(J I D P T S)
ISSN(O):2581-6934

Predicting and Classifying Liver Functionality Disease through Machine Learning Methods

Dr V Ramesh[1]
[1]Assistant Professor Department of Computer Science,
[1]Kakatiya Government College, Hanumakonda, Telangana.

Abstract: Among the most prevalent harmful diseases in the world, liver disease has been on the rise in recent decades and is predicted to keep rising. Researchers found it extremely difficult to predict the disease using massive medical data sets. They are using machine learning strategies like clustering and classification to tackle this issue. The primary goal of the research is to employ classification algorithms to predict a patient's likelihood of getting liver disease. The contrast of these algorithms is based on how effectively they identify data and how quickly they run. Taking into account these performance variables, the algorithm which functions as a more effective classifier is chosen. A liver disease that continues for more than six months is referred to be a chronic condition of the liver. Therefore, the proportion of patients who get the disease will be used to provide both positive and negative information. Classifiers have been employed to process liver disease percentages, and the result is shown as a confusion matrix. When a training data set is available, we offer several classification strategies that can significantly enhance classification performance. Subsequently, excellent, and poor values are identified using a machine learning classifier.

Keywords: Machine learning, Naive Bayes, Random Forest

INTRODUCTION

The medical field will discover this method useful in predicting illnesses. Using the patient's blood reports, the model will help evaluate if the liver the status is normal or abnormal. The medical companies will find this patient information helpful during their treatment. The present models use an array of algorithms for machine learning that provide results with lower accuracy and are unable to manage big data bundles. In the process of training and assessing the liver datasets, there is a noticeable lack of performance. While these earlier algorithms have proven adequate, further work has to be done to improve their prediction rate to improve the accuracy of liver disease detection [1].The hardest structure in the body of a person, the liver is an efficient gland because, among its many roles, it produces and secretes bile. The rib cage covers the liver which is situated in the upright part of the abdomen. The two main lobes are overflowing with tiny lobules. There exist two different supply bases for the liver cells. The portal vein delivers nutrients from the intestines, whereas the hepatic artery carries blood with plenty of oxygen pushed from the heart. Because classification algorithms may determine an individual's probability of developing liver failure based on certain traits or qualities, they are frequently designed to be used for liver disease forecasting. The F-Tree technique has the highest level of precision of all the algorithms examined, compared to the solutions now in utilize, making it an excellent choice for predicting diseases of the liver. Liver disease classification frequently employs feature selection when combined with fuzzy K-means classification techniques. These methods can assist in identifying important features that help differentiate between various liver diseases [2].The study compared the results after analyzing the performance of these innovative algorithms utilizing evaluates such as data correctness, data effectiveness, and correction rate. Compared the multilayer perceptron technique to the other algorithms examined in the research, the results indicated that it attained the highest level of accuracy. In order to make the distinction between cirrhosis, hepatitis, and non-liver diseases, researchers used Bayesian classification in this work. The liver patient dataset was divided into numerous disease classes utilizing both the FT tree and Naive Bayes algorithm design. The accuracy and execution time of each method were evaluated. According to the research study, the FT tree approach was not as effective in terms of execution time as the Naive Bayes algorithm.

Related Work

Authors examined the efficacy of several categorization techniques in the identification of liver conditions in this work. Decision Trees, Naive Bayes, Multilayer Perceptron's, K-Nearest Neighbors, Random Forests, Random Forests, and Logistic Regressions featured amongst these methods. The authors used standards such as accuracy, recall, sensitivity, and specificity to evaluate their performance. Out of all the algorithms examined, Naive Bayes provided the best results in terms of precision, based to the results. In addition, it was discovered that the Random Forest and Logistic Regression algorithms provided good results when recall had been taken through consideration. In our current day, machine learning is growing more prevalent. The use of machine learning (ML) for backing pharmacological and therapeutic diagnostics is increasing. Most of them share, however, a fast increasing access to enormous amounts of data [3]. A comprehensive overview of the primary challenges and commonly employed solutions to machine learning problems in medicine can be found in the text that follows. By selecting just a few of the input's beneficial features and removing the other aspects, feature selection aims to lessen the amount of work required. This work utilized a combination of sampling and oversampling methods to balance the dataset for accurate liver disease prediction. J48, Multilayer Perception, Random Forest, Multiple Linear Regression, Support Vector Machines, and Genetic Programming were among the classification methods which were employed. The most successful Random Forest algorithm for predicting diseases of the liver was one which employed oversampling at greater rates [3].

Methodology

Artificial Neural Network

A neural network incorporating back propagation was developed. There were a total of ten input neurons in the input layer in this network. The total amount of features in the dataset can be represented by the number of inputs. Rectified Linear Unit activation function is employed in the input layer. There is just a single layer in the output layer that makes use of the sigmoid activation function. In order to get the necessary recognition rate needed to identify a patient's liver condition. To get the desired optimal outcome, the neural network models' parameters must be changed. The learning rate, momentum rate, and hidden neurons are these variables. These parameters are all included in the neural networks that back propagate. The system's learning rate is its processing power, while the learning speed is defined by the system's momentum rate [4]. In order to achieve the best result, the network's hidden neuron count needs to be adjusted. The number of neurons essential at the hidden layer is determined by experiment to identify the optimal neurons that can properly represent the characteristics in the input dataset and yield the desired best outcome. By changing the neurons, the amount of neurons needed in the hidden layer has been experimented tested. Because of its simplicity in derivatives and kind switching capacity, the sigmoid function was used in the output layer [5].

Support Vector Machine

It attempts to identify the optimal hyper plane for splitting the data into several groups. A possible approach for putting SVM in use with Python is using the scikit-learn module. Upon been sorted, the data is divided into training sets and test sets. Twenty out of every hundred is the testing data which is utilized. Sixty-one out of every two hundred is the fixed set utilized in algorithm preparation. An assortment of hyper planes can be generated employing an SVM.The hyper plane with the largest distance to the closest training data point for every known as the functional margins achieves a decent separation since, usually, a greater margin results in a smaller classifier generalization error[6]. Decision boundaries known as hyper planes aid in the classification of the data points. The classes of the data points that lie on each side of the hyper plane are different. Additionally, the quantity of features impacts the hyper plane's size. A hyper plane that achieves an adequate separation, known as a functional margin, is the one that is closest to the nearest data point of any class. An increased margin indicates a lower classifier reduction fault.

Random Forest

Utilizing data mockups, the random forest generates decision trees and extracts the projected results from each tree. In the end, the best explanation is chosen by voting. Following been sorted, the data is separated into training sets and test sets. Twenty of every hundred is the testing data set. One hundred and sixty out of every two hundred is the fixed set used for algorithm setup. The data is divided into several groups and subgroups by the application. If an individual develops lines connecting subgroups into a group, lines joining data points through a subgroup, and so on. The erection seems to resemble a tree [7]. Predicting the Accuracy of Liver Disease using Machine Learning Specialists throughout the globe are increasingly considering modeling and using medical data sets. The primary goal is to reduce the amount of time needed between the liver test and the production of the report and outcome. They employed a few machine learning methods, including random forests, ANNs, decision trees, and naive Bayes. Then, the TP (true positives), FN (false negatives), FP (false positives), and TN (true negatives) anomalies are found using the Pearson correlation [8]. This is done in order to determine the algorithms' affectability, specificity, and accuracy. Using predetermined formulae, the generated words are utilized to calculate the sensitivity, affectability, specificity, and accuracy. A single number representing either a zero or a one will be the result; the binary zero will indicate that the patient's liver is healthy, and the binary one will indicate that the patient's liver is unwell. The data entered by the user will be recorded and used to retrain the algorithm. The algorithm will be trained and skilled to advance with accuracy thanks to these extra tabulated values. The data set was gathered by the authors from the UCI machine. Charts show the outcomes of each of these algorithms.

Segmentation of Liver using CT Scan and Finding Disease

Some organs are undetectable to the human retina when using normal X-ray equipment. These factors reinforce the benefits of using CT scans, which are superior to X-rays in terms of their ability to display structural details. The images taken during the CT scan will have exact resolution. They suggested segmenting the Scan picture, determining the liver location, and then masking it off from the backdrop using WTA. The final stage involves calculating the impacted area percentage. The CT scan is a very accurate technology that should be used in order to determine the severity or progression of the liver disease. In the medical profession, this CT scan is often utilized to get data about the humanoid development. The initial step is to carefully examine the images to identify various components [9]. The Erode and Dilate algorithm is employed to handle images in the next phase. Here, viewing point values are adjusted. In addition, they undertake WTA treatment to separate them from the liver area. The liver area and the non-liver area are the two outputs provided by the WTA. They call the good that is produced "cropping." The obtained duplicate is subsequently altered to create a binary where this organ is white and the rest of the substance is black. Subsequently, median filtering is applied to reduce noise while smoothing up the texture.

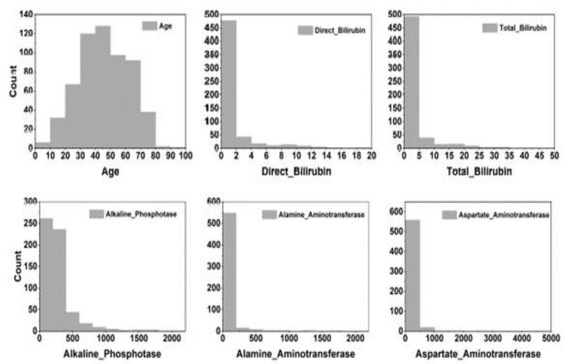


Figure 1: Histogram of various patients' frequency distributions related to attribute

Results

Following that, the dataset is pre-processed to remove deficiencies as well as empty cells with the goal to improve the precision of the liver disease prediction. Following that, we construct a confusion matrix to get greater clarity on the amount of accurate and inaccurate predictions. In the past, a number of classification and prediction processes were used, and the accuracy was verified by implementing, if feasible, combinations of several techniques. Creating a code that achieves 90% exactitude is our goal. The benefits include enhanced accuracy, early risk prediction, and better classification. The correlation between each characteristic that is shown in the data set appears in Figure.2. The greater relationship or relationship between the values of a particular characteristic and the second attribute is apparent in a lighter color separating them on the graph. By employing the dataset values for training the model and predicting the dataset value to test it, the accuracy of each model is determined [10]. Accuracy is determined by the number of the model's predictions come true. When compared to other models, Random Forest has the best accuracy, shown in Figure.3. Additionally, we are able to discover about the outliers in that typical by examining the box plot. The IQR (interquartile range) look at is used in the box plot to determine the aberrations.

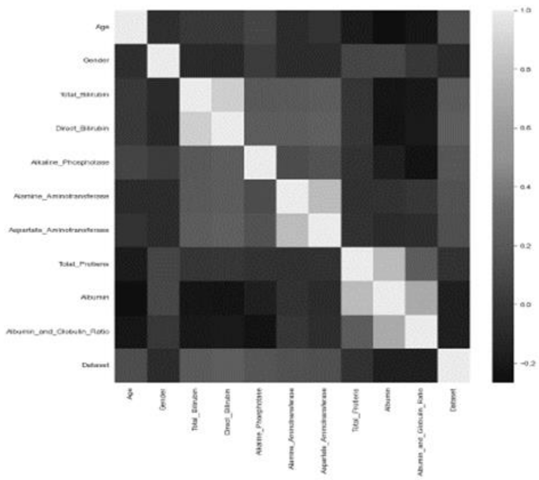


Figure 2: The attribute correlation graph

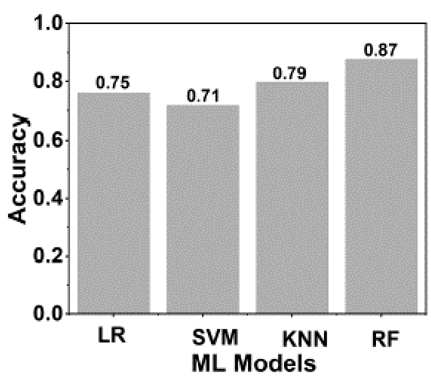
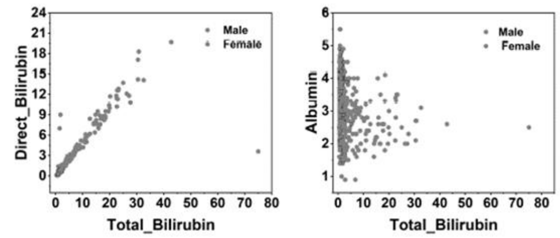


Figure 3: Each machine learning technique used for classification is measured with alternatives in regard to accuracy.

Researchers and scientists are deliberately enhancing medical diagnostics with machine learning models and classification algorithms almost anywhere on the planet, and results are improving. The present study uses Support Vector Machines, K-Nearest Neighbor, and Logistic Regression to predict liver disease. As everybody knows, the liver is the biggest internal organ in the body and is responsible for several critical processes like the synthesis of glycogen, the generation of bile, triglycerides and cholesterol, and blood clotting factors and proteins [11]. A decline in function often needs to impact far more than 75% of the liver tissue. Therefore, it is essential to



detect the condition at an early stage in order that it may be treated before it develops.

Figure 4: Scatter plots showing the correlation between albumin & total and direct bilirubin

Conclusion

The use of prediction and classification algorithms, which lessens the stress on physicians, was made possible by the liver patient data set. We recommended using machine learning methods to assess the patient's overall liver health. A chronic liver disease has lasted for at least six months. Our suggested classification schemes can greatly improve classification performance when a training data set is available. Subsequently, proficient and weak values are determined using a machine learning classifier. As a result, the recommended classification model's outputs demonstrate accuracy in forecasting the final result. Our exploration will concentrate mainly on utilizing Machine-learning techniques to predict liver disease. The accuracy of liver disease prediction and classification models will likely be improved in the future by incorporating more diverse data sources, combining multiple machine learning techniques, and training machine learning models to predict a person's probability of developing liver disease based on their unique characteristics.

REFERENCES

[1] K. M. Sazzadur Rahman, F.M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy(2019). "A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms".

[2] Dutta, K., Chandra, S., & Gourisaria, M. K. (2022). Early-Stage detection of liver disease through machine learning algorithms.

[3] Khan Idris, Sachin Bhoite (2019), "Applications of Machine Learning for Prediction of Liver Disease"

[4] Panduranga Vital Terlapu (2021), "Intelligent Liver Disease Prediction (ILDP) System Using Machine Learning Models"

[5] Joloudari, J. H., Saadatfar, H., Dehzangi, A., &Shamshirband, S. (2019), Computeraided decision-making for predicting liver disease using PSO-based optimised SVM with feature selection.

[6] 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.

[7] 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

[8] 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.

[9] 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.

[10] R. Kalaiselvi, K. Meena, V. Vanitha, ―Liver Disease Prediction Using Machine Learning Algorithms‖, International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Oct 2021.

[11] M.Ardra Meghana Simon, N. Kalyan Saradhi, P.Sahithi, R. Sai Kiran, Mrs. A. Aparna, ―Liver Disease Prediction using Machine Learning‖, A Journal Of Composition Theory, Volume XIV, Issue VI, JUNE 2021.

[12] S. Ramana, S. C. Ramu, N. Bhaskar, M. V. R. Murthy and C. R. K. Reddy, "A Three-Level Gateway protocol for secure M-Commerce Transactions using Encrypted OTP," 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2022, pp. 1408-1416, doi: 10.1109/ICAAIC53929.2022.9792908.

[13] N.Bhaskar, S.Ramana, & M.V.Ramana Murthy. (2017). Security Tool for Mining Sensor Networks. International Journal of Advanced Research in Science and Engineering, BVC NS CS 2017, 06(01), 16–19. ISSN Number: 2319- 8346

[14] Karunakar Pothuganti, (2018) ‘A comparative study on position based routing over topology based routing concerning the position of vehicles in VANET’, AIRO International Research Journal Volume XV, ISSN: 2320-3714 April, 2018 UGC Approval Number 63012.

[15] Ramana, S., Bhaskar, N., Murthy, M.V.R., Sharma, M.R. (2023). Machine Learning for a Payment Security Evaluation System for Mobile Networks. In: Rajakumar, G., Du, KL., Rocha, Á. (eds) Intelligent Communication Technologies and Virtual Mobile Networks. ICICV 2023. Lecture Notes on Data Engineering and Communications Technologies, vol 171. Springer, Singapore. https://doi.org/10.1007/978-981-99-1767-9_26

[16] N. Bhaskar, S. Ramana and G. M. Kumar, "Internet of Things for Green Smart City Application Based on Biotechnology Techniques," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083965.

[17] K. Pothuganti, B. Sridevi and P. Seshabattar, "IoT and Deep Learning based Smart Greenhouse Disease Prediction," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2021, pp. 793-799, doi: 10.1109/RTEICT52294.2021.9573794.

[18] Ahmad and K. Pothuganti, "Smart Field Monitoring using ToxTrac: A Cyber-Physical System Approach in Agriculture," 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 723-727, doi: 10.1109/ICOSEC49089.2020.9215282.