International Journal of Computer Science & Communication (ISSN: 0973-7391)
Volume 13 • Issue 2 pp. 26-32 March 2022 - Sept 2022 www.csjournals.com

Prediction and Analysis of Liver Disorder Disease using Machine Learning Algorithms

Jaswinder Singh[1], Kirti Kangra[2]
Department of Computer Science and Engineering
Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India
jaswinder_singh_2k@rediffmail.com, kirtikangra98@gmail.com

ABSTRACT

The liver is vital because it regulates the majority of blood chemical levels and excretes bile, which assists the liver remove waste products. Disease prediction is delicate because any mistake could result in the incorrect person being treated or the wrong patient not being treated. In health-related research, machine learning techniques are commonly used. In this study, current methodologies were investigated in order to identify the most effective predictive algorithms. Machine Learning algorithms viz. Support Vector Machine (SVM), Nave Bayes (NB), Decision Tree (DT), Random Forest (RF), Artificial Neural Network (ANN), Logistic Regression (LR), and k-Nearest Neighbors (k-NN) were also analyzed on open-source WEKA software against the two datasets. The BUPA Dataset and the ILP Dataset are two similar structured datasets that we used. RF had a 73% accuracy rate on the BUPA dataset, whereas LR had a 72% accuracy rate on the ILP dataset. RF had a 0.76 ROC curve against the BUPA dataset, while LR and SVM had a 0.74 ROC curve against the ILP dataset. Our findings will aid health-care organizations in comprehending the value and application of predictive algorithms in the prediction of liver disease.

Keywords: ***Liver Disorder Disease, Machine Learning Algorithms***

I. INTRODUCTION

The liver is involved in many biological functions, including protein synthesis and blood clotting, as well as cholesterol, glucose, and iron metabolism. “Cirrhosis, alcohol misuse, hepatitis A, B, C, D, and E, infectious mononucleosis, nonalcoholic fatty liver disease, and iron overload are just a few of the diseases and ailments that can affect the liver [1]”. The problem is that it is an organ capable of functioning even if it is partially injured. Every year, around 10 lakh new patients with liver disease are diagnosed in India. Another annoyance is the scarcity of specialist doctors. As a result, the necessity for automated and precise systems to classify healthy versus unhealthy people is critical for human existence [2].

Recently, machine learning (ML) techniques have been employed for discovering patterns and generalizing for prediction. EHRs are generated to diagnose diseases. EHRs are then used by doctors to diagnose diseases. Data may be hidden in EHRs. Diagnostic errors in certain EHRs can lead to incorrect prediction. “Traditional decision making in healthcare institutions is mainly relied on the instincts and talents of doctors, rather than the amount of data hidden in EHRs”. Medical decision support systems (MDSS) must be generated with the use of ML techniques in order to solve this dilemma. Many emerging approaches to disease detection rely on MDSS. The consequences of a badly designed MDSS can be disastrous and lead to undesirable outcomes. Hospitals can lower their costs by using a MDSS that is appropriately developed and analyzed.

A. Statistic of liver disease in India

“Liver disease is the tenth most common cause of death in India, as reported by the World Health Organization”. Every fifth Indian may be affected by liver disease [3]. According to the most recent WHO data, 264,193 people died from liver disease in India in 2018, accounting for 3.0% of all deaths. According to the World Health Organization's Global Health Observatory data, India's liver disease burden is high, with 23 fatalities per 100,000 people related to cirrhosis [4]. In terms of liver disease, India ranks 62nd in the globe [5].

B. STATEMENT OF PROBLEM

To find an appropriate tradition classifier that can help in diagnosis of Liver disease.

1) OBJECTIVE

I. General Objective

- The general objective of this paper to predict the liver disorder using Machine Learning technique.

II. Specific Objective

- To apply classification analysis and find the values of different classifiers.
- To compare the values of different evaluation parameters to find the best among them.

This work is consisting of multiple Segments. The related research on employing ML algorithms to predict liver disease is covered in Section II.

II. RELATED WORK

In [6] R. A. and S. K. Reddy experimented using DT, RF, LR supervisied ML algorithms to predict Liver Disease. Dataset aquired from Keggal and R- Software was used to perform the experiment. Results showed an accuracy of 65.9% for DT, 68% for LR and 76.7% for RF. RF outperform than others.

In [7] P. K. Gantayat *et al.* worked on machine learning algoriths : KNN and RF. Dataset was taken from UCI ML Repository. 10 fold cross validation was used to split data. Accuracy for the both classifiers were 70%, 68% respectively.

In [8] H. Subhani and S. Badugu discussed about SVM, NB, C4.5 , RF, J48, MLP and Bayesian Network ML algorithms to pedict Liver Disease. According to their efficiency rate, SVM and KNN are excellent classifiers for patients with liver disease because of their high accuracy.

In [9] A. D. Praveen *et al.* worked to predict Liver disease using different ML algorithms such as KNN, SVM, RF, NB, and AdaBoost. “The data set was obtained from hospital and clinical centers of Andhra Pradesh, India. According to performance study, the KNN and AdaBoost models outperform other experimental models in predicting liver disease, with an accuracy of 100%”.

In [10] G. S. Harshpreet Kaur proposed a method that combined three ML algothms namely: LR, DT, and KNN classifiers. The data was acquired from Kaggle database of Indian liver patient (ILP) records. Python was used to implement the suggested model, and the results showed that the accuracy was 77.58 percent.

In [11] J. Singh, Sachin Bagga , Ranjodh Kaur purposed an Intellient system that can predict Liver disease using machine learning techniques. On the Liver Patient dataset, numerous classifiers, including LR, SVM, RF, NB, J48, and KNN, were tested to determine their accuracy. Dataset named ILP Dataset was taken from the UCI Repository. WEKA tool was used for experiment.

In [12] M. Fathi *et al.* expeimented using different SVM Kernal tricks using 10 cross validation. Two datasets BUPA and ILPD were used for the experiment. In the pre-processing phase data is normalized and sorted performed.After that different SVM tricks were applied for prediction. Results showed accuracy 90.9% and F1-score 94% for ILPD dataset and 92.2% and 94.3% for BUPA dataset, respectively.

In [13] M. A. Kuzhippallil and C. Joseph compared various classification models and visualization techniques to predict liver disease with feature selection. Isolation Forest was used to remove outliers. KNN, LR, DT, RF, MLP, XGBoost, Gradient Boost, and LightGBM Classifier were among the ML classifiers employed. The optimal features necessary for liver disease prediction are retrieved using a genetic algorithm paired with XGBoost.

In [14] C. C. Wu *et al.* worked for the Liver Disease Prediction use RF, NB, ANN, and LR with 10 fold-cross validation. “The dataset was obtained from New Taipei City Hospital Banqiao Branch. R software (Version 3.4.2) and Weka (V.3.9) were used for the experiment”. Results showed that accuracy of RF, NB, ANN, and LR 87.48, 82.65, 81.85, and 76.96%. RF outperform than others.

In [15] R. Reza *et al.* discussed various machine learning algorithms such as SVM, non- linear SVM with RBF kernel and KNN. Dataset was carried from UCI ML repository. In this study two dataset were used.

In [16] N. Nahar *et al.* described the use of ensemble methods to detect Liver disease. Dataset was downloaded from UCI ML repository. Five ensemble algorithms, AdaBoost, LogitBoost, BeggRep, BeggJ48, and RF, were used and their accuracy, RMSE TPR, FPR, and ROC curves were compared. The LogitBoost algorithm outperforms than others, with an accuracy of 71.53%.

In [17] A. S. Singh and A. Chowdhury experimented using LR, KNN and SVM for Liver Disease prediction. The dataset was taken from the ILP Dataset downloaded from UCI. The experiment was carried out using the Python . According to the results of the study, LR was the strong predictor of liver disease.

In [18] Muthuselvan used NB, J48, Random Tree, K-star machine learning algorithms for Liver disease. “Data was collected from the Andhra Pradesh’s North East area in India. WEKA software was used for experiment. The accuracy of the NB algorithm for the liver disease dataset was 60.6%, K-star 67.2%, J48 71.2% and the Random Tree algorithm was 74.2%”. Random Tree performed better than others.

In [19] V. J. Gogi and M. N. Vijayalakshmi discussed the Liver Disease predition using ML techniques namely: SVM, LR and DT were used. The dataset aquired from lab reports of 574 patients.Experiment was carried out using MATLAB2016. The result of the experiments showed an accuracy of 95.8% for LR, 82.7% for SVM and 94.9 % for DT.

In [20] L. Alice Auxilia performed experiment using various ML algorithms for predicting the Liver Disease with UCI dataset. Various ML algorithms SVM, DT, RF, ANN, NB were used. This experiment was performed using R-Studio environment with R programming language.

Table 1 Related work

ML Algorithm	References
SVM	[8],[9],[11],[12],[15],[17],[19],[20]
DT	[6],[8],[10],[11],[13],[18]–[20]
RF	[6]–[9],[11],[13],[14],[16],[20]
LR	[6],[10],[11],[13],[14],[17],[19]
KNN	[7],[9]–[11],[13],[15],[17]
ANN	[14],[20]
NB	[8],[9],[11],[18],[20]
MLP	[8],[13]
AdaBoost	[9],[16]

Table 1 demonstrate ML algorithms used by numerous researchers. We can learn about the most commonly used algorithms from this table, which will be used for further dataset analysis.

A. DATA SET

For this experiment two similar kind of datasets are used: 1) BUPA Liver Disorders 2) Indian Liver Disease Dataset. Datasets are download from UCI machine learning repository [21][22].

1) BUPA Liver Disorders

There are 345 rows and 7 columns in the data set. The data set's seven columns correspond to the following information: “1) mcv mean corpuscular volume 2) alkphos alkaline phosphotase 3) sgpt alamine aminotransferase 4) sgot aspartate aminotransferase 5) gammagt gamma-glutamyl transpeptidase 6) drinks number of half-pint equivalents of alcoholic beverages drunk per day 7) selector field used to split data into two sets”.

2) Indian Liver Disease Dataset

There are 167 non-liver patient records and 416 liver patient records in this data collection. 10 variables in this collection of data are “age, gender, total Bilirubin, direct Bilirubin, total proteins, albumin, A/G ratio, SGPT, SGOT and Alkphos”.

This data can be used to predict whether a patient will suffer liver disease in the future. If the answer to the target variable is yes, then liver disease is present. If the answer is No, then liver disease does not exist.

Table 2 Description of Datasets

Properties	BUPA Liver Disorders Dataset	Indian Liver Disease Dataset
Number of Attributes	7	10
Number of Instances	345	583
Missing Values	0	0

Table 2 provide an overview of datasets how many attributes, instances and missing values present.

B. Software

WEKA tool is used to verify the anticipated strategy [23]. Many machine learning methods are available for machine learning analysis in the WEKA tool. Java code or a data record can be used as the input for the classifier. Many capabilities are available in WEKA, including classification, clustering and visualization, association rule analysis, regression and data pre-processing, to name just a few.

III. EXPERIMENTAL RESULT

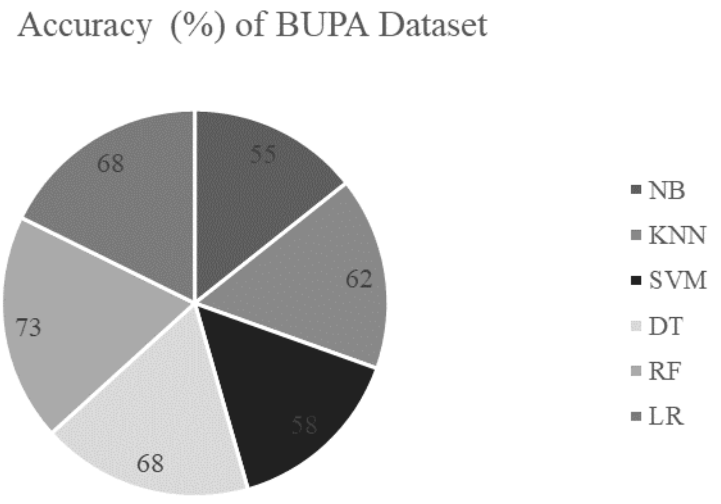
This study's main objective is to evaluate whether a patient has liver disease based on the dataset's properties. Two liver disease datasets were downloaded from UCI for the investigation. SVM, DT, RF, LR, KNN, and NB are the most often used prediction algorithms, according to Table 1. Because there are no missing values in the datasets, the preprocessing phase can be skipped.

The results against the BUPA dataset are shown in Table 2. With scores of 73 percent, 0.72, and 0.76, RF outperforms the others in accuracy, recall, and ROC curve (Table 3). In terms of accuracy, NB received a score of 55 percent, KNN 62%, DT scored 68%, and LR received a score of 68% (Figure 1**)**. After RF, DT and LR outperforms others. With 0.75, SVM outperforms in terms of Precision. Following RF, LR outperforms the other. As a result, both RF and LR can be utilized for the BUPA dataset.

Table 3 BUPA Dataset Analysis

	Accuracy (%)	Precision	Recall	Roc curve	Kappa value
NB	55	0.60	0.55	0.64	0.15
KNN	62	0.63	0.62	0.63	0.24
SVM	58	0.75	0.58	0.50	0.008
DT	68	0.68	0.68	0.66	0.34
RF	73	0.72	0.73	0.76	0.43
LR	68	0.67	0.68	0.70	0.32

Figure 2 Accuracy of BUPA Dataset

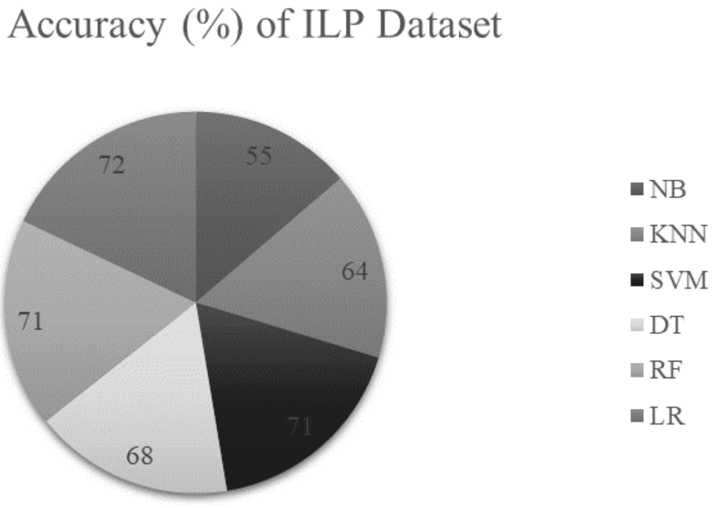


The results against the ILP dataset are shown in Table 3. With 72% accuracy, 0.72 recall, and 0.74 ROC curve, LR outperforms others. In terms of accuracy, NB scored 55%, KNN 64%, SVM 71%, DT 68%, and RF 71%. SVM and RF both perform better after LR in terms of accuracy (Table 4). NB received a higher precision score of 0.79. (Figure 2). With a value of 0.74 both LR and RF performs better than others. As a result, we can conclude that LR can be utilized to predict ILP dataset. The accuracy and Kappa values are out of range for SVM. So, for ILPD SVM is not a good choice.

Table 4 ILP Dataset Analysis

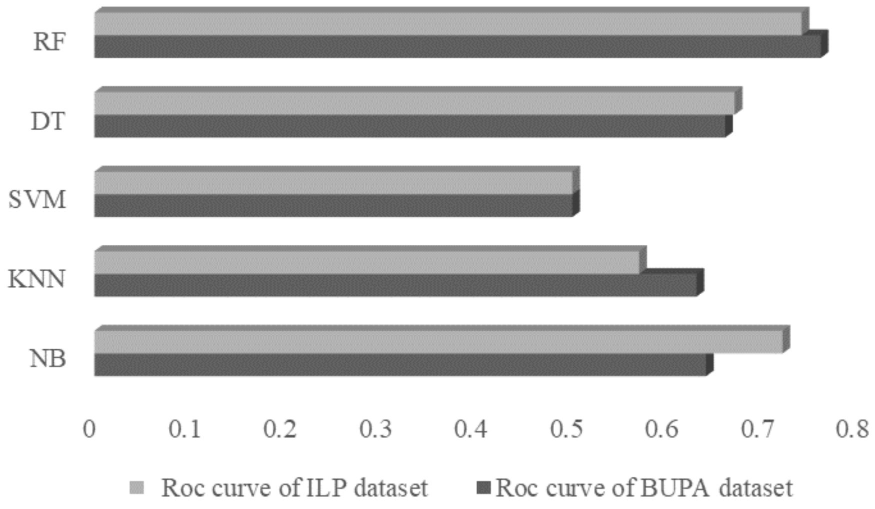
	Accuracy (%)	Precision	Recall	Roc curve	Kappa value
NB	55	0.79	0.55	0.72	0.24
KNN	64	0.66	0.64	0.57	0.17
SVM	71	?	0.71	0.50	0
DT	68	0.66	0.69	0.67	0.17
RF	71	0.68	0.71	0.74	0.18
LR	72	0.69	0.72	0.74	0.21

Figure 3 Accuracy of ILP Dataset



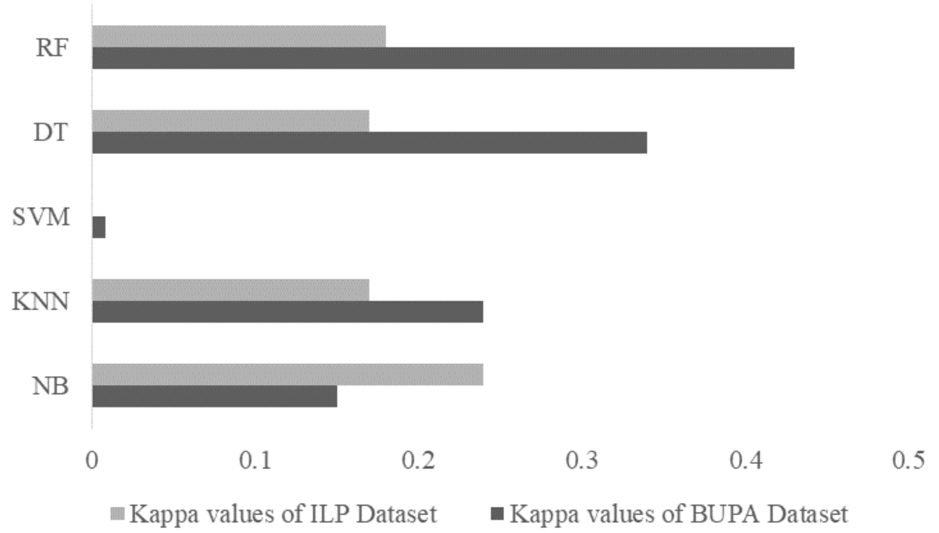
In above paragraphs we have individually discussed the results of both datasets. By comparing both dataset we can say that RF can be used to detect Liver disease. For BUPA dataset RF perform better with **73%, 0.76** in conditions of accuracy and ROC curve. For ILPD, LR perform better with **72%, 0.74**.

Figure 4 ROC curve comparison of BUPA & ILP



By comparing their kappa values from figure 4 we can conclude that all the selected classifiers are not strong enough to predict Liver disease. So, we can say that there should be a hybrid model that can help in prediction of liver disease.

Figure 5 Comparison of Kappa values



IV. CONCLUSION AND FUTURE WORK

Disease prediction is a critical task that can be accomplished effectively using ML. The liver is a vital organ in the human body. This disease can lead to other fetal diseases, thus early detection can save a human life. In this study, SVM, KNN, NB, DT, and RF ML algorithms were used on the similar structured liver datasets, BUPA and ILPD. WEKA software was used to conduct the experiment. In terms of accuracy, we found that RF works best for BUPA dataset and LR works best for ILP dataset. For both datasets, RF works better for ROC curves. LR and RF can be utilized on both datasets. Using the results of the experiment, this research can assist health institutions in diagnosing liver disease. In the future, a hybrid model could be used to produce more accurate results.

REFERENCES

[1] “Liver Disease Symptoms, Treatment, Stages, Signs, Types, Diet.” https://www.medicinenet.com/liver_disease/article.htm (accessed Jul. 24, 2021).

[2] S. Kumari, M. Singh, and K. Kumar, “Prediction of Liver Disease Using Grouping of Machine Learning Classifiers,” *Lect. Notes Networks Syst.*, vol. 175, no. 4, pp. 339–349, 2021, doi: 10.1007/978-3-030-67187-7_35.

[3] “liver disease: Is liver disease the next major lifestyle disease of India after diabetes and BP? - Times of India.” https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/is-liver-disease-the-next-major-lifestyle-disease-of-india-after-diabetes-and-bp/articleshow/58122706.cms (accessed Jul. 06, 2021).

[4] “Liver Disease in India.” https://www.worldlifeexpectancy.com/india-liver-disease (accessed Jul. 06, 2021).

[5] “LIVER DISEASE DEATH RATE BY COUNTRY.” https://www.worldlifeexpectancy.com/cause-of-death/liver-disease/by-country/ (accessed Jul. 06, 2021).

[6] R. A. and S. K. Reddy, “Prognosticating Liver Debility Using Classification Approaches of Machine Learning,” 2021, doi: doi.org/10.1007/978-981-15-7234-0_3.

[7] P. K. Gantayat, Sachikanta Dash, Bhabani P Mishra, Shiba Ch Barik, Sambit Mohanty, “Liver Disease Prediction Using Machine Learning Algorithm,” 2021, doi: 10.1007/978-981-16-0171-2_56.

[8] S. C. Satapathy, K. S. R. K. Shyamala, D. R. Krishna, and M. N. F. Editors, *Advances in Decision Sciences , Image Processing , Security and Computer Vision*, *International Conference on Emerging Trendsin Engineering(ICETE-* 2020),vol. 2.

[9] A. D. Praveen, T. P. Vital, D. Jayaram, and L. V. Satyanarayana, “Intelligent Liver Disease Prediction (ILDP) System using Machine Learning Models,” *Lecture Notes in Electrical Engineering* 2021 pp. 1–15.

[10] G. S. Harshpreet Kaur, “The Diagnosis of Chronic Liver Disease using Machine Learning Techniques,” *Inf. Technol. Ind.*, vol. 9, no. 2, pp. 554–564, 2021, doi: 10.17762/itii.v9i2.382.

[11] J. Singh, Sachin Bagga , Ranjodh Kaur, “ Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques,” *Procedia Comput. Sci.*, vol. 167, pp. 1970–1980, 2020, doi:

10.1016/j.procs.2020.03.226.

[12] M. Fathi, Mohammadreza Nemati, Seyed Mohsen Mohammadi, Reza Abbasi-Kesbi, “A Machine Learning Approach based on Data Required in Liver Disease,” vol. 32, no. 2, pp. 1–9, 2020, doi: 10.4015/S1016237220500180.

[13] M. A. Kuzhippallil and C. Joseph, “Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients,” pp. 778–782, 2020, doi: 10.1109/ICACCS48705.2020.9074368.

[14] C. C. Wu *et al.*, “Prediction of fatty liver disease using machine learning algorithms,” *Comput. Methods Programs Biomed.*, vol. 170, no. March, pp. 23–29, 2019, doi: 10.1016/j.cmpb.2018.12.032.

[15] R. Reza, G. Hossain, A. Goyal, S. Tiwari, and A. Tripathi, “Automatic Liver Disease Diagnosis and Prediction through Machine Learning Algorithms,” 2019.

[16] N. Nahar, F. Ara, M. A. I. Neloy, V. Barua, M. S. Hossain, and K. Andersson, “A Comparative Analysis of the Ensemble Method for Liver Disease Prediction,” *ICIET 2019 - 2nd Int. Conf. Innov. Eng. Technol.*, pp. 23–24, 2019, doi: 10.1109/ICIET48527.2019.9290507.

[17] A. S. Singh and A. Chowdhury, “Prediction of Liver Disease using Classification Algorithms,” *2018 4th Int. Conf. Comput. Commun. Autom.*, no. December, pp. 1–3, 2018, doi: 10.1109/CCAA.2018.8777655.

[18] S. Muthuselvan, S. Rajapraksh, K. Somasundaram, and K. Karthik, “Classification of liver patient dataset using machine learning algorithms,” *Int. J. Eng. Technol.*, vol. 7, no. 3.34 Special Issue 34, pp. 323–326, 2018, doi: 10.14419/ijet.v7i3.34.19217.

[19] V. J. Gogi and M. N. Vijayalakshmi, “Prognosis of Liver Disease : Using Machine Learning Algorithms,” *2018 Int. Conf. Recent Innov. Electr. Electron. Commun. Eng.*, no. July 2018, pp. 875–879, 2020, doi: 10.1109/ICRIEECE44171.2018.9008482.

[20] L. A. Auxilia, “Liver Disease,” *2018 2nd Int. Conf. Trends Electron. Informatics*, no. Icoei, pp. 45–50, 2018.

[21] “UCI Machine Learning Repository: Liver Disorders Data Set.” https://archive.ics.uci.edu/ml/datasets/liver+disorders (accessed Jul. 24, 2021).

[22] “UCI Machine Learning Repository: ILPD (Indian Liver Patient Dataset) Data Set.” https://archive.ics.uci.edu/ml/datasets/ILPD+(Indian+Liver+Patient+Dataset) (accessed Jul. 24, 2021).

[23] “Weka 3 - Data Mining with Open Source Machine Learning Software in Java.” https://www.cs.waikato.ac.nz/ml/weka/ (accessed Jul. 11, 2021).