



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijrset@gmail.com

www.ijrset.com

A Comparison of Machine Learning Algorithms for Predicting Liver Disease

¹Dr. Gurukesava Dasu Gopisetty, ²Tanikonda Koteswara Rao, ³Surabathuni Ramchand,
⁴Pathella Naga Gopi Adithya, ⁵Pusuluri Yaswanth

¹Professor, Department of Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur,
Andhra Pradesh, India.

^{2,3,4,5}BTech Information Technology, KKR & KSR Institute of Technology and Sciences, Guntur,
Andhra Pradesh, India.

ABSTRACT: The liver is a vital internal organ of the human body that performs two primary functions: blood filtration and digestion. The inhalation of contaminated air, narcotics, excessive alcohol consumption, and obesity all contribute to liver damage. It causes liver encephalopathy, aberrant nerve function, bloody coughing or vomiting, renal failure, liver failure, and many other symptoms. The diagnosis of this illness is very expensive and difficult. Early-stage prediction will prevent it from being too late and help save lives. There are various types of machine learning for the diagnosis of liver disease, such as Logistic Regression, KNN, random forest, and they differ in accuracy, precision, and sensitivity. According to the analysis's findings, the Logistic regression had the most accuracy. Additionally, the primary goal of the current work is to use clinical data for the prediction of liver disease. Through our analysis, we explore several ways to describe such data.

KEYWORDS: Liver Disease, Liver Disease Prediction, Liver Diagnosis, Machine Learning and its algorithms, Logistic Regression, K-Nearest Neighbour, Random Forest.

I. INTRODUCTION

The liver, the largest internal organ, aids in detoxification and digestion, with a remarkable ability to regenerate itself within 7 to 14 days. Liver disease contributes to 2.95% of global deaths, ranking India 63rd. Liver ailments like cirrhosis and liver cancer are major health concerns globally, causing approximately 2 million deaths annually. Machine learning in biomedicine enhances disease detection and prognosis, offering impartial decision-making and cost-effective treatments. Techniques such as logistic regression, KNN, and RF are employed for liver disease classification, evaluated through accuracy, precision, recall, and f-1 score. Receiver operating characteristic (ROC) analysis is utilized to compare performance. Liver disease diagnosis is challenging, but early detection using machine learning algorithms can mitigate risks. The biomedical field's focus on improving algorithm performance is crucial for effective liver disease management. Through advancements in machine learning, early diagnosis and intervention for liver diseases can be enhanced, reducing morbidity and mortality rates globally.

The motivation for comparing machine learning algorithms in predicting liver diseases stems from the urgent need to improve diagnostic accuracy and prognosis in this critical health domain. With the complexity of liver diseases and the limitations of traditional diagnostic methods, machine learning offers a promising solution by analyzing vast clinical data to identify patterns and biomarkers for early detection. By evaluating different algorithms, researchers aim to uncover the most effective approaches, fostering innovation in diagnostic methodologies and ultimately improving public health outcomes for individuals globally.

Liver diseases is to systematically evaluate and compare the performance of various machine learning techniques in accurately diagnosing and predicting the progression of liver diseases using clinical data. This involves identifying the most effective algorithms for different types of liver conditions, assessing their predictive capabilities, and determining their potential for early detection and prognosis. The objective is to enhance diagnostic accuracy, optimize treatment strategies, and ultimately improve patient outcomes in the management of liver diseases on a global scale

II. LITERATURE REVIEW

[1] This research focuses on developing an intelligent Clinical Decision Support (CDS) framework for early diagnosis and management of liver cancer using legacy Ultrasound (US) image data. The proposed framework involves multiple stages, beginning with image acquisition and pre-processing, followed by image enhancement using an algorithm. The second stage focuses on cancer detection and segmentation, utilizing various image segmentation techniques. The third stage involves disease classification, matching investigated objects with a disease dictionary. The final stage predicts cancer progression using an array of US images. The experiment was conducted on a liver cancer dataset with 200 patients, consisting of 120 benign and 80 malignant cases. The study underscores the urgency of developing early cancer detection models and the potential of computer-aided diagnosis in medical imaging. The proposed framework draws inspiration from road sign detection and extends its application to liver cancer diagnosis and progression prediction, demonstrating promising results for efficient disease management.[1]

[2] In this article, the authors explore the application of data mining in the medical field, specifically focusing on predicting diabetes and liver disorders. They introduce two novel fuzzy approaches to enhance prediction accuracy. The first approach involves using the fuzzy c-means clustering algorithm for data clustering, followed by classification using the K-nearest neighbor (KNN) algorithm. The second approach employs fuzzy c-means clustering followed by classification using fuzzy KNN. Both methods are tested on PIMA diabetes and liver disorder datasets, demonstrating more precise models than those found in existing literature. The second approach, in particular, yields better results. The introduction of fuzzy algorithms positively impacts the outcome of diabetes and liver disorder prediction models, resulting in a remarkable increase in classification accuracy. The article highlights the importance of disease prediction models, especially given the alarming global statistics on diabetes-related deaths and new cases. The paper concludes with a promise of further analysis and comparison of results with existing methods.[2]

[3] This research paper focuses on the application of cluster analysis, a data summarization technique, in the medical domain to aid in the prediction of liver disorders. The increasing use of medical information systems and the growth of medical databases necessitate efficient computer-assisted analysis to handle large datasets. The study proposes a classification-based technique for predicting liver disorders, aiming to achieve better accuracy than traditional classification algorithms. The results from the proposed method are reported to be promising and accurate. The paper emphasizes the significance of data mining techniques, particularly clustering, in medical diagnosis, offering insights into building resource classification schemes for medical datasets. The primary goal is to assist physicians in disease diagnosis through efficient and reliable training samples, utilizing machine learning classifiers based on clustering methods. The related work section discusses the current landscape of data analysis methods in medicine, highlighting the role of clustering in handling massive datasets without assigned labels. Overall, the paper underscores the potential of cluster analysis in improving medical diagnosis through insightful pattern recognition.[3]

[4] The study identifies four crucial parameters (age, AST, albumin, and platelet count) as the most important features for predicting advanced fibrosis. These parameters exhibit statistically significant relationships (P -value < 0.0001) and accepted correlation coefficients ($|r| > 0.1$) with the presence of advanced fibrosis. The Particle Swarm Optimization (PSO) model achieves the highest correlation coefficient of 0.28, while the Decision Tree (ADT) model attains the highest accuracy of 84.4% and the highest Area Under the Receiver Operating Characteristic (AUROC) of 0.76. The proposed machine learning models are suggested as acceptable, safe, and cost-effective alternatives for predicting advanced fibrosis compared to relatively risky tools like liver biopsy in chronic hepatitis C virus patients. This implies that these models could serve as valuable tools in clinical settings for assessing and managing advanced liver fibrosis in a less invasive manner.[4]

III. PROPOSED WORK

A. Data Collection

We get a dataset from the UCI Machine Learning for this experiment. Additionally, the initial dataset was gathered in Andhra Pradesh, India's northeast. Data from 583 liver patients, 75.64% of whom are men and 24.36% of whom are women, are included in this dataset. This dataset has 11 distinct parameters, however we only used 10 of them for further analysis and 1 of them as a target class and the following are the 11 parameters (data columns) along with the non-null count and data types.



Fig :1



Fig: 2

Parameters Table

Data columns	Non-null count	Data type
Age	583 non-null	int64
Gender	583 non-null	object
Total bilirubin	583 non-null	float64
Direct bilirubin	583 non-null	float64
Alkaline phosphatase	583 non-null	int64
Alanine aminotransferase	583 non-null	int64
Aspartate aminotransferase	583 non-null	int64
Total proteins	583 non-null	float64
Albumin	583 non-null	float64
Albumin and globulin rate	579 non-null	float64
Dataset	583 non-null	int64

Table 1: Parameters description

B. Data Preprocessing

In this study, we examined the data of 583 liver patients, 416 of whom were liver patients and 167 of whom were not. Figure 1 depicts the ratio of total liver patients. Furthermore, 441 male samples and 112 female samples were taken for analysis from the liver patient's dataset (Fig. 2). There are some associated characteristics in the heatmap in Fig. 3. These columns have a low correlation in some cases. In order to improve our ability to predict liver illness, we therefore removed some of the features

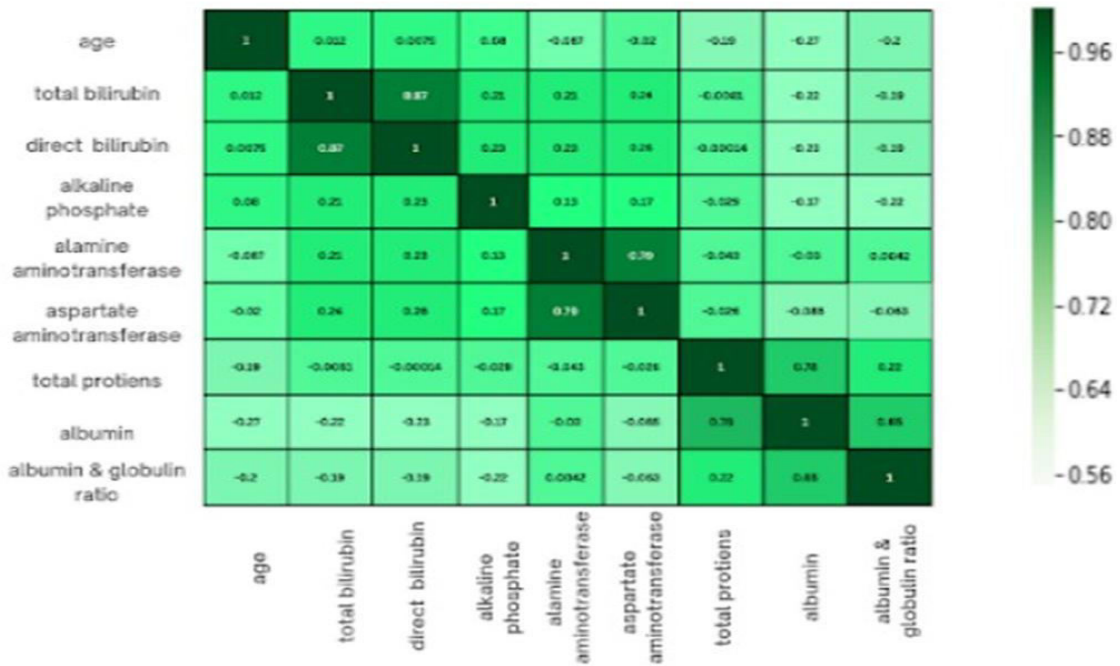


Fig 3: Heat map for the liver dataset's correlated columns.

A. Representation of Classical Algorithms

- 1) **Logistic Regression:** Logistic regression is a statistical method used to predict binary outcomes by examining the correlation between independent variables. It's commonly employed in various fields, such as politics and education, to forecast outcomes like election results or college admissions. By analyzing factors like GPA, SAT scores, and extracurricular activities, logistic regression rates new cases' likelihood of falling into one of two outcome categories based on historical data. This technique is essential in machine learning, allowing algorithms to categorize incoming data based on past information, and it becomes more accurate as more relevant data is included. For instance, in predicting liver disease, logistic regression can classify patients based on gender and other variables, effectively determining the most influential factors for classification. With its ability to handle different types of data and classify new observations using both continuous and discrete datasets, logistic regression remains a fundamental approach in machine learning.
- 2) **KNN:** The K-NN algorithm places the new instance in the category that is most similar to the available categories based on the assumption that the new case and the available cases are KNN: One of the simplest machine learning algorithms based on the supervised learning method is K-Nearest comparable. When a new data point is added, the K-NN algorithm classifies it based on how similar the existing data is to it and stores it all. This means that utilizing the K-NN method, fresh data can be quickly and accurately categorized into a suitable category. Although the K-NN technique can be applied to classification and regression problems, it is most frequently utilized for classification issues. Since K-NN is a non-parametric technique, it makes no assumptions about the underlying data. It is also known as a lazy learner algorithm since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to perform an action when classifying data. The KNN method simply saves the information during the training phase, and when it receives new data, it categorises it into a category that is quite similar to the new data.
- 3) **Random forest:** Random forests, also known as random decision forests, are an ensemble learning technique for classification, regression, and other assignments that generates the class that is the mean forecast (regression) or method of the classes (classification) of the individual decision trees. When a new set of samples is input into a random forest, each decision tree in the forest makes a prediction on this set of samples separately and integrates

the prediction results of each tree to get a final result. This is done after the random forest has established a large number of decision trees in accordance with a specific random rule. This technique works by developing a large number of decision trees during training time. For decision trees' tendency to overfit their training set, random decision forests are the best option. There is an immediate relationship between the united trees and the result that the forest of trees can produce. To obtain forecasts that are more accurate many decision trees created are used as weak classifiers. Multiple weak classifiers combine to construct a robust

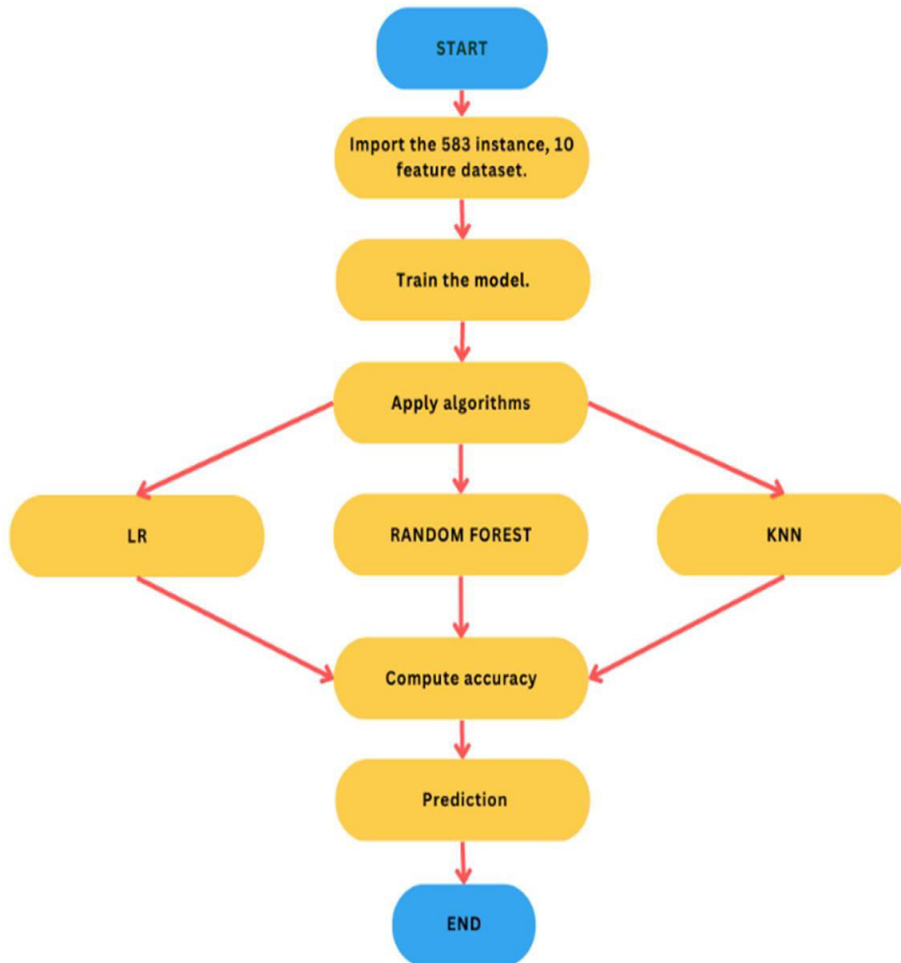


Fig 4: Flow chart of proposed model

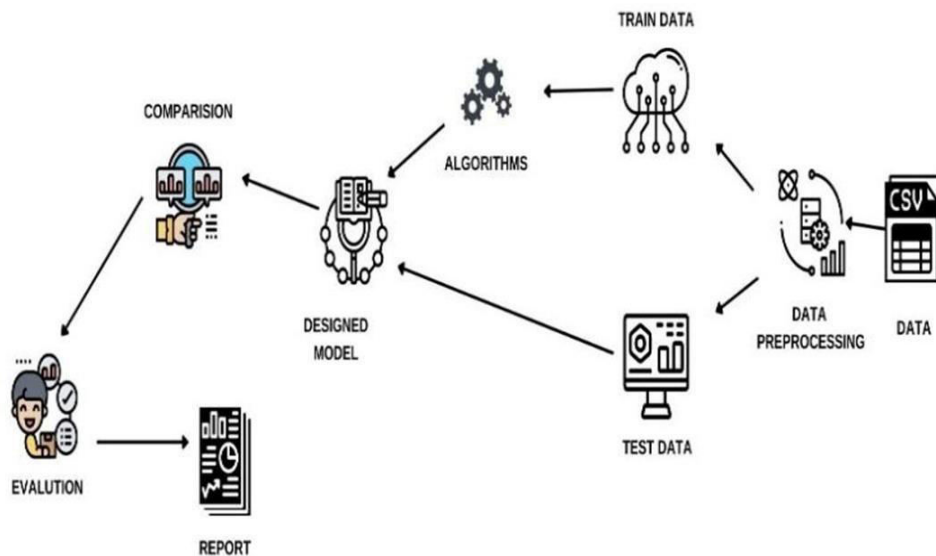


Fig 5: Machine learning model Architecture.

IV. RESULT

A. Classification of Evaluation Techniques

Our work, we employed certain genuine estimations to assess the test execution of various categorization algorithms. The various evaluation methodologies, such as accuracy, sensitivity, specificity, precision, and the f1 measure, were used to examine the performance of the classification systems. As a result, the confusion matrix determines the exhibition assessment variables. In this case, true positive (TP): The wrongly identifies a patient as having liver disease. True Negative (TN): The conclusion of the prediction correctly rejects the patient's allegation of liver illness. False Negative (FN): The prediction result mistakenly eliminates the possibility that a patient has liver disease. The precision of the prediction model offers the difference between sound and patient capacity ratio. To establish classification precision, one must first determine the true positive, true negative, false positive, and false prediction properly identifies in a patient as having liver illness. False Positive (FP): The forecast negative.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)}$$

This effectability test successfully distinguishes patients with liver illness. It mostly demonstrates the test's positive results. It is also known as recall and positive rate.

$$\text{Sensitivity} = \frac{(TP)}{(TP+FN)}$$

The negative repercussions of the condition are underlined in detail. It indicates the amount of the patients' missing sickness. It is also known as the "real negative rate" (TNR).

$$\text{Sensitivity} = \frac{(TN)}{(FP+TN)}$$

Positive predictive value is another name for precision. It displays the percentage of favourable outcomes that classifier algorithms accurately anticipated

$$\text{Precision} = \frac{(TP)}{(TP+FP)}$$

F1 combines accuracy and recall to determine the model's precision. It provides the model's FP and FN proportions.

$$F1 = \frac{2 * (R * P)}{(R + P)}$$

B. Analysis of the result

In this experiment, we examined three machine learning classifiers for the categorization of liver disease dataset using various techniques. After employing oversampling methods, KNN's accuracy value was 70.28%. RF received 74.28%. The highest result was 76.57% for LR. The LR classification algorithm outperforms the other classifiers for predicting liver disease based on these measurement criteria.



Statistical results comparison

ALGORITHM S	ACCURAC Y	PRECISIO N	RECAL L	F1_SCOR E
LR	76.57%	75.00%	99.50%	87.16%
RF	74.28%	84.40%	77.90%	80.32%
KNN	70.20%	83.10%	84.90%	84.32%

Table 2: Statistical results

Liver Disease Prediction Application Interface

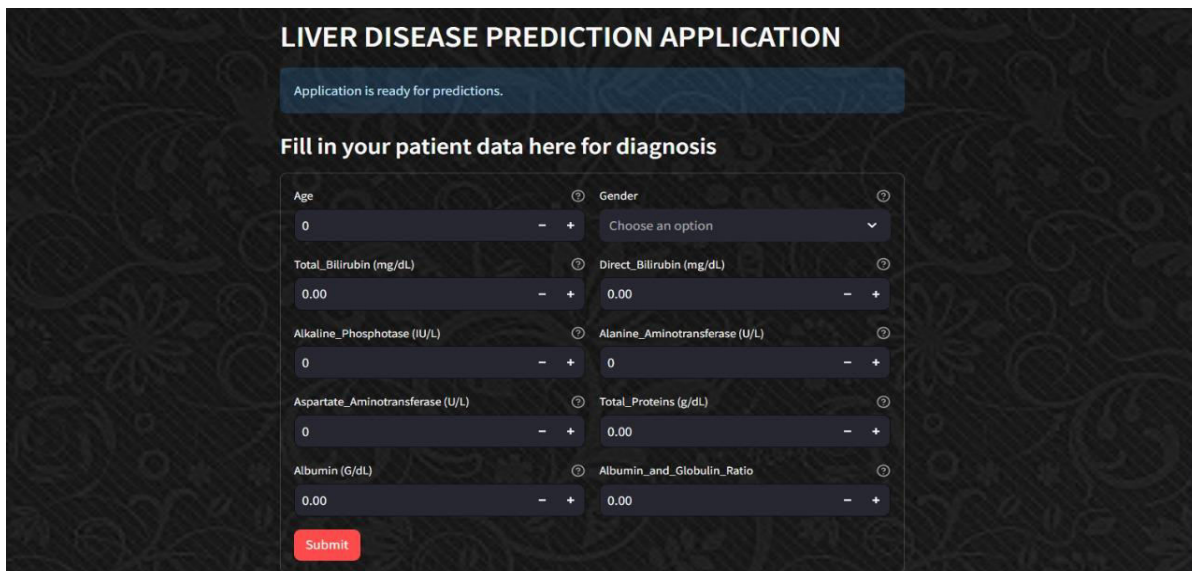


Fig 6: Intake of inputs Interface



LIVER DISEASE PREDICTION APPLICATION

Application is ready for predictions.

Fill in your patient data here for diagnosis

Age	45	Gender	Male
Total_Bilirubin (mg/dL)	1.70	Direct_Bilirubin (mg/dL)	0.80
Alkaline_Phosphotase (IU/L)	315	Alanine_Aminotransferase (U/L)	12
Aspartate_Aminotransferase (U/L)	38	Total_Proteins (g/dL)	6.30
Albumin (G/dL)	2.10	Albumin_and_Globulin_Ratio	0.50

Submit

Diagnosis:

The supplied values suggest that the patient has a liver disease. It is suggested to provide critical emphasis on diagnosing further symptoms of the patient.

Fig 7: Result case1 patient has a liver disease

LIVER DISEASE PREDICTION APPLICATION

Application is ready for predictions.

Fill in your patient data here for diagnosis

Age	75	Gender	Female
Total_Bilirubin (mg/dL)	0.80	Direct_Bilirubin (mg/dL)	0.20
Alkaline_Phosphotase (IU/L)	120	Alanine_Aminotransferase (U/L)	20
Aspartate_Aminotransferase (U/L)	29	Total_Proteins (g/dL)	7.60
Albumin (G/dL)	4.80	Albumin_and_Globulin_Ratio	1.30

Submit

Diagnosis:

The supplied values suggest that the patient does not have a liver disease.

Fig 7: Result case2 patient does not have a liver disease

V. CONCLUSION

In conclusion, this research focused on developing a liver disease diagnosis system using supervised machine learning classifiers. Among the six classifiers evaluated, the LR classifier exhibited the highest accuracy in predicting liver disease. The study suggests the potential of machine learning algorithms, particularly LR, in early diagnosis and decision support for liver infections. Furthermore, the findings highlight the importance of exploring additional algorithms to further enhance prediction accuracy. Overall, this research has significant implications for improving liver disease diagnosis, particularly in low-resource settings, and underscores the promising role of machine learning in advancing medical research and therapeutic interventions.

REFERENCES

1. Ali et al., "Intelligent image processing techniques for cancer progression detection, recognition and prediction in the human liver," 2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE), Orlando, FL, USA, 2014, pp. 25-31, DOI: 10.1109/CICARE.2014.7007830.
2. N. Chetty, K. S. Vaisla and N. Patil, "An Improved Method for Disease Prediction Using Fuzzy Approach," 2015 Second International Conference on Advances in Computing and Communication Engineering, Dehradun, India, 2015, pp. 568-572, DOI: 10.1109/ICACCE.2015.67.
3. M. S. P. Babu, M. Ramjee, S. Katta and Swapna K, "Implementation of partitioned clustering on ILPD dataset to predict liver disorders," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2016, pp. 1094-1097, DOI: 10.1109/ICSESS.2016.7883256.
4. Somaya Hashem¹, Gamal Esmat², Wafaa Elakel², Shahira Habashy³, Safaa Abdel Raouf¹, Mohamed Elhefnawi⁴, Mohamed ElAdawy³, Mahmoud ElHefnawi¹, 5, "Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients", DOI 10.1109/TCBB.2017.269084, IEEE 2017.
5. k. Thirunavukkarasu, A. S. Singh, M. Irfan and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 2018, pp. 1-3, DOI: 10.1109/CCAA.2018.8777655.
6. S. Hashem et al., "Comparison of Machine Learning Approaches for Prediction of Advanced Liver Fibrosis in Chronic Hepatitis C Patients," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 3, pp. 861-868, 1 May-June 2018, DOI: 10.1109/TCBB.2017.2690848
7. Veena G. S, D Sneha, Deepti Basavaraju and Tripti Tanvi, "Effective Analysis and Diagnosis of Liver Disorder", DOI: 10.1109/ICCSP.2018.8524347, IEEE, 2018.
8. L. Alice Auxilia, "Accuracy Prediction using Machine Learning Techniques for Indian Patient Liver Disease", DOI: 10.1109/ICOEI.2018.8553682, IEEE, 2018.
9. M. Ramaiah, P. Baranwal, S. B. Shastri, M. Vanitha and C. Vanmathi, "Analytical Comparison of Machine Learning Techniques for Liver Dataset," 2019 Innovations in Power and Advanced Computing Technologies (I-PACT), Vellore, India, 2019, pp. 1-5, DOI: 10.1109/i-PACT44901.2019.8960009.
10. Nazmun Nahar, Ferdous Ara, Md. Arif Istiek Nelay, "A Comparative Analysis of the Ensemble Method for Liver Disease Prediction", DOI: 10.1109/ICIET48527.2019.9290507, IEEE, 2019.
11. S. Ambesange, V. A, R. Uppin, S. Patil and V. Patil, "Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques," 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bengaluru, India, 2020, pp. 98-102, DOI: 10.1109/CCEM50674.2020.00030.
12. Aleksandar Jeremic¹, Dejan Nikolic², Milena Kostadinovic³, and Milena Santric Milicevic⁴, "Predicting the Assisted Living Care Needs Using Machine Learning and Health State Survey Data", DOI: 10.1109/EMBC44109.2020.9175661, IEEE, 2020.
13. Maria Alex Kuzhippallil, Carolyn Joseph, Kannan A, "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients", DOI: 10.1109/ICACCS48705.2020.9074368, IEEE, 2020. [new]
14. Dr. P. Hamsagayathri, Mr. S. Vigneshwaran "Symptoms Based Disease Prediction Using Machine Learning Techniques", DOI: 10.1109/ICICV50876.2021.9388603, IEEE, 2021.
15. C. Geetha, Dr. AR. Arunachalam, "Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms", DOI: 10.1109/ICCCI50826.2021.9402463, IEEE 2021.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details