Grenze International Journal of Engineering and Technology, January Issue

Leveraging Machine Learning for Early Detection and Prevention of Liver Disease Mortality in India

Harshita Rajpoot[1], Aaditi Raj[2] and Sachin Bhoite[3]
[1-3]Department of Computer Science and Applications, Dr. Vishwanath Karad MIT World Peace University, Pune, India
Email: harshitakrish23@gmail.com, aaditiraj01@gmail.com, sachintenjuly@gmail.com

***Abstract*— The most common causes of liver illness that results in death in India are alcohol-related liver disease and viral hepatitis, notably hepatitis B and C. Cirrhosis, liver failure, and hepatocellular cancer can develop as a result of these illnesses and can be fatal. Fatigue, stomachache, edema, and jaundice are typical symptoms. Confusion and bleeding are additional late- stage signs. The prevention of liver disease-related mortality in India can be greatly helped by early detection, hepatitis immunization, and lifestyle changes. There are several factors that contribute to liver disease in India, some of the most important ones being cirrhosis of the liver, chronic hepatitis B and C infections, excessive alcohol intake, and non-alcoholic fatty liver disease (NAFLD).**

These illnesses can cause serious symptoms like jaundice, exhaustion, abdominal discomfort, and, in more severe cases, liver failure. If left untreated, liver failure can be deadly and frequently results in death. Remarkably, India now reports 268,580 liver-related deaths each year, which constitutes 3.17% of all deaths and represents 18.3% of the 2 million liver-related deaths globally. We used different Machine Learning classification techniques, first we did data cleaning and visualized the data, split the data, standardized the data and applied PCA, performed logistic regression with PCA, logistic regression without PCA, Support Vector Machine, Gradient Boosting Classifier, Decision Tree Classifier, Random Forest Classifier without PCA, Random Forest Classifier with PCA, Random Forest Classifier with Feature Selection, Random Forest Classifier with PCA and Feature Selection.

***Index Terms*— Liver Disease, Liver related deaths, Machine Learning classification techniques.**

I. INTRODUCTION

In India, the growing acknowledgment of liver diseases as significant public health concerns is underscored by the fact that the country accounted for 18.3% of the two million liver disease-related deaths worldwide in 2015. Compared to China, the other Asian nation with a sizable population, the contribution of cirrhosis and its consequences, commonly known as chronic liver diseases (CLDs), as causes of death in India has been steadily rising since 1980 [1]. Alcoholic and non-alcoholic liver diseases, as well as hepatitis B and C, are most likely the main causes of mortality from liver cancer and cirrhosis. In India, tropical illnesses such as dengue fever, leptospirosis, malaria, and scrub typhus are frequently associated with liver damage.

Moreover, tuberculosis with hepatic involvement, hydatid cysts, and amebic liver abscesses are more prevalent in India [2]. Liver illnesses account for 2.4% of all deaths in India, making them the tenth most common cause of death, according to the World Health Organization (WHO).

Grenze ID: 01.GIJET.11.1.546_1

It is estimated that between 10 and 15 percent of people have liver disease, with rural areas having a higher frequency [3]. Liver- related deaths in India have reached a staggering figure of 268,580 (3.17% of all deaths) per year [4]. In this research paper for predicting liver diseases using machine Learning techniques (MLT), which included Logistic Regression with PCA, Logistic Regression without using PCA, Support Vector Machine, Gradient Boosting Classifier, Decision Tree Classifier, Random Forest Classifier without PCA, Random Forest Classifier with PCA, Random Forest Classifier without PCA and Feature Selection. Since classification algorithms may determine a patient's likelihood of having liver disease based on specific characteristics, they are frequently designed to be utilized for liver disease forecasting. traits or attributes. In light of the available options, it was discovered that the most accurate algorithm is the Random Forest classifier without using PCA and Feature Selection. Among the tested algorithms, it is a good option for predicting the liver diseases The main contribution of this paper is:
A framework proposed for liver disease prediction. The framework explains how this research is being conducted and every step is mentioned in the framework along with its workflow.
Rest of the paper is as follows:

- Section II covers the Literature review.
- Section III describes Proposed framework - The framework proposed by the researchers is explained in this section.
- Section IV explains Methodology – All the steps performed in this research are deeply explained in this section
- Section V explains Result – The result and outcomes are analyzed and shown in this section
- Section VI explains Conclusion – The overall conclusion of this research is discussed in this section.

II. Literature Review

Among the key fields that use data mining is biomedical science. This field of research is highly delicate as it involves human lives. Data mining has been employed in predicting, preventing, and treating patients [5]. Advancements have led to the proliferation of numerous machines in our lives. One of the most prominent applications where computers, being the most commonly used machines, prove valuable is in knowledge extraction through Machine Learning. The liver is the largest internal organ in the body and plays a crucial role in blood circulation. It can be affected by parasites and viruses, leading to inflammation and impaired function. [6]. It can continue to perform its usual functions even when partially damaged. However, early diagnosis of liver disease is crucial [7]. In the research paper - Prediction of Liver Diseases Based on Machine Learning Technique for Big Data (2018), the authors used Machine Learning techniques like SVM, Boosted C5.0, and Naive Bayes to predict liver diseases using big datasets. Their model achieved high accuracy, sensitivity, and specificity, with performance ranging from 90.50% to 95.70%. And the Conclusion to this research was authors developed Machine Learning techniques using a hybrid classifier for liver disease diagnosis, improving performance and enabling experts to identify and prescribe further medical examinations. Future work could incorporate fast dataset techniques like Apache Hadoop or Spark [8]. This paper presents a system for analyzing liver illness using the Indian Liver Patient Dataset PSO features selection methods. The algorithms used include J48, MLP, Random Forest, Bayesnet Classification, and SVM. J48 Classification and Bayes Net provide better results. Future work aims to categorize feature selection algorithms into four groups and further improve the liver region's management. The system can also identify cardiac disorders using the heart dataset and disease classification [9]. The paper “Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease (2019).” The study proposes liver infection forecasting as a solution to liver disease, using incalculable order systems and five classifiers: Naive Bayes, support vector machines, logistic regression, K Nearest Neighbour and Random Forest. The goal is to identify the most effective algorithm for predicting liver infection using numerous studies on a range of disorders in recent years. Examining recent research in this area, particularly in the biomedical field, reveals numerous studies utilizing data mining for Machine Learning techniques. Data mining systems are crucial in healthcare for early diagnosis and treatment of diseases. Various Learning algorithms, including K- Nearest Neighbour, Support Vector Machines, Logistic Regression, Navi Bayes, and Random Forest, were used to predict individuals with chronic liver disappointment infection [10]. The paper “Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease (2021)”. This research outlines a method for estimating liver disease risk using blood test results, employing Machine Learning algorithms to predict outcomes. The authors developed a system that uses accurate blood test reports to forecast liver disease risk using the most precise model. This paper investigates liver disease prediction using various classification algorithms, including Logistic Regression, Decision Tree, Random Forest,

KNNeighbor, Gradient Boosting, Extreme Gradient Boosting, and LightGB. Results show Random Forest, Light GB, and Ada booting algorithms provide greater accuracy, making Light GB suitable for liver disease forecasting[11]. The paper “Liver Disease Prediction using Machine Learning Classification Techniques(2022)” various algorithms to predict liver disease risk using historical and categorised data from liver disease patients. The algorithms include Logistic regression, Decision Tree, Random Forest, KNNeighbor, Gradient Boosting, Extreme Gradient Boosting, and LightGB. The study found that these algorithms can predict liver disorders with 90% or more accuracy, with 100% accuracy for the SVM model. Future work could include an iOS or Android app and a website to help a significant portion[12]. In the paper Liver Disease Prediction and Classification using Machine Learning Techniques(2023) The research suggests using Machine Learning algorithms to analyze a patient's chronic liver condition using both positive and negative statistics. Classifiers are used to process percentages of liver disease, and the model's outputs demonstrate accuracy in predicting outcomes, based on available training data sets and classification strategies. The author recommends using Machine Learning methods to assess a patient's liver health, using chronic liver disease as a percentage. Classifier processing improves classification performance with available training data. Future work involves using deep Learning techniques to anticipate liver illness, expanding data sources, combining multiple techniques, and training models based on individual characteristics. Comprehensible models can aid medical workers in making informed decisions and providing better treatments for patients[13]. Machine Learning (ML) is a subfield of artificial intelligence that allows machines to learn without explicit instructions. Random forest model performed well in admission prediction system developed by Dhrvesh K et al. in their paper[14]. In this study about individual explanatory features which can be used by the students to analyse and improve their performance are not provided [15].

III. Proposed Framework

After data is being collected from source UCI ML Repository University of California, School of Information and Computer Science. Data is initially preprocessed. Data are preprocessed by replacing missing values with mean & get rid of infinite values. There are training and testing sets for the database after the data preprocessing phase. ML gives the expertise to anticipate the features impacting the company based on traditional data [16]. The deep CNNs offers promising prospects for efficient and accurate auto object identification in biodiversity research [17]. The algorithms comparison conducted in the paper primarily focuses on the results of validation testing to make decisions regarding the selection of algorithms for detection purposes and which is a unique[18]. Machine learning models have become extremely effective tools for managing and forecasting agricultural crop diseases such as tomato, corn, grape etc [19] and Machine Learning Framework for Implementing Alzheimer’s Disease[20].
are used. The Process is completed when correct findings that are compared across several Machine Learning algorithms are generated. The research paper's goal is to investigate, and the following subsections provide a quick overview of this research process. The Framework is shown below in Figure 1.

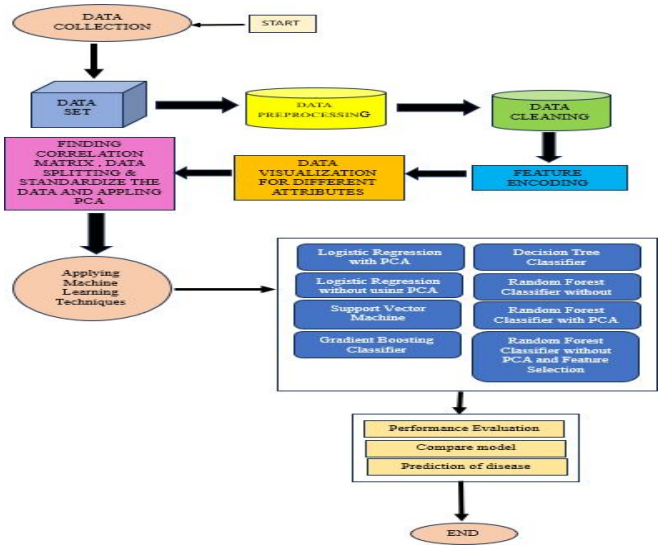

Figure 1: Proposed flowchart for predicting liver disease

The above figure 1 explains all over working of the framework proposed by researchers. It shows all the Machine Learning Techniques performed from data collection, data preprocessing, data cleaning, and all the Machine Learning Algorithms applied and performance analysis of model.

IV. METHODOLOGY

A. Data collection

The researchers analyzed 583 Indian liver patient records from the UCI repository. The dataset includes 142 females, 441 males, with ages above 89 recorded as 90 years old. The data description is given below in Table 1.

B. Data Preprocessing

Data Cleaning is essential for Machine Learning algorithms to be as precise and effective as possible. For accurate data representation and Machine Learning classifiers that must be properly trained and evaluated, data preparation is necessary.

C. Data Cleaning

In order to avoid numerical instability and guarantee correct computations, handling infinite numbers is essential in data analysis and Machine Learning algorithms. It aims to improve the quality and reliability of the data by replacing missing values and handling infinite values accordingly, making making the data suitable for further analysis and modeling tasks.

D. Feature Encoding

The 'Gender' column is first encoded by authors. In order to replace the string "Male" with 1 and any other gender (in this case, assuming "Female") with 0, we utilised the apply() method and a lambda function. This encoding collapses gender information into a binary format where 1 denotes "Male" and 0 denotes other genders, simplifying gender information. The 'Dataset' column is encoded in the subsequent operation. The value 2 is mapped to 0 (signifying no liver illness) and the value 1 is mapped to 1 (signifying the presence of liver disease) in a mapping dictionary that is created using the map() method. With this encoding, the target variable is reduced to a binary value, where 1 denotes the presence of liver disease and 0 denotes its absence.

TABLE I: FEATURES OF DATASET

S.no	Attributes	Description
1	Age	Age in years
2	Gender	if a person is male or female
3	Total Bilirubin	Total Billirubin in mg/dL
4	Direct Bilirubin	Conjugated Billirubin in mg/dL
5	Alkaline Phosphotase	ALP in IU/L
6	Alamine Aminotransferase	ALT in IU/L
7	Aspartate Aminotransferase	AST in IU/L
8	Total Protiens	Total Proteins g/dL
9	Albumin	Albumin in g/dL
10	Albumin and Globulin Ratio	A/G ratio
11	Dataset	Patient with liver disease or no disease

E. Data Visualization

The first visualisation is in Figure 2. It counts the instances of each gendercategory using the 'value_counts()' method on the 'Gender' column. The next step is to make a bar plot, where the y-axis is the count of each gender and the x-axis is the gender categories (such as "Male" and possibly other genders).

The second visualisation, Figure 3, is identical to the first but concentrates on the 'Dataset' column. The ' Dataset' column likely indicates whether each patient has liver disease (1) or not (0), and the value_counts() method counts the instances of each category there. The distribution of patients with liver disease (1) and without liver disease (0) is then shown in a bar plot with blue bars.

The third visualization, Figure 4 is a pair plot where authors used Seaborn library to create the pair plot. In this illustration, the diagonal contains histograms for each individual variable, and every pair of numerical variables from the dataset called "data" is compared using scatter plots. The 'hue='Gender'' parameter uses the 'Gender' column to colour the data points, giving information about how the variables relate to one another while taking gender classifications into account.

The fourth visualization, Figure 5 Using Python's Seaborn module, create a scatter plot. In the beginning, a new figure and axis are made with the precise dimensions of 8x6 inches. The dataset's "Albumin" column serves as the x-axis values for the scatter plot, and its ratio along with globulin ratio on y-axis.

F. Correlation Matrix

It is generally simpler to consider the link between variables when developing an effective dataset analysis. How closely two variables move in respect to one another is determined by a statistic called correlation. Using Python's Seaborn package, the correlation matrix for the dataset 'data' is computed and represented as a heatmap as shown in Figure 9. To create a correlation matrix, the correlation coefficients between each numerical column in the dataset are first determined. The correlation between any two variables is represented by each member in the matrix. Correlations range from -1 (perfectly negative correlation) to 1 (perfectly positive correlation),with 0 denoting no correlation.

G. Data Splitting

In Machine Learning, partitioning data is crucial for training a model on one subset and validating its performance on another to assess generalizability. The target variable (y) is retrieved from the 'Dataset' column, and features (X) are extracted from the DataFrame, excluding the last column. Using scikit-learn's train_test_split function, 20% of the data is set aside for testing (X_test and y_test), and 80% for training (X_train and y_train). Parameters ensure the split is reproducible, yielding the same split each time the code is run.

H. Standardize the data and applying PCA

Using the StandardScaler from scikit-learn, standardisation entails rescaling the dataset's features to have a mean of 0 and a standard deviation of 1. For many Machine Learning methods, especially those based on distances or gradients, standardising the data is crucial since it assures that all characteristics are on the same size. This procedure stops larger-scale features from controlling the Learning process. Standardisation is followed by PCA. A dimensionality reduction method called PCA seeks to save the most crucial pieces of information in the data while transforming the original characteristics into a lower-dimensional space. Principal components, the modified features, are orthogonal and uncorrelated. This procedure can facilitate training, lessen model complexity, and occasionally enhance model performance. The provided code snippet does not demonstrate the PCA transformation, however it is normally carried out using the scikit-learn PCA class following standardisation.

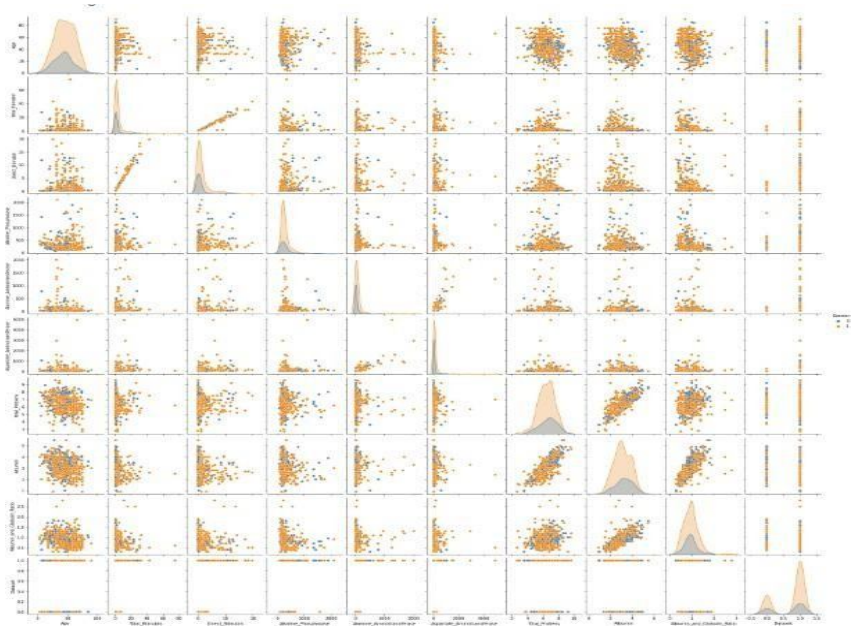

Figure 2. Pair plot of patients differentiated by ‘Gender’ variable

This figure 2 represents a pair plot, which contains histograms for each individual variable. The gender coloumn gives information about how the variables relate to one another while taking gender classifications into account.

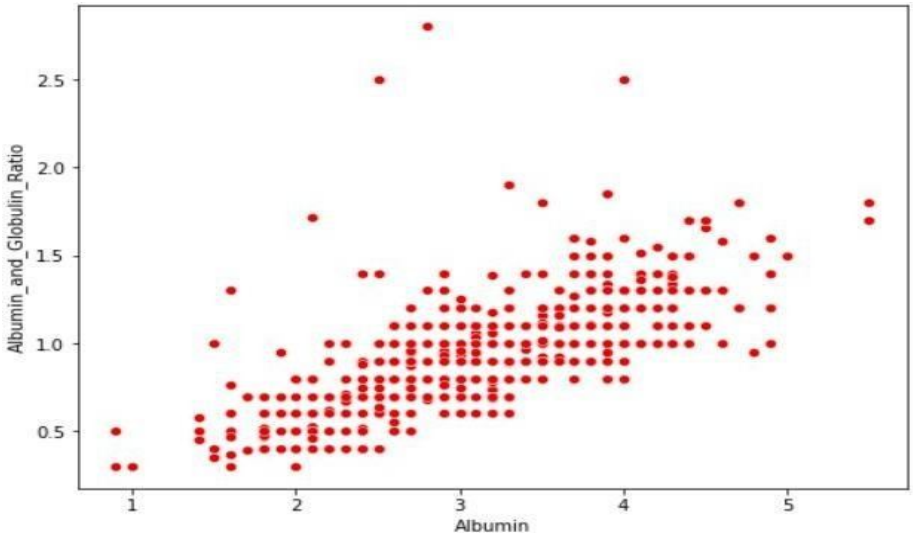

Figure 3: Scatter plot of Patients' Albumin and Albumin and Globulin Ratio

A Figure 3 representing "Albumin" column which serves as the x-axis values for the scatter plot, and its ratio along with globulin ratio on y-axis.

I. Dimensionality reduction using PCA

To determine the explained variances of each main component, Principal Component Analysis is applied to the standardised training data (X_train_std). The percentage of the dataset's overall variance that each component contributes to explaining is given in the explained_variance_ratio_ property. Then, using np.cumsum(explained_variances),the cumulative variance is determined. The variable threshold is set at 0.95, meaning that 95% of the total variance in the dataset is what is intended to be retained. When the cumulative variance first crosses the threshold (0.95) or is equal to it, the algorithm locates the index of the cumulative explained variances array at that location. This index shows how many principal components must be kept in order to preserve at least 95% of the variance. Then, PCA is used once more with the chosen number of components (num_components). Through this process, the data's dimensionality is decreased while maintaining the majority of its crucial details. The X train pca and X test pca files contain the transformed principal components and can be utilised for additional analysis or
modelling. Dimensionality reduction, which can result in more effective and precise Machine Learning models, is crucial for managing high-dimensional datasets.

J. Logistic Regression with PCA

Principal Component Analysis (PCA)-derived reduced-dimensional data is used to train and assess a logistic regression model. The model is trained on lower-dimensional data (X_train_pca) and predictions are made on test data (X_test_pca). Using scikit-learn's accuracy_score, the model's accuracy is 75%.

K. Logistic Regression without PCA

A Logistic Regression model is trained without PCA, using the original high-dimensional data. The maximum number of iterations (max_iter) is set at 900. Predictions are made on the high-dimensional test data (X_test), with an accuracy score of 75%.

L. Support Vector Machine (SVM)

A Support Vector Machine (SVM) classifier with a linear kernel is trained on full-dimensional data (X_train and y_train). Predictions are made on the test data (X_test), achieving an accuracy score of 74% using scikit-learn's accuracy_score.

M. Gradient Boosting Classifier

A Gradient Boosting Classifier is trained using PCA-reduced data with hyperparameters: n_estimators = 100, learning_rate = 0.1, max_depth = 3, random_state = 42. The model is trained on X_train_pca and evaluated on X_test_pca, achieving 78% accuracy.

N. *Decision Tree Classifier*

A Decision Tree Classifier with max_depth=3 and random_state=42 is trained on PCA-reduced data (X_train_pca). The model's accuracy is assessed on the test data (X_test_pca), ensuring straightforward model performance and avoiding overfitting.

O. Random Forest Classifier without PCA

Without utilising PCA to reduce its dimensionality, the original, full-dimensional dataset was subjected to a Random Forest Classifier. The model's accuracy is then assessed, giving a gauge of how well it performs in categorising unobserved data. Utilising the original, full-dimensional test data (X_test and y_test), the trained Random Forest Classifier is put to the test. By contrasting the model's predictions with the actual test labels (y_test), the score method determines the accuracy of the model. The percentage of samples in the test dataset that were properly classified is known as accuracy. The accuracy of Random Forest Classifier without using Principal Component Analysis is 76%.

P. Random Forest Classifier with PCA

a Random Forest Classifier is applied to the PCA- derived reduced-dimensional data. Using the accuracy_score function from scikit-learn, the Decision Tree Classifier's accuracy is determined by contrasting its predictions (y_pred) against the actual test labels (y_test). The accuracy of Decision Tree Classifier is 74%. Features' spatial dimensions, and the model gains knowledge from these reduced features. The reduced-dimensional test data (X_test_pca and y_test) are used to evaluate the trained Random Forest Classifier. By contrasting the model's predictions with the actual test labels (y_test), the score method determines the accuracy of the model. The percentage of samples in the test dataset that were properly classified is known as accuracy. The accuracy of Random Forest Classifier using PCA is 76%.

Q. Random Forest Classifier with Feature Selection

Here a Random Forest Classifier is used to evaluate feature importance and selects a subset of the most relevant features from the original dataset (X). Feature selection is crucial for improving model performance, reducing overfitting, and speeding up the training process by focusing on the most informative features while discarding less relevant ones. The modified dataset with only the chosen features is contained in the selected_features variable, and the shape of this variable illustrates the reduced dimensionality following feature selection.

R. Random Forest Classifier with PCA and Feature Selection

The original, full-dimensional dataset represented by X and y is used to develop and train a Random Forest Classifier model. The accuracy rating shows the percentage of correctly predicted samples in the dataset used for both training and testing. The accuracy for Random Forest Classifier with Principal Component Analysis and Feature Selection is 100%.

V. RESULT

The accuracy results for the Machine Learning Classification techniques applied in this research is given below in Table 2.

TABLE II: COMPARISON WITH EXISTING SYSTEM

S.No	Authors	MLT Used	Data Set USED	Accuracy (%)
1	Engy A. El-Shafeiy et. al[8]	Boosted C5.0, SVM, and Naïve Bayes (NB)	Mansoura Central Hospital data and Egyptian Liver Research Institute	97.20 %
2	M. Banu Priya et. al[9]	J.48, MLP, SVM, Random Forest, Bayes net	Indian liver patient data from UCI Repository	95.04%
3	Muktevi Srivenkatesh[10]	logistic regression, Naïve Bayes, SVM, Random Forest, K-NN	Liver infection forecast data	76.27%
4	Rakshith D B et. al[11]	KNN, Naïve Bayes	UCI ILPD dataset	70%

VI. Conclusion

The analysis of numerous demographic and medical parameters is involved in the prediction of liver disorders using Machine Learning techniques to help with diagnosis and prognosis. For classification tasks relating to liver illnesses, various Machine Learning methods, including Logistic Regression, Support Vector Machine(SVM), Decision Trees, Gradient Boosting, and Random Forest, were used in this work. Additionally, dimensionality was reduced and the most crucial features were identified using feature selection techniques like PCA and SelectFromModel, which increased the effectiveness and accuracy of the model. The outcomes of these methods showed that Machine Learning has a strong potential for diagnosing liver disorders. Particularly SVM and Random Forest demonstrated encouraging results, demonstrating their effectiveness in managing the complexity of medical datasets. In order to reduce the amount of features while maintaining critical data, which is vital for both model performance and interpretability, feature selection approaches, particularly PCA, have proven useful. It's crucial to remember that the accuracy and dependability of predictions are substantially influenced by the choice of features, the calibre of the dataset, and the choice of the right algorithms. For more dynamic and precise predictions, future research might concentrate on investigating additional attributes, utilizing more sophisticated Machine Learning methods, and incorporating real-time data. To ensure the models are accurate as well as clinically relevant and interpretable, collaboration between medical specialists and data scientists is crucial. This will enable more effective decision-making in the diagnosis and management of liver illnesses.

References

[1] Dipankar Mondal, Kaushik Das, et al.: Epidemiology of Liver Diseases in India, Published online 2022 Jan 28
[2] Paul K.Thuluvath, et al.: An Introduction to Liver Disease in India, Published online 2021 Aug 20
[3] https://www.zeebiz.com/trending/news-world-liver-day-2023-april-19-fatty-liver-obesity-diseases-231262
[4] Prof Mohammad Sultan Khuro, Liver Diseases in India: Hope and despair, Great Kashmir.com, todays paper
[5] Kavakiotis, I., et al.: Machine Learning and data mining methods in diabetes research. Comput.Struct. Biotech. J. 15, 104–116 (2017)
[6] Pandey, B., Singh, A.: Intelligent techniques and applications in liver disorders. Survey, January 2014.
[7] Takkar, S., Singh, A., Pandey, B.: Application of Machine Learning algorithms to a welldefined clinical problem: liver disease.. Int. J. E-Health Med. Commun. (IJEHMC) 8(4), 38–60 (2017)
[8] Engy El-Shafeiy, et al.: Prediction of Liver Diseases Based on Machine Learning Technique for Big Data. (AMLTA2018) (pp.362-374), Jan 2018
[9] M. Banu Priya, et al.: Performance Analysis of Liver Disease Prediction Using Machine Learning Algorithms. (IRJET) Volume - 05 Issue - 01, Jan 2018
[10] Muktevi Srivenkatesh, Performance Evolution of Different Machine Learning Algorithms for Prediction of Liver Disease. (IJITEE), Volume – 9, Issue – 2, December 2019
[11] Rakshith D B, et al.: Liver Disease Prediction System using Machine Learning Techniques. (IJERT), Volume – 10, Issue – 06, June 2021
[12] Ketan Gupta, et al.: Liver Disease Prediction using Machine Learning Classification Techniques. (IEEE), April 2022
[13] Srilatha Tokala, et al.: Liver Disease Prediction and Classification using Machine Learning Techniques. (IJACSA) Volume – 14, 2023
[14] Dhruvesh K, Rashmi P, Rushabh S , Sachin B (2019) Engineering college admission preferences based on student performance. Int J Comput Appl Technol Res 8(09):379–384. ISSN: 2319-8656
[15] Shubham Khandale and Sachin Bhoite, "Campus Placement Analyzer: Using Supervised Machine Learning Algorithms", IJCATR, vol. 8, no. 09, 2019, ISSN 358-362.
[16] A. Kulkarni, D. Bhandari and S. Bhoite, "Restaurants rating prediction using machine learning algorithms", International Journal of Computer Applications Technology and Research, 2019
[17] T. Surabhi, B. Sachin and C. Advait, "Deep Convolutional Neural Networks for Automated Butterfly Species Recognition and Classification", 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 778-783, 2023.
[18] Vishwakarma S., Bhoite S., “Classification of Corn Leaf Disease using Resnet18, Alexnet and VGG16” (2024), 2, pp. 1333 – 1339 https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209080343&partnerID=40&md5=ca87b09b0b6ed92e379996e153238a50
[19] Amondkar M., Bhoite S. “Machine Learning Approach for Onion Leaf Disease Detection: A Case Study in Maharashtra, India”(2024), 2, pp. 1249 – 125 https://www.scopus.com/inward/record.uri?eid=2-s2.0-85209153844&partnerID=40&md5=8ba5e43379c396de1b19bc5b3999f11f
[20] Gufran Ahmed Ansari, et al.: Machine Learning Framework for Implementing Alzheimer’s Disease. (IEEE) August 2020