Machine Learning Techniques’, International Journal of Recent Technology and Engineering, Vol 8, pp.944-950,2019.

[3] Theresa Princy R,J. ‘Thomas,Human heart Disease Prediction System using Data Mining Techniques’, International Conference on Circuit Power and Computing Technologies,Bangalore,2016.

[4] Fahd Saleh Alotaibi, ‘Implementation of Machine Learning Model to Predict Heart Failure Disease’, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 6, 2019.

[5] Greenspan H, Van Ginneken B, Summers RM. ‘Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique’. IEEE Trans Med Imaging. p.p 35:1153,vol.9,2016.

[6] Shen D, Wu G, Suk HI. ‘Deep learning in medical image analysis’, Annu Rev BiomedEng.,p.p19:221,vol.48 ,2017.

[7] R.Bharathi, T.Abirami,” Energy Efficient Clustering with Disease Diagnosis Model for IoT based Sustainable Healthcare Systems”, Sustainable Computing: Informatics and Systems, 23 September 2020, https://doi.org/10.1016/j.suscom.2020.100453

[8] S.Krishnan J. Geetha S, ‘Forecast of Heart Disease Using Machine Learning Algorithms,First International Conference on Innovations in Information and Communication Technology ,2019.

[9] Nabaouia Louridi , Meryem Amar ,Bouabid El Ouahidi, ‘Identification of Cardiovascular Disease using Machine Learning Algorithms’, 7th International Mediterranean Congress of Telecommunications,2019.

[10] Sonam Nikhar, A.M. Karandikar ‘Prediction of Heart Disease Using Machine Learning Algorithms’ in International Journal of Advanced Engineering, Management and Science (IJAEMS) , vol-2, June-2016.

Biographies

K. S. Neelu Kumari has received a Ph.D. degree in the field of VLSI from Anna University in 2016, M.E. Applied Electronics from Kongu Engineering College in 2006 and B.E Electronics and Instrumentation Engineering from Bharathidasan University in 2002. She is currently working as Associate professor, department of Electronics and Communication Engineering in P. A. College of Engineering and Technology. Her research interests include VLSI, Nanotechnology and machine learning.

D.V.Anushka, S.Kalaiarasi, N.Umavathi has a student ,UG Degree Scholer in the field of Electronics and Communication Engineering in P.A College of Engineering and Technology.