PREDICTING CRITICAL LIVER FAILURE USING SUPERVISED MACHINE LEARNING ANALYSIS

Vishwan T[1], Santhosh Sivan E[2], Elakkiya Priya S[3], Maheswari M[4]

[1] *Student, Computer science and engineering, Anand Institute of Higher Technology, Chennai, India*
[2] *Student, Computer science and engineering, Anand Institute of Higher Technology, Chennai, India*
[3] *Assistant Professor, Computer science and engineering, Anand Institute of Higher Technology, Chennai, India*
[4] *Assistant Professor, Computer science and engineering, Anand Institute of Higher Technology, Chennai, India*

ABSTRACT

The liver's purpose is to filter blood as it circulates through the body, turning nutrients and medications ingested from the digestive system into ready-to-use molecules. The liver also removes poisons and other chemical wastes from the blood and prepares them for excretion. Liver failure occurs in the cells of the liver of an individual. Nowadays, machine learning is used in healthcare systems to forecast illness progression. Data is the most critical prerequisite for artificial intelligence. The freshly obtained dataset of liver failure is utilized to construct a machine learning project for predicting this disease. Pre-processing techniques like univariate and bivariate analysis will be done before giving them to the algorithm. The data is visualized for a better understanding of the features, and a classification model is built based on them. Existing classification Algorithms are applied in the dataset and are compared with each other, based on their performance metrics such as accuracy, F1 score, recall the best one among them is deployed for the web application.

Keywords: - *Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), and Decision Tree (DT), Machine Learning (ML), Features, Classifiers.*

1. INTRODUCTION

The liver is responsible for a variety of vital processes, including the production of blood proteins that help in clotting, the transfer of oxygen, and the maintenance of the immune system. Making bile, a chemical required to aid digestion, Assisting the body in storing sugar (glucose) as glycogen, getting rid of toxic compounds in the bloodstream, such as drugs and alcohol, as well as breaking down saturated fat and generating cholesterol. Liver failure is a potentially fatal condition that needs rapid medical intervention. The liver is the body's biggest organ, and it is responsible for digestion and the elimination of toxins. The leading causes of mortality are liver disorders and cirrhosis.

Machine learning has had a tremendous influence on the biomedical area for the prediction and detection of liver disease. Machine learning provides a guarantee for improved illness identification and prediction, which has sparked interest in the biomedical industry, as well as increasing the objectivity of the decision-making process. Supervised learning is essentially synonymous with categorization. The labelled examples in the liver training data set provide supervision in the learning. One of the most "death-dealing diseases on the globe" is liver disease. It takes a team of experts to assess the damage. Medical problems may be readily solved utilizing machine learning techniques, and the cost of diagnostics can be decreased. The major goal of this study is to forecast findings

more accurately and lower the cost of diagnostics in the medical industry. As a result, we employed several categorization algorithms to determine whether individuals had liver disease or not.

2. RELATED WORKS

For the prediction of liver disease, a hybrid classification system was presented. Furthermore, data sets are gathered from the Kaggle database of Indian liver patient records. The proposed model integrates three classifier algorithms: Logistic Regression, Random Forest, and KNN. The suggested method was implemented in Python using the Spyder tool, and the results were evaluated in terms of accuracy, precision, and recall. [1]. By comparing various data mining classification methods, they predicted and analyzed liver problem disorders to improve performance accuracy. By analyzing the performance accuracy of each classification algorithm, three classification algorithms were developed to predict liver condition disorders. SVM, NB, and C4.5 decision tree classifiers are used. K-fold cross-validation is used to split data depending on the Test set used to test the model and the Training set used to train the data. [2]. The objective is to examine decision tree algorithms for diagnosing liver illness, such as J48, LMT, Random Tree, Random Forest, REPTree, Decision Stump, and Hoeffding Tree. Decision tree methods are used to examine the liver dataset and assess their performance against seven performance parameters [3]. Six machine learning techniques have been applied including LR, KNN, DT, SVM, NB, RF and the performance of these techniques were estimated on various perspectives such as accuracy, precision, recall, f-1 score. Moreover, the performance was compared using the receiver operative characteristic (ROC) [4]. They applied five sort of classifiers that is Naïve Bayes, logistic regression, support vector machines, Random Forest, K Nearest Neighbor for the examination of liver malady. The classification exhibitions are assessed with 5 distinctive by and large execution measurements, i.e., precision, kappa, mean absolute error (MAE), Root mean square error (RMSE), and F measures [5]. Thus, by referencing the work done in the related articles, we were finally able to do a analysis with the help of our dataset [6] for developing our front-end web application for the detection of Liver failure.

3. EXISTING SYSTEM

The diagnosis of liver failures is done using a B-mode ultrasound (BUS) imaging technique. They attempted to enhance BUS-based computer assisted diagnosis for liver failures by transferring knowledge from multi-view CEUS pictures, which included the arterial phase, portal venous phase, and delayed phase. This supervised TL used a support vector machine plus (SVM+), a specially developed transfer learning (TL) classifier for paired data with shared labels. It was also the first time that a nonparallel hyper plane based SVM+ (NHSVM+) was presented to increase TL performance by transferring per-class knowledge from the source domain to the target domain.

4. PROPOSED SYSTEM

The proposed method is analysing existing machine learning models based on the past data of liver failure like the features and target column is identified first using our domain knowledge related to health care. Then dataset is visualized for better understanding of features and then the dataset is split into two parts normally in 7:3 ratio where the data is used for training and testing. Different algorithms accuracies is measured and compared using their performance metrics and the best algorithm is applied on the trained data to get better understanding of the features and a classification model is built based on their learning.

5. IMPLEMENTATION

The system is divided into four sections. First module in the system is Data Pre-Processing, Validation techniques were performed to tackle the missing value, duplicate value in the dataset. The second module is Data Analysis and Visualization, this was very helpful in exploring and getting to know the dataset and helped us with identifying patterns, corrupt data, outliers, and much more. The third module is training and testing the machine learning model, we trained and tested four classification classifiers with our dataset and calculated their performances. The fourth module is a web application having an easy to use user interface to collect liver report values from the user.

5.1. Machine Learning

It is an Artificial Intelligence application that allows computers to learn and develop without being explicitly programmed. It is the process of teaching a machine to use its prior expertise to solve a problem. A significant volume of data may be processed, analyzed, and abstracted by the machine. Its distinct and intelligent behavior

enables it to identify correlations and insights that are not easily obvious to the human eye, allowing it to draw abstractions from experience.

5.2 System Architecture

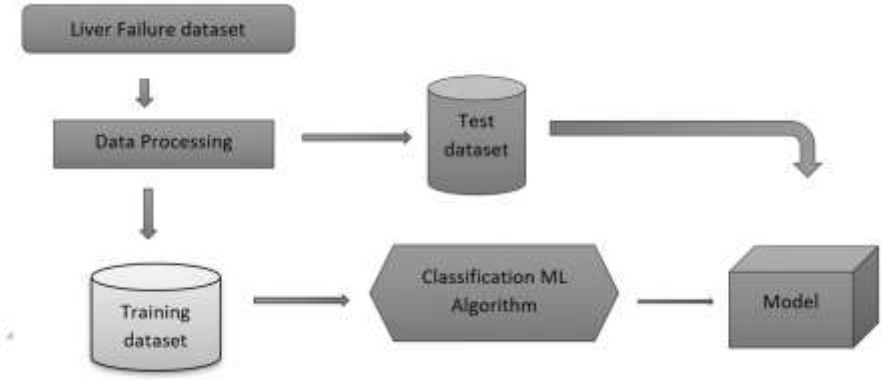

Fig -1: Architecture Diagram for the System

5.3. Dataset

Dataset was obtained from the Kaggle website (www.kaggle.com). It is a repository with approximately 50,000 publicly accessible datasets. Labels that indicate the presence or absence of Liver Failure are noted as present or absent. The dataset is an open-source.csv data collection that anybody may utilize. It started with 584 patients data in rows and characteristics in 11 columns, but after pre-processing, it was reduced to 579. There were 420 men and 179 women.

5.4. ML Techniques

To educate the system, machine learning techniques like as Supervised Learning and Unsupervised Learning are utilized, which are further broken into methodologies such as Classification, Regression, and Clustering. This strategy is entirely reliant on the type of dataset used to train the model, which might be tagged, unlabeled, or large. These various machine learning methods are employed in a wide range of applications. Machine learning algorithms employ computer approaches to immediately comprehend data without depending on a preconceived equation as a model.

5.4.1. Logistic Regression

A logistic regression model predicts a dependent data variable by examining the connection between one or more existing independent variables. Logistic regression is a statistical analysis approach that predicts a binary result, such as yes or no, based on past observations of a data set.

5.4.2. Decision Tree

Decision Tree Analysis is a generic, predictive modelling method with applications in a variety of fields. In general, decision trees are built using an algorithmic technique that discovers alternative ways to segment a data set depending on distinct circumstances. The objective is to build a model that predicts the value of a target variable using basic decision rules learned from data attributes.

5.4.3. Random Forest

Random forest is a supervised learning method that may be used for classification and regression. However, it is mostly employed to solve categorization difficulties. Similarly, the random forest algorithm generates decision trees from data samples, then receives predictions from each of them, and eventually votes to choose the best option. It is an ensemble approach that outperforms a single decision tree because it eliminates over-fitting by averaging the

results. It also gives a good method for dealing with missing data, as well as a decent forecast without the need for hyper-parameter adjustment. In addition, at the node's splitting point in every random forest tree, a subset of characteristics is chosen at random.

5.4.4. Support Vector Machine

The Support Vector Machine is a popular Supervised Learning approach that may be applied to classification and regression problems. An SVM model is essentially a representation of several classes in a multidimensional hyperplane. The primary objective of SVM is to partition datasets into classes in order to find the maximum marginal hyperplane (MMH), which may be performed in two steps. First, SVM will construct hyperplanes that best separate the classes repeatedly. The hyperplane that best divides the classes will then be chosen.

6. RESULT AND DISCUSSION

Several studies were conducted in attempt to develop a viable approach of predicting Liver Failure in advance. The dataset was examined in order to select the best machine learning technique. Logistic Regression, Decision Tree, Random Forest, and Support Vector Machine were the algorithms chosen for assessment.

Table -1: Performance of Algorithms for Liver Failure

Algorithm	Accuracy (%)	Sensitivity	Specificity	TP Rate	TN Rate	FP Rate	FN Rate
Logistic Regression	69.41	0.9	0.16	0.067	0.788	0.211	0.932
Decision Tree	63.95	0.71	0.46	0.211	0.426	0.573	0.788
Random Forest	69.08	0.83	0.3	0.129	0.629	0.37	0.87
Support Vector Machine	71.37	1	0	0	1	0	1

Chart -1: Architecture Diagram for the System

From the experiments made in our liver failure dataset were the whole dataset was split in the 7:3 ratio where 70% was used for training the model and remaining 30% for testing the model we found that the Accuracy result of Logistic Regression was 70.86 %, Decision Tree Classifier was: 65.72 %, Random Forest Classifier was: 68.23 %, Support Vector Machine Classifier was: 71.37 %. Since SVM had the Highest Accuracy we implemented the Liver predictor web application with it. The result was an interactive website which predicts the failure of liver based on the input.

Fig -2: User Interface for Liver Predictor

Fig -3 Liver Predictor Result for Affected Individual

Fig -4 Liver Predictor Result for Normal Individual

7. CONCLUSION

Data cleaning and processing, missing value analysis, exploratory analysis and assessment were all part of the analytical process. By comparing four algorithms with our liver failure dataset we found that SVM gives the best accuracy of 71.37% among all others. As a result, we used it in our liver failure predictor project. Based on the patient's health report, this application can assist in the detection of liver failure.

8. REFERRENCES

[1]. Shaheamlung, Golmei, and Harshpreet Kaur. "The diagnosis of chronic liver disease using machine learning techniques." INFORMATION TECHNOLOGY IN INDUSTRY 9.2 (2021): 554-564.

[2]. Kefelegn, Shambel, and Pooja Kamat. "Prediction and analysis of liver disorder diseases by using data mining technique: survey." International Journal of pure and applied mathematics 118.9 (2018): 765-770.

[3]. Nahar, Nazmun, and Ferdous Ara. "Liver disease prediction by using different decision tree techniques." International Journal of Data Mining & Knowledge Management Process 8.2 (2018): 01-09.

[4]. Rahman, AKM Sazzadur, et al. "A comparative study on liver disease prediction using supervised machine learning algorithms." International Journal of Scientific & Technology Research 8.11 (2019): 419-422.

[5]. Muktevi Srivenkatesh. "Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Diseases." International Journal of Innovative Technology and Exploring Engineering 9.2 (2019): 2278-3075

[6]. Dataset source - This dataset was downloaded from the UCI ML Repository: http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science.