DISEASE DETECTION USING MACHINE LEARNING IN HUMAN BEINGS

[1]Dr. R. Rambabu, [2]Dr. D. Naga Purnima, [3]G. Swarnalatha

[1]Professor & HOD, Department of Computer Science & Engineering, Rajamahendri Institute of Engineering & Technology, Rajamahendravaram, A.P, India
[2]Professor& HOD, Department of Mathematics, Rajamahendri Institute of Engineering & Technology, Rajamahendravaram, A.P, India
[3]Associate Professor, Department of Computer Science & Engineering, Rajamahendri Institute of Engineering & Technology, Rajamahendravaram, A.P, India

ABSTRACT: Timely and precise evaluation of any health-related issue is crucial for both sickness prevention and treatment. When dealing with a significant issue, the conventional diagnostic approach might not be the most effective. A more accurate diagnosis than the traditional approach may be possible with the development of a medical diagnosis system based on Machine Learning (ML) algorithms for the prediction of any disease. We have used many ML algorithms in the creation of a disease prediction system. Over 230 illnesses were included in the dataset that was processed. The diagnosis system provides an output representing the possible disease that a person may have based on their age, gender, and symptoms. Out of all the methods, the weighted hybrid (Support Vector Machine and Decision Tree) algorithm produced the best results interms of accuracy and senstivity.

KEYWORDS: Machine Learning (ML), Disease Prediction, Symptoms

I. INTRODUCTION

Medicine and healthcare are some of the most crucial parts of the economy and human life. There is a tremendous amount of change in the world we are living in now and the world that existed a few weeks back. Everything has turned gruesome and divergent. In this situation, where everything has turned virtual, the doctors and nurses are putting up maximum efforts to save people's lives even if they have to danger their own [1]. There are also some remote villages which lack medical facilities. Virtual doctors are board-certified doctors who choose to practice online via video and phone appointments, rather than in person appointments but this is not possible in the case of emergency. Machines are always considered better than humans as, without any human error, they can perform tasks more efficiently and with a consistent level of accuracy. A disease predictor can be called a virtual doctor, which can predict the disease of any patient without any human error [2].

Also, in conditions like COVID-19 and EBOLA, a disease predictor can be a blessing as it can identify a human's disease without any physical contact. Some models of virtual doctors do exist, but they do not comprise the required level of accuracy as all the parameters required are not being considered. The primary goal was to develop numerous models to define which one of them provides the most accurate predictions[3]. While ML projects vary in scale and complexity, their general structure is the same. Several rule-based techniques were drawn from machine learning to recall the development and deployment of the predictive model. Several models were initiated by using various Machine Learning (ML) algorithms that collected raw data and then bifurcated it according to gender, age group, and symptoms. The data-set was then processed in several ML models like Fine, Medium and Coarse Decision trees, Gaussian Naive Bayes, Kernel Naive Bayes, Fine, Medium and Coarse KNN, Weighted KNN, Subspace KNN, and RUSBoosted trees. According to ML models, the accuracy varied. While

processing the data, the input parameters data-set was supplied to every model, and the disease was received as an output [4].

Machine Learning is the domain that uses past data for predicting. Machine Learning is the understanding of computer system under which the Machine Learning model learn from data and experience[5]. The machinelearning algorithm has two phases: 1) Training & 2) Testing. To predict the disease from a patient's symptoms and from the history of the patient, machine learning technology is struggling from past decades. Healthcare issues can be solved efficiently by using Machine Learning Technology. We are applying complete machine learning concepts to keep the track of patient's health. ML model allows us to build models to get quickly cleaned and processed data and deliver results faster. By using this system doctors will make good decisions related to patient diagnoses and according to that, good treatment will be given to the patient, which increases improvement in patient healthcare services[6]. To introduce machine learning in the medical field, healthcare is the prime example. To improve the accuracy of large data, the existing work will be done on unstructured or textual data. For the prediction of diseases, the existing will be done on linear, KNN, Decision Tree algorithm.

II. LITERATURE SURVEY

Tapas RanjanBaitharu, Subhendu Kumar PaniA et. al., [7] Pulmonary cancer is an illness marked by unchecked cell cycle in lung, and it is the most common cause of mortality for both men and women. In the process of KDD (knowledge Discovery in Data), clustering is a crucial step. There are numerous possible benefits for it. The learning sample has a significant impact on how well classifications work. As a result, categorization systems better in terms of forecast or diagnostic quality, need less

computational power to develop models .Since they pick up knowledge more quickly, and easier to comprehend. Using information on lung cancer in various settings, a quantitative examination of data categorization quality is offered. Comparing common algorithms prediction abilities numerically.

LilikAnifah, Haryanto, RinaHarimurti et. al., [8] Gray Level Co-event Matrices (GLCM) again on Artificial Neural System have been proposed for cancer detection. The Tumor image archive Collection, which is made up of CT images, that is used to obtain the pulmonary information. Imaging or before, extraction, edge detection, and identification of cancer progression using a three-layer neural network back-propogation. The outcome demonstrated that framework can more accurately distinguish among healthy lung tissue and lung cancer.

Lynch C M, Abdollahi B, Fuqua J D et. al., [9] automated training and testing classifiers have been used to forecast the lifespan of lung cancers.

Krishnaiah,V., Narsimha,G., Subhash Chandra,N et. al., [10] offered a strategy for almost detecting and correctly diagnosing the illness, aiding the physician in preserving the service user. The chance of a someone developing emphysema can be predicted utilizing common bowel cancer signals such age, race, whistling, chest tightness, and pain in the neck, chest, or arms.

K. C. Howlader, and A. Mazumder M. S. Satu, et al. [11] utilizing 116 records and 34 variables gathered from three Dhaka, Bangladesh, hospitals. Seven different methods were used to categorize the diseases.J48, Heoffding Tree, SMO, IBk, KStar, and Naive Bayes are among them. With a low accuracy of 70.83 percent,

naive Bayes performed the best in the developed framework.

Deeanna Kelley et.al [12] performed a work, to analyze and detect heart disease. In this the algorithm used was Naive Bayes algorithm. In Naïve Bayes algorithm they used Bayes theorem. Hence Naive Bayes has a very power to make assumption independently. The used data-set is obtained from a diabetic research institutes of Chennai, Tamilnadu which is leading institute. There are more than 500 patients in the dataset. The tool used is Weka and classification is executed by using 70% of Percentage Split. The accuracy offered by Naive Bayes is 86.419%.

Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador, et.al [13] described the Prediction for similarities of disease by using ID3 algorithm in television and mobile phone. This paper gives a programmed and concealed way to deal with recognize designs that are covered up of coronary illness. The given framework utilize information min-ing methods, for example, ID3 algorithm. This proposed method helps the people not only to know about the diseases but it can also help's to reduce the death rate and count of disease affected people.

DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem, et.al, [14] described the disease Predicting system using data mining techniques. In this paper they talk about MAFIA (Maximal Frequent Item set algo-rithm) and K-Means clustering. As classification is important for prediction of a disease. The classification based on MAFIA and K-Means results in accuracy.

G. Singh and L. Samavedham, et.al [15] describe a methodology that has the potential to be translated into first-line diagnostic tool for NDs. We also demonstrate the applicability of this methodology for diagnosing PD subjects in early stages of the disease, i.e., subjects in age of 31–60 years.

III. METHODOLOGY

In this section disease prediction using machine learning is observed.
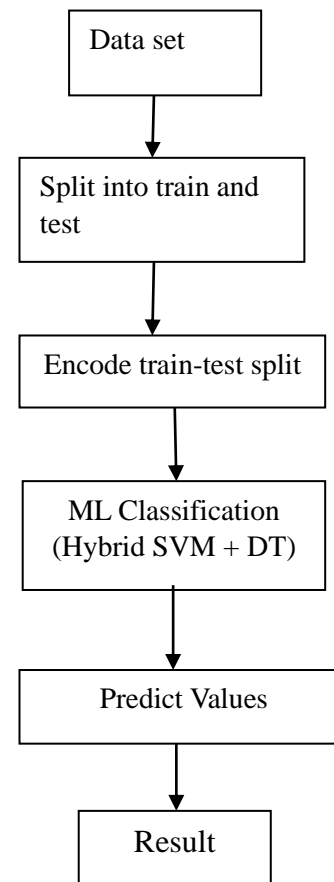

Fig.1: Block Diagram Of Disease Prediction

The dataset of 6 attributes with mere 100 records collected from Enam Medical Diagnosis Centre, Savar, Dhaka, Bangladesh. They build the site to store manual data into MySQL format and to use MySQL data to convert into CSV data format so that we can use the CSV data file in our python programming. The Cleveland heart dataset from the UCI machine learning. Analyze the dataset taken from the Cleveland heart disease database. After storing all the manual data into our website database, we checked the noisy data and eliminated them from our

dataset.This analysis dataset was taken from the web, which basically supports many datasets that are open source and accessible in a variety of data formats. The data source is achieved with various attributes, where the data set is used to create an efficient model to classify and predict the severity of disease in critically ill patients.

After gathering multiple records, pre-processing of heart disease data occurs. Six patient records out of a total of 303 in the sample had missing data. Following the removal of those six records from of the dataset, pre-processing was conducted on the remained 297 patient information. For only the dataset's properties, a multiclass variables and a binary classification are implemented. To determine whether disease is present or not, the multi-class variable is used. The value is set to 1 when the patient has disease; otherwise, it is set to 0 when the patient doesn't have disease. By converting diagnosis values from medical records, data is pre-processed. The data pre-processing outcomes for the 297 patient records show that 137 patient information with a value of 1 have problems, while 160 patient information with a value of 0 have no disease.

The feature selection and modeling keep on repeating for various combinations of attributes. From among the 13 attributes of the data set, two attributes pertaining to age and sex are used to identify the personal information of the patient. The remaining 11 attributes are considered important as they contain vital clinical records. Clinical records are vital to diagnosis and learning the severity of disease.As previously mentioned, this experiment makes use of a variety of (ML) techniques, including NB, LR, DT, KNN, and SVM. The experiment was conducted again using all 13 attributes and ML techniques. The performance analysis of this hybrid system is evaluated in terms of

Accuracy, and Sensitivity. To know whether disease is present or not.

IV. RESULT ANALYSIS

This section discusses the hybrid Machine Learning for analysing the results and predicting disease.

Table 1: Performance Analysis

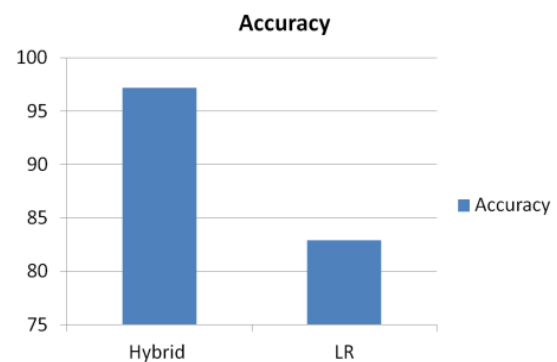| ML classifiers	Accuracy	Sensitivity
Logistic Regression	82.9	86.1
Hybrid SVM + DT	97.2	94.3

Fig. 2: Accuracy Comparison Graph

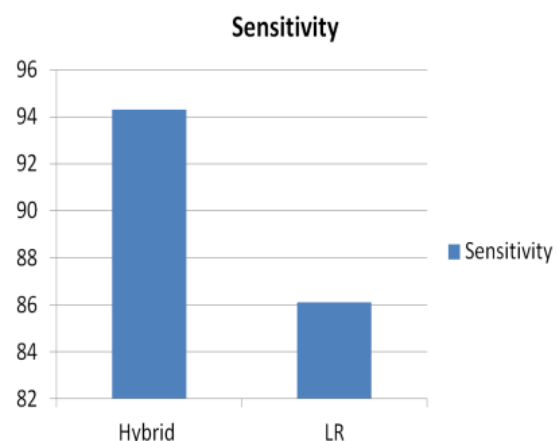Fig.2 shows comparative graph of accuracy for LR and hybrid model for prediction of disease.

Fig.3: Sensitivity Comparison Graph

The comparative graph of the sensitivity for the hybrid model for disease prediction and the LR is shown in Fig. 3.

V. CONCLUSION

This investigation has to be extended further in order to concentrate on real-world datasets instead of only theoretical techniques and simulations. Combining features from Support Vector Machines (SVM) and Decision Trees (DT) is the recommended Hybrid Machine Learning technique. It has been demonstrated that hybrid machine learning is incredibly accurate in predicting disease. In this research, they suggest a hybrid method that uses machine learning techniques to identify important highlights, hence increasing the accuracy of the disease prediction.

VI. REFERNECES

[1] A. Mir, S.N. Dhage, in 2018 Fourth International Conference on Computing Communication Control and Au- tomation (ICCUBEA) (IEEE, 2018), pp. 1{6

[2] Y. Khourdi_, M. Bahaj, Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization, Int. J. Intell. Eng. Syst. 12(1), 242 (2019)

[3] S. Vijayarani, S. Dhayanand, Liver disease prediction using svm and naïve bayes algorithms, International Journal of Science, Engineering and Technology Research (IJSETR) 4(4), 816 (2015)

[4] S. Mohan, C. Thirumalai, G. Srivastava, Effective heart disease prediction using hybrid machine learning techniques, IEEE Access 7, 81542 (2019)

[5] T.V. Sriram, M.V. Rao, G.S. Narayana, D. Kaladhar, T.P.R. Vital, Intelligent parkinson disease prediction using machine learning algorithms, International Journal of Engineering and Innovative Technology (IJEIT) 3(3), 1568 (2013)

[6] A.S. Monto, S. Gravenstein, M. Elliott, M. Colopy, J. Schweinle, Clinical signs and symptoms predicting in- uenza infection, Archives of internal medicine 160(21), 3243 (2000)

[7] Tapas RanjanBaitharu, Subhendu Kumar PaniA, Comparative Study of Data Mining Classification Techniques using Lung Cancer Data, International Journal of Computer Trends and Technology (IJCTT)–volume 22 Number 2– April 2015.

[8] LilikAnifah, Haryanto, RinaHarimurti, "Cancer lung detection on CT Scan image using ANN backpropagation based gray level co occurrence matrix feature." 978-1-5386-3172-0/17/ 2017 IEEE .

[9] Lynch C M, Abdollahi B, Fuqua J D, et al., " Prediction of lung cancer patient survival via supervised machine learning classification techniques", International Journal of Medical Infomatics. Volume 108, December 2017, Pages 1-8.

[10] Krishnaiah,V., Narsimha,G., Subhash Chandra,N., Diagnosis of LungCancer Prediction System Using Data Mining Classification Techniques, et al,/(IJCSIT)International Journal of Computer Science and Information Technologies, Vol. 4 (1), 2013.

[11] K. C. Howlader, M. S. Satu, and A. Mazumder, "Performance Analysis of Different Classification Algorithms that Predict Heart Disease Severity in Bangladesh," 2017 International Journal of Computer Science and Information Security (IJCSIS), vol. 15, no. 5, pp. 332-340, May, 2017, doi: 10.1109/CEEICT.2016.7873142.

[12] Deeanna Kelley "Heart Disease: Causes, Prevention, and Current Research" in JCCC Honors Journal

[13] Ponrathi Athilingam, Bradlee Jenkins, Marcia Johansson, Miguel Labrador "A Mobile Health Intervention to Improve Self-Care in Patients With Heart Failure: Pilot Randomized Control Trial" in JMIR Cardio 2017, vol. 1, issue 2, pg no:1

[14] DhafarHamed, Jwan K. Alwan, Mohamed Ibrahim, Mohammad B. Naeem "The Utilisation of Machine Learning Approaches for Med-ical Data Classification" in Annual Conference on

New Trends in Information & Communications Technology Applications - march-2017

[15] G. Singh and L. Samavedham, "Unsupervised learning based feature extraction for differential diagnosis of neurodegenerative diseases: a case study on early-stage diagnosis of Parkinson disease," Journal of neuroscience methods, vol. 256, no. 30, pp. 30-40, Dec. 2015.