AN AUTOMATED DIAGNOSTIC SYSTEM FOR HEART DISEASE PREDICTION BASED ON CNN-LSTM

Ashwini L. Mahalkar
Department Of Computer Science & Engg.
Deogiri Institute Of Engg.& Management Studies, Aurangabad.

Prof. Smita S. Ponde
Department Of Computer Science & Engg.
Deogiri Institute Of Engg.& Management Studies, Aurangabad.

Abstract: in this paper a heart disease prediction model for the prediction of occurrence of heart disease is presented. It aimed towards identifying the best classification algorithm for identifying the possibility of heart disease in a patient. This work is justified by performing a comparative study and analysis using three deep learning classification algorithms name CNN, DNN and LSTM are used at different levels of evaluations. This will provide researchers and medical practitioners to establish a better understanding and help them identify a solution to identify the best method for predicting the heart diseases. Results show that LSTM is better among three having accuracy of 95% in prediction of heart disease.

Keywords: LSTM, CNN, Prediction, Heart Disease, feature extraction

I. INTRODUCTION

Present-days heart disease is treated as one of the major cause of human death in the world. 10% of the total death occurs in the world is due to heart disease only. Hence the disease has become one of the biggest concerns in various countries of the world. As per University of Rochester's Medical centre view the main source for heart condition are overweight, lack of physical activity, fatness, consumption of malnutrition and tobacco. As heart condition is widely accepted because the major source of death hence medical analysis of heart condition becomes a daily need for each person ..

The most complicated and sophisticated task within the field of medical sciences is that the prediction of heart condition. Heart is considered to be the most vital organ of the human body. There is an intense need in predicting the extent and seriousness of heart condition that provide an accurate treatment to the patients. Heart diseases are often mentioned various conditions that cause abnormal functioning of heart, which can involve blood vessels, arteries etc. Effective diagnosis of heart disease results in an appropriate treatment to a patient. This requires a deep study of cardiovascular analysis of the patient that includes symptoms such as chest pain chest tightness, chest pressure, and discomfort in breathing, numbness etc. The cardiovascular diagnosis involves certain decisions to be taken based upon the health history and therefore the clinical trial results of an individual . The process of decision making is a challenging task to the medical practitioners which has to be done accurately and efficiently where a mere negligence may lead to the life risk of a patient.

Heart diseases are often managed effectively with a mixture of lifestyle changes, medicine and, in some cases, surgery. With the proper treatment, the symptoms of heart condition are often reduced and therefore the functioning of the guts improved. The predicted results are often wont to prevent and thus reduce cost for surgery and other expensive. The overall objective of my work are going to be to predict accurately with few tests and attributes the presence of heart condition . Attributes considered form the first basis for tests and provides accurate results more or less. Many more input attributes are often taken but our goal is to predict with few attributes and faster efficiency the danger of getting heart condition. Decisions are often made supported doctors' intuition and knowledge instead of on the knowledge rich data hidden within the data set and databases. This practice results in unwanted biases, errors and excessive medical costs which affects the standard of service provided to patients. Data mining holds great potential for the healthcare industry to enable health systems to systematically use data and analytics to spot inefficiencies and best practices that improve care and reduce costs. According to the opportunities to improve care and reduce costs concurrently could apply to as much as 30% of overall healthcare spending. The successful application of knowledge mining in highly visible fields like e-business, marketing and retail has led to its application in other industries and sectors.

Deep learning may be a more popular machine learning method. It is not only when applying it in image classification tasks but also uses normal tabular data.[4] The main objective of this work is to identify the key patterns and features from the medical data using the classification algorithms and then to select the most relevant attributes for silent heart attack diagnosis. The use of CNN-LSTM will further enhance the accuracy of the results.

II . LITERATURE REVIEW

Several experiments are conducted on medical data sets using multiple classifiers and features selection techniques.

There's little research on the classification of the heart disease dataset. Many of them show good classification accuracy a number of them are given in table 1

L.Ali et al., [1] proposed an automatic diagnostic system for the diagnosis of cardiovascular disease. The proposed diagnostic system used χ 2 statistical model for features refinement and DNN for classification. The strength of the proposed diagnostic system was evaluated using six different evaluation metrics including accuracy, sensitivity, specificity, MCC, AUC and ROC charts. Moreover, the performance of the proposed method was compared with other well-known machine learning models and with other methods discussed within the literature. From the experimental results, we will safely conclude that the proposed diagnostic system can improve the standard of deciding during the diagnosis process of heart condition. The proposed method achieved higher detection accuracy for HF disease, but this study didn't investigate the time complexity of the proposed hybrid diagnostic system. In future studies, it'll be investigated because it is taken into account a very important think about clinical application.

Yogita solanki et al., [2] presents various issues associated with healthcare & different machines learning algorithms the target of this research was to form a model for heart condition detection that deals with non-linear features, an enormous amount of knowledge and supply more accurate result than existing research. The dataset is trained by Deep Neural Network, this neural network have 13 attributes as a input and it represents 3 hidden layers during which first layer have 100 neuron from which 40% neuron are dropout for preventing over fitting, second layer have 50 neurons with dropout of 30% and third hidden layer have 20 neurons with dropout of 20with keras package then validate the model. This deep learning with keras package has increased the speed of prediction. This model performs better than the prevailing approach with 85.72% precision, 88.24% recall, 12.1% loss and 89.15% accuracy.

S.Nandini et al., [3] during this paper, they need presented a system which is suitable for real-time heart diseases prediction and may be utilized by the users who have coronary disease. Different from many other systems it's ready to both monitor and prediction. The diagnosis system of the system is ready to predict the center disease by using ML algorithms and therefore the prediction results are supported the center disease dataset instance. On the opposite hand, the system is extremely inexpensive, we used amped pulse sensor and send the information to mobile via Arduino suite microcontroller. For checking the variances and lift the alarm if the user's pulse rate rise than the traditional rate of the heart. To prove the effectiveness of the system we've got distributed experiments for both monitoring and diagnosis system . we ran experiments with some popular algorithms like KNN, Decision Tree, Random Forest, Naive Bayes, SVM, Logistic Regression. The experiment was dispensed with the holdout test and therefore the accuracy of the proposed system was 89% achieved with the Random forest.

Sumit et al., [4] during this paper, they organized in a logical sequence that usually follows your methodology section. Compare to alternative algorithms and improvement, it's tried smart results for prediction. During this paper, we tend to deploy a deep learning neural networks (DNN) exploitation Talos improvement. Talos improvement is fresh improvement techniques in DNN. Talos give higher accuracy (90.76%) to alternative optimizations. It's applied on the guts malady datasets and ascertains the great prediction. Exploitation the Talos improvement they produce a Keras model and deploy it.

Anjan et al., [5] Data collection is applied using numerous sources that are primary factors answerable for any kind of cardiovascular disease and thereby employing a structure the database is made. The research focuses on establishing SHDP (Smart cardiovascular disease Prediction that takes into consideration the approach of NB (Naive Bayesian) classification and AES (Advanced Encryption Standard) algorithm for resolving the problem of cardiopathy prediction. it's revealed that in relevance accuracy, the prevailing technique surpasses the Naive Bayes by yielding an accuracy of 89.77%in spite of reducing the attributes. AES yields in high security performance evaluation as compared to PHEA (Parallel Homomorphic Encryption Algorithm).

Vincy et al., [6] they developed decision network using Naive Bayesian classification algorithm and Laplace smoothing technique. The system predicts whether a patient have heart disease or not. Laplace smoothing technique makes more accurate results than Naive Bayes alone to predict patients with cardiovascular disease. The system is expandable within the sense that more number of records or attributes are often incorporated. Presently the users can use either 13 attributes prediction or 6 attributes prediction if they don't know the results of fluoroscopy test, thallium test etc. But the one with 13 attribute is more accurate since it's 86% accuracy. The system can even incorporate other data processing techniques for prediction. this technique can function a training tool to teach nurses and medical students to diagnose patients with cardiovascular disease .

S. Mohan et al., [7] involved to find the numerous features using machine learning techniques to extend the accuracy of heart condition prediction. the mixture of features was applied within the standard classifiers to live the performance of the prediction. The hybrid random forest with a linear model (HRFLM) provided higher accuracy within the center condition prediction. In pre-processing, the multiclass variables and binary classification were applied within the attributes of the dataset to test the presence or absence of cardiopathy. The age is a crucial factor for cardiopathy prediction, which has been neglected within the paper. From this literature survey, it's clear that the present methods have

some limitations. to deal with these limitations, a replacement deep learning-based approach is proposed to enhance the performance of heart condition prediction....

Bindhika et al., [8] In this paper, they proposed a method for heart disease prediction using machine learning techniques, these results showed a geat accuracy standard for producing a better estimation result. By introducing new proposed Random forest classification, they find the problem of prediction rate without equipment and propose an approach to estimate the heart rate and condition. Sample results of heart rate are to be taken at different stages of the same subjects, they find the information from the above input via ML Techniques. Firstly, they introduced a support vector classifier based on datasets.

Rupali Patil [9] Developed Decision Support in Heart Disease Prediction System using both Naive Bayesian Classification and Jelinek-mercer smoothing technique. The system extracts hidden knowledge from a historical heart condition database. Jelinek-mercer smoothing technique is that the simpler than naive bayes to predict patients with heart condition. This model could answer complex queries, each with its own strength with simple model interpretation and a simple access to detailed information and accuracy. The system is expandable within the sense that more number of records or attributes is often incorporated and new significant rules are often generated using underlying data processing technique. Presently the system has been using 13 attributes of medical diagnosis. It also can incorporate other data processing techniques and extra attributes for prediction.

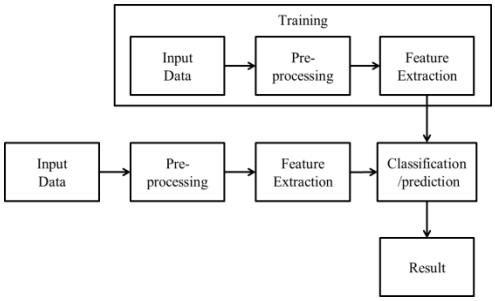
Table 1 techniques applied for prediction of heart disease

Year	Author	Database	Tool	Feature extraction algorithm	Classification and prediction algorithm	Accuracy
2019	Ali et al [1]	Cleverland heart disease dataset	MATLAB	x^2statistical model	DNN	93.33%
2018	Yogita et al [2]	Cleveland database of UCI repository,	Python	ANN	DNN	89.50%
2018	S.Nandhin et al [3]	Cleveland database of UCI repository, MongoDB	Python, WAN	-	Decision tree	75.82
					Random forest	89.0%
					SVM	87.91%
2020	Sumit et al [4]	Heart Disease UCI dataset	WEKA	DNN	Hyper parameter optimization(Talos)	90.78%
2019	Repaka et al [5]	UCI dataset	CAD	AES	Naïve Bayes	89.77%
2017	Vincy Cherian et al [6]	Cleveland Heart Disease database	-	Laplace Smoothing	Naïve Bayes	86%
2019	Mohan et al [7]	UCI dataset	R Studio Rattle	Linear method + Random Forest	hybrid HRFLM	88.7%
2020	Bindhika et al [8]	UCI repository	Python	-	hybrid random forest with a linear model	92%
2014	Rupali Patil [9]	Cleveland Heart Disease database	MATLAB	Jelinek-Mercer (JM) smoothing	Naïve Bayes	78%

III. PROPOSED SYSTEM

The heart disease prediction has been carried out by many existing methods which are based on machine learning and clustering methods, but still those existing methods lack in efficiency. The objective of this research is to improve the efficiency of heart disease prediction using deep learning method. The CNN-LSTM method analyses the data in two

directional ways that tend to improve the performance of prediction. The LSTM is a deep learning method that analyses the data effectively and retrieves the key feature which are required for prediction. Fig.1 shows the block diagram of the proposed method



.

Fig 1 Block diagram of proposed system

the preprocessing block denotes the normalization process performed on feature vectors. After normalization, data is partitioned into training and testing sets. To avoid bias, feature ranking and selection is performed on training data.. The training data with reduced features is applied to neural network for training purpose. The performance of the trained neural network is evaluated using testing data. Features are extracted from pre-processed data. The extraction and matching of features is based on these measures. Besides the simple point feature a more advanced type of feature is also presented. Feature extraction technique is used to extract the features by keeping as much information as possible from large set of data .

Classification is a process of categorizing a given set of data into classes; it can be performed on both structured and unstructured data. The process starts with predicting the class of given data points. The classes are often referred to as target, label or categories. The classification predictive modeling is the task of approximating the mapping function from input variables to discrete output variables. The main goal is to identify which class/category the new data will fall into. Classification is performed using CNN- LSTM algorithm. Input to classifier is input features and trained data. Classifier compares input features with trained data and predicts heart disease.

IV. CONCLUSION

Heart disease is the leading cause of death for both women and men, so it requires an effective prediction technique for early diagnosis. Several data mining and machine learning techniques have been developed to improve the performance of the prediction. This research aimed to improve the performance of heart disease prediction using the CNN-LSTM model. The developed model attained highest precision value compared to other data processing techniques, because the proposed model can analyze data during a bi-directional manner and also investigates the linear relationship between the extracted features.

The proposed method achieved the classification accuracy of 95.04%, which is high compared to other techniques..

REFERENCES

[1] Liaqat Ali, Atiqur Rahman, Aurangzeb Khan, Mingyi Zhou, Ashir Javeed, Javed Ali Khan, "An Automated Diagnostic System for Heart Disease Prediction Based on x^2Statistical Model and Optimally Configured Deep Neural Network", Digital Object Identifier 10.1109/ACCESS.2019.2904800.

[2] Yogita Solanki1, Sanjiv Sharma "Analysis and Prediction of Heart Health using Deep Learning Approach",2019, International Journal of Computer Sciences and Engineering. Vol.7(8), Aug 2019, E-ISSN: 2347-2693.

[3] S.Nandhini, Monojit Debnath, Anurag Sharma, Pushkar, "Heart Disease Prediction using Machine Learning", .International Journal of Recent Engineering Research and Development (IJRERD) ISSN: 2455-8761 Volume 03 – Issue 10, October 2018, PP. 39-46 39

[4] Sumit Sharma, Mahesh Parmar, "Heart Diseases Prediction using Deep Learning Neural Network Model", International Journal of Innovative Technology and Exploring Engineering (IJITEE) Volume-9 Issue-3, January 2020.

[5] Repaka, A. N., Ravikanti, S. D., & Franklin, R. G., "Design And Implementing Heart Disease Prediction Using Naives Bayesian", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). doi:10.1109/icoei.2019.8862604

[6] Vincy Cherian, Bindu M.S., "Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace Smoothing Technique", International Journal of Computer Science Trends and Technology (IJCST) – Volume 5 Issue 2, Mar – Apr 2017 Page 68

[7] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). "Effective Heart Disease Prediction using Hybrid Machine Learning Techniques", IEEE Access, 1–1. doi:10.1109/access.2019.2923707

[8] Galla Siva Sai Bindhika1, Munaga Meghana, Manchuri Sathvika Reddy, Rajalakshmi, "Heart Disease Prediction Using Machine Learning Techniques", International Research Journal of Engineering and Technology (IRJET) Volume: 07 Apr 2020 Page 5272

[9] Ms. Rupali R. Patil, "Heart Disease Prediction System using Naive Bayes and Jelinek-mercer smoothing", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2014, 6787