INTERNATIONAL JOURNAL OF ADVANCED INNOVATIVE TECHNOLOGY IN ENGINEERING

Published by Global Advanced Research Publication House Journal

Home page: www.ijaite.co.in

# Analysis And Prediction of Liver Cirrhosis Using Machine Learning

[1]Ms. M. G. Muthal, [2]Dr. A. B. Gadicha

[1,2]Department of Computer Science & Engineering, P.R. Pote Patil College of Engineering & Management Amravati, Maharashtra, India

[1]mohinimuthal@gmail.com, [2]ajjugadicha@gmail.com

**Article History**

Received on: 15 April 2022

Revised on: 12 May 2022

Accepted on: 31 May 2022

**Keywords:** Liver Cirrhosis, Machine Learning, SVM, Random Forest, Decision Tree, Naive Bayes

**e-ISSN:** 2455-6491

**Production and hosted by**
www.garph.org

## ABSTRACT

The death rate in India is high due to Liver Cirrhosis as a result of a bad lifestyle, storage of food, uncontrolled blood sugar, obesity, smoking, consumption of alcohol, and inhaling of harmful gases. Earlier detection can reduce death rates and it also helps the doctors to give the proper treatment to the patients. The liver Cirrhosis datasets are analyzed by using Machine learning algorithms for accurate disease diagnosis. This paper proposed the four machine learning models such as SVM, Random Forest, Decision Tree, and Naive Bayes for analysis and prediction of liver cirrhosis. This work gathers 200 pictures of two separate classes, i.e., healthy liver, and unhealthy liver, using an image source. This image source is MedPix1, a free open-access archive of digital photographs that medical schools, medical practitioners, and academics can use. The methodology actually has a good result for the various classifiers but still can be improved. It is found that the Support vector machine and random forest classifiers give a 100% classification accuracy score followed by decision trees which are 84.61. The results achieved are relatively good when compared.

## 1. INTRODUCTION

In the human body after the skin, the largest organ is the liver. The weight of the liver of an adult is approximately three pounds. The liver is located on the right side under the right lung and is protected by the rib cage. It is segregated into right and left lobes. It is like a chemical factory. The liver plays many functions to remain healthy and it converts food into proteins and bile that are essential for digestion. Ingested and toxic substances from the blood are eliminated. It breaks many nutrients from the intestine and controls the excretion of cholesterol producing quick energy at times of need with the help of vitamins, sugars, and minerals stored in the liver. The cell is the basic element that builds up the tissues in the body. The normal property of the cells is to grow and to divide into new cells. Whenever the cell becomes old or damaged new cell takes place of it. Sometimes the procedure goes wrong. The body does not produce new cells on the other hand nodules or tumors are formed by the tissues of old or damaged cells. Benign and malignant are the two categories of liver tumor. Benign is safer than a malignant tumor.

**Benign tumors:** Benign tumors are rarely harmful to life. Unlike malignant, if it is once removed it does not generally grow again. It does attack tissues around them and does not extend to another part of the body.

**Malignant tumors:** Malignant are harmful to life. It is very unsafe and grows again even if is detached from the body. It can be fatal for the stomach or intestine. It spreads and reaches another area of the body. Liver cancer is of two types primary and secondary. Whenever tumor (malignant) starts in the liver itself, it is called primary liver cancer. Secondary liver cancer starts in another part of the body and spreads in the liver. Tumor in hepatocytes is called hepatocellular carcinoma (HCC). It is primary liver cancer. Around 75-90 percent of liver cancer is HCC. There are other primary liver cancers such as cholangiocarcinoma or bile duct cancer, mixed HCC and cholangiocarcinoma tumor of mesenchymal tissue, sarcoma and hepatoblastoma, a rare malignant tumor found in children.

Liver diseases cause millions of deaths every year. Liver Cirrhosis and Viral hepatitis alone causes 1.34 million deaths every year. Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. An early diagnosis of liver problems will increase patient's survival rate. Liver failures are at a high rate of risk among Indians. It is expected that by 2030 India may become the World Capital for Liver Cirrhosis Diseases. The widespread occurrence of liver infection in India is contributed due to a deskbound lifestyle, increased alcohol consumption, and smoking. There are about 100 types of liver infections. With such alarming figures, it is necessary to have concerned about tackling these diseases. After all, one cannot expect a developed and prosperous nation, with unhealthy youths.

The main motivation to conduct the study presented in this study was the fact that due to the development of improved solutions for prediction risk of bleeding and thus a faster and more accurate diagnosis of complications in cirrhotic patients, mortality of cirrhosis patients caused by bleeding of varices fell at the turn in the 21th century. Due to this fact, additional research in this field is needed.

Cirrhosis is defined as the histological development of regenerative nodules surrounded by fibrous bands in response to chronic liver injury, that leads to portal hypertension and end-stage liver disease. Recent advances in the understanding of the natural history and pathophysiology of cirrhosis, and in the treatment of its complications, result in improved management, quality of life, and life expectancy of cirrhotic patients. At present, liver transplantation remains the only curative option for a selected group of patients, but pharmacological therapies that can halt progression to decompensated cirrhosis or even reverse cirrhosis are currently being developed. This concise overview focuses on the diagnosis, complications, and management of cirrhosis, and novel clinical and scientific developments based on machine learning and deep learning.

The contribution of this paper is

- The main objective of this research is to use machine learning algorithms for classification to identify liver patients from healthy individuals.
- Design a machine learning model for predicting liver disease.
- Increased convenience for predicting a liver disease
- Improve the accuracy of liver Cirrhosis detection.

## 2. RELATED WORK

Jagdeep Singh et. al. (2020), proposed work is done on Indian liver patient dataset from the University of California Irvin dataset. The different attribute like age, direct, bilirubin, gender, total bilirubin, alkphos, sgpt, albumin, globulin ration and sgot etc, of the liver patient dataset, are used to predict the liver disease risk level [1].

Xieyi Pei et. al. (2021), demonstrated that ML models give more accurate predictions, the best accuracy reached to 0.9415 in the XGBoost model. Feature importance analysis not only confirmed some well-known FLD risk factors, but also demonstrated several novel features for predicting the risk of FLD, such as hemoglobin. Classification models were used in the detection of the disease. These models include logistic regression (LR), random forest (RF), artificial neural networks (ANNs), k-nearest neighbors (KNNs), extreme gradient boosting (XGBoost), and linear discriminant analysis (LDA). Predictive accuracy was assessed by area under curve (AUC), sensitivity, specificity, positive predictive value, and negative predictive value [2].

Aixia Guo et. al. (2021), proposed the deep learning model for predicting longer term mortality among patients with liver cirrhosis more reliably than the MELD-Na variables alone using common EHR data variables. Our findings suggest that newer informatics methods might benefit patients who are inadequately triaged by the MELD-Na score. Future work should validate this methodology in actual patient data and incorporate competing the competing risk of transplant to avoid mortality.

Rashid Naseem et. al. (2020), proposed the model for early prediction of liver disease utilizing various machine learning approaches. However, this study compares ten classifiers including A1DE, NB, MLP, SVM, KNN, CHIRP, CDT, Forest-PA, J48, and RF to find the optimal solution for early and accurate prediction of liver disease. The datasets utilized in this study are taken from the UCI ML repository and the GitHub repository. The outcomes are assessed via RMSE, RRSE, recall, specificity, precision, G-measure, F-measure, MCC, and accuracy. The exploratory outcomes show a better consequence of RF utilizing the UCI dataset. Assessing RF using RMSE and RRSE, the outcomes are 0.4328 and 87.6766, while the accuracy of RF is 72.1739% that is also better than other employed classifiers. However, utilizing the GitHub dataset, SVM beats other employed techniques in terms of increasing accuracy up to 71.3551%. Moreover, the comprehensive outcomes of this exploration can be utilized as a reference point for further research studies that slight assertion concerning the enhancement in extrapolation through any new technique, model, or framework can be benchmarked and confirmed.

Aleksandar Aleksic et. al. (2021), proposed one ensemble data mining methodology, as the most modern in the field of prediction, for integrating on one new way the two most commonly used techniques in prediction, classification with precede attribute number reduction and multiple logistic regression for calibration. The method was evaluated in the study, which analyzed the occurrence of variceal bleeding for 96 patients from the Clinical Center of Nis, Serbia, using 29 data from clinical to the color Doppler. Obtained results showed that proposed method with such big number and different types of data demonstrates better characteristics than individual technique integrated into it.

Nakao et al. (2018) proposed a nationwide survey for acute liver failure to illuminate current status of hepatic patient in particular city. The advantage is that the clinical feature of patient could be derived and compared with the existing surveys. The drawback is that it is necessary for patient to recover the liver failure as earlier as possible. The experimental result of proposed system achieves better result than existing.

Kant and Ansari (2016) proposed an initial centroid selection method for k-mean algorithm by utilizing Atkinson index to estimate the inequality efficiency and accuracy of datasets for liver patient. K-mean clustering is a popular approach. The drawback is that it is sensitive as the limited number of outliers result could be influenced. A clustering is a method where the dataset had been decomposed into several clusters. The other drawback is that the random selections of initial centroids are sensitive. To overcome drawbacks, the k-means clustering algorithm had been utilized. The advantage is that provides effective clustering in k-means. The experimental result of proposed system achieves accurate result than existing.

Pakhale and Xaxa (2016) proposed a survey of liver disease classification to categorize several techniques such as decision tree, support vector machine, naïve bayes and artificial neural network. A data mining plays a significant part in health care industry. The advantage was that predicted classification was accurate. The disadvantage is that it is necessary to protect classified dataset of liver patient within specified period of time. The experimental result of proposed system attains better performance than existing.

Abdel-Basset et al. (2018) proposed a comprehensive quadratic assignment problem to provide appropriate significant application such as scheduling, production and communication process. The advantage is that bound of hard problem provides simplicity and less computational cost. The disadvantage is that if the problem size got increased, then it is impossible to resolve polynomial time within particular period of time. The main aim was to derive complexity and nature of problem. A quadratic assignment problem had been deliberated as major combinatorial optimization assortment problems. The experimental result of proposed system achieves better optimization problem than existing.

Desir et al. (2020) proposed a capacity constrained assortment problem-based Markov chain to select a subsection of problems are delivered from substitutable so as to increase assorted revenue exhibited from random substitution behaviour of patient. The advantage is that normal Markov chain model provides constant approximation factor for both cardinality and capacity constrained assortment optimization problem. The disadvantage is that it is essential for user to provide optimal approximation ratio than other method. The experimental result of proposed system achieves better result than existing.

Sumida et al. (2019) proposed a revenue utility assortment problem to increase linear combination among firm and user expected utility. Based on the multinomial logit method, delivered assorted problem could be maximized by uni-modular matrix. A parametric linear method had been utilized to produce number of users in assortment problem. The advantage is that user maintains better quality of assorted solution. The disadvantage is that it is required to estimate assorted value within certain period of time. The experimental result of proposed system achieves feasibility than existing.

Zhang et al. (2020) offered incapacitated and capacitated assortment problems through paired combinatorial logit method to identify number of maximum expected products acquired from individual user. An assortment had been transformed into corresponding problem in order to identify fixed point value function by resolving nonlinear integer program. The advantage of proposed system is that NP hard problem provides approximate non-linear integer program and randomized rounding. It is flexible because normal correlation structure between user utility. The disadvantage is that multinomial method does not hold any value than choice models. The experimental result of the proposed system achieves better optimized result than existing.

Singh and Pandey (2018) suggested an integrated intelligent diagnosis system for principle component analysis and k-nearest neighbor to scrutinize dataset for liver disorder patients. Liver played an important part for human to perform different functions such as chemical detoxification, protein production, drug metabolizing, blood clotting, glucose storage, cholesterol production and bilirubin clearance. There are also other early symptoms caused by liver disease are nausea, fatigue, weight loss, poor appetite and upper right quadrant abdominal pain. To overcome this drawback, linear and quadratic principal component analysis had been developed. The advantage is that it provides efficient ability for resolving different problems through feature extraction methods. A computer aided diagnostic system had been implemented as a fruitful work for researching clinical process for liver patients. The other drawback is that assessment of liver disorder at initial phase became very problematic. A k-nearest neighbor network was utilized to differentiate between sick and healthy liver patients. The experimental result of proposed system achieves highly extracted classification rate than existing.

.

**Figure 1: Flow of the Facial Expression Recognition System**

| Author | Proposed Methodology | Advantage | Disadvantage |
|---|---|---|---|
| Jagdeep Singh et. al. (2020) | proposed work is done on Indian liver patient dataset from the university of California Irvin dataset. | To develop an intelligent liver disease prediction software based on feature selection and classification. To achieve a good accuracy level. | The number of relations become inconvenient. |
| Xieyi Pei et. al. (2021), | Implement the XG-Boost model, physicians can efficiently identify FLD in general patients; this would help in prevention, early treatment, and management of FLD. | It demonstrated that ML models give more accurate predictions, the best accuracy achieved. | Only appropriate features for liver disease disorder must be selected |
| Aixia Guo et. al. (2021) | Proposed the machine and deep learning models for predicting mortality of liver cirrhosis. | Advanced informatics techniques showed promise for risk prediction in patients with cirrhosis. | It requires time to understand patient details. |
| Rashid Naseem et. al. (2020) | Proposed the model for early prediction of liver disease utilizing various machine learning approaches such as A1DE, NB, MLP, SVM, KNN, CHIRP, CDT, Forest-PA, J48, and RF to find the optimal solution for early and accurate prediction of liver disease. | Reduced the error rate.<br>Overcome the overfitting problem<br>Better performance of optimal classifiers on both datasets concerning each assessment measured | Requires computation time |
| Aleksandar Aleksic et. al. (2021) | Proposed one ensemble data mining approach for prediction, classification of bleeding in liver cirrhosis. | The proposed ensemble model can be useful and extensible to other hospitals in the world treating this illness, the liver cirrhosis and its consequences as the bleeding of varices. | Many samples are incorrectly classified |
| Nakao et al. (2018) | A nationwide survey for acute liver failure to illuminate current status of hepatic patient in particular city. | The clinical feature of patient could be derived and compared with the existing surveys | It is necessary for patient to recover the liver failure as earlier as possible. |
| Kant and Ansari (2016) | An initial centroid selection method for k-mean algorithm by utilizing Atkinson index to estimate the inequality efficiency and accuracy of datasets for liver patient. | It provides effective clustering in k-means | The random selections of initial centroids are sensitive. |
| Pakhale and Xaxa (2016) | A survey of liver disease classification to categorize several It is necessary to techniques such as decision tree, support vector machine, Naïve bayes and artificial neural network. | Predicted classification was accurate. | It is necessary to protect classified dataset of liver patient within specified period of time. |
| Abdel-Basset et. al. (2018) | A comprehensive quadratic assignment problem to provide appropriate significant application production and communication process. | Bound of hard problem provides simplicity and less computational cost | If the problem size got increased, then it is impossible to resolve polynomial time within particular period of time. |
| Desir, A et. al. (2020) | Assortment optimization problem under Markov chain to select a subsection of problems that are delivered from substitutable liver patients. | Normal Markov chain model provides constant approximation factor for assortment optimization problem. | It is essential for optimal user to provide optimal approximation ratio than other method. |

## 3. PROPOSED METHODOLOGY

In this paper, the methodology of analysis of the classification process on collected data is discussed. The following figure shows the process for classification in which the liver cirrhosis data sets have been taken. Firstly, the set is used for data preprocess (such as cleaning, and missing values), and then supervised filters are used for feature extraction. Then classifiers such as SVM, Random Forest, Decision tree, Naive Bayes, etc. are applied using the python tool. The best classifier is then chosen based on the performance and error evaluation.

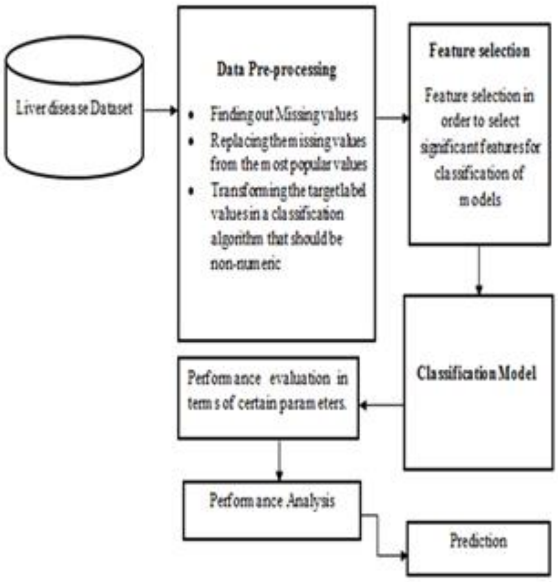

Figure 1: Prediction of liver disease using Machine learning

### A. *Dataset Description*

This is difficult and expensive to procure data sets containing actual patients' diagnostic image samples rather than digitally generated samples. This work gathers 200 pictures of two separate classes, i.e., healthy or normal liver and unhealthy or abnormal liver, using an image source. This image source is MedPix1, a free open-access archive of digital photographs that medical schools, medical practitioners and academics can use. This provides a wide variety of Liver CT scans to conduct this research. Originally developed by the Departments of the Radiology and Biomedical Informatics of Uniformed Services University, Bethesda, Maryland, USA, the National Library of Medicine is a free electronic reference image archive. Table 4.1 explains the image distribution into two classes. All the images are three-dimensional of various sizes and are processed in the following steps. The data preparation steps are addressed in the following section, considering all the information relating to image data.

Table 1: Image Distribution into Two Classes

| Sr. No. | Image Name | Number of Images |
|---|---|---|
| 1 | Normal or Healthy | 48 |
| 2 | Unhealthy Or Abnormal | 152 |

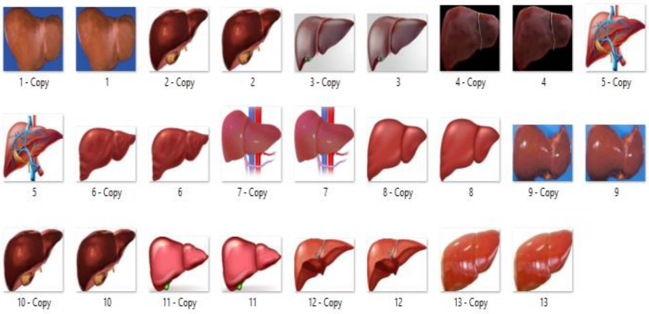

Figure 2: Sample dataset image of Normal or healthy liver

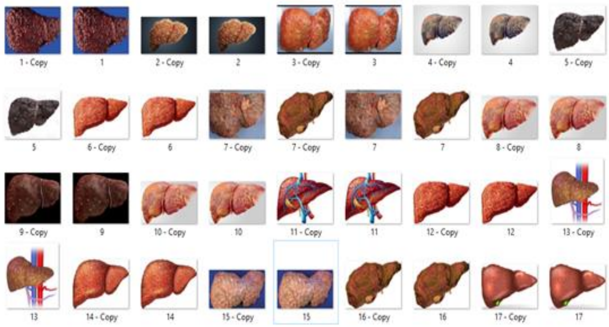

Figure 3: Sample dataset image of Unhealthy or Abnormal liver

The prediction of liver diseases includes several steps. These steps are:

- Data preprocessing
- Feature selection
- Data classification
- Performance analysis
- Prediction

### B. *Data Pre-Processing*

At first, the pre-processing of data is carried out once it is collected. The dataset consists of 200 CT scan images belonging to two classes. Some of the downloaded images were in different formats;

therefore, all the images' types were transformed to ".jpg" to maintain consistency and uniformity of data. The approach followed in this research splits the data into training and testing datasets such as 70% of the total images are used for training and the remaining 30% for testing the model.

In this step, several tasks are performed. And the liver disease target data are categorized into two groups i.e., group 1 represents the presence of cirrhosis liver disease and group 2 represents the healthy liver of the patient.

0-Healthy Liver
1-Cirrhosis Liver

### C. *Feature Selection*

One of the main segments in chronic liver disease prediction is the selection of important features of liver disorder.

### D. *Data Classification*

Four Machine learning algorithms such as SVM, Random Forest, Decision Tree, and Naive Bayes will be implemented for predicting liver diseases. The function of the collected items assigns to the target class or category. The classification aims to get the target class to predict accurately of liver diseases. After data pre-processing, features are inserted into a machine learning model.

### E. *Performance Evaluation*

Different classification criteria including False Positive Rate, True Positive Rate, Accuracy, Train Accuracy, Validation accuracy, and ROC are computed for performance evaluation of the classification. Performance analysis: In this step, the performance of the classification model is analyzed.

### F. *Performance Analysis*

In this step, the performance of the classification model is analyzed.

**Prediction:** In this step, the mapping of selected features is carried out onto the training model for classifying the given features so that liver disease can be predicted. To generate predictions, a specialist doctor labels the gathered liver disease dataset. The classification is developed as a multi-class issue and the classification of medical data is carried out into different classes. Hence, every class refers to a certain category of chronic liver disorder. This process can find the probabilities of a patient suffering from a liver disease based on the selected important features.

### G. *Classification Algorithms*

This section explains the architecture used for the Support Vector Machine, Random Forest, Decision tree, Naive Bayes models of machine learning.

**Support Vector Machine:** The support Vector Machine or SVM algorithm is a simple yet powerful Supervised Machine Learning algorithm that can be used for building both regression and classification models. SVM algorithm can perform really well with both linearly separable and non-linearly separable datasets. Even with a limited amount of data, the support vector machine algorithm does not fail to show its magic. The objective of the SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. In short, the hyperplane is (n-1)- D plane for n features.

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed Support Vector. Since these vectors support the hyperplane, hence called a Support vectors.

**Random Forest:** An RF is established with a numeral of decision trees, and every tree acquires its position arrangement effect by utilizing dissimilar classification. This method permits the evaluation of the sampling allocation utilizing the random sampling technique, also particularly appropriate for some minute models. The essential procedure of classification on basis of RF is as follows. Figure 4.4 depicts the flowchart for Random Forest.

The unique training illustration set is developed, in which the number of cases is X also the amount of contribution character is Y. This illustration will be the training set for increasing the tree.

2. A secondary training set is arbitrarily created by sampling with the substitution bootstrap technique for n tree times; hence, the subordinate training set for the RF with numeral n tree is created.

3. Ahead of choosing the characters (features) for every non-leaf node (internal node), this technique at random chooses a definite number of characteristics from all distinctiveness, utilizes them as divide characteristics of the existing decision tree, and chooses the optimal one to divide the nodes. The number of characters tried at

every division is indicated by mtry, mtry ≤ M.

4. Expect to prune, the tree expansion is increased.

5. The created trees are joint with RF. every tree in the RF transmits an entity choice for the mainly accepted group, and the classifier result is resolute by a mass choice of the trees.

6. Considering that set S comprises k sorts of attribute principles and every kind of attribute principle creates one sub-node.

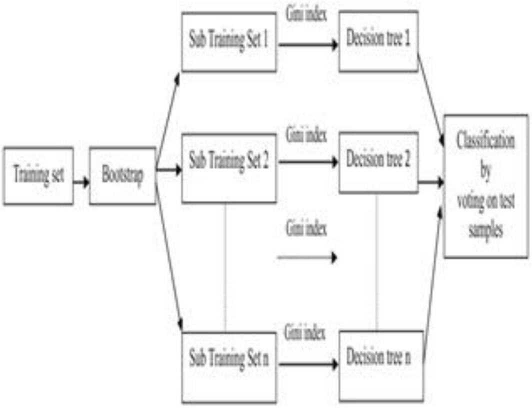

Figure 4: RF classification algorithm Flowchart

**Decision Tree:** The decision tree is a classifier algorithm in the structure form of a "tree". Decision Trees are simple, but very commonly used methods for moving the inductive logic into a programming environment. It works with discrete-valued parameters. The basic intuition about the inductive philosophy on which the decision tree algorithms are based is that a "good" decision tree to be constructed with learning characteristics should be small as possible.

**Naive Bayes:** Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features. They are among the simplest Bayesian network models but coupled with kernel density estimation, they can achieve higher accuracy levels. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

## 4. RESULT ANALYSIS

In this section of the research, the datasets used for Analysis and Prediction of Liver Cirrhosis and the results obtained in each implementation phase such as pre-processing, feature extraction, classification, and predicted results using the various machine learning classifiers have been presented. These results have been discussed and related implications have also been discussed. The proposed research work has been developed using python and anaconda software tools. In the thesis, the overall research objective for a novel and robust Analysis and Prediction system for Liver Cirrhosis has been realized in multiple phases.

### *A. Performance Metrics*

This section presents the experimental results and performance metrics for different models. Accuracy is one of the common performance metrics. It is the measure of all the correctly identified cases. It is mostly used when all the classes are equally important. Accuracy is the proportion of correctly classified examples to the total number of examples, while the error rate is incorrectly classified instead of correctly.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP is True-positive, TN is True-negative, FP is False-positive and FN is False-negative. The six models have yielded different accuracy scores, which are displayed in table 2.

Table 2: Accuracy Score of various Models

| Sr. No. | Classifiers | Accuracy in % |
|---|---|---|
| 1 | SVM | 100 |
| 2 | Random Forest | 100 |
| 3 | Decision Tree | 84.61 |
| 4 | Naive Bayes | 69.23 |

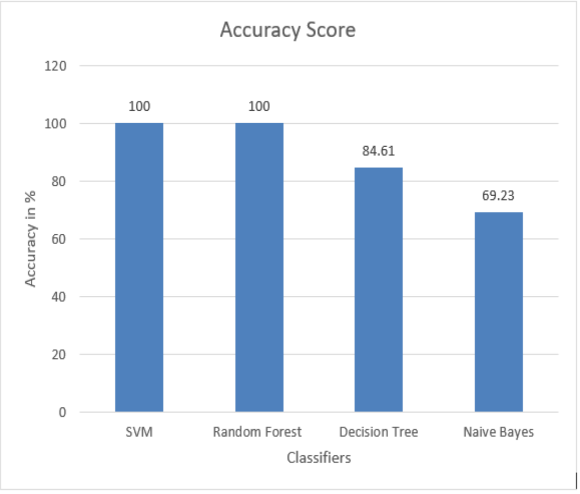

Figure 5: Bar graph of the accuracy measures of different models

Figure 5 shows the Bar graph of the accuracy measures of different models. The methodology actually has a good result for the various classifiers but still can be improved. It is found that Support vector machine and random forest classifiers gives the 100% classification accuracy score followed by decision tress which is 84.61. The results achieved are relatively good when compared.

### *B. Confusion Matrix*

A confusion matrix is a very popular measure used while solving classification problems. It can be applied to binary classification as well as for multiclass classification problems. Confusion matrices represent counts from predicted and actual values. The output “TN” stands for True Negative which shows the number of negative examples classified accurately. Similarly, “TP” stands for True Positive which indicates the number of positive examples classified accurately. The term “FP” shows False Positive value, i.e., the number of actual negative examples classified as positive; and “FN” means a False Negative value which is the number of actual positive examples classified as negative. One of the most commonly used metrics while performing classification is accuracy.

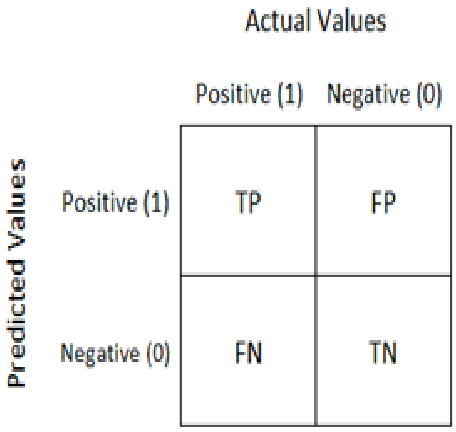

Figure 6: Representation of confusion matrix

### *C. Confusion Matrix and AUC Graph of Classifiers*

| | |
|---|---|
| 4 | 0 |
| 0 | 9 |

Figure 7: Confusion Matrix of SVM

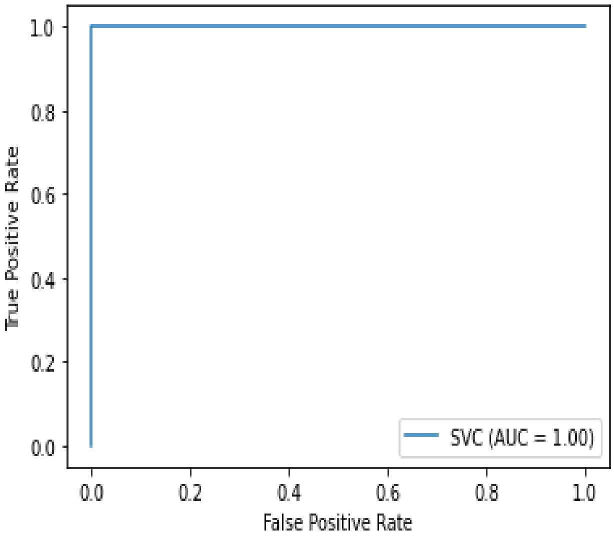

Figure 8: AUC Graph of SVM

| | |
|---|---|
| 4 | 0 |
| 0 | 9 |

Figure 9: Confusion Matrix of Random Forest

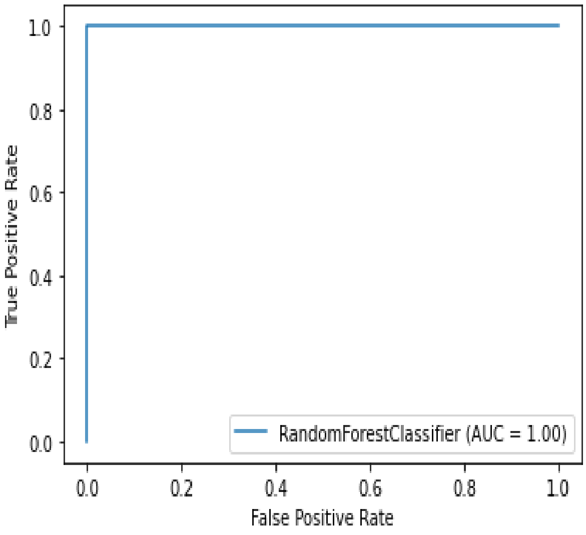

Figure 10: AUC Graph of Random Forest

| | |
|---|---|
| 4 | 0 |
| 2 | 7 |

Figure 11: Confusion Matrix of Decision Tree

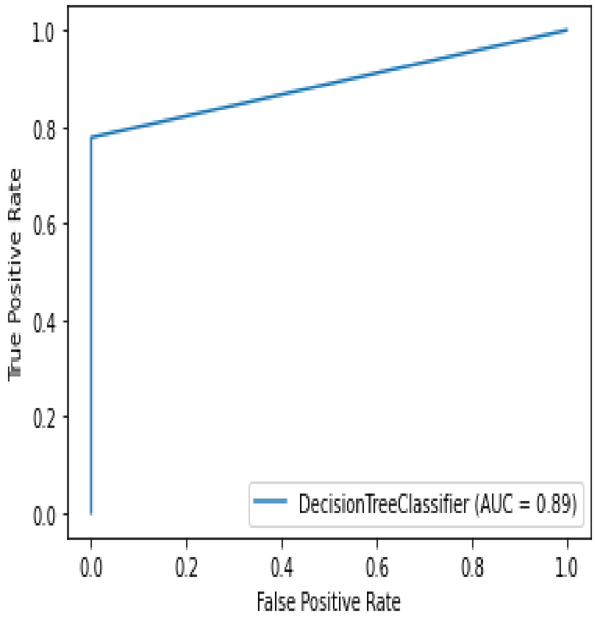

Figure 12: AUC Graph of Decision Tree

|  |  |
|---|---|
| 3 | 1 |
| 3 | 6 |

Figure 13: Confusion Matrix of Naïve Bayes

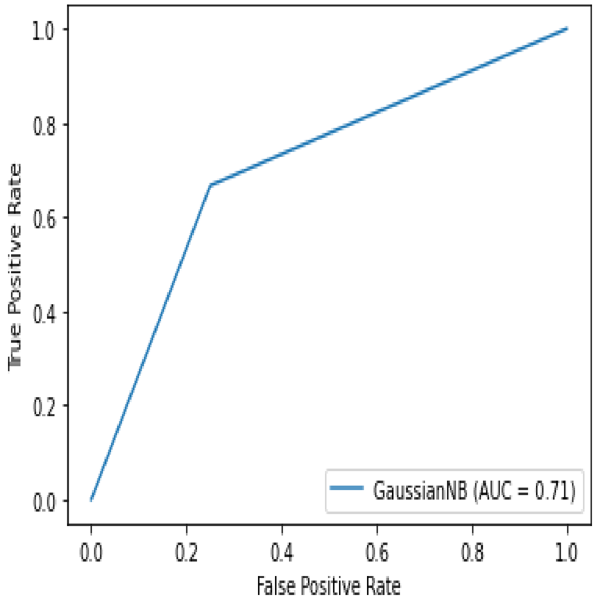

Figure 14: AUC Graph of Naïve Bayes

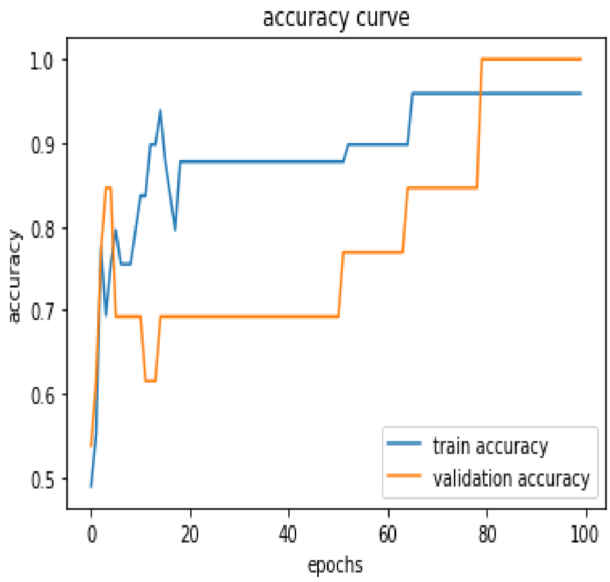

Figure 15: Training and Testing accuracy of model

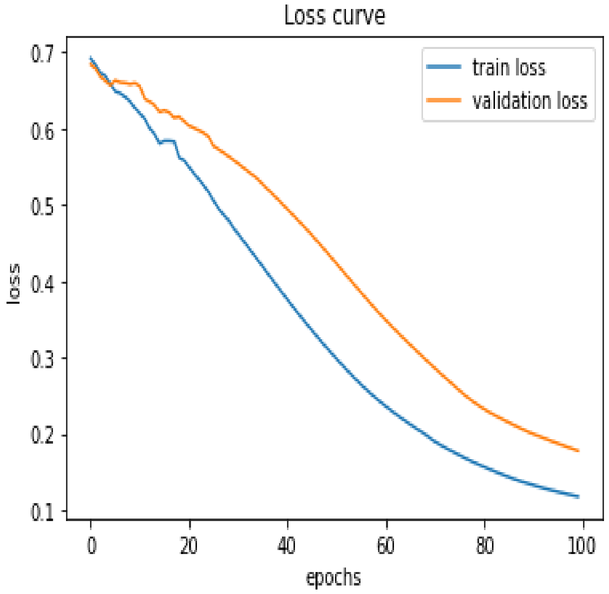

Figure 16: Losses of model

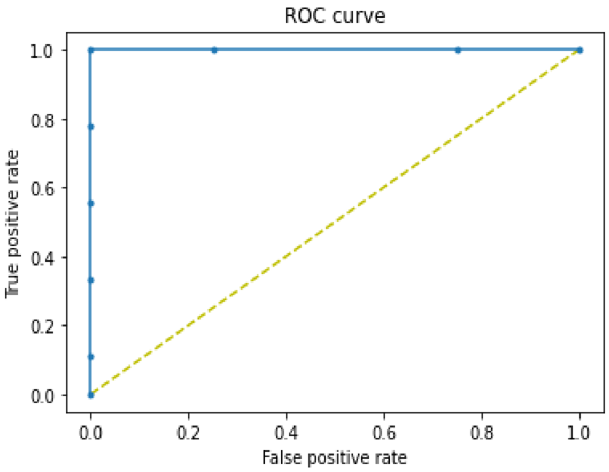

Figure 17: ROC of Model

## 5. CONCLUSION AND FUTURE SCOPE

Today's scenario is that the liver cancer diagnosis is still tedious. Its growth is fast the survival of the patient is not promised. It is difficult to diagnose Liver Cirrhosis in the initial stage later on it becomes very dangerous. The classification algorithms are helpful for the early detection of Liver Cirrhosis. In this paper, machine learning approaches have been discussed and the implementation of different machine learning classifiers. The discovered facts related to Liver Cirrhosis analysis entail that Liver Cirrhosis detection can be governed by the fundamental concept of the geometric distribution. The work proposes that the investigation related to the significance of parameters responsible for Liver Cirrhosis can be realized in the light of the machine learning model. Liver Cirrhosis detection can also be realized based on the fundamental principle of information gain. The reliability and mean time to failure of the Liver Cirrhosis testing system can be realized in the light of parallel system configuration. The factors leading to Liver Cirrhosis can be analyzed on the basis of machine learning Algorithms. The effect of alcohol consumption leading to Liver Cirrhosis can be sensed using the concept learning approach. Liver Cirrhosis can also be analyzed in the light of neural networks and neuro-associator. The performance of the machine learning model is best described in terms of its accuracy. All the classifier gives a minimum error rate with high accuracy.

In future work, consider the diagnosis of liver disease on the thermogram images. Fuzzy rules and fuzzy logic will be used to detect the liver abnormalities in thermo gram images and as suggested by the examiner proposed method can be implemented for other liver datasets the performance measure shall be computed in the future and the time complexity of the proposed work can be compared with other algorithms for the usefulness of the algorithm in future.

## CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

## FUNDING SUPPORT

The author declares that they have no funding support for this study.

## REFERENCES

[1] Singh, J., Bagga, S., & Kaur, R. (2020). Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques. Procedia Computer Science, 167, 1970–1980. doi:10.1016/j.procs.2020.03.226

[2] Pei, X., Deng, Q., Liu, Z., Yan, X., & Sun, W. (2021). Machine Learning Algorithms for Predicting Fatty Liver Disease. Annals of Nutrition and Metabolism, 77(1), 38–45. doi:10.1159/000513654

[3] Guo A, Mazumder NR, Ladner DP, Foraker RE (2021) Predicting mortality among patients with liver cirrhosis in electronic health records with machine learning. PLoS ONE 16(8):e0256428. https://doi.org/10.1371/journal.pone.0256428

[4] Rashid Naseem, Bilal Khan, Muhammad Arif Shah, Karzan Wakil, Atif Khan, Wael Alosaimi, M. Irfan Uddin, Badar Alouffi, "Performance Assessment of Classification Algorithms on Early Detection of Liver Syndrome", Journal of Healthcare Engineering, vol. 2020, Article ID 6680002, 13 pages, 2020. https://doi.org/10.1155/2020/6680002

[5] Nakao, M., Nakayama, N., Uchida, Y., Tomiya, T., Ido, A., Sakaida, I., Yokosuka, O., Takikawa, Y., Inoue, K., Genda, T. et al. (2018), 'Nationwide survey for acute liver failure and late-onset hepatic failure in japan', Journal of gastroenterology 53(6), 752-769.

[6] Kant, S. and Ansari, I. A. (2016), 'An improved k means clustering with atkinson index to classify liver patient dataset', International Journal of System Assurance Engineering and Management 7(1), 222-228

[7] Pakhale, H. and Xaxa, D. K. (2016), 'A survey on diagnosis of liver disease classification', Int J Eng Tech 2(3).

[8] Abdel-Basset, M., Manogaran, G., Rashad, H. and Zaied, A. N. H. (2018), 'A comprehensive review of quadratic assignment problem: variants, hybrids and applications', Journal of Ambient Intelligence and Humanized Computing pp. 1-24

[9] Desir, A., Goyal, V., Segev, D. and Ye, C. (2020), 'Constrained assortment optimization under the markov chain–based choice model', Management Science 66(2), 698-721.

[10] Sumida, M., Gallego, G., Rusmevichientong, P., Topaloglu, H. and Davis, J. (2019), Revenue-utility trade off in assortment optimization under the multinomial logit model with totally unimodular constraints, technical report, Tech. rep., Cornell University, Ithaca, NY.

[11] Zhang, H., Rusmevichientong, P. and Topaloglu, H. (2020), 'Assortment optimization under the paired combinatorial logit model', Operations Research.

[12] Singh, A. and Pandey, B. (2018), An efficient diagnosis system for detection of liver disease using a novel integrated method based on principal component analysis and k-nearest neighbor (pca-knn), in 'Intelligent Systems: Concepts, Methodologies, Tools, and Applications', IGI Global, pp. 1015-1030

[13] Singh, A. and Pandey, B. (2018), An efficient diagnosis system for detection of liver disease using a novel integrated method based on principal component analysis and k-nearest neighbor (pca-knn), in 'Intelligent Systems: Concepts, Methodologies, Tools, and Applications', IGI Global, pp. 1015-1030

[14] Rajathi, G. I. and Jiji, G. W. (2019), 'Chronic liver disease classification using hybrid whale optimization with simulated annealing and ensemble classifier', Symmetry 11(1), 33

[15] Yu, L., Liou, I. W., Biggins, S. W., Yeh, M., Jalikis, F., Chan, L.-N. and Burkhead, J. (2019), 'Copper deficiency in liver diseases: a case series and pathophysiological considerations', Hepatology communications 3(8), 1159-1165.

[16] Saima Afrin, F. M. Javed Mehedi Shamrat, Tafsirul Islam Nibir, Mst. Fahmida Muntasim, Md. Shakil Moharram, M. M. Imran, Md Abdulla, "Supervised machine learning based liver disease prediction approach with LASSO feature selection", Bulletin of Electrical Engineering and Informatics, Vol. 10, No. 6, December 2021, pp. 3369-3376

[17] Mostafa, F.; Hasan, E.; Williamson, M.; Khan, H. Statistical Machine Learning Approaches to Liver Disease Prediction. Livers 2021, 1, 294-312. https://doi.org/10.3390/livers1040023

[18] Keerthana, PSM and Phalinkar, Nimish and Mehere, Riya and Bhanu Prakash Reddy, Koppula and Lal, Nidhi, A Prediction Model of Detecting Liver Diseases in Patients using Logistic Regression of Machine Learning (March 28, 2020). Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020, Available at SSRN: https://ssrn.com/abstract=3562951 or http://dx.doi.org/10.2139/ssrn.3562951

[19] A. Sivasangari, B. J. Krishna Reddy, A. Kiran and P. Ajitha, "Diagnosis of Liver Disease using Machine Learning Models," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 627-630, doi: 10.1109/I-SMAC49090.2020.9243375.

[20] C. Geetha and A. Arunachalam, "Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9402463.

[21] K. Thirunavukkarasu, A. S. Singh, M. Irfan and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-3, doi: 10.1109/CCAA.2018.8777655