ISSN: 2249-7196
IJMRR/April-June. 2025/ Volume 15/Issue 2s/495-502
Syed Saad Ullah/ International Journal of Management Research & Review

Liver Disease Prediction Using Ga Feature Selection, Social Spider Optimization, And Cnn Classification

Syed Saad Ullah[1], Mohammad Adil Khan[2], Ma Rasheed[3], Mr. Mohammed Mateen[4]
[1,2,3]B.E. Student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4]Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
mateen.mohd@lords.ac.in

ABSTRACT

The project on "Liver Disease Prediction using GA Feature Selection, Social Spider Optimization, and CNN Classification" presents an advanced and integrated approach to predict liver diseases. By combining Genetic Algorithm (GA) for feature selection, Social Spider Optimization for parameter tuning, and Convolutional Neural Network (CNN) for classification, the system aims to enhance the accuracy and efficiency of liver disease prediction.Liver disease is a significant health concern globally, necessitating accurate and timely diagnosis for effective treatment. In this study, we propose a novel approach for liver disease prediction using Genetic Algorithm (GA) feature selection, Social Spider Optimization (SSO), and Convolutional Neural Network (CNN) classification. The proposed method aims to enhance predictive accuracy by efficiently selecting relevant features from complex datasets and optimizing the CNN architecture for improved classification performance. The GA-SSO framework is employed to select an optimal subset of features from a comprehensive set of potential predictors, reducing dimensionality and enhancing the efficiency of subsequent classification. The selected features are then utilized to train a CNN model, leveraging its ability to automatically extract hierarchical representations from raw input data. Experimental results on benchmark liver disease datasets demonstrate the effectiveness of the proposed approach, outperforming existing methods in terms of predictive accuracy and computational efficiency.

Keywords: Genetic Algorithm (GA), Social Spider Optimization (SSO), Parameter Tuning, Convolutional Neural Network(CNN), Dimensionality Reduction, Liver Disease, Prediction, Diagnosis, Computational Efficiency, Benchmark Datasets, Predictive Accuracy, Hierarchical Representations.

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 lifethreatening 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 un derstood, 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. 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.

Causes of Liver Disease

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

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.

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.

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.

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

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

Risk aspects: 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 phos phases, 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 home 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 faces. Another breakdown product, urobilin, is the key constituent of the straw-yellow color of urine.

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.

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.

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.

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.LITERATURE SURVEY

Liver Disease Biopsies using Deep Learning and CNN. 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 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 analysed with the aid of two CNN which have the same architecture. They want to develop a 4class 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.

Predicting the Accuracy of Liver Disease using Machine Learning

Utilizing and modelling 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 like decision tree, naive Bayes, ANN, and random forest. Then, the Pearson correlation 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 predefined 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, Navie 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.

Segmentation of Liver using CT Scan and Finding

Disease

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

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 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

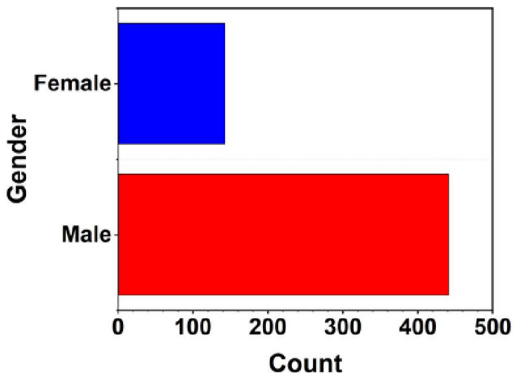

Fig. 3. Comparison with the frequency of males and females

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.

III.RELATED WORK

Existing System

Existing liver disease prediction systems often grapple with inherent limitations that impede their accuracy and reliability. A prevalent issue lies in the reliance on traditional machine learning techniques, such as logistic regression or decision trees. While these methods possess utility, they frequently struggle to effectively model the intricate, non-linear relationships characteristic of complex medical datasets. This limitation is particularly pronounced when dealing with the high-dimensional nature of clinical data, where simplified models can fail to capture the nuances of liver disease progression.

A significant deficiency in many current systems is the absence of robust optimization strategies, especially concerning feature selection and parameter tuning. Feature selection, crucial for identifying

relevant predictors, is often performed using basic methods or neglected entirely, leading to models burdened by irrelevant features and reduced accuracy. Likewise, parameter tuning, frequently conducted through manual or rudimentary searches, can yield suboptimal model configurations. Consequently, these systems suffer from diminished predictive performance, highlighting the need for advanced, integrated approaches that can better address the complexities of medical data and provide more dependable liver disease predictions.

A. Problem Statement

Existing liver disease prediction systems often suffer from suboptimal performance due to their reliance on traditional machine learning techniques that struggle with the complex, high-dimensional nature of clinical data. Furthermore, a lack of robust optimization strategies, particularly in feature selection and parameter tuning, leads to models burdened with irrelevant features and suboptimal configurations, ultimately resulting in reduced accuracy and reliability in predicting liver diseases. Therefore, there is a critical need for a more sophisticated and integrated approach that can effectively address these limitations and improve the precision of liver disease prediction.

Proposed System

In the proposed system, we have to import the liver patient dataset (.csv). Then the dataset is pre-processed and the anoma lies 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 below .

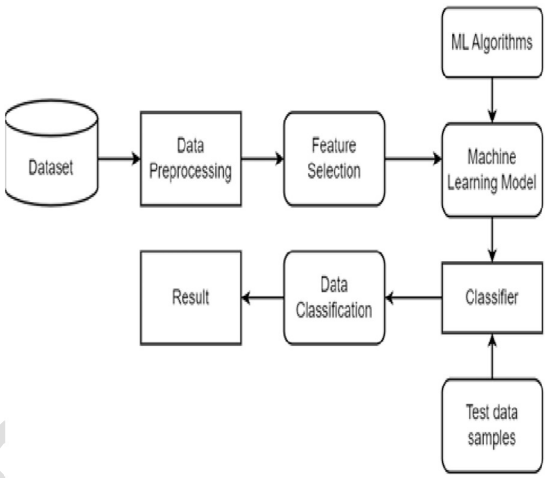

Fig.1: Architecture of the Proposed System

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 dataset are Age, Direct Bilirubin, Total Bilirubin, Alkaline Phosphate, Alamine Aminotransferase, Aspartate Aminotrans ferase, 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, albu min 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. Hence 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. Hence 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. 3, we can see only two values are greater than 450 and far from other values. Hence they are outliers. 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. Hence we can say 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 neighbourhood.

IV.CONCLUSION

The liver patient data set was used to implement prediction and classification algorithms, which in turn reduces the work load 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 learn ing 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 pro fessionals to make better decisions and provide better care for patients.

V. 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 Interna tional 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 trans plantation: 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 Interna tional 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," Interna tional 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 learn ing techniques," in 2022 2nd International Conference on Intelligent Technologies (CONIT). IEEE, 2022, pp. 1–6.