Liver Disease Prediction and Classification using Machine Learning Techniques

Srilatha Tokala, Koduru Hajarathaiah, Sai Ram Praneeth Gunda, Srinivasrao Botla, Lakshmikanth Nalluri, Pathipati Nagamanohar, Satish Anamalamudi, Murali Krishna Enduri
Department of Computer Science and Engineering, SRM University-AP, Amaravati, India

***Abstract*—Recently liver diseases are becoming most lethal disorder in a number of countries. The count of patients with liver disorder has been going up because of alcohol intake, breathing of harmful gases, and consumption of food which is spoiled and drugs. Liver patient data sets are being studied for the purpose of developing classification models to predict liver disorder. This data set was used to implement prediction and classification algorithms which in turn reduces the workload on doctors. In this work, we proposed apply machine learning algorithms to check the entire patient's liver disorder. Chronic liver disorder is defined as a liver disorder that lasts for at least six months. As a result, we will use the percentage of patients who contract the disease as both positive and negative information We are processing Liver disease percentages with classifiers, and the results are displayed as a confusion matrix. We proposed several classification schemes that can effectively improve classification performance when a training data set is available. Then, using a machine learning classifier, good and bad values are classified. Thus, the outputs of the proposed classification model show accuracy in predicting the result.**

***Keywords*—Machine learning algorithms; classification model; classifier; liver disease**

I. Introduction

The liver is the most imperative structure in a human build. Insulin is broken down by the liver. The liver breaks bilirubin with glucuronidation, which further helps its defecation into bile [1]. It is also accountable for the breaking down and excretion of many unwanted products. It shows a noteworthy role in altering toxic materials. It shows a noteworthy role in collapsing medicinal products. It's named Drug metabolism. The weight would be 1.3 kg. The liver consists of 2 immense portions namely the privileged portion, and the left estimate. The gallbladder is located below the liver, near the pancreas. The Liver along with these organs helps to consume and give nutrition. Its job is to help the flow of the wounding materials in the stream of blood from the stomach, before passing it to whatsoever is left of the body. Liver sicknesses are triggered when the working of the liver is affected or any injury has happened to it [2].

The development of liver disorders [3] is complicated and varied in character, influenced by a number of variables that determine disease susceptibility. Sex, ethnicity, genetics, environmental exposures (viruses, alcohol, nutrition, and chemicals), body mass index (BMI), and coexisting diseases like diabetes are among them. A high mortality rate is associated with liver problems, which are life-threatening diseases. The usual urine and blood tests are the first step in the prognosis of liver disorders. A LFT (liver functions test) is recommended for the patient based on the symptoms seen [4].

Liver disease is a significant health issue affecting millions of people globally. Early detection and accurate classification of liver diseases can lead to better patient outcomes and reduce the burden on the healthcare system. One-third of adults and an increasing proportion of youngsters in affluent nations suffer from non-alcoholic fatty liver disease (NAFLD) [5], a growing health issue. The abnormal buildup of triglycerides in the liver, which in some people causes an inflammatory reaction that can lead to cirrhosis and liver cancer, is the first sign of the condition. While there is a significant correlation between obesity, insulin resistance, and non-alcoholic fatty liver disease (NAFLD), the pathophysiology of NAFLD remains poorly understood, and treatment options are limited. However, machine learning techniques have demonstrated encouraging results in predicting and categorizing liver diseases based on patient data. By utilizing sophisticated algorithms to analyze and learn from large datasets, these techniques can identify patterns and anticipate outcomes. The employment of machine learning techniques in liver disease prediction and classification is a dynamic area of research, with continual advancements being made to enhance accuracy and decrease healthcare costs.

A. Overview of Liver Disease

Liver disease refers to an abnormality in the liver's function, resulting in illness [4]. The liver is responsible for many vital functions within the body, and if it becomes damaged or infected, the loss of these functions can have a significant impact on overall health. Hepatic disorder is another term used to describe liver disease [6]. This umbrella term encompasses a range of possible complications that prevent the liver from performing its assigned roles. Even if only a quarter of the liver is still functioning and the rest is damaged, this organ's efficiency will be greatly reduced. The liver is the biggest hard structure in the human build and is well thought-out as a gland because, amid its many roles, it creates and secretes bile. The liver is stood at the upright part of the abdomen and the rib cage shelters it. It has two core lobes that are thru with small lobules. The liver cells have two dissimilar bases of a blood source. The hepatic artery transfers heart-driven blood abundant in oxygen, while the portal vein provides nutrients from the intestines. Generally, the vein's job is to bring the blood from all other organs to the heart, but the portal vein permits nutrients from the digestive region to go into the liver for treating and purifying the former to flow into the general circulation. The portal vein proficiently transports the chemicals that liver cells require to yield the proteins, cholesterol, and glycogen needed for usual body actions.

B. Causes of Liver Disease

There are numerous activities that prompt liver maladies [7]. The classifications are:

1) Infection: The liver can become infected by parasites and viruses, which can lead to inflammation or edoema and compromise liver function. The virus that typically results in liver damage is spread through blood or sperm and is primarily brought on by tainted food, contaminated water, or contact with an infected person. Hepatitis A, C, and B are liver infections that can affect people.

2) Immune system abnormality: The body's immune system is administered by certain ailments, to attack other body parts. The liver is also affected. These diseases could be Autoimmune hepatitis. In addition, it could be Primary biliary cholangitis, and Primary sclerosing cholangitis.

3) Inheritance: A rare gene genetically inherited from either of your parents can cause a buildup of various compounds in the liver, which can cause liver damage. Wilson's disease, Hemochromatosis, and alpha-1 antitrypsin deficiency are three examples of genetic liver illnesses.

4) Cancer and other progressions: Cancers that have may reason liver diseases are Liver adenoma, Bile duct cancer, and Liver cancer.

5) Others: The general reasons are prolonged alcohol abuse, fat buildup in the liver (NAFLD), certain drugs or over-the-counter treatments, and certain herbal mixes.

6) Risk apects: Factors that might raise the risk of liver diseases are excessive usage of liquor, being overweight, diabetes of type, tattoos, piercings of the body, drug injection with used needles, transfusion of blood, exposure to foreign blood, unprotected intercourse, exposure to chemicals, and inheritance.

C. Chemicals Compounds in Liver

Chemicals such as Bilirubin, Albumin, Alkaline phosphatase, Aspartate aminotransferase, and globulin are existent in the liver and perform a vital role in the daily operations of the healthy liver.

1) Bilirubin: Bilirubin is a yellowish complex that arises in the usual catabolic trail that breaks down heme in vertebrates. Bile and urine emit it. Raised volumes of bilirubin in the body cause diseases. The bilirubin is accountable for the yellow shade of cuts and the yellow staining in jaundice disease. Its following breakdown products, like stercobilin, are accountable for the brown color of feces. Another breakdown product, urobilin, is the key constituent of the straw-yellow color of urine.

2) Alkaline phosphatase: In beings, alkaline phosphatase is existent in all tissues all over the body but is mainly focused in the liver, intestinal mucosa, bile duct, bone, kidney, and placenta. In the serum, two kinds of alkaline phosphatase isozymes prevail skeletal and liver. In childhood, most of the alkaline phosphatase is of the skeletal source. Most of the mammals including humans have these types of alkaline phosphatases:

- ALPI: It is intestinal having a molecular mass of 150 kDa.
- ALPL: It is tissue-nonspecific mainly present in the liver, kidney, and bone.
- ALPP: It is placental and is also known as Regan isozyme.
- GCAP: It is a germ cell.

3) Aspartate aminotransferase: AST is a kind of enzyme. AST levels are higher in the heart and liver. AST is found in the kidneys and muscles, although in less amounts. It is very low in human blood. When muscle or liver cells are injured, the AST is released into the bloodstream. The AST test will therefore be useful for tracking or identifying liver damage or dysfunction.

4) Albumin: They're globular proteins. Serum albumins are common and are the most imperative protein of blood. It binds thyroxine (T4), water, cations like Ca2+ and Na+, hormones, fatty acids, bilirubin, and pharmaceuticals. Its core part is to govern and normalize the oncotic pressure of the blood. It binds several fatty acids, cations, and bilirubin.

5) Globulin: They are protein globules. They are heavier than albumin at the molecular level. It will not dissolve in pure water but will solvate in dilute salt solutions. The liver produces some globulins. Globulin absorption in fit human blood is around 2.6-3.5 g/dL. There are several different types of globulins, including beta, alpha 1, alpha 2, and gamma globulins. Any unfitting amounts of these chemicals produced in the kidney can reason an imbalance and cause liver diseases. These are considered features. There are n number of kinds of liver illnesses and these are grounded based on the proportion of these chemicals stashed.

II. Machine Learning and Liver Disease

Classification algorithms are often defined to be used for forecasting the liver disease as they can help predict whether a patient has the disease or not based on certain features or characteristics. Based on the existing solutions, it was found that the F-Tree algorithm shows the highest accuracy among the algorithms tested, making it a suitable choice for forecasting liver disease. Feature selection along with the fuzzy K-means classification methods are commonly used in the classification of liver diseases. These methods can help identify important features that can be used to distinguish between different types of liver disorders. As the same attribute values may be present in different liver disorders, using fuzzy-based classification can help improve the performance of the classification process by taking into account the degree of similarity between instances [8], several classification algorithms, including $J48$, SVM, RF, and MLP, were used to classify liver diseases. The study evaluated the performance of these state-of-the-art algorithms using metrics such as data accuracy, data effectiveness, and correction rate, and compared the results. The findings indicated that the multilayer perceptron algorithm achieved the highest accuracy compared to the other algorithms examined in the research. In this work [9], researchers employed Bayesian classification to distinguish between various liver diseases, including cirrhosis, hepatitis, and non-liver diseases. They used both Naive Bayes and FT

tree techniques to categorize the liver patient dataset into different illness subtypes, and evaluated the performance of these methods in terms of accuracy and execution time. Their analysis revealed that Naive Bayes algorithm outperformed the FT tree algorithm in terms of execution time.

In this work [10], authors explored the effectiveness of several classification methods in diagnosing liver diseases. These methods included Naive Bayes, Decision Tree, Multilayer Perceptron, K-Nearest Neighbors, Random Forest, and Logistic Regression. To evaluate their performance, the authors used metrics such as precision, recall, sensitivity, and specificity. The results showed that Naive Bayes achieved the highest precision compared to the other algorithms examined. Additionally, the Logistic Regression and Random Forest algorithms were found to have good results when recall was considered.

In [11], to create a model for analysing liver illness, the WEKA tool was employed. To create their suggested model, the Naive Bayes, Decision Tree (DT), and $J48$ algorithms were employed. The algorithms' accuracy and execution time were measured, and the results were compared to those of the available options. The findings demonstrated that the $J48$ and DT algorithms outperformed the Naive Bayes algorithm in terms of accuracy.

In [12], uses a dataset of Indian patients with liver disease and implements it using various classification algorithms, including LR, K-NN, and SVM. Additionally, confusion matrix was used to evaluate the algorithms against one another. Based on their correctness, the experimental analysis of these algorithms was examined and evaluated. According to the findings, LR and KNN techniques can both achieve a sizable degree of accuracy, although LR has a high sensitivity. With this, it can be inferred that LR is an appropriate strategy for disease prediction. A novel feature model with the classification methods Random Forest, SVM, $J48$, Bayesian Net, and MLP was proposed in [13], [14], [15].

Three distinct procedures are used to examine and implement a comparative study for the prediction of liver disease. On the dataset containing records of liver diseases, the normalisation method min-max was examined and tested in the initial phase. The second step uses the PSO feature selection approach to choose the dataset's necessary components for predicting liver illness. The implementation of the classification algorithms and an accuracy-based performance evaluation of the methods were done in the third phase. The experimental study leads to the conclusion that the $J48$ method outperformed when PSO feature selection was used. In this work [16], aimed to balance the dataset for accurate prediction of liver diseases by using a combination of sampling and oversampling techniques. Several classification algorithms were used, including $J48$, Multilayer perceptron, Random Forest, Multiple linear regression, Support Vector Machines, and Genetic programming. The Random Forest algorithm with oversampling at higher rates demonstrated the best performance in predicting liver diseases. Similarly, In this work [17] utilized various classification algorithms such as Naive Bayes, FT tree, $J48$, SVM, RF, MLP, K-NN, LR, and $C4.5$ to classify liver diseases. Different techniques such as feature selection, fuzzy K-means classification, normalization, PSO feature selection, oversampling, and undersampling were also employed to enhance the performance of these algorithms. The performance of these algorithms was evaluated based on different factors such as accuracy, execution time, precision, recall, sensitivity, specificity, and F-measures. The results of these studies indicate that different algorithms perform differently depending on the dataset and application [18]. Therefore, it is crucial to evaluate the performance of multiple algorithms and techniques to identify the best approach for a specific dataset and application. Although the outcomes have been found to differ depending on the dataset employed and the specific algorithm used, some algorithms such as Multilayer perceptron, Random Forest, Support Vector Machines, and Artificial Neural Network generally perform better than others in predicting liver diseases.

ML has a multiplied widespread routine in the modern ages. Using ML as a means of aid in therapeutic and pharmacological diagnostics is increasing. However, the common thing is rapidly increasing access to huge amounts of data. The following content gives a thorough summary of the major defies and prevalent resolutions to ML challenges [19] in medicine. Feature selection is to decrease the struggle by selecting a subsection of the useful features in the input and disposing of the residual features.

It is clear from the above statement that various studies have been conducted using different machine learning algorithms to predict liver diseases and evaluate their performance. Different algorithms such as Naive Bayes, Decision Trees, Random Forest, Logistic Regression, K-NN, SVM, Artificial Neural Networks, and C4.5 have been used in these studies. The performance of these algorithms has been evaluated based on factors such as accuracy, precision, recall, F1-score, specificity, and execution time. It is also noted that different studies have different conclusion on which algorithm performs best, it depends on the dataset and evaluation metrics used.

A. Logistic Regression

Its new name is the Logit model [20]. It is used to simulate the likelihood of a specific class or event prevailing, such as pass/fail, healthy/sick, alive/dead, or win/lose. It's a mathematical model. In its most basic form, it employs a logistic strategy to advance a binary dependent variable, while several complicated expansions exist. A binary logistic model [21] will include a variable that is dependent and has two possible values, such as pass/fail, which represents an indicator variable, and the two values are regarded as 0 and 1.

B. Support Vector Machine

It targets to find an ideal hyperplane that splits the data into diverse classes. A way of implementing SVM [22] in python is by making use of the scikit-learn package. The data is sorted out and is disjointed into test data and training sets. The testing data is set as twenty out of every hundred. The set for algorithm preparation is fixed as one hundred and sixty out of every two hundred. An SVM forms either a set of hyperplanes. It is built in an immeasurable-extension space. A hyperplane that has the farthest distance to the nearest data point of any class achieves a good separation, called a functional margin. In essence, a wider margin implies a smaller simplification defect for the classifier.

C. Convolutional Neural Networks

CNN is a portion of DNN [23]. It is broadly purposed for the exploration of imageries. They are also known as shift invariant ANN or space invariant ANN. They have a mutual-bulks architecture. Here, the network develops a mathematical process termed convolution. They are solely, neural networks that practice convolution as a replacement for general matrix multiplication in a minimum of one of their layers.

D. MLP Classifier

MLP [24] is a portion of (ANN). Also called, feed-forward ANN. An MLP encompasses the bottom of three sheets of nodes. The three sheets are the first input, the mid-hidden, and the final output layers. Further, each other node is doubled as a neuron that makes use of a nonlinear activation, with an exception of the input node. MLP practices an administered learning system that is known as backpropagation, for training. MLP [25] can be distinguished from linear perceptron because of its manifold layers and non-linear instigation.

E. Random Forest

The random forest [26] creates decision trees on data mockups and gains the prediction from all the formed decision trees. Finally, it elects the best elucidation by voting means. The data is sorted out and is disjointed into test data and training sets. The testing data is set as twenty out of every hundred. The set for algorithm preparation is fixed as one hundred and sixty out of every two hundred. The program fragments the data into many groups and subgroups. If an individual draws lines between data points in a subgroup, lines connecting subgroups into a group, and so on. The erection appears to be tree-like. The hyperplane that maximizes the distance to the nearest data point in the training set provides a reliable separation between classes.

III. Literature Review

A. Liver Disease Biopsies using Deep Learning and CNN

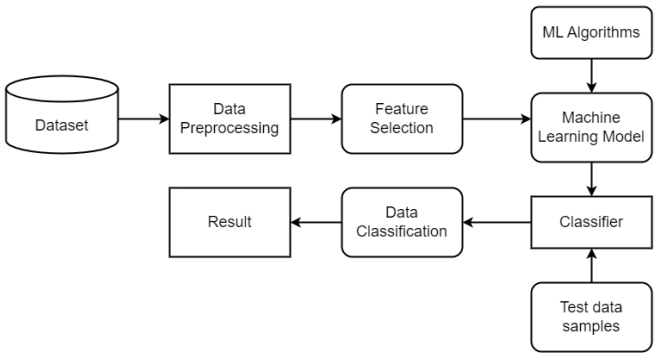

Fig. 1. Processing of ML algorithms

The author sought to implement a completely involuntary tool for diagnosing liver disease by using liver biopsy images. The author considered using biopsy images because there is a respectable chance of differentiating an unhealthy and healthy liver using these images. The projected tactic is to use image study and deep learning and further determine an efficient CNN [27] architecture and further execution. NAFLD is common. An investigation stated that almost forty percent of all liver illnesses around the globe are caused by NAFLD. The existence of NAFLD in the liver can be found by the sign of hepatic steatosis, and also other reasons for fat build-up, such as major consumption of alcohol, lasting use of steatogenic medicines, and genetic problems. In this proposed method, the biopsy images of the liver are taken and the hepatic structures existent in these are analyzed with the aid of two CNN [28] which have the same architecture. They want to develop a 4-class detection system, which detects sinusoids, ballooned hepatocytes, veins, and fat droplets initially. In the concluding phase, this detection system is united with each other to complete the methodology. It then calculates the fat and ballooning ratio. The found ratio will aid in concluding the patient's condition. They made use of seven hundred and twenty liver biopsy images. Six hundred and twenty of these images are used for testing the algorithm, sixty are used for validation and the rest of the forty images are used for testing. These images originally are of $10,000 * 10,000$ pixels or above. The area in which the tissue is extant is hauled out from each image and the resultant images are $64 * 64$ pixels.

B. Predicting the Accuracy of Liver Disease using Machine Learning

Utilizing and modeling medicinal data sets are now considered by specialists across the globe. The main plan here is to shorten the time gap in the middle of testing the liver, and generating the report and final result. They used some Machine Learning algorithms [29] like decision tree, naive Bayes, ANN, and random forest. Then, the Pearson correlation [30] to find the anomalies such as TP (true positives), FN (false negatives), FP (false positives), and TN (true negatives) is applied. This is done to find the precision, specificity, and affectability of the algorithms used. The produced words are used to compute the sensitivity, affectability, specificity, and accuracy using pre-defined equations. The author intended to create an interface in which the user could enter the patient's report as input. The algorithms are then skill-trained using the allocated data set, and the output of the user input information is determined. The output will be a single number that is either a zero or a one, with the binary one indicating that the patient's liver is sickly and the binary zero indicating that the patient's liver is healthy. The user-given input data will be logged and this data will further train the algorithm again. These additional tabulated values will train and skill the algorithm to progress with precision. UCI machine is the site from where the authors obtained the data set. The outcoming results of these respective algorithms are charted. These grids show outcomes, which are encouraging Accuracy not cited. They made use of ML algorithms- ANN, Naïve Bayes, and Decision tree to make the model. There are numerous types of liver ailments, the authors considered the general liver diseases to ease the process and get a precise result.

C. Segmentation of Liver using CT Scan and Finding Disease

The authors lit up an idea by making the use of Abdominal CT, liver disease can be perceived. Some organs cannot be perceived through standard X-Ray equipment. These are the

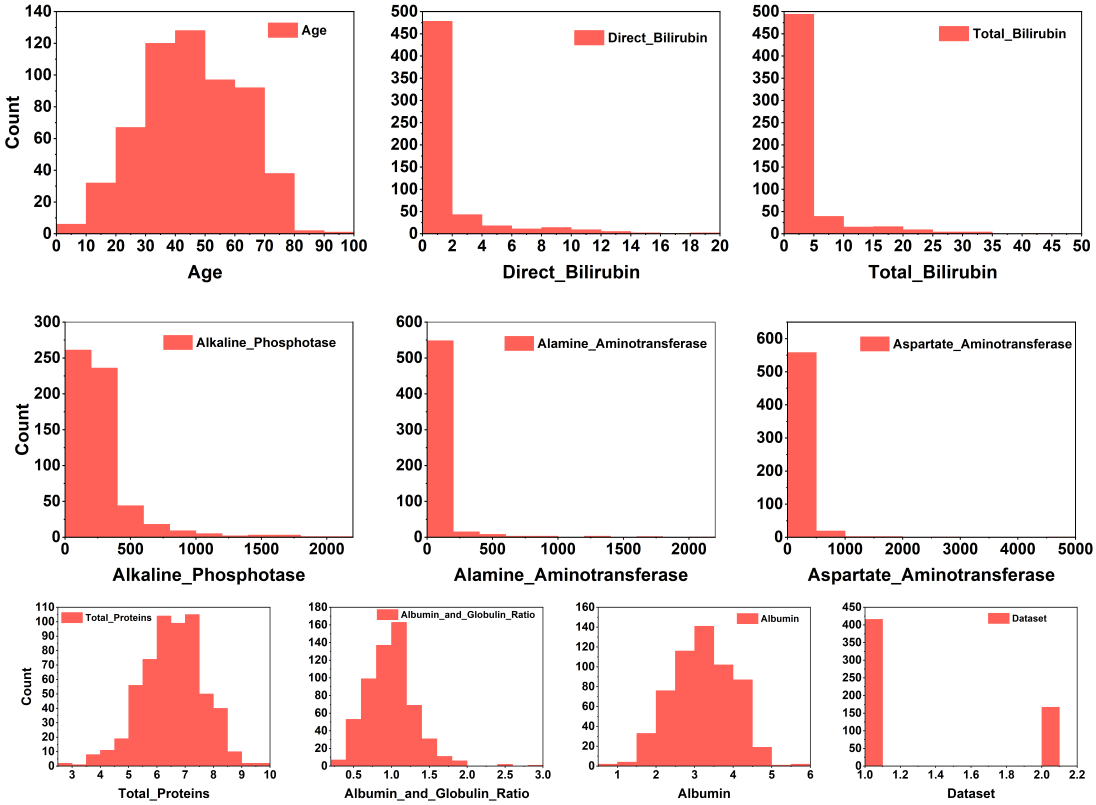

Fig. 2. Histogram for frequency distributions of various patients based on attribute

TABLE I. ACCURACY FOR ML METHODS

ML Models	Accuracy
Logistic Regression	0.76
Support Vector Machine	0.72
Nearest Neighborhood	0.80
Random Forest	0.88

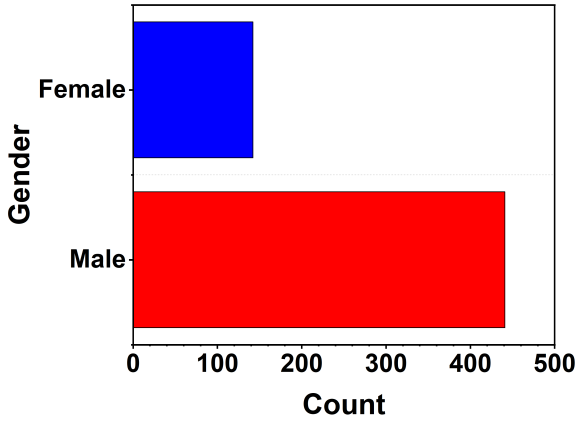

Fig. 3. Comparison with the frequency of males and females

conditions that strengthen the motives to use CT Scans as they can show the structures better than an X-Ray. These CT scan-produced images will have an accurate resolution. They proposed to use WTA to segment the Scan image, identify the liver placement, and differentiate it from the background. In the ending step, the percentage of the area affected is calculated. With the intention of perceiving the progress or sternness of the liver disease, one should use a highly precise technique, that is CT Scan. This CT Scan is extensively used in this medical field to attain info about the humanoid build. In the initial step, the imageries are scrutinized to find different parts. In the following step, the images are handled using the Erode and Dilate algorithm. Viewing point values are adjusted here. Further, they are processed with WTA to segment the liver area. The WTA gives two outputs namely the liver area and the non-liver area. The output yielded is referred to as cropping. Then the gathered copy is adapted into a binary where this organ is white and the rest is black. Then median filtering is done to reduce the noise and smoothens the texture. Then the damaged region of the liver is tallied using a formula. The authors affirm that the typical precision of image breakdown is around eighty-one percent and the typical precision of liver breakdown is around 92 percent. The author used the WTA to distinguish the liver. Another method suggested is to use the binary threshold to isolate the liver area and the diseased area. The closing process in this paper is to measure the fraction of diseased spaces in the liver.

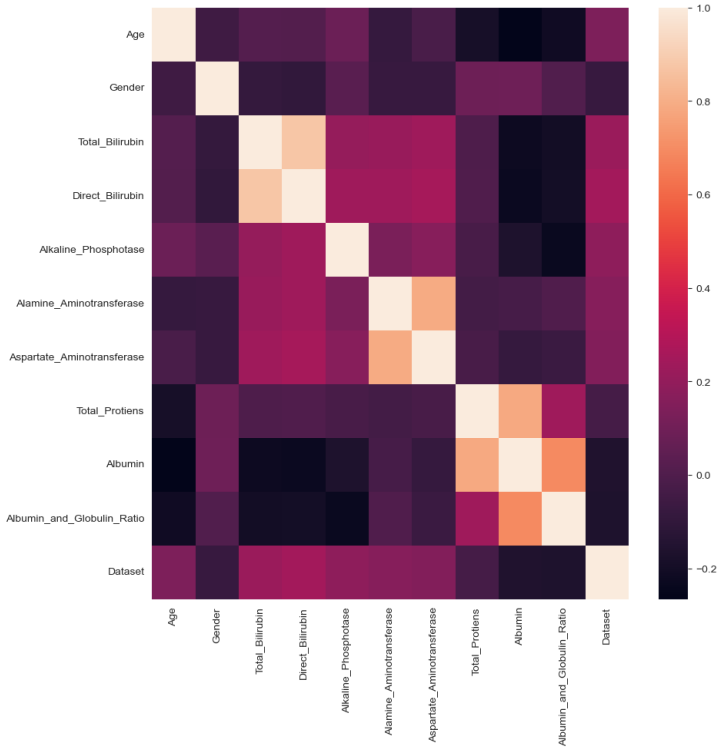

Fig. 4. The correlation graph between each attribute

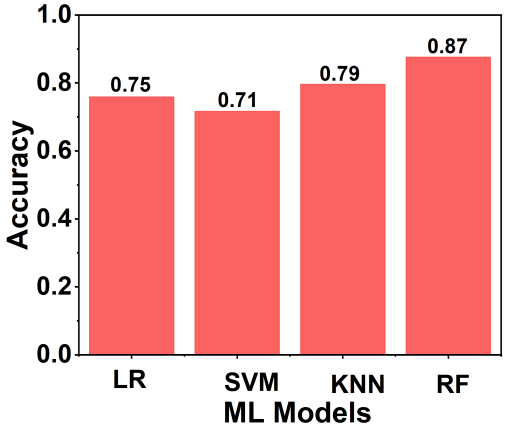

Fig. 5. The accuracy of each machine learning technique used for classification is compared with others

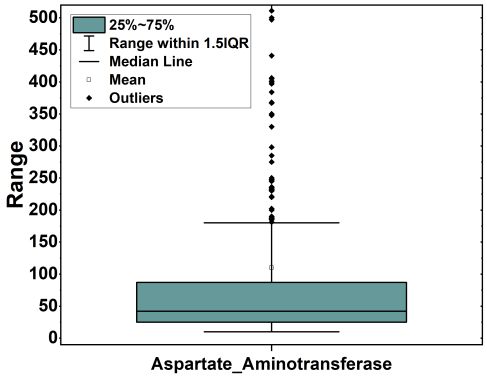

Fig. 6. The box plot represents about the outliers present in that attribute

D. Liver Disease Prediction using Classification Algorithms

The proposed process by the authors shows that machine learning can not only predict the disease but also can recognize hidden patterns for diagnosis and decision-making. The everyday growing cases of liver disorders are considered to be a common problem around the world. The goal of this thesis is to provide competent findings in identifying liver disease using classification algorithms [31]. Logistic Regression, K-Nearest Neighbor, and Support Vector Machines are the algorithms utilized for this type of job. The algorithms called Classification algorithms [32] are predominantly being used in predicting diseases among machine learning algorithms. ML techniques are now very helpful in the healthcare sector for the prediction of diseases from medical databases. In almost every continent there are researchers and scientists who are rigorously using machine learning models [33] with classification algorithms to strategically enhance medical diagnostics and are showing better results. Logistic Regression, K-Nearest Neighbor, and Support Vector Machines are used in this paper to predict liver disease. We all know that the liver is the body's largest internal organ, performing vital functions such as producing blood clotting factors and proteins, producing triglycerides and

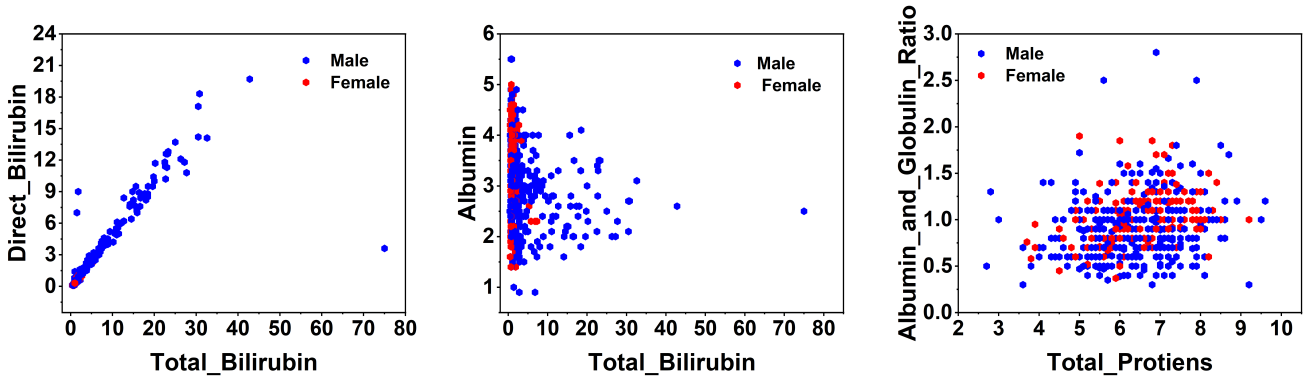

Fig. 7. Scatter plots for Direct_Bilirubin vs Total_Bilirubin, Albumin vs Total_Bilirubin, Albumin_and_Globulin_ Ratio vs Total_Proteins

cholesterol, glycogen synthesis, and bile production. In most cases, a decrease in function must affect much more than 75% of the liver tissue. So it's crucial to notice at an untimely stage at which the disease can be treated before it progresses to the severe stage

IV. Proposed System

In the proposed system, we have to import the liver patient dataset (.csv). Then the dataset is pre-processed and the anomalies and full-up empty cells in the dataset are removed, so that we can further improve the effective liver disease prediction. Then we construct a Confusion matrix for accomplishing an enhanced lucidity of the no of correct/incorrect predictions. Formerly, several classification and prediction procedures and if possible, combinations of different algorithms are implemented and check the accuracy. Our objective is to develop a code that delivers an exactitude of 90%. The advantages are improved classification, early prediction of risks, and improved accuracy. The block diagram of the overall system is shown in Fig. 1.

V. Results

Using various methods, we begin our study in this part with the data-processing stage and go on to feature extraction, classification, and prediction analysis. The attributes used in the datset are Age, Direct Bilirubin, Total Bilirubin, Alkaline Phosphate, Alamine Aminotransferase, Aspartate Aminotransferase, Total Proteins, Albumin, and Globulin Ratio, Albumin, Dataset (where data set is the class label). Each histogram tells us about the frequency distributions for various patients in that particular attribute, shown in Fig. 2.

We represented attributes count or frequency of the patients based on the age, direct_bilirubin, total_bilirubin, alkaline_phosphotase, alamine_aminotransferase, aspartate_aminotransferase, total_proteins, albumin_and_globulin_ratio, albumin, and data set. The Fig 4 shows the correlation between each attribute used in data set are plotted. The lighter the color between two attributes in the graph the higher the values of one attribute are dependent or correlated on the second attribute. From the Fig. 4 we can say that direct_bilirubin and total_bilirubin are highly correlated.

The Fig. 3 the frequency of males used in the dataset is compared with frequency of females. The Data set contains a total of 441 males and 142 females. The accuracy of each model is obtained by training the model with the dataset values and testing it by predicting the dataset value. The number of correct predictions done by the model gives us accuracy. From Fig. 5, we can say that Random Forest has the highest accuracy compared to other models. The box plot is plotted which tells us about the outliers present in that attribute. The box plot identifies the outliers using IQR (Inter Quartile Range) method. From Fig. 6, we can see only two values are greater than 450 and far from other values. Hence they are outliers. From Fig. 7 scatter plots for Direct_Bilirubin vs Total Bilirubin, Albumin vs Total_Bilirubin, Albumin and Globulin Ratio vs Total Proteins are plotted. Scatter plots are a valuable tool for visualizing the relationship between variables, with dots representing the data points. They are commonly employed to illustrate the associations between variables and how changes in one variable impact another. From the Fig. 7, we can see for Direct_Bilirubin vs Total Bilirubin the scatter plot is like a straight line which indicates they are highly related. Table I shows the accuracy of the ML method results. The random forest method gives good accuracy than LR, SVM, and nearest neighborhood.

VI. Conclusion

The liver patient data set was used to implement prediction and classification algorithms, which in turn reduces the workload on doctors. We suggested employing machine learning techniques to examine the patient's total liver condition. A liver condition that has persisted for at least six months is considered chronic. We will thus utilise the proportion of people who get the condition as both positive and negative data. A confusion matrix is used to represent the outcomes of classifier processing of percentages of liver disease. When a training data set is available, our proposed classification schemes can significantly enhance classification performance. Then, using a machine learning classifier, good and bad values are classified. Thus, the outputs of the proposed classification model show accuracy in predicting the result.

The extent of our work is that we will apply deep learning techniques to predict liver disease. Some of the future directions are improve the accuracy of liver disease prediction

and classification models is to include more diverse data sources, improving liver disease prediction and classification is to combine multiple machine learning techniques, machine learning models can be trained to predict the likelihood of liver disease in individuals based on their unique characteristics. Another important direction in liver disease prediction and classification using machine learning is to develop models that are explainable. This means that the models should provide clear and interpretable insights into the factors that contribute to liver disease. Explainable models can help healthcare professionals to make better decisions and provide better care for patients.

REFERENCES

[1] A. Arjmand, C. T. Angelis, A. T. Tzallas, M. G. Tsipouras, E. Glavas, R. Forlano, P. Manousou, and N. Giannakeas, "Deep learning in liver biopsies using convolutional neural networks," in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2019, pp. 496–499.

[2] L. A. Auxilia, "Accuracy prediction using machine learning techniques for indian patient liver disease," in *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2018, pp. 45–50.

[3] A. Spann, A. Yasodhara, J. Kang, K. Watt, B. Wang, A. Goldenberg, and M. Bhat, "Applying machine learning in liver disease and transplantation: a comprehensive review," *Hepatology*, vol. 71, no. 3, pp. 1093–1105, 2020.

[4] S. Sontakke, J. Lohokare, and R. Dani, "Diagnosis of liver diseases using machine learning," in *2017 International Conference on Emerging Trends & Innovation in ICT (ICEI)*. IEEE, 2017, pp. 129–133.

[5] J. C. Cohen, J. D. Horton, and H. H. Hobbs, "Human fatty liver disease: old questions and new insights," *Science*, vol. 332, no. 6037, pp. 1519–1523, 2011.

[6] F. Himmah, R. Sigit, and T. Harsono, "Segmentation of liver using abdominal ct scan to detection liver desease area," in *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*. IEEE, 2018, pp. 225–228.

[7] M. B. Priya, P. L. Juliet, and P. Tamilselvi, "Performance analysis of liver disease prediction using machine learning algorithms," *International Research Journal of Engineering and Technology (IRJET)*, vol. 5, no. 1, pp. 206–211, 2018.

[8] T. R. Baitharu and S. K. Pani, "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset," *Procedia Computer Science*, vol. 85, pp. 862–870, 2016.

[9] U. R. Acharya, S. V. Sree, R. Ribeiro, G. Krishnamurthi, R. T. Marinho, J. Sanches, and J. S. Suri, "Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm," *Medical physics*, vol. 39, no. 7Part1, pp. 4255–4264, 2012.

[10] N. Nahar and F. Ara, "Liver disease prediction by using different decision tree techniques," *International Journal of Data Mining & Knowledge Management Process*, vol. 8, no. 2, pp. 01–09, 2018.

[11] A. Naik and L. Samant, "Correlation review of classification algorithm using data mining tool: Weka, rapidminer, tanagra, orange and knime," *Procedia Computer Science*, vol. 85, pp. 662–668, 2016.

[12] A. N. Arbain and B. Y. P. Balakrishnan, "A comparison of data mining algorithms for liver disease prediction on imbalanced data," *International Journal of Data Science and Advanced Analytics (ISSN 2563-4429)*, vol. 1, no. 1, pp. 1–11, 2019.

[13] M. A. Kuzhippallil, C. Joseph, and A. Kannan, "Comparative analysis of machine learning techniques for indian liver disease patients," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2020, pp. 778–782.

[14] K. R. Asish, A. Gupta, A. Kumar, A. Mason, M. K. Enduri, and S. Anamalamudi, "A tool for fake news detection using machine learning techniques," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*. IEEE, 2022, pp. 1–6.

[15] M. K. Enduri, A. R. Sangi, S. Anamalamudi, R. C. B. Manikanta, K. Y. Reddy, P. L. Yeswanth, S. K. S. Reddy, and G. A. Karthikeya, "Comparative study on sentimental analysis using machine learning techniques," *Mehran University Research Journal of Engineering and Technology*, vol. 42, no. 1, pp. 207–215, 2023.

[16] M. Islam, C.-C. Wu, T. N. Poly, H.-C. Yang, Y.-C. J. Li *et al.*, "Applications of machine learning in fatty live disease prediction," in *Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth*. IOS Press, 2018, pp. 166–170.

[17] S. Mohanty, P. K. Gantayat, S. Dash, B. P. Mishra, and S. C. Barik, "Liver disease prediction using machine learning algorithm," in *Data Engineering and Intelligent Computing: Proceedings of ICICC 2020*. Springer, 2021, pp. 589–596.

[18] C. Liang and L. Peng, "An automated diagnosis system of liver disease using artificial immune and genetic algorithms," *JOURNAL OF MEDICAL SYSTEMS*, vol. 37, no. 2, 2013.

[19] R. A. Khan, Y. Luo, and F.-X. Wu, "Machine learning based liver disease diagnosis: A systematic review," *Neurocomputing*, vol. 468, pp. 492–509, 2022.

[20] A. S. Abdalrada, O. H. Yahya, A. H. M. Alaidi, N. A. Hussein, H. T. Alrikabi, and T. A.-Q. Al-Quraishi, "A predictive model for liver disease progression based on logistic regression algorithm," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 7, no. 3, pp. 1255–1264, 2019.

[21] F. E. Harrell, Jr and F. E. Harrell, "Binary logistic regression," *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis*, pp. 219–274, 2015.

[22] E. M. Hashem and M. S. Mabrouk, "A study of support vector machine algorithm for liver disease diagnosis," *American Journal of Intelligent Systems*, vol. 4, no. 1, pp. 9–14, 2014.

[23] Z. Yao, J. Li, Z. Guan, Y. Ye, and Y. Chen, "Liver disease screening based on densely connected deep neural networks," *Neural Networks*, vol. 123, pp. 299–304, 2020.

[24] M. Abdar, N. Y. Yen, and J. C.-S. Hung, "Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees," *Journal of Medical and Biological Engineering*, vol. 38, no. 6, pp. 953–965, 2018.

[25] T. Bikku, "Multi-layered deep learning perceptron approach for health risk prediction," *Journal of Big Data*, vol. 7, no. 1, pp. 1–14, 2020.

[26] T. A. Assegie, R. Subhashni, N. K. Kumar, J. P. Manivannan, P. Duraisamy, and M. F. Engidaye, "Random forest and support vector machine based hybrid liver disease detection," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 3, pp. 1650–1656, 2022.

[27] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[28] J. Murphy, "An overview of convolutional neural network architectures for deep learning," *Microway Inc*, pp. 1–22, 2016.

[29] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[30] I. Cohen, Y. Huang, J. Chen, J. Benesty, J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.

[31] M. Ghosh, M. Raihan, M. Sarker, M. Raihan, L. Akter, A. K. Bairagi, S. S. Alshamrani, and M. Masud, "A comparative analysis of machine learning algorithms to predict liver disease." *Intelligent Automation & Soft Computing*, vol. 30, no. 3, 2021.

[32] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 48, no. 3, pp. 128–138, 2017.

[33] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in *Emerging technology in modelling and graphics*. Springer, 2020, pp. 99–111.