JOURNAL OF BASIC SCIENCE AND ENGINEERING

NEW ANALYSIS LIVER DISEASE DIAGNOSIS USING LOGISTIC REGRESSION

[1]Allu Aneesha [2]Dr. Prasuna Grandhi [3]Madhuri Draksharam
[1]M. Tech Scholar, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[2]Associate Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
[3]Assistant Professor, Dept. of CSE, St. Ann's College of Engineering & Technology, Chirala.
e-mail: grandhiprasuna@gmail.com

ABSTRACT: Liver disease has become a prominent global health concern, with conditions like cirrhosis and liver cancer ranking among the leading causes of mortality worldwide. We are going discuss how to predict risk of liver disease for a person, based on the blood test report results of the user. The major goal of this study is to employ classification algorithms to distinguish between liver patients and healthy people. Chemical components present in the human body, as well as tests such as SGOT and SGPT, determine whether a person is a patient, or whether they need to be diagnosed. 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. In this regard, this study provides an extensive review of the progress of applying Artificial Intelligence in forecasting and detecting liver diseases and then summarizes related limitations of the studies followed by future research. These are decision trees, random forests, and logistic regressions. Accuracy, specificity, sensitivity, and the area under the receiver operating characteristic (ROC) curve are the metrics that are used in order to assess the performance of these models. The programming language which was used is python and machine learning Sklearn was used to build the model using classification algorithms like Logical regression, SVM. A data-driven technique that makes use of supervised learning algorithms is presented in this research report as a method for estimating the likelihood of developing liver disease. Subsequentaluation of the Five algorithms for machine learning, restated. The goal of this research is to analyse prediction algorithms in order to relieve doctors of their workload. Thus, the outputs of the proposed classification model show accuracy in predicting the result.**Keywords:** Liver disease, Confusion matrix, Use case diagram, back propagation algorithm, Random Forest, Data Mining, Deep Learning, support vector machine.

1. INTRODUCTION

Worldwide Factors such as air pollution, poor dietary habits, excessive alcohol consumption, and misuse of medications contribute to the rising prevalence of liver diseases each year this spectrum of conditions, including cirrhosis, liver cancer, and fatty liver disease, poses significant threats to personal health and overall well-being [1]. In rural areas the intensity is still manageable but in urban areas, and especially metropolitan areas the liver disease is a very common sighting nowadays. Liver diseases cause millions of deaths every year. Viral hepatitis alone causes 1.34 million deaths every year[2]. These technologies will assist physicians in making accurate patient decisions, and with the use of automatic classification tools for liver illnesses the patient wait at

liver experts such as endocrinologists will be reduced [3]. 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. Other vital functions of the liver include bile production, protein production, storing and releasing glucose, processing haemoglobin, blood cleaning, immune factor production, clearing bilirubin, it is the primary and most crucial body organ, and the maintenance of its health is essential for improved overall health [4]. Analyze the efficacy of the random forest, decision tree, and the logistic regression supervised learning algorithms in determining the likelihood of liver disease [5]. Evaluate the prediction models created by the three algorithms with respect to accuracy, sensitivity, specificity, and area under the ROC curve [6]. The aim of this study was to examine a machine learning-based model capable of accurately identifying patients afflicted with liver disease. Utilizing a dataset sourced comprising Indian liver patient records [7] the study seeks to explore the effectiveness of five distinct Machine learning techniques such as Logistic Regression, Random Forest, Support Vector Machines, and Gradient Boosting, reformulated. and K-Nearest Neighbors [8]. 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 [9].

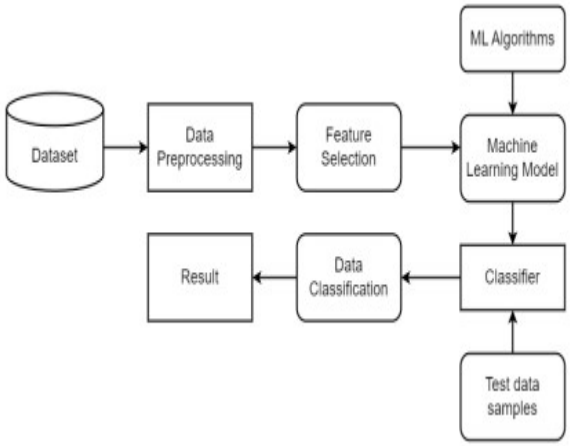

Figure.1. Processing of ML algorithms

2. RELATED WORKS

The application of the K-Nearest Neighbors (KNN) classification algorithm for Classification of large-scale medical health data which could potentially be utilized for liver disease prediction by leveraging relevant medical data [10]. The Forecasting liver disease utilizing gradient boosting machine learning methodologies, highlighting the importance of feature scaling to enhance the accuracy and effectiveness of the predictive models [10]. Spider is an open-source cross-platform integrated development environment (IDE) for scientific programming in the Python language includes support for interactive tools for data inspection and embeds Python-specific code quality assurance and introspection instruments and Rope [11]. Liver Patient Dataset is used, which is based on Indian patient and Random Forest (RF) algorithm is used to predict the disease with

different pre-processing techniques. Data set is checked for skewness outliers and imbalance using univariate and bivariate analysis and then suitable algorithms used to remove outliers [12]. The set for algorithm preparation is fixed as one hundred and sixty out of every two hundred to SVM forms either a set of hyper planes It is built in an immeasurable-extension space. A hyper plane that has the farthest distance to the nearest data point of any class achieves a good separation, called a functional margin [13]. The primary goal of this study is to apply six different supervised machine learning classifiers to develop a reliable method of identifying people who suffer from chronic liver disease [14]. Machine learning is being employed in this research because of the large amounts of data being used to make predictions about the future it has been shown in the past that assessment outcomes may be somewhat subjective [15].

3. SYSTEM ARCHITECTURE

Initially, we will take clinical data (Indian Liver Patient Dataset) as input for liver disease diagnosis model. Indian Liver Patient Dataset (ILPD) was obtained forms the UCI Machine Learning Repository [16]. Data pre-processing is the process of transforming raw data into an understandable format in data mining as we cannot work with raw data. Pre-processing of data is primarily to check the data quality [17]. The hyper plane that maximizes the distance to the nearest data point in the training set provides a reliable separation between classes. Automatic systems can be used to overcome these limitations of conventional disease diagnosis methods. The main aim of clinical decisions is to provide a correct and timely diagnosis [18]. Evaluate supervised learning algorithms for their potential use in forecasting the risk of liver disease in low-resource areas.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

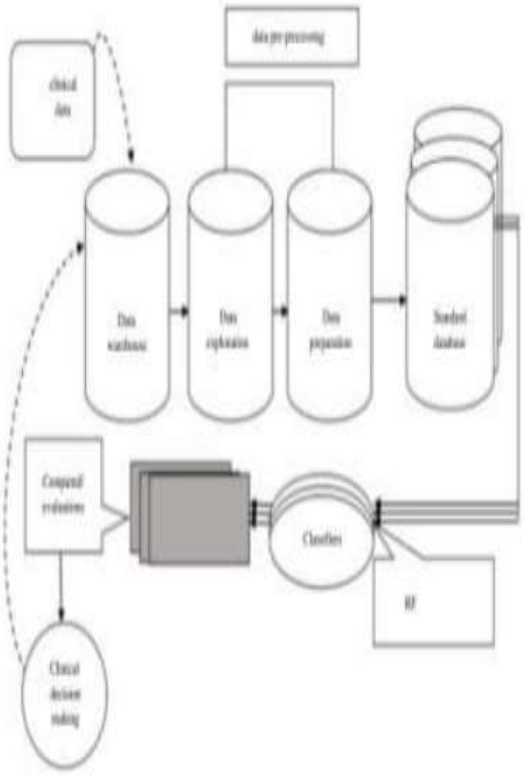

Figure 2 System Architecture

4. PROPOSED SYSTEM

In this proposed framework we use dataset which consists of Indian liver patient data use of this dataset carry out pre-processing and feature selection on this particular dataset. At this point we have enormous number of features in dataset; hence feature selection is a vital part in our Machine Learning model results [19]. Researchers construct hybrid models which combine diverse machine learning algorithms and approaches to enhance the precision of their predictions. The quality of the models is evaluated using performance [20]. The algorithm operates on the fundamental principle of similarity, where similar items are proximate to each other. KNN functions by identifying the K nearest training sample and making predictions [21]. The system asks you to enter your details including age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT. Values of last eight parameters mentioned here, can be known by blood test report of the user [22]. It represents flow of process which we have implemented to develop the prediction system

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure. 3. Work flow Proposed Model

5. MATERIALS AND METHODS

Class balancing and ranking features in the balanced data we will describe the dataset we utilized and the primary stages of the selected strategy for forecasting the risk of liver sickness. Finally detail the ML models that were used in order to make sense of the experimental result. [23]. The logistic regression model itself simply models probability of output in terms of input and does not perform statistical classification though it can be used to make a classifier, for instance by choosing a cutoff value and classifying inputs [25]. This approach involves determining the distance between each sample and the unknown sample. Subsequently, the K known samples that exhibit the highest similarity to the unknown sample are selected.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

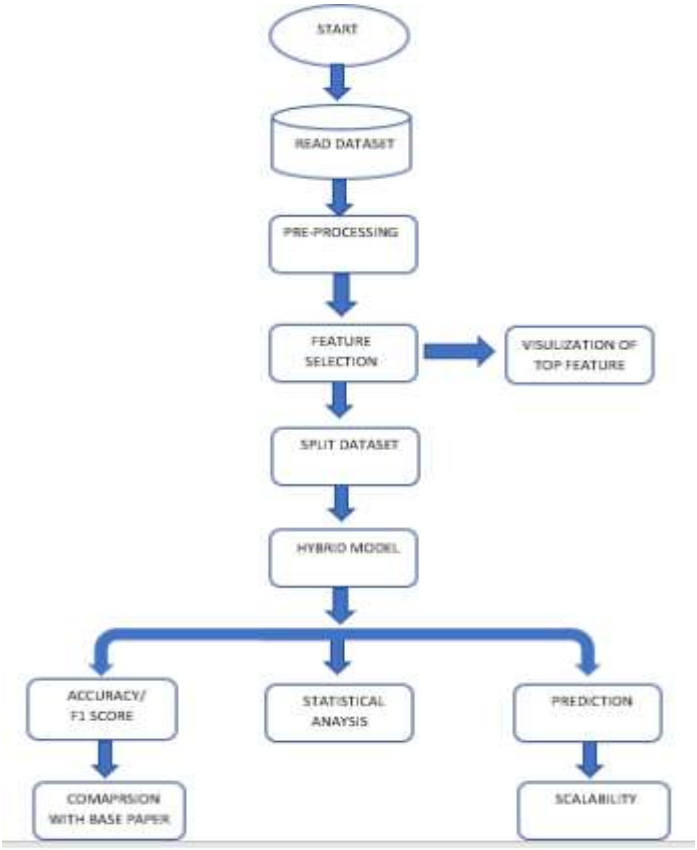

Figure. 4.: Flowchart of Proposed approach

6. CLASSIFICATION ALGORITHM

Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification accuracy. Different data mining algorithms like Logistic Regression, Support Vector Machines (SVM), K Nearest Neighbour (KNN) [23].

Algorithm:

Steps in Logistic Regression: We will follow these steps in order to successfully implement the Logistic Regression using Python.

Ster1: Data pre-processing

Step2: Fitting of logistic regression to the training set

Step 3: Predicting the test result

Step4: Testing the accuracy of the result (the creation of the confusion matrix)

Step5: Step of visually representing the result of the test set.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

7. ANALYSIS AND 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. It will also help in evaluating the models on a large-scale real-time data, and thereby making modifications if required according to the feedback of physicians, will yield a robust diagnostic model. The concepts of deep learning can be combined with the ML algorithms in the development of predictive models. In order to effectively spend healthcare resources to stop the course of diseases and improve patient outcomes, it is necessary to precisely identify persons at high risk. In conclusion, our research shows that supervised learning algorithms, and the random forest algorithm in particular, can accurately forecast the risk of liver disease from patient data. The findings highlight the potential public health uses of data-driven methods to illness risk prediction. The effectiveness of a hybrid model kind of binary classification model, may be seen by plotting its Receiver Operating Characteristic (ROC) curve. The hybrid-score of 88.09 indicates a specific threshold or cutoff value chosen for classifying instances as either positive or negative based on the hybrid model's predictions

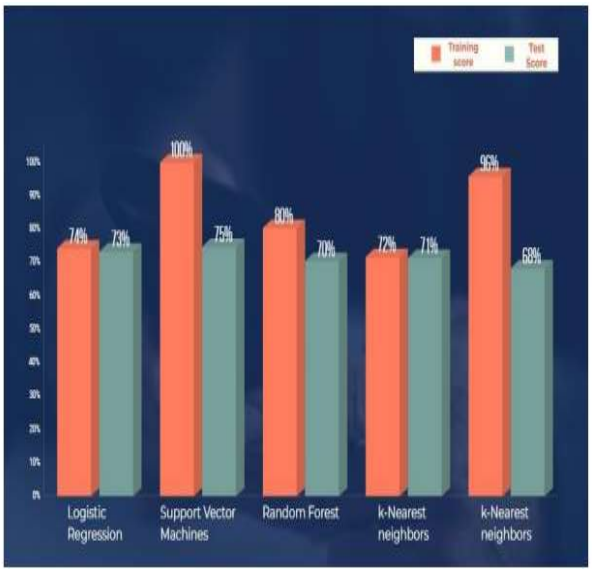

Figure 5. Comparison of model scores between the training set and test set

8. CONCLUSIONS AND FUTHER WORK

Additionally, data pre-processing steps were conducted to address missing values, standardize features, rename columns, and encode categorical variables. To facilitate model training and assessment, the dataset was partitioned into training and testing datasets. Although people are becoming more conscious of health nowadays and are joining yoga classes, dance classes; still the sedentary lifestyle and luxuries that are continuously being introduced and enhanced; the problem is going to last long. Despite the truth that people are becoming more health-conscious and enrolling in yoga and dancing classes, the sedentary lifestyle and facilities that are continuously being delivered and improved will proceed to be an issue. 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. But further data proving its validity and efficiency is required for its constant use by physicians. Furthermore, it would be beneficial to examine the performance of the model on other liver disease datasets. The current study utilized a specific dataset, but evaluating the model's performance on diverse datasets from different populations or regions can provide a better understanding of its generalizability and robustness.

9. REFERENCES

[1] T.G. Cotter, M. Rinella, Nonalcoholic fatty liver disease 2020: the state of the disease, Gastroenterology, 158, 1851-1864 (2020).

[2] S. Lee, H. Huang, M. Zelen, Early detection of disease and scheduling of screening examinations, Statistical Methods in Medical Research, 13, 443-456 (2004).

[3] S. Samarpita and R. N. Satpathy, Applications of Machine Learning in Healthcare: An Overview,2022 1st ICIDeA, Bhubaneswar, India, 51-56 (2022).

[4] K. Sellamuthu, S. P, P. K and R. S, Liver Disease Prediction using Logistic Regression, 2022 8thICSSS, Chennai, India, 01-06 (2022). https://doi.org/10.1109/ ICSSS54381.2022.9782179

[5] C.C. Wu, W.C. Yeh, W.D. Hsu. et al. Prediction of fatty liver disease using machine learning algorithms, Computer Methods and Programs in Biomedicine, 170, 23-29 (2019).

[6] Ratna Raju Mukiri, B. Suresh Kumar , B.V.V.Siva Prasad, Effective Data Collaborative Strain using Rectree Algorithm,SUSOM- 2019.

[7] Abderrahmane Ed-daoudy*, Khalil Maalmi, ―Realtime machine learning for early detectionof heart disease using big data approach‖, 2019.

[8] RahmaAtallah, Amjed Al-Mousa, ―Heart Disease Detection Using Machine LearningMajority Voting Ensemble Method‖, 978-1-7281-2882-5/19/$31.00 IEEE, 2019.

[9] Gautam Chitnis,Vidhi Bhanushali, Aayush Ranade, Tejasvini Khadase, Vaishnavi Pelagade, Jitendra Chavan, ―A Review of Machine Learning Methodologies for Dental Disease Detection‖, IEEE India Council International Subsections Conference (INDISCON), 2020.

[10] YedilkhanAmirgaliyev, Shahriar Shamiluulu, Azamat Serek, ―Analysis of Chronic Kidney Disease Dataset by Applying Machine Learning Methods‖, 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

[15] Abdalrada A S, Yahya O H, Alaidi A H M, Hussein N A, Alrikabi H T and Al-Quraishi T 2019 A predictive model for liver disease progression based on logistic regression algorithm Period. Eng. Nat. Sci. 7 1255–64
[16] Arbain A N and Balakrishnan B Y P International Journal of Data Science and Advanced Analytics
[17] Singh A K 2019 A Comparative Study on Disease Classification using Machine Learning Algorithms SSRN Electron. J. 114 1–10
[18] Haque M R, Islam M M, Iqbal H, Reza M S and Hasan M K 2018 Performance Evaluation of Random Forests and Artificial Neural Networks for the Classification of Liver Disorder Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018 1–5
[19] Arshad I, Dutta C, Choudhury T and Thakral A 2018 Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques Proc. 2018 Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2018 163–8
[20] Mala K, Sadasivam V and Alagappan S 2015 Neural network based texture analysis of CT images for fatty and cirrhosis liver classification Appl. Soft Comput. J. 32 80–6
[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