iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53305

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Diagnosis of Chronic Liver Disease using Machine Learning Techniques

Kalyani Biju[1], Maneesha B Manoj[2], Snehagee K S[3], Sreelakshmi P P[4], Hima Mohan[5]
[1,2,3,4]*Student,* [5]*Assistant Professor, Department of Computer Science and Engineering, Sree Buddha College of Engineering, Elavumthitta, Kerala, India*

Abstract: ***Liver disease is one of the deadly diseases. This issue has been increasing all over the world. Liver disease is a process that may cause the destruction and regeneration of the liver parenchyma. Many diseases and conditions and also use of some drugs may also cause liver diseases. So early detection and treatment can recover the disease in its early stage. Machine learning algorithms like Logistic Regression, KNN and Random Forest algorithms can be applied to detect the liver related diseases. Four stages of liver diseases such as healthy liver, fatty liver, liver fibrosis and liver cirrhosis are detected in this system. Liver disease prediction involve various steps such as pre-processing, feature extraction and classification. Datasets are collected from the Kaggle database of Indian liver patient records***
Keywords: Machine Learning, Liver Diseases, KNN, Random Forest, Logistic Regession

I. INTRODUCTION

Early diagnosis of liver disease is very important in order to prevent from serious damage to the liver and safe from liver transplant. The liver is a large organ found at the top of the abdomen (tummy area) on the right side. It is made up of cells; blood vessels and bile ducts. The main cells in the liver are known as hepatocytes. Chronic liver disease starts when the scar tissue takes place of healthy tissue in the liver of the human body. The liver separates nutrients and waste as they move through your digestive system. It also produces bile, a substance that carries toxins out of your body and aids in digestion. The term “liver disease” refers to any conditions that can affect and damage liver. According to a report, chronic liver disease was the reason for 65% of liver disease increase in deaths between 1999 and 2016. The liver performs major filtering systems in our body. The main functions of the liver include removal of waste from the body, generation of amino acids, generation of blood clots, creation of new proteins, and production of bile for digestion. Moreover, it metabolizes medicines into the ingredient required for the body's healthiness. The growth of liver disorder causes muscle loss, itching, weight loss, and kidney failure in human beings. The improper functioning of the liver due to chronic liver disease causes ill effects in the entire body. These effects can be the reason for jaundice, high blood pressure, and a swollen abdomen, etc. There is the possibility of developing even more severe symptoms in case liver disease turns extreme.

Machine learning algorithms can be applied to predict liver disease. Logistic Regression, KNN, Random Forest can be used to predict the liver disease. Four stages of liver disease such as fatty liver, liver fibrosis, liver cirrhosis and healthy liver can be predicted using machine learning algorithms. Prediction of liver disease include several steps such as data pre-processing, feature selection, classification, performance evaluation, performance analysis and prediction.

1) In data pre-processing, the collected data is pre-processed because the data may contain missing values or some redundant values. The missing values are replaced with the missing values or closest value to their feature. The dataset is divided into training set and testing set.
2) Several features of the patient such as age, gender and also some other clinical features are also collected in the feature selection step.
3) Classification is an important process. The features are inserted in a classification model after data pre-processing. Classification aims to get the target class to predict accurately for all case data.
4) Performance evaluation involve computation of different performance measures such as accuracy, precision and recall
5) In performance analysis, each classification model performance is analyzed.
6) Prediction is done by giving several features to the trained model. The classification of data is carried out into different classes. Hence every class refers to a certain category of liver diseases. This process can predict the corresponding liver disease of a patient.

II. LITERATURE SURVEY

A. *Pathogenesis And Prevention Of Hepatic Steatosis*

In the system Hepatic Steatosis is detected by the accumulation of Triacylglycerol in liver through blood. Hepatic Steatosis or fatty liver is defined as intrahepatic TAG of at least 5% of liver weight or 5% of hepatocytes containing lipid vacuoles in the absence of a secondary contributing factor such as excess alcohol intake, viral infection, or drug treatments. This article summarizes the mechanism involved in the accumulation of triacylglycerols in the liver. The advantage of using this system is low cost the procedure is very safe and not risky.

B. *Consideration Of Ultrasound Scanning Approaches In Nonfatty Liver Disease Assessment Through Acoustic Structure Qualification*

In this system Acoustic Structure Quantification (ASQ) based on statistical analysis of ultrasound echoes technique used to detect the hepatic steatosis. Non-alcoholic fatty liver disease (NAFLD) is a risk factor for hepatic fibrosis and cirrhosis. Hepatic fat fractions of 70 living donors were assessed with magnetic resonance spectroscopy. A clinical Ultrasound machine equipped with a 3-MHz convex transducer was used to scan each participant using the intercostal, epigastric and subcostal planes raw data for estimating raw data for estimating two ASQ Parameter. ultra sound are excellent for finding nodular liver surface, round edges and hypo echoic nodules in liver. Intercostal imaging is an appropriate scanning approach for ASQ analysis of the liver.

C. *Presence And Significance Of Microvesicular Steatosis In Non-Alcholic Fatty Liver Disease*

The presence of microvesicular steatosis was associated with higher grades of steatosis, ballooning cell injury and presence of megamitochondria.

The system is based on Multiple Logistic Regression Technique. The relationship between microvesicular steatosis and various histological features that characterize. NAFLD was tested by multiple logistic regression, after controlling for age, gender, race, body mass index and diabetes. The advantage of using the system is easier to implement, interpret and very efficient to train. The system need lot of data are needed for training.

D. *Non-Invasive Means Of Measuring Hepatic Fat Content*

Hepatic steatosis affects 20% to 30% o the general adult population in the western world. Currently, the technique of choice for determining hepatic fat deposition and the stage of fibrosis is liver biopsy. It may also be subject to sampling error. Non- Invasive techniques such as ultrasound, Computerized tomography (CT), magnetic resonance imaging (MRI) and proton magnetic spectroscopy (1M MRS) can detect hepatic steatosis, but currently cannot distinguish between simple steatosis and steatohepatitis or stage the degree of fibrosis accurately.

The advantage of using this system is Ultrasound, CT, MRI, HMRS method grade hepatic fat content accurately. Due to ionizing radiation the system limits its use in longitudinal studies and in children.

E. *Diagnois Accuacy Of Fibroscan And Factors Affecting Measurnments*

Evaluating Liver steatosis and fibrosis is important for patients with non-alcoholic fatty liver disease. Although liver biopsy and pathological assessment is he gold standard for these conditions, this technique has several disadvantages. The evaluation of steatosis and fibrosis using ultrasound B-mode imaging is qualitative and subjective. The liver stiffness measurements (LSM) and controlled attenuation parameter (CAP) determined using Fibroscan are the evidence- based non-invasive measures of liver fibrosis and steatosis respectively. The advantage of using this system is very useful and cost effectiveness. But the system has lower accuracy rate.

F. *Quantitative Ultrasound Aproaches For Diagnosis And Monitoring Hepatic Steatosis In Non-Alcholic Fatty Liver Disease*

Nonalcoholic fatty liver is a major global health concern with increasing prevalence, associated with obesity and metabolic syndrome. Recently, quantitative ultrasound-based imaging techniques have dramatically improved the ability of ultrasound to detect and quantify hepatic steatosis.

These newer ultrasound techniques posses man inherent advantages similar to conventional ultrasound such as universal availability, real-time capability and relatively low cost along with quantitative rather than a qualitative assessment of liver fat. The system is highly operator dependent.

G. Hepatic Steatosis Assessment Using Quantitative Ultrasound Parameter Imaging Based Scatter Envelope Statistics

Hepatic steatosis is a key manifestation of non-alcoholic fatty liver disease (NAFLD). Early detection of hepatic steatosis is of critical importance. Currently, Liver biopsy is he clinical golden standard for hepatic steatosis assessment. However, liver biopsy is invasive and associated with sampling errors. QUS envelope statistics techniques, both statistical model-based and non-model-based have shown potential for hepatic steatosis characterization. Ultrasound imaging which is based on the amplitude of the envelope of beamformed radiofrequency (RF) signals is frequently used in clinical routine for the assessment.

H. Quantitative Ultrasound Approaches For Diagnosis And Monitoring Hepatic Steatosis In Nonalcholic Fatty Liver Disease

Non-alcoholic fatty liver disease (NAFLD) is the leading cause of advanced liver diseases. Fat accumulation in the liver changes the hepatic microstructure and the corresponding statistics of ultrasound backscattered signals. Acoustic structure quantification (ASQ) is atypical model-based method for analyzing backscattered statistics. The advantage of using this system is ultra sound entropy imaging improved the performance of NAFLD evaluation. But this system is highly operator dependent.

I. Transient Elastography-Based Profiles In A Hospital Based Pdiatric Population In Japan

One of the most important indications for the use of fibroscan methodology in the evaluation of non-alcoholic fatty liver disease (NAFLD) which can vary in presentation from simple steatosis to non-alcoholic steatohepatitis (NASH) a leading cause of chronic liver disease in both adults and children. Degree of obesity and poor cooperation of young children should be weighted and risk factor for measurement failure.

J. Effect Of Fatty Infiltration Of The Iver On The Shannon Entropy Ultrasound Backscattered Signals

In this system clinical ultrasound image data were used for constructing entropy parametric images, which were compared with sonographic fatty sores to explore the effects of fatty liver on the signal uncertainty of ultra sound signals. Information entropy is suggested to be combined with standard ultrasound B-mode scanners for the routine examination of fatty liver disease.

III.METHODOLOGY

This work focus on the predictive analysis of liver related disorders. This research work consists of the basic three pillars. These three pillars include: pre-processing, feature extraction and classification.

1) Input data values: The dataset is taken as input from the Kaggle database. The dataset has 582 instances and 10 attributes and one target (Age, Gender).
2) Data Pre-Processing: In the initial stage , apply to remove missing and redundant values from the datasets.
3) Division of Input Data: In the second step, the partitioning of whole data is carried out into two sets of training and testing. The first set of training occupies 70% part of the overall data while the rest of the part will signify the data for testing purposes.
4) Classify Data: To predict final live or non liver disease technique of voting classification is applied is applied in this work. The voting classification is the combination of logistic regression, decision tree and KNN classifiers.

A voting classifier is a machine learning estimator that trains various base models or estimators and predicts on the basic of aggregating the findings of each base estimator. It is a technique that may be used to improve model performance , ideally achieving better performance than any single model used in the ensemble. There are two approaches to the majority vote prediction for classification ; they are hard voting and soft voting. Hard voting predict the class with the largest sum of votes from model. Soft voting predict the class with the largest summed probability from models. The models used in the ensemble must agree with their predictions.

Logistic regression is a statistical method used for classify dataset that includes one or more independent variables for determining result. The classification output represents the value of one among two possible results. This approach distinguishes the software modules as detected or non-defected. This algorithm also makes use of some metrics information for classifying the software elements similar to the voting classification is the combination of logistic regression, decision tree, and KNN classifiers. Logistic regression is a statistical method used for classifying dataset that includes one or more independent variables are 0 more independent variables can be also interpreted using this model. KNN is an upfront classifier. This classifier takes into account all possible outcomes. This classifier make use of the similarity approach for classifying a pattern. The classification of patterns is carried out on the basis of the voting majority of its neighboring patterns. A class is allowed to every pattern with the maximum number of corresponding votes and 1. This model can be used for data analysis. The association amid one dependent binary variable and one or based on a distance factor in the classification outcome.

Random Forest builds multiple decision tree and merge them together for more accuracy and stable prediction.

a) Prepare training set based on all classifier.
b) Apply voting classification for prediction.
c) Analyze performance in terms of Accuracy, precision and recall.

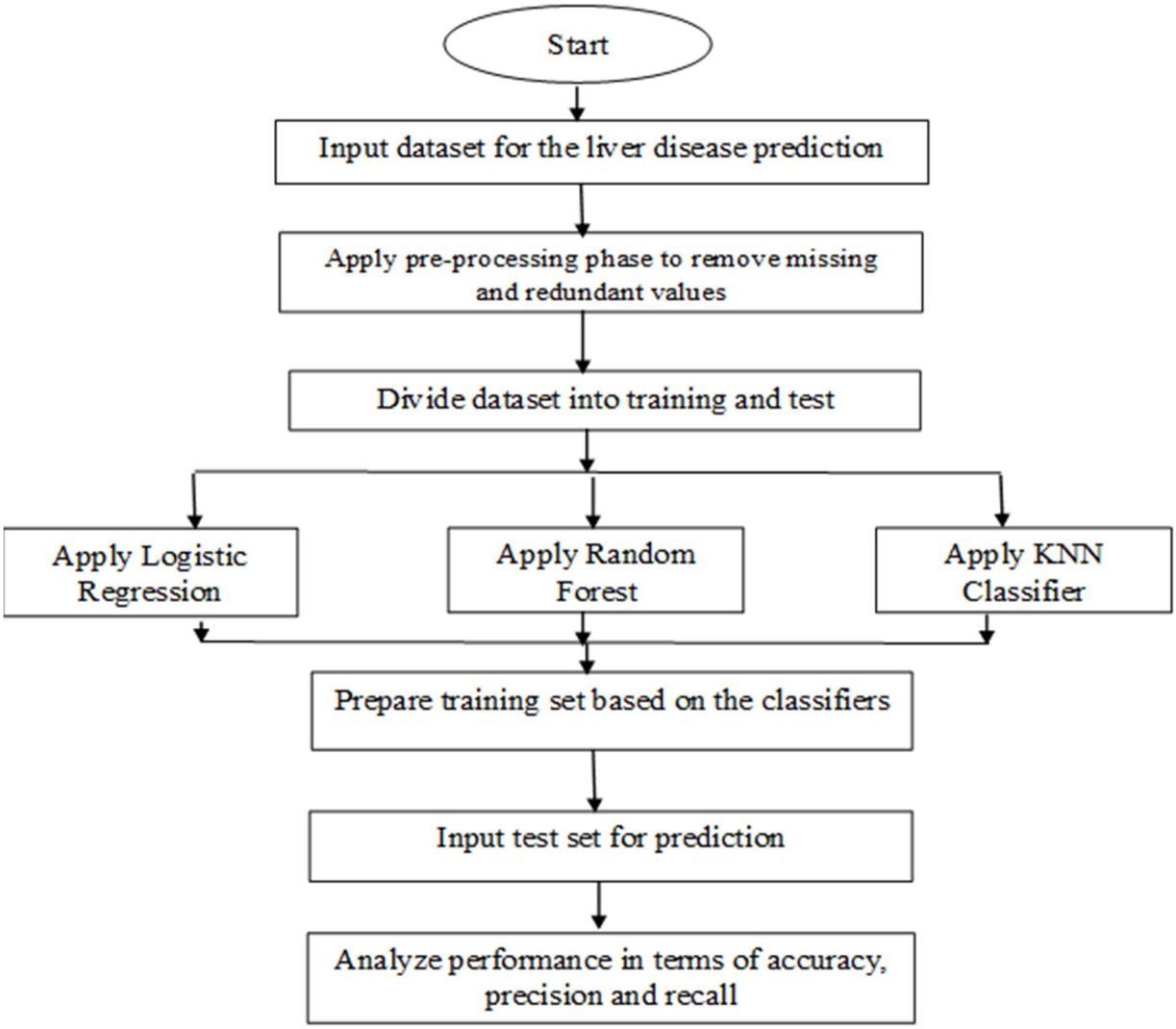


Fig 1 System Architecture

IV. RESULTS

In this system, machine learning algorithms are used for prediction of liver disease. This prediction model involves three classifiers such as Logistic Regression, Random Forest, and KNN classifiers. Dataset is from Kaggle database of Indian liver patient records. Different evaluation parameters such as accuracy, precision and recall are considered for the prediction model. Accuracy, precision and recall of the three different classifiers are also considered.

A. Accuracy

Accuracy is the degree of closeness between a measurement and its true value. Accuracy score is calculated by dividing the number of correct predictions by the total prediction number.

Accuracy= (TP + TN) / (TP +TN + FP + FN)

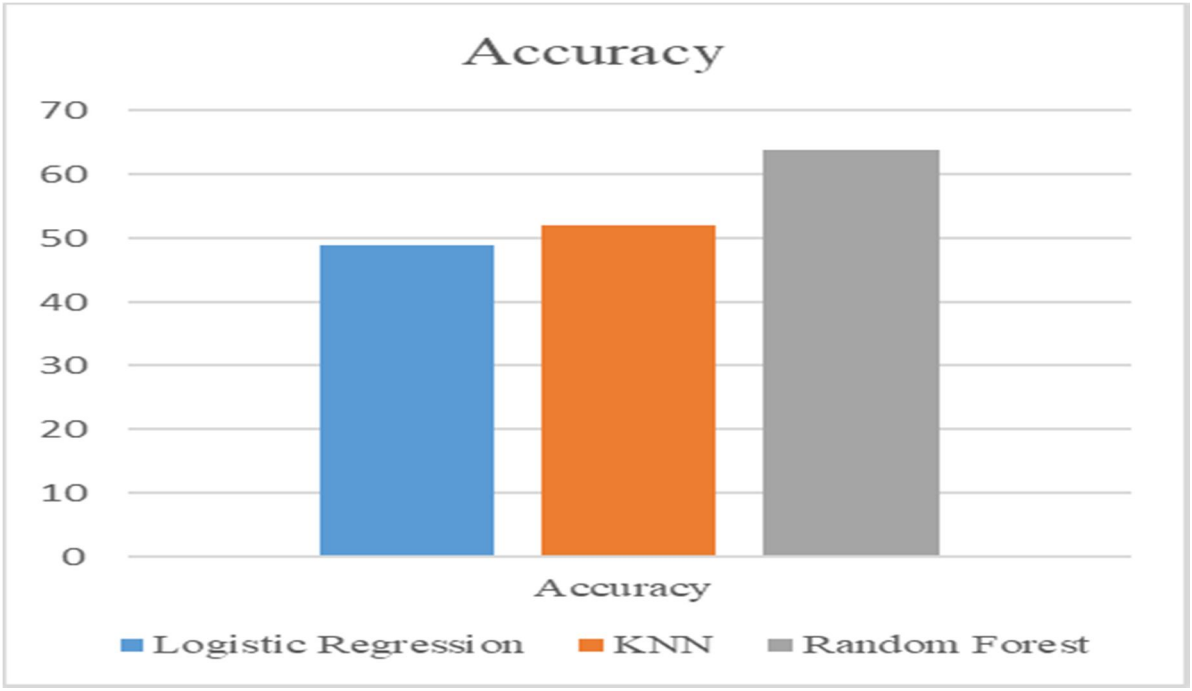


Fig 2 Accuracy

B. Precision

Precision is a metric that measures the accuracy of positive predictions. It is the number of true positive predictions divided by the number of true positive predictions plus false positive predictions.

Precision= (TP) / (TP + FP)

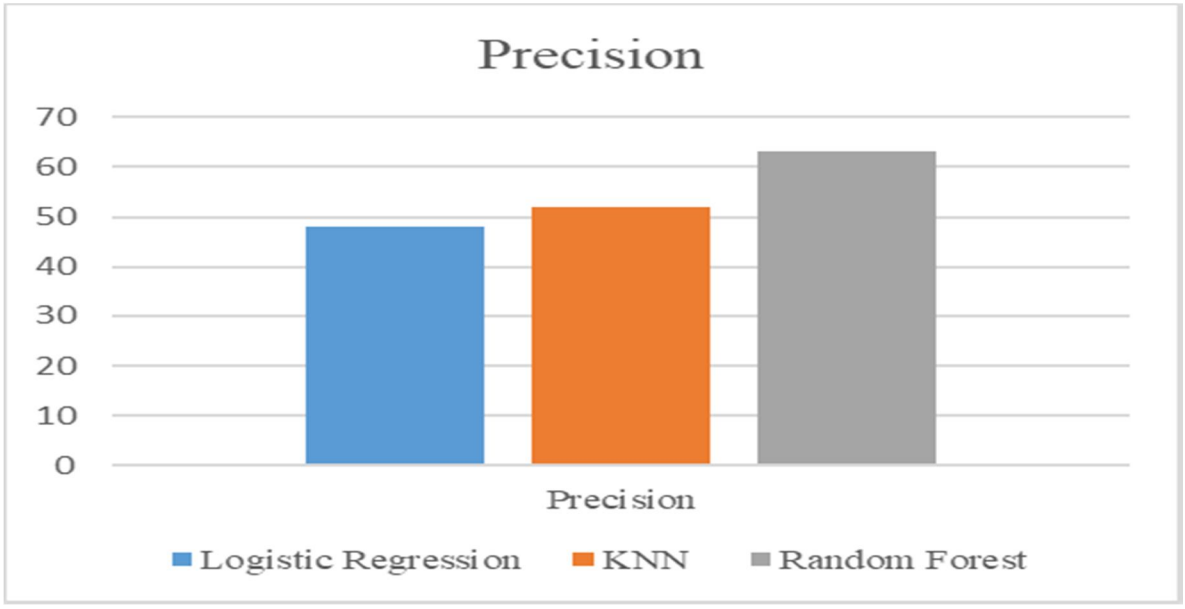


Fig 3 Precision

C. Recall

Recall measures the completeness of positive predictions. It is the number of true positive predictions divided by the number of true positive predictions plus false negative predictions.

Recall = (TP) / (TP + FN)

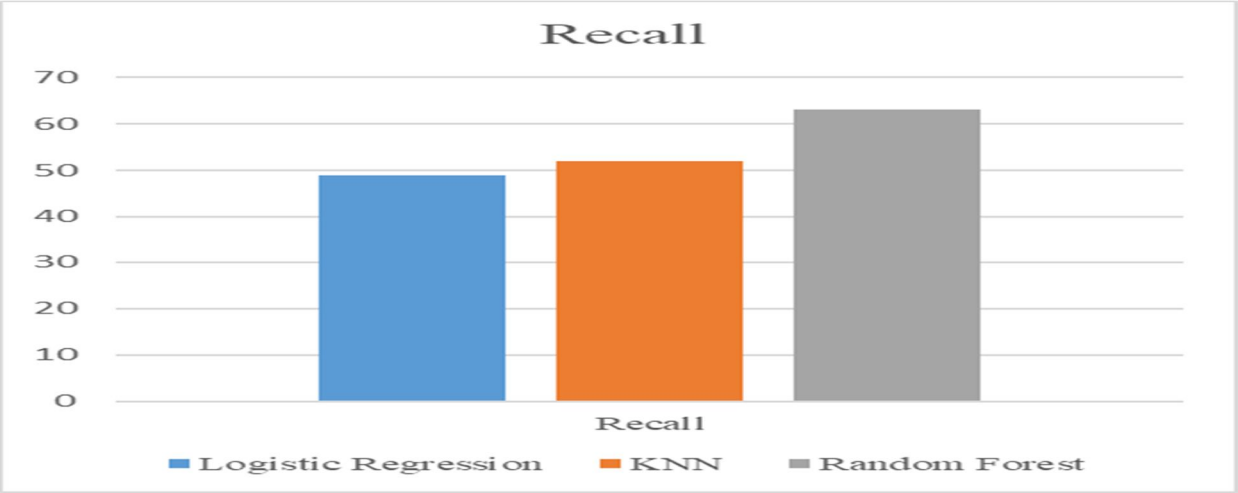


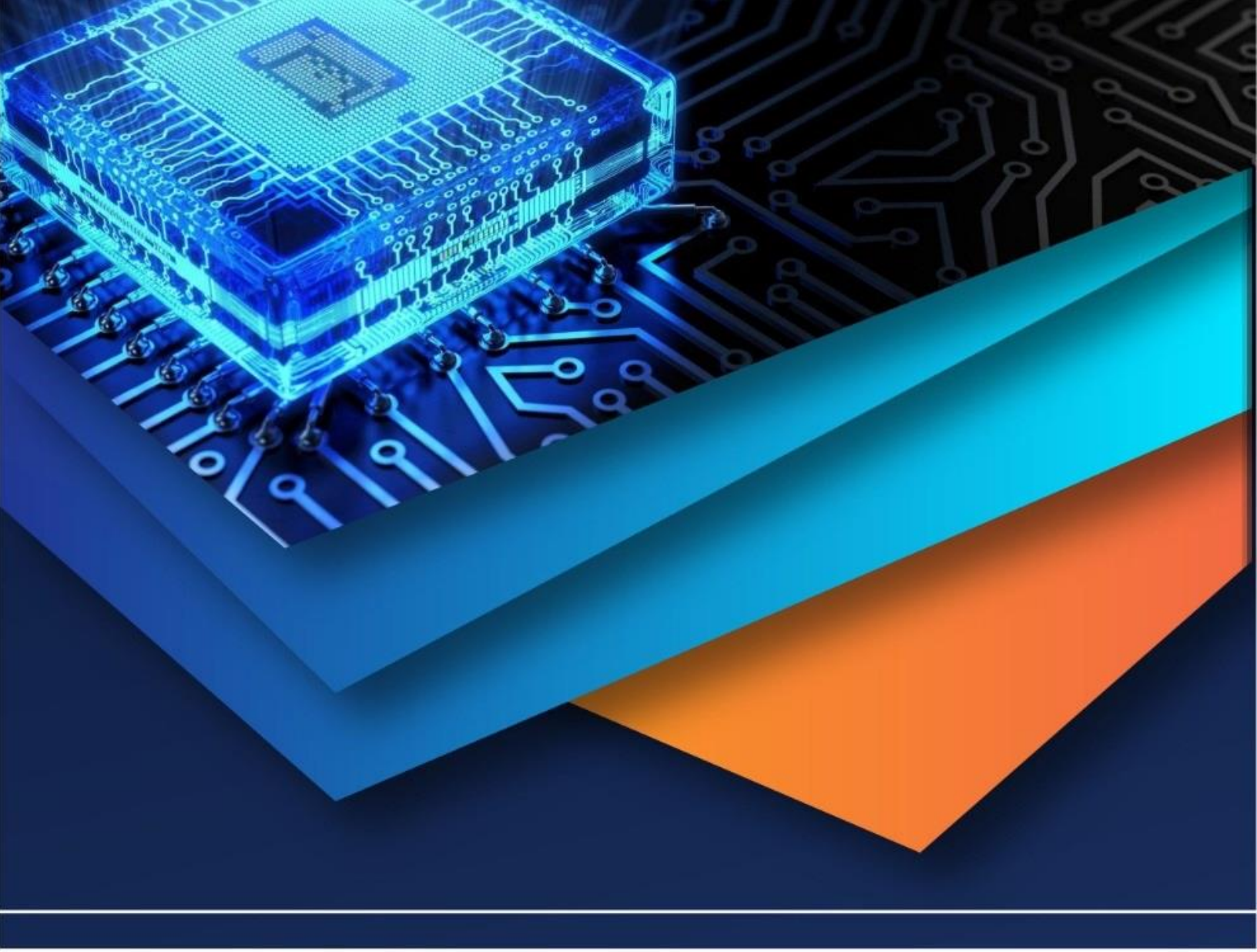
Fig 4 Recall

V. CONCLUSION

Chronic Liver disease is also considered to be one of the deadly diseases, so early detection and treatment can recover the disease easily. 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. The prediction is employed to obtain the result of liver diseases with accuracy. The discovery of associations among independent and dependent variables is done with the help of prediction as suggested by its name. The hidden knowledge of liver disease is recognized and extracted using a historical liver disease database. The complex queries are responded to diagnose liver disease. Thus, practitioners can make intelligent clinical decisions. To conclude, the approach of the predicting analysis helps in predicting future possibilities by current data. The proposed model improved by applying a combination of three classifiers, Logistic regression, Random forest, and KNN algorithm. This will improve the life span of a patient suffering from Chronic Liver Disease (CLD) in early stages. The data for this purpose is taken from the Kaggle database of Indian liver patient records. This work considers different parameters in terms of accuracy, precision, and recall for analyzing the efficiency of the newly devised prediction model. The python is employed for the implementation of the suggested model and the result proved regarding accuracy that is achieved 65 percent.

REFERENCES

[1] H.-J. Chan et al.: Ultrasound Sample Entropy Imaging: New Approach for Evaluating Hepatic Steatosis and Fibrosis .
[2] F. Nassir, R. S. Rector, G. M. Hammoud, and J. A. Ibdah, ‘‘Pathogenesis and prevention of hepatic steatosis,’’ Gastroenterol. Hepatol., vol. 11, no. 3, pp. 167–175, Mar. 2015
[3] S. R. Mehta, E. L. Thomas, J. D. Bell, D. G. Johnston, and S. D. Taylor-Robinson, ‘‘Non-invasive means of measuring hepatic fat content,’’ World J. Gastroenterol., vol. 14, no. 22, pp. 3476–3483, Jun. 2008
[4] A. M. Pirmoazen, A. Khurana, A. El Kaffas, and A. Kamaya, ‘‘QUS approaches for diagnosis and monitoring hepatic steatosis in NAFLD,’’ Theranostics, vol. 10, no. 9, pp. 4277–4289, 2020
[5] Z. Zhou, Q. Zhang, W. Wu, S. Wu, and P.-H. Tsui, ‘‘Hepatic steatosis assessment using quantitative ultrasound parametric imaging based on backscatter envelope statistics,’’ Appl. Sci., vol. 9, no. 4, p. 661, Feb. 2019.
[6] S. Oeda, K. Tanaka, A. Oshima, Y. Matsumoto, E. Sueoka, and H. Takahashi, ‘‘Diagnostic accuracy of FibroScan and factors affecting measurements,’’ Diagnostics, vol. 10, no. 11, p. 940, Nov. 2020 .
[7] Y.-H. Lin et al., ‘‘Considerations of ultrasound scanning approaches’’ Ultrasound Med. Biol., vol. 45, no. 8, pp. 1955–1969, Aug. 2019.
[8] S. Tandra et al., ‘‘Presence and significance of microvesicular steatosis in NAFLD’’ J. Hepatol., vol. 55, no. 3, pp. 654–659, Sep. 2011.
[9] P.-H. Tsui and Y.-L. Wan, ‘‘Effects of fatty infiltration of the liver on the Shannon entropy of ultrasound backscattered signals,’’ Entropy, vol. 18, no. 9, p. 341, Sep.
[10] K. Pu et al., ‘‘Diagnostic accuracy of (CAP) as a non-invasive test for steatosis in NAFLD,’’ BMC Gastroenterol., vol. 19, no. 1, p. 51, Apr. 2019.
[11] Y. Cho et al., ‘‘Transient elastography-based liver profiles in a hospitalbased pediatric population in Japan,’’ PLoS ONE, vol.10,no.9,Sep.2015, Art.no.e0137239.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY