Liver Disease Prediction Using Machine Learning Algorithms

HARSHA TALELE[1], KAJAL SHAH[2], TANUJA PATIL[3], JYOTI SHIROLE[4], POOJA BHAVSAR[5]
[1, 2, 3, 4, 5] *Student, Department of Computer Engineering, KCES's College of Engineering & Management, Jalgaon, Maharashtra, India*

Abstract*— Diseases related to the liver and heart are becoming more and more familiar with time. With continuous technological advancements, these will only increase in the future. Although people are becoming more conscious of health nowadays and are joining yoga classes and dance classes, still the sedentary lifestyle and luxuries that are continuously being introduced and enhanced, the problem is going to last long. Population ageing and the increase of chronic conditions incidence and prevalence produce a higher risk of hospitalization or death due to liver disease. This is exceptionally high for patients with multi-morbidity, leading to significant resource consumption. The most crucial challenge is to identify possible high-risk patients to improve health care service provision and also to reduce costs. Nowadays, population health management, based on intelligent models, can be used to assess the risk and identify these "complex" patients infected with liver disease

Indexed Terms*— DT Algorithms, KNN Algorithms, Liver Disease, MLP Classifier, SVM Algorithms

I. INTRODUCTION

Liver illnesses have become one of the leading causes of death worldwide in recent decades, and they have also become a life-threatening disease. Chronic diseases account for roughly 59 per cent of global mortality and 46 per cent of global conditions, according to the WHO, and chronic diseases claim the lives of around 35 million people worldwide. Problems with the liver are frequently not discovered until it is past the point of no return, as the liver continues to function even when partially injured. Early detection could potentially save a person's life. The need of this project is to describe a framework for an iterative approach to event discovery of high-risk patients that is based upon a clinical data repository and natural language processing techniques. The paper is also open to new challenges and future changes to other advanced technologies in getting the prediction to be correct. There is a considerable risk of data misuse when different organizations use the data to find out otherwise, as minute changes in the parameter will lead to varying rates in the prediction. To avoid any such risk and secure the user data, there is an urgent need to ensure the data.

The rest of the paper is organized as follows. Proposed algorithms are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

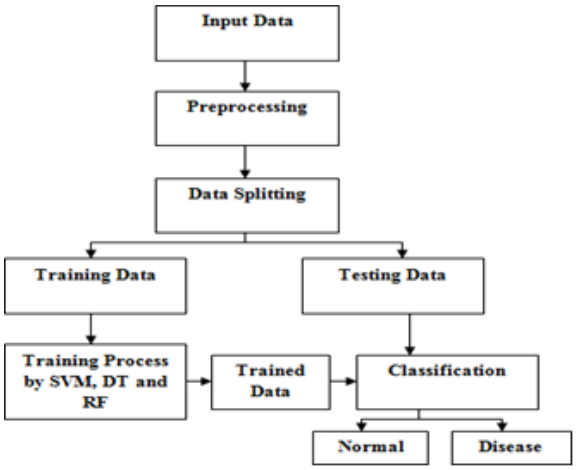

Fig1.1. Process for Liver Disease Predicition

II. LITERATURE SURVEY

vasan durai, Dinesh , katheredy (2017)[1] explained the patient data sets are analyzed for the predictability of the subject to have a liver disease based purely on a widely analyzed classification model. Since there are pre-existing processes to analyze the patient data and the classifier data, the more important facet here is to predict the same the conclusive result with a higher rate of accuracy.

V. V. Ramaligan, A. Pandian, R. Ragavendran[2] proposed that Human body had a versatile number of

parts which are playing a key role for the healthier life of any human being. Some of those parts are brain, heart, liver, kidneys, etc. The currently developed application is a graphical user interfacing unit with which the doctors or medical practitioners can utilize to identify the symptoms and give as input and can identify the problems of liver diseases at various stages.

Joel Jacob, Joseph chkalakal mathew, Johns mathew, Elizabeth Issac[3] proposed that the medical records of the patients as a vast source of data are applied to the data mining techniques to extract the valid dataset to predict the liver disease. The classification algorithms have been widely used in the decision-making process. RNN being a text classifier of deep learning technique with the advantage of processing in multiple loops in a sequential manner to obtain best performances measured by the factor of accuracy has been proposed in this study.

Alice Auxilia in year 2018 proposed "Accuracy Prediction Using Machine Learning ". [4] The utilization of medicinal datasets has pulled in the consideration of specialists around the world. Machine Learning methods have been broadly utilized as a part of creating choice emotionally supportive networks for ailments forecast through an arrangement of therapeutic datasets. Indian Liver Patient's datasets demonstrate that proposed technique amazingly enhances the illnesses expectation precision The point of this task is to some degree diminish the time delay caused because of the superfluous forward and backward transporting between the healing Centre and the pathology lab

A.Siyasangari Baddigam Jaya Krishna Reddy,Annamareddy Kiran,P.Alitha in year 2022 proposed of a system"Diagnosis Of Liver Disease Using Machine Learning Models ".the 21st-century, the issue of liver disease has been increasing all over the world. As per the latest survey report, liver disease death toll has been rise approximately 2 million per year worldwide. The overall percentage of death by liver disease is 3.5% worldwide. Chronic Liver disease is also considered to be one of the deadly diseases, so early detection and treatment can recover the disease easily.

III. PROPOSED ALGORITHM

K-NEAREST NEIGHBOUR ALGORITHM

The k-nearest neighbour algorithm is a powerful non-parametric technique for density estimation and classification. It is an instance-based learning or lazy learning algorithm, where the approximation happens only locally, and all the computation is done after sort.
Advantages:

1. Some Noise reduction techniques can improve the accuracy of the algorithm.
2. Case-retrieval Nets can improve the run-time performance for large data sets.
3. Ease of implementation and debugging.

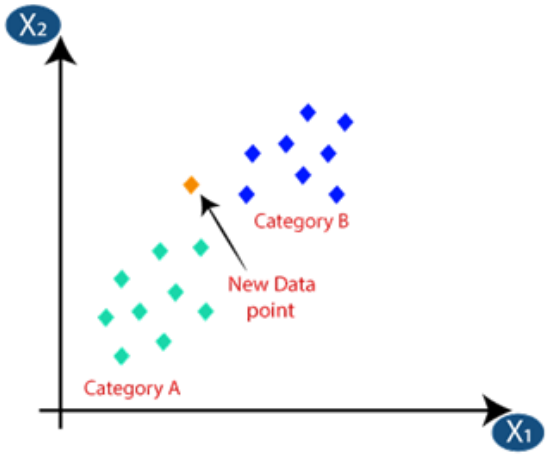

Fig 2.1 K-nearest neighbor algorithm

SUPPORT VECTOR MACHINE

A support vector machine is a discriminative classification algorithm defined by separatinhyhyperplanese. It is a supervised learning algorithm in which the training data set is labelled which outputs optimal hyperplane categorising new datasets.

Advantages:
1. Effective in large scale regression problems.
2. Effective in high dimensional space.
3. Effective In cases number of samples is more minor norr than the number of dimensions.
4. Memoryeefficiencyt due to the use of support vectors with slight modifications.

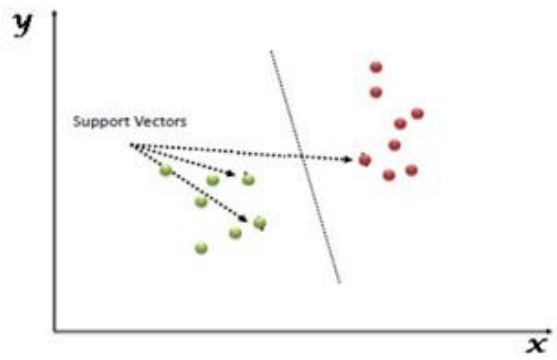

Fig. 2.2 Support Vector Machine

RANDOM FORESTS

Random forests are an ensemble learning method for classification and regression. The ensemble learning method constitutes multiple models to increase the performance of the models. Random forests generate numerous decision trees in training time, and the classification is done by taking the mode of all the classes output from individual trees.

Advantages:

1. Efficient for large datasets.
2. Handles many input variables without deletion.
3. Estimation of essential variables for classification.
4. Effective accuracy is maintained when ample data is missing.

Disadvantages:

1. Random forests are prone to the over-fitting problem for datasets with noisy classification/regression tasks.

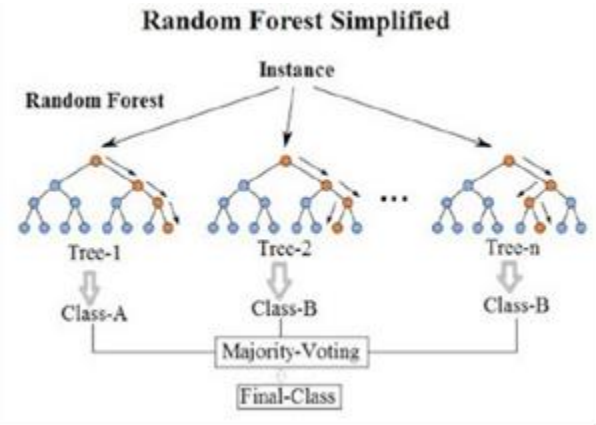

Fig 2.3 Random Forest Working Structure

MLP CLASSIFIER

Multi-layer perceptron(MLP) is a supervised learning algorithm that learns a function

F(.): m -> R o by training on a dataset, where is the number of dimensions for input and is the number of sizes for output. MLP classifier relies on an underlying Neural Network to perform the classification task

advantages:

1. Can be applied to complex non-linear problems.
2. works well with larextensiveput data.
3. provides quick predictions after training.
4. The same accuracy ratio can be achieved even with more minor data.

Disadvantages:

It is unknown to what extent each independent variable is affected by the dependent variables. Computations are complex and time-consuming.

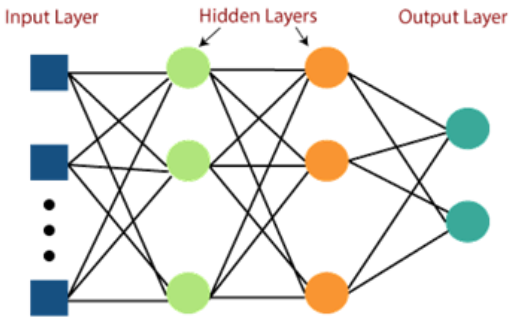

Fig.2.4 MLP Classifier structure

- Confusion Matrix

A confusion matrix is an N x N matrix used for evaluating the performance of a classification model, Where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification

Model is performing and what kinds of errors it is making

Predicted Class

		Yes	No	
Actual Class	Ye s	True Positives(TP)	False Positives(FP)	P

	No	False Negatives(F N)	True Negatives(T N)	N
		P Complement	N Complement	P+ N

Table No. 1 Components of Confusion Metrix

The methods used for the calculating the parameters:

$$\text{Precision} = \frac{\text{Total Relevant Retrieved}}{\text{Total Retrieved}}$$

$$\text{Recall} = \frac{\text{Total Relevant Retrieved}}{\text{Total Relevant}}$$

$$\text{Accuracy} = \frac{\text{Total Relevant}}{\text{Total Records}}$$

IV. EXPECTED RESULT

Our goal is to get prediction on the basis of given datasets of people whether the person is having the chronic disease or liver disease symptoms or not. With using many different algorithm we are trying to predict which algorithm will play vital role in predicting the disease. This system will be very useful for many hospitals and even professional doctors to easily detect the disease. Also, general user can use this system for their finding out the disease. This system will change the way and can be early as possible as it will lead to save the person's life. This whole work is focused on how we can predict the disease by given datasets so that will help in preventing and curing the disease of the patients.

We will expect as a result, prediction of the disease will be given by four of the algorithms that is MLP classifier, Svm, DT and Knn algorithms. So, if we see in terms of accuracy level the Svm Algorithms gives the best accuracy and if we see in terms of false positive parameter MLP classifier stands out in that. The experimental results suggest that the MLP and SVM algorithm can give best results in prediction, it also provides the benefits of efficiency and protection of the algorithm.

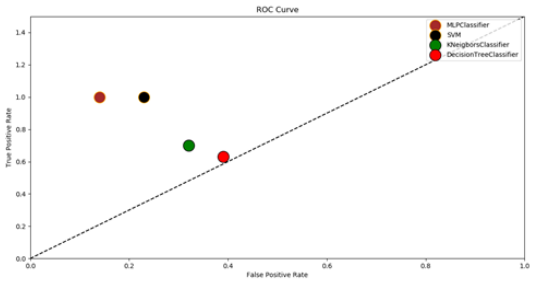

Fig.1.2 Predicted Graph result of All Classifier

V. CONCLUSION

Population ageing and the increase of chronic conditions incidence and prevalence produce a higher risk of hospitalization or death due to liver disease. This is exceptionally high for patients with multi-morbidity, leading to significant resource consumption. The most critical challenge is to identify possible high-risk patients to improve health care service provision and also to reduce costs. The project's primary focus is implementing machine learning algorithms Naïve Bayes, C5.0, and Random Forest to predict the risk of hospitalization or death starting from the administrative and socio-economic datasets. Thus the training would be provided on the training dataset, which would help predict results for the testing phase dataset provided.

REFERENCES

[1] Vasan Durai, Dinesh, Kalthireddy, "Liver disease prediction using machine learning", International Journal of Advance Research, Ideas and Innovations in Technology, (2017)

[2] V.V. Ramalingam, A.Pandian, R. Ragavendran, "Machine Learning Techniques on Liver Disease A Survey", International Journal of Engineering & Technology, 7 (4.19) (2018) 485-495.

[3] Joel Jacob, Joseph Chakkalakal Mathew, Johns Mathew, Elizabeth Issac, "Diagnosis of Liver Disease Using Machine Learning Techniques" International Research Journal of Engineering and Technology (2018).

[4] L. Alice Auxilia, "Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease", Proceedings of the 2nd

International Conference on Trends in Electronics and Informatics (ICOEI 2018).

[5] Ain Najwa Arbain, B. Yushalinie Pillay Balakrishnan, "A Comparison of Data Mining Algorithms for Liver Disease Prediction on Imbalanced Data", International Journal of Data Science and Advanced Analytics, 2018

[6] Nazmun Nahar and Ferdous Ara, "Liver disease prediction by using different Decision tree techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2, March 2018

[7] Shambel Kefelegn, Pooja Kamat, "Prediction and Analysis of Liver Disorder Diseases by using Data Mining Technique: Survey", International Journal of Pure and Applied Mathematics 2018