Machine Learning Algorithms for Detecting Liver Disease: The Review

V.S. Mane
Department of Computer Science, Shivaji University, Kolhapur

S.T. Bhosale
Asst. Prof., V. P. Institute of Management Studies & Research, Sangli

ABSTRACT
Liver disease is the cause of death and effect of disease is very high on human. many researchers are studied on these predicting detect liver disease with help of medical report blood test, biopsy test, platelets, common liver function test like Alanine Transaminase, Aspartate Transaminase, Alkaline phosphatase, Albumin and total proteins, Bilirubin, Gamma-glutamyl transferase, L-lactate dehydrogenase, Prothrombin time.

Upcoming days liver disease increase day by day as per various research paper study. Machine learning algorithms are very useful for biomedical domain or medical field. Mainly they are used for detecting disease. In this review paper studied and overview of some different research paper. And these different research papers give a description of machine learning algorithm like logistic regression, random forest, decision tree, Support Vector Machine, K-nearest, Naïve Bayes that are used for predicting detecting different liver disease.

Keywords
Liver Disease, Machine Learning Algorithms, Deep learning algorithms

1. INTRODUCTION
The liver is the largest organ of our human body. Liver function is digesting food and releasing the toxic substances s of the body. A toxic substance called as ammonia. Some viruses and misuse of alcohols are affected to body. Because of that liver can damaged and human to lead can cause serious health problems or complications. liver disease is one of the most health issues in world. 2 million peoples died yearly in world because of liver disease. Machine learning algorithms is most useful for biomedical field of liver disease detection and prediction. ML gives better and accurate prediction with low cost. They also increase the objectivity of the decision-making process [1].

Disease diagnosis is the process of identify disease with help of patient's symptoms or condition or signs. Some signs or symptoms are non-specific they are most challenging and detection of disease is most significant point of treatment. ML algorithms can help the predict disease based previous data [2].

Nowadays data mining techniques are finding best patterns are applied on medical dataset for predicting disease. The main object of the research work is predicting liver disease like Cirrhosis, Hepatitis, alcoholic, non-alcoholic, liver cancer from Liver Function Test (LFT) dataset using ML Classification algorithms [3].

The basic classification of liver disease is the acute and chronic. liver disease symptoms are nausea, vomiting, right upper quadrant abdominal pain, fatigue, weakness, back pain, swelling, weight loss, fluid in abnormal cavity, general itching, pale stool, enlarged spleen and gallbladder, yellowing of the skin and eyes. Sometimes patients have no symptoms. Tests such as imaging tests and liver function tests can check for damage and help to diagnosis diseases [4].

The liver is large, meaty organ that are present on the right side of belly. Weight is 3-pound, reddish brown in color. Liver has two sections called right, left lobes. The liver main function is straining the blood coming from digestive track, before passing it to rest of the body. The liver also makes proteins they are important for blood clotting and other functions [5].

Liver diseases are classified based on their aetiology and effect on the liver. The aetiology involve infection, injury, exposure to drugs or toxic substances, a process, or a genetic abnormality. The above causes hepatitis, cirrhosis, and stones that can increase in size and cause blockages, fatty infiltration and, liver cancer. Genetic disease can also cause when to much iron or copper builds up in the body. Extra copper or iron is poisons to body. [7]

1.1 Types of Liver Disease
i. **Caused by viruses**- like Hepatitis A, Hepatitis B, Hepatitis C.

ii. **Caused by drugs, Poisons or too much alcohol**- like Fatty **liver** disease, Non-alcoholic fatty liver disease and Cirrhosis

iii. **Liver cancer**.

iv. **Inherited disease (Genetic) disease**- Hemochromatosis, Hyperoxaluria and Wilson disease.

Liver disease have different effects on the liver depending on the stage of the disease. In general, liver disease is characterized by chronic inflammation and damage to liver cells, which can lead to the buildup of scar tissue (fibrosis) and eventually cirrhosis. Here are some effects of liver disease on the liver at different stages:

Early-stage liver disease: In the early stages of liver disease, there no significant effects on the liver other than inflammation and mild damage to liver cells. The liver still functions normally, and possible to recover the damage with early treatment.

Fatty liver disease: In fatty liver disease, the liver becomes infiltrated with fat (fat build up in liver), which can cause inflammation and damage to liver cells. this can lead to the development of fibrosis and cirrhosis. Fatty liver disease is also called as hepatic steatosis.

Fibrosis and cirrhosis: As liver disease progresses to fibrosis and cirrhosis, the liver becomes increasingly damaged and scarred. This can lead to a decline in liver function and a range of complications, including diabetes, hypertension (high blood pressure in the veins that supply the liver), ascites (fluid buildup in the abdomen), hepatic encephalopathy (brain dysfunction caused by liver failure), and an increased risk of liver cancer.

Liver cancer: In advanced stages of liver cancer, the cancer cells can infiltrate and damage healthy liver tissue, leading to further liver dysfunction and an increased risk of complications. Two types of liver cancer. Primary and secondary cancer, primary cancer starts in your liver. Secondary cancer spreads to another part of body.

The effects of liver disease on the liver can depending on the underlying cause of the disease, as well as individual factors such as age, gender, genetics, and lifestyle factors such as alcohol consumption and diet. Early diagnosis and treatment can help slow down or even reverse the progression of liver disease, so it's important to seek medical attention if experience any symptoms of liver disease or have any concerns about liver health.

Liver disease detect with the help of medical imaging technique such as Ultrasound (US), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). This review paper aimed to design and develop a predicting model for fatty liver disease using the technique of machine learning approach.

1.2 Machine Learning Types

ML is defined technology is used for train machines to perform automatically different actions like prediction, detection, recommended based on manually dataset without human intervention. Algorithm is defined a set of rules, statistical techniques and mathematical theorems used to extract meaningful information from dataset.

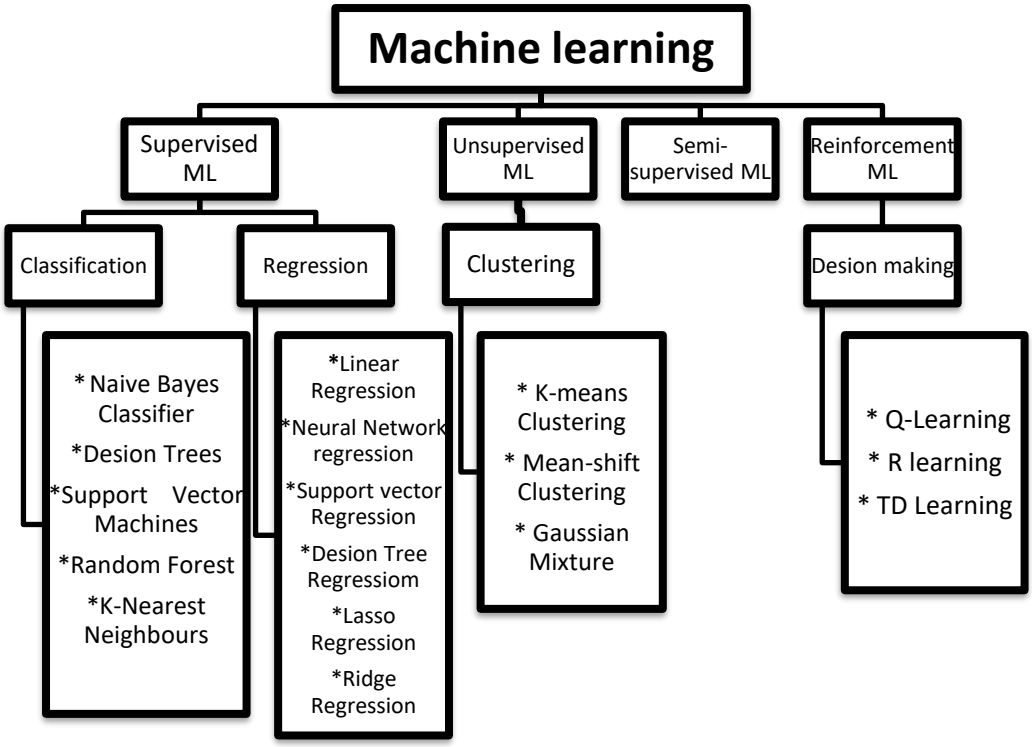


Fig. 1 Types of Machine Learning Algorithms

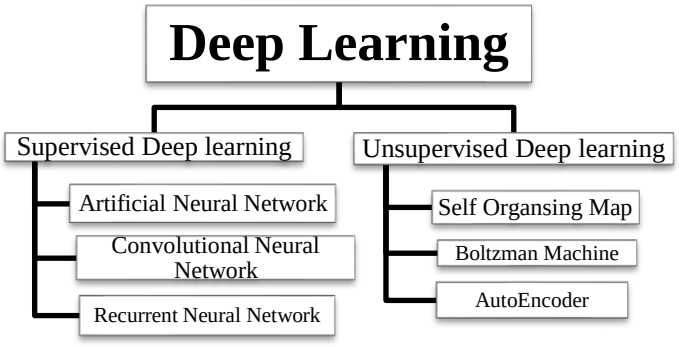


Fig. 2. Types of Deep Learning Algorithms

Explanation definition of each type of machine learning and deep learning. Supervised learning means the algorithm learns to predict data from input data and models are trained us labelled data. Supervised algorithms learning from the training dataset behavior like a teacher supervising the learning process. Unsupervised learning, the algorithm has only input data and learns the original hidden structure from the input data. Unsupervised learning in which models are not supervised using training dataset. Sometimes unlabeled data and need to extract hidden patterns from the dataset that time use

unsupervised machine learning algorithms. semi-supervised learning depends on supervised and unsupervised means labelled and unlabeled data i.e., input data have label data, and some unlabeled data. In reinforcement learning, the system learns through the environment and perform action well. this tool is used for biomedical field, such as detect diseases [2].

Deep Learning is a subset of machine learning that contains multiple layers that holding multiple levels of perception, such that each layer receives information from the previous layer and passes the result to the next layer.[2] There are different deep learning method like recurrent neural networks (RNN), deep auto-encoders, and convolutional neural networks (CNN). RNN is a deep learning algorithm with internal memory to keep the recent information.

2. DESCRIPTION OF THE ALGORITHMS

2.1 Logistics Regression (LR)

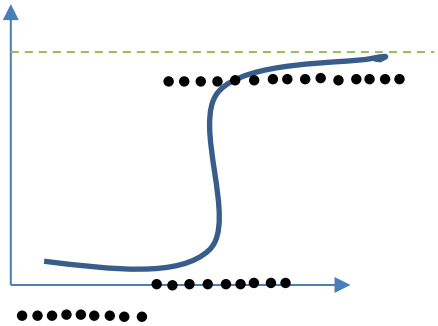
Logistic Regression is supervised classification algorithm. Logistic Regression can perform with any numerical factors. Regression processes the connection between the element factors by surveying probabilities (p) utilizing an underlying logistic function [1]

Logistic regression predicts output as categorical (yes or no) or discrete value (0 or 1). It gives probabilistic value between 0 and 1. Logistic linear equation [1,2]

$$P=\frac{1}{1+e^{(b0+b1x1+\cdots\ldots+bnxn)}}$$

Therefore, logistic function or sigmoid function represent value greater than 1 or less than 0 which is not possible as per the hypothesis of logistic regression.

$$0 \leq h\theta(x) \leq 1$$



Predicted Y lies within 0 and 1 range

Fig.3. Logistic Regression sigmoid Curve

2.2 Decision Tree

A decision tree is a popular machine learning algorithm used for both classification and regression tasks. It is a graphical representation of all the possible outcomes of a decision based on certain conditions. The tree is made up of nodes and branches, with each node representing a condition or a decision, and each branch representing the possible outcomes or the results of the decision.

Decision trees are built by recursively splitting the dataset into subsets based on the values of different features until a stopping criterion is reached, such as reaching a certain depth or when all the data points in a subset belong to the same class. The splitting process involves selecting the feature that provides the most information gain, which means that it separates the data points into the most distinct and homogeneous groups.

Once a decision tree is built, it can be used to predict the outcome for a new data point by following the branches of the tree based on the values of the features of that data point. Decision trees are popular because they are easy to interpret and can handle both categorical and continuous data. However, they can be prone to overfitting and may not perform well on datasets with high dimensionality or with many missing values. [1,2,4]

2.3 Random Forest (RF)

Random Forest (RF) is a type of supervised machine learning algorithm used for classification, regression, and other tasks. It is an ensemble learning method that combines multiple decision trees to improve predictive accuracy and reduce overfitting.

In a random forest, each decision tree is trained on a randomly sampled subset of the training data and a randomly selected subset of the features. This means that each tree in the forest is trained on a different subset of the data, and the trees are therefore different from each other. During prediction, the output of all the trees is combined to generate the final prediction.

The key advantages of random forests are that they are relatively easy to use, can handle high-dimensional data, and are robust to noisy and missing data. They also provide measures of feature importance, which can be used to understand the relative importance of different features in the data.

One potential disadvantage of random forests is that they can be slow to train on large datasets, particularly when there are a large number of trees in the forest. Additionally, the resulting model may be difficult to interpret, as it involves multiple decision trees with complex interactions between them. [1,2,5]

2.4 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm. SVM used for classification and regression analysis. The main contain of SVM is to find a hyperplane that separates the data into two classes with maximum margin. The margin is the separation between each class's nearest data points and the hyperplane.

In SVM, the training data is represented set of feature vectors in a high-dimensional space. The hyperplane is then constructed by finding the optimal separating hyperplane that maximizes the margin between the two classes. This hyperplane is selected it can generalize well to unseen data.

There are two main types of SVM: linear SVM and nonlinear SVM. Linear SVM is used when the data can be separated using a straight line or a hyperplane, nonlinear SVM is used when the data cannot be separated using a straight line or a hyperplane. Nonlinear SVM uses a technique called the kernel trick to map the data into a higher dimensional space where it

SVM has particular advantages over other classification algorithms, such as high accuracy, robustness to outliers, and the ability to handle high-dimensional data. Yet, particularly for huge datasets, it can be computationally expensive.

SVM is a powerful and versatile algorithm that can be used for a vast range of classification and regression tasks in machine learning. [1,2,3,5]

2.5 K-Nearest Neighbors

K-Nearest Neighbors (KNN) is a machine learning algorithm. KNN used for classification and regression tasks. KNN is a

non-parametric and lazy learning algorithm, KNN does not make any assumptions about the underlying data distribution and does not have a training phase. KNN stores all the data and makes predictions based on the similarity between new data points and the stored data.

In KNN, the "k" represents the number of nearest neighbors used to classify or predict the output for a new data point. The algorithm calculating the distances between the new data point and all the stored data points, and then selecting the k-nearest neighbors based on this distance metric. The most common distance metrics used in KNN are Euclidean distance and Manhattan distance.

In classification tasks, the algorithm assigns the new data point to the class that is most frequent its k-nearest neighbors. In regression tasks, the algorithm predicts the value of the target variable by taking the average or weighted average of the values of its k-nearest neighbors.

KNN has different advantages, such as simplicity and interpretability, but it also has some limitations. limitations is that it can be computationally expensive, especially when working with large datasets. KNN is sensitive to irrelevant and noisy features, and it requires a proper choice of the distance metric and the value of k for optimal performance. [1,2]

2.6 Naive Bayes

Naive Bayes (NB) is a classification algorithm that is based on Bayes' theorem of mathematics. NB is a probabilistic algorithm that is commonly used in machine learning for text classification, spam filtering, and sentiment analysis.

The algorithm is called "naive" because NB supposed that the features used to classify the data are independent of each other, which is not always true. NB can still be a powerful tool for classification tasks.

To apply the algorithm, one first needs to train the model on a dataset with known labels. During training, the algorithm calculates the probability of each feature given each class label. This information is used to calculate the probability of a new observation belonging to each class label. The class with the highest probability is assigned as the predicted label for the new observation.

NB is a relatively simple and computationally efficient algorithm that can work with small datasets. it may not be the best choice for more complex problems where the features are highly correlated or where the underlying assumptions of the algorithm are not met. [1,2,3]

2.7 Artificial Neural Network

An artificial neural network (ANN) is a computational model inspired by the structure and function of the biological neural networks found in the human brain. comprise of interconnected nodes (neurons) organized into layers that process and transmit information.

Each neuron receives input signals from other neurons or external sources, processes using a non-linear activation function, and generates an output signal that is transmitted to other neurons. The connections between neurons are weighted, which allows the network to learn from data by adjusting these weights based on the error between the predicted and actual outputs.

ANNs are commonly used in machine learning and artificial intelligence applications, such as image recognition, natural language processing, and predictive analytics. They are highly flexible and used for both supervised learning (network is trained on labeled data) and unsupervised learning (where the network identifies patterns and relationships in unlabeled data).

Different types of ANNs, including feedforward neural networks, convolutional neural networks, and recurrent neural networks, each of which has its own unique characteristics and applications.

3. MACHINE LEARNING PROCESS

Machine learning process is the flow of work create and execute the model. They are also called machine learning life cycle. They main purpose of the life cycle is find the solution of problems.

3.1 Data Collection

a dataset from UCI Machine Learning Repository. dataset was collected from the northeast of Andhra Pradesh, India. Data set is built on numerical and nominal data types.

3.2 Attributes Information

1. Age: Age of the patient
2. Gender: Gender of the Patients
3. TB: Total Bilirubin
4. DB: Direct Bilirubin
5. Alkphos: Alkaline Phosphotase
6. Sgpt: Alamine Aminotransferase
7. Sgot: Asparatate Aminotransferase
8. TP: Total Proteins
9. ALB: Albumin
10. AG Ratio: Albumin and Globulin Ratio
11. Selector field used to split the data into two sets (labeled by the experts)

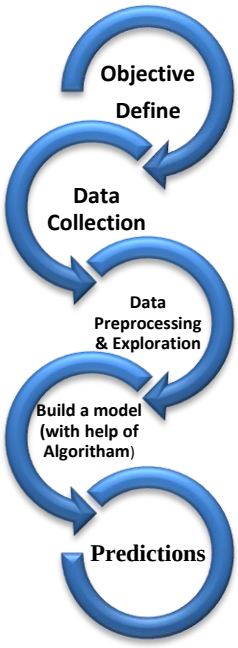


Fig.4. Process for building model

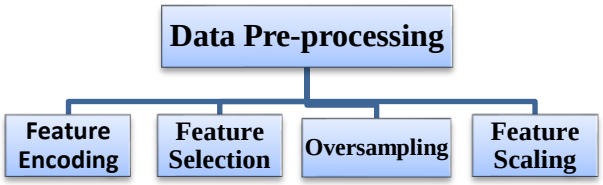


Fig.5. Data Pre-processing

3.3 Data Pre-processing & Exploration

This dataset contains 583 (416 liver patient and 167 non liver patient) liver patient's data whereas 441 male patients and 142 are female patients. [1,2,3,4,5]

Data Pre-processing is the improve the prediction accuracy. If the dataset has missing values, outliers, irrelevant attributes they show low accuracy.This study is improving the accuracy and reduce the error rate. In the pre-processing encoding of categorical feature values, removal of irrelevant features, handling of outliers and oversampling techniques.[5]

1)Label Encoding: Gender contains categorical values Male and Female. Label encoder is used to convert categorical to numerical value means male as 1 and female as 0.

2)Feature Selection: Feature selection is the process of selecting a subset of relevant features or variables that are most useful for a given problem. means it involves choosing the most informative features from a dataset that can improve the performance of a machine learning model.

There are several techniques used for feature selection, including:

a) **Filter methods**: Filter methods evaluate the relevance of features based on statistical measures such as correlation or mutual information.
b) **Wrapper methods**: Wrapper methods evaluate the performance of a machine learning model using different subsets of features, and select the subset that gives the best performance.
c) **Embedded methods**: Embedded methods combine feature selection with the model training process, such as regularization techniques like Lasso or Ridge regression.

Feature selection technique depends on the type of data, the size of the dataset, and the specific problem at hand. Feature selection helps to reduce the complexity of a model, improve its accuracy, and make it more interpretable. it can also lead to a loss of information and should be used carefully to avoid overfitting.

3)Oversampling: Oversampling is a technique used in data analysis and machine learning to address class imbalance, which occurs when one class of data is significantly underrepresented compared to other classes. In oversampling, the minority class is artificially increased by duplicating existing examples or generating new examples based on existing ones, until the class distribution is more balanced.

Oversampling performed using various algorithms, such as Random Oversampling, Synthetic Minority Over-sampling Technique (SMOTE), and Adaptive Synthetic Sampling (ADASYN). These algorithms aim to create new synthetic data points that are similar to the existing ones in the minority class, but with some variations to introduce diversity.

Oversampling helps improve the performance of machine learning models by providing more training examples for the minority class, reducing the risk of overfitting to the majority class, and improving the model's ability to generalize to new data. Oversampling increase the risk of overfitting to the minority class, and it can not be effective if the minority class is too small or too different from the majority class.

4)Feature Scaling: Feature scaling is the process of standardizing the range of features or variables in a dataset. It is step in different machine learning algorithms. Feature Scaling helps to improve the performance of the model by ensuring that all features are on the same scale.

In machine learning, the term "feature" refers to the independent variables or input variables that are used to predict the dependent variable or output variable. These features have different scales, units, and ranges. For example, consider a dataset that includes two features: age and income. Age may range from 0 to 100, while income may range from 0 to 1 million. the range of the two features is significantly different, and the machine learning algorithm may give more importance to the feature with the larger range, which can affect the accuracy of the model.

Feature scaling is performed to normalize the range of the features to a similar scale, typically between 0 and 1 or -1 and 1. Different methods for scaling features, including standardization and normalization. Standardization involves subtracting the mean of the feature and dividing by its standard deviation, normalization involves scaling the feature to have a minimum and maximum value of 0 and 1.

Feature scaling is an important pre-processing step in machine learning that can help improve the accuracy and performance of models.

4. MEASUREMENT OF TECHNIQUE

Measures the test execution of various classification algorithm. The classification methods evaluated various evolution procedures i.e., accuracy, sensitivity, specificity, and precision and f1 measure. The exhibition evaluation variables evaluated by the confusion matrix. The result of prediction [5,6]

True Positive (TP): correctly identifies that a patient has liver disease.

False Positive (FP): incorrectly identifies that a patient has liver disease.

True Negative (TN): correctly rejects that a patient has liver disease.

False Negative (FN): incorrectly rejects that a patient has liver disease

$$\textbf{Accuracy}= \frac{\text{True Positive}+ \text{True Negative}}{\text{True Positive}+\text{False Positive} +\text{True Negative} +\text{False Negative}}$$

$$\textbf{Sensitivity} = \frac{\text{True Positive}}{(\text{True Positive}+\text{False Negative})}$$

$$\textbf{Specificity}= \frac{\text{True Negative}}{\text{False Positive} + \text{True Negative}}$$

$$\textbf{Precision}= \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$

$$\textbf{F1}= \frac{2*Recall*Precision}{(Recall+Precision)}$$

5. ANALYSIS OF THE RESULT

5.1 "A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms"

Table.1 The performance comparison of six supervised machine learning techniques

	KNN	NB	RF	SVM	DT	LR
Classification Accuracy	0.62	0.53	0.74	0.64	0.69	0.75
Precision	0.72	0.36	0.85	0.69	0.77	0.91
Sensitivity	0.76	1	0.81	0.88	0.79	0.78
f-1	0.74	0.53	0.83	0.77	0.78	0.84
Specificity	0.35	0.46	0.5	0.21	0.48	0.47

The six-machine learning classifier for the liver disease dataset. LR highest accuracy of 75% and NB lowest performance 53%. precision, LR highest score 91% and NB lowest 36%. sensitivity, SVM highest value 88% and KNN lowest 76%. Logistics Regression was also the best performer in terms of f1 measure 83% and NB obtained the lowest 53%. specificity DT highest value 48% and LR the lowest 47%. According this measurement Logistic Regression is more accurate other techniques for predicting disease. [1]

5.2 "Liver Disease Prediction using SVM and Naïve Bayes Algorithms"

Table. 2 The performance comparison of SVM and Naïve Bayes Algorithms

Algorithms	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	TP Rate	Precision	F Measure
Naïve Bayes	61.28	38.72	0.612	0.558	0.251
SVM	79.66	20.34	0.796	0.766	0.331

The accuracy measure for the Naïve Bayes and SVM classification algorithms. An experimental result shows the performance of SVM is better than Naïve Bayes algorithm. the SVM classifier is best algorithm because of its highest classification accuracy and comparing the execution time, the Naïve Bayes classifier needs minimum execution time. [3]

5.3 "Liver Disease Prediction by Using Different Decision Tree Techniques"

Table. 3 The performance comparison of Decision Tree Techniques

Techniques	Tree Size	ACC (%)	MAE	PRE	REC	FME	Kappa Statistics	Time
J48	65	65.69	0.3678	0.651	0.657	0.654	0.158	0.11
LMT	1	69.47	0.4116	0.632	0.695	0.628	0.065	0.88
Random Forest		69.30	0.3464	0.667	0.693	0.674	0.186	0.5
Random Tree	267	66.55	0.3382	0.662	0.666	0.663	0.183	0.01
REP Tree	27	66.13	0.3800	0.630	0.691	0.629	0.067	0.03
Decision Stump	Single level	70.67	0.4392	0.499	0.707	0.585	0.379	0.01
Hoeffding Tree	1	69.75	0.4091	0.634	0.700	0.619	0.0501	0.12

In this paper decision tree algorithm such as J48, LMT, Random Forest, Random tree, REP Tree, Decision Stump and Hoeffding Tree this algorithm detect the liver disease at an earlier stage. These algorithm gives various result based on Accuracy, Mean Absolute Error, Precision, Recall,

Kappa statistics and Runtime. From the analysis, Decision Stump outperforms better than other algorithms and its accuracy is 70.67%. The application of Decision tree in predicting liver disease.[4]

5.4 "Diagnosis of Liver Disease Using Machine Learning Techniques"

Table. 4 The performance comparison of ML Techniques

Classification Algorithm	Accuracy	Precision	Sensitivity	Specificity
Logistic Regression	73.23	78.57	88	30.62
KNN	72.05	80.98	83.78	44.04
SVM	75.04	77.09	79	71.11
ANN	92.8	93.78	97.23	83

The four-machine learning algorithms are used i.e., SVM, Logistic Regression, KNN and Artificial Neural Network. Performance evaluation was based on certain performance metrics. ANN is highest accuracy with an accuracy of 98%. it discovered that ANN proved highly efficient. A GUI, which can be used as a medical tool by hospitals and medical staff was implemented using ANN. [9]

5.5 "Diagnosis of Liver Disease using Machine Learning Models"

Table. 5 The performance comparison of ML Techniques

Method	SVM	DT	RF
Accuracy	0.9518	0.8795	0.9277
Precision	0.8095	0.7333	1.0000
Recall	1.000	0.6470	0.6470

In this paper, the algorithm of Support Vector Machines (SVM), Decision Tree (DT) and Random Forest (RF) predict liver disease with better precision, accuracy and reliability. [10]

5.6 "Diagnosis of liver diseases using machine learning"

Table. 6 The performance comparison of ML Techniques

Algorithm	Accuracy	Precision	Sensitivity	Specificity
SVM	71	64.1	71.5	88.3
Back Propogation	73.2	65.7	73.3	87.7

This study examines two techniques for identifying patient parameters and the expression of the genome in an effort to enhance the diagnosis of liver illnesses. The study also outlines drawbacks and examines the computational techniques that can be used to the aforementioned methodology. It suggests ways to increase these algorithms' effectiveness.

6. CONCLUSION

Machine algorithms are mostly used in analyzing medical information, genetic records, and medical images of patients.

Different research papers and as per studies have proved the machine learning models are used in the biomedical domain. These models can search hidden patterns, features from medical datasets that provide early prediction and detect of diseases. Machine learning algorithms plays better role in detecting liver disease. In this study, find the research gap of the researches that use ML algorithms in liver disease. This paper explained review of different machine learning algorithm are used for detecting liver disease. This review studied many algorithms SVM, K-nearest neighbors, random forest, and the decision trees gives good accuracy for detecting disease.

7. ACKNOWLEDGMENT

The authors would like to thank the V.P. Institute of Management Studies & Research, Sangli for providing the infrastructure to carry out the research. This work is supported by Shivaji University, Kolhapur.

8. REFERENCES

[1] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain, "A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms"International journal of scientific & technology Research volume 8, 2019.

[2] Ibrahim Mahmood Ibrahim, Adnan Mohsin Abdulazee, "The Role of Machine Learning Algorithms for Diagnosing Diseases", JASTT, Vol. 02, No. 01, pp. 10 – 19 (2021)

[3] Dr. S. Vijayarani, Mr.S. Dhayanand,"Liver Disease Prediction using SVM and Naïve Bayes Algorithms", International Journal of Science, Engineering and Technology Research (IJSETR) Volume 4, Issue 4, April 2015

[4] Nazmun Nahar1 and Ferdous Ara, "Liver Disease Prediction by Using Different Decision Tree Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.8, No.2,2018

[5] M. Banu Priya, P. Laura Juliet, P.R. Tamilselvi "Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET),2018

[6] Chieh-Chen Wu, Wen-Chun Yeh, Wen-Ding Hsu, Md. Mohaimenul Islam, Phung Anh (Alex) Nguyen , Tahmina Nasrin Poly, Yao-Chin Wang, Hsuan-Chia Yang, Yu-Chuan (Jack) Li ,"Prediction of fatty liver disease using machine learning algorithms",2019

[7] Elias Dritsas, Maria trigka, "Supervised Machine Learning Models for Liver Disease Risk Prediction", 2023

[8] Neha Tanwar1 and Khandakar Faridar Rahman, "Machine Learning in liver disease diagnosis: Current progress and future opportunities", IOP Publishing,2021

[9] Hands-On Machine Learning with Scikit-Learn and TensorFlow, by Aurélien Géron Released March 2017, Publisher(s): O'Reilly Media,

[10] Introduction to Machine Learning with Python: A Guide for Data Scientists - Andreas C, Andreas C. Müller, Sarah Guido "O'Reilly Media, Inc.", 2016

[11] Lee, J. H., Joo, I., Kang, T. W., Paik, Y. H., Sinn, D. H., Ha, S. Y., ... & Bang, W. C. (2020). Deep learning with ultrasonography: automated classification of liver fibrosis using a deep convolutional neural network. European radiology, 30, 1264-1273.

[12] Sontakke, S., Lohokare, J., & Dani, R. (2017, February). Diagnosis of liver diseases using machine learning. In 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI) (pp. 129-133). IEEE.

[13] Wu, C. C., Yeh, W. C., Hsu, W. D., Islam, M. M., Nguyen, P. A. A., Poly, T. N., ... & Li, Y. C. J. (2019). Prediction of fatty liver disease using machine learning algorithms. Computer methods and programs in biomedicine, 170, 23-29.

[14] Ma, H., Xu, C. F., Shen, Z., Yu, C. H., & Li, Y. M. (2018). Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China. BioMed research international, 2018.