Liver Disease Prediction

Dr. Ritu Rani
Dept of Artificial Intelligence and Data Science
Indira Gandhi Delhi Technical University for Women

riturani@igdtuw.ac

Diya (05601012023)
CSE Department

Indira Gandhi Delhi Technical University for women

diya056btcse23@igdtuw.ac.in

Divya Khosya(05301012023)
CSE Department
Indira Gandhi Delhi Technical University for women

divya053btcse23@igdtuw.ac.in

Abstract - Approximately one million people around the world lose their lives every year owing to different disorders of the liver. The International Food Policy Research Institute published its Global Food Policy Report for the year 2024 on May 29. More than 38% of Indians replace healthy food with junk and processed foods, according to that report, which slow down metabolism and has several adverse effects on the liver.

Traditional diagnostic methods are also available; however, these can be extremely expensive. Early detection of the disease is important, at a stage when it can be treated. Machine learning has huge impact, such that technology development is transforming the way conditions are diagnosed early in healthcare.

This thesis investigates the datasets of liver patients available for constructing classification models and predicting liver disease. Therefore, in this thesis, development of a feature model and comparative analysis are done to improve the prediction accuracy at different phases of Indian liver patients.

The LDP is introduced in this present study to predict liver disease that could be useful by health professionals, stakeholders, students, and researchers. It compared the performances of five algorithms, namely Support Vector Classifier, K-Nearest Neighbours, Logistic Regression, and Classification and Regression Trees.

Based on the results, the best result obtained is from Logistic regression at 76.99%.

Keywords—machine learning, supervised learning, knn, logistic regression, *decision tree classifier, SVM, random forest classifier, and gradient boosting classifier*

I. INTRODUCTION

Large and fleshy, the liver hugs the right side of the abdomen. It weighs approximately three pounds, is red-brown, and has a rubbery texture. The liver itself is split into two large sections: right and left lobes. Both the under regions of the liver-being the pancreas, intestines, and the gallbladder-are the places through which food is chemically broken down, absorbed, and metabolized. The alarming increase in the use of junk foods is destroying people's health. Liver disorders have increased manifold as a result of this trend and it is posing a severe threat to people's health. Liver disease is any disturbance of the liver [2].

Common liver disorder

- Fatty liver is a reversible condition characterized by excessive accumulation of triglyceride fat in liver cells and can occur due to high intake of alcohol or other reasons.
- Hepatitis is generally a viral infection that develops from viruses transmitted through contaminated substances or by direct contact with infected body fluids. It is a serious health hazard.
- Cirrhosis of the liver is a severe and potentially fatal condition where there is extensive loss of cells of the liver, which becomes shrunken, leathery, and hard. The liver has a capacity for regeneration; however, its cell losses are overtaking replacement.
- Liver cancer risk is heightened in the setting of cirrhosis or specific viral hepatitis infections. The liver also has a tendency to develop secondary (metastatic) malignancies from other organs.

Diagnosis of liver disease may be difficult as symptoms often do not appear until the disease has progressed to a chronic stage. Inflammation of the liver, scarring of the liver known as cirrhosis, and liver cancer or failure are the three stages at which liver diseases can be detected. With tremendous strides in the biological sciences, the health sector has been able to make tremendous strides in the prediction of disease with the use of machine learning approaches. Since machine learning algorithms may deal with voluminous data and use several techniques such as clustering, association, and classification, they would therefore be a possible option. These methods will hugely improve the accuracy of illness prediction. We are mainly using six Machine Learning Algorithims which are K-Nearest Number, Logistic Regression, Decision Tree

Classifier, SVM, Random Forest Classifier, and Gradient Boosting Classifier.

II. SYSTEM METHODOLOGY

There could be many kinds of reasons for liver disease, and mainly some important factors need to be kept in mind while diagnosing the disease. The total bilirubin, the sum of all the bilirubin formed by the breakdown of worn-out red blood cells in the body, and the direct bilirubin, an important constituent of bile-a digestive fluid manufactured by the liver-are two essential components which help in diagnosing the disease of the liver. Determining and keeping track of these elements are at times crucial for the liver to maintain its health.

A. Alkaline phosphatase: This protein is responsible for the release of catalyst-acting enzymes that enable the liver to manufacture the bile juice

B. Alanine aminotransferase: This essential enzyme is highly rich in the liver as well as present in other tissues, such as plasma

C. Aspartate aminotransferase: This essential enzyme is utilized by the liver during the manufacturing process of the bile juice.

D. Total Proteins: This refers to the overall amount of proteins present in the body.

E. Albumin is one such crucial protein that makes up the majority of the total amount of proteins present in the human body.

The study's objective was achieved with the UCI Repository's ILPD - Indian Liver Patient Dataset comprising 583 instances/rows with eleven unique features.

III. LITERATURE REVIEW

They often refer to classification algorithms as if they are being used for predicting liver disease, which might determine the likelihood of a patient having liver disease based on several traits or qualities.

P. Rajeswari, G. Sophia Reena, et al. (2010) proposed a classification based on disorders in the liver. The training dataset, which had 345 instances and 7 unique features, was developed by collecting data from the UCI repository. The aim of the research work is to develop a predictive model for liver disease. The objective is to tell whether a patient has liver disease by using some input features obtained from medical records. The algorithms used are FT tree algorithms and K Star algorithms, and overall, the FT tree algorithm performed in speed with an accuracy of 97.10% when tested on datasets related to liver illness. It has been determined that the FT Tree algorithm is hence more accurate than other algorithms. The study's objective was achieved with the UCI Repository's ILPD - Indian Liver Patient Dataset - which comprises 583 instances/rows with eleven unique features.

Sa'diyah Noor Novita Alfisahrin, and Teddy Mantoroet al. (2013) have proposed to identify whether the patients have liver disease based on the 10 important features of liver disease using the decision trees NB Tree and Naive Bayes. According to the findings, NB Tree algorithm has the maximum accuracy. On the other hand, the Naïve Bayes method computes the quickest. However, for a while, it was intended to do better.

S. Dhamodharan [2014] has proposed that there are many disorders in the liver that need a physician's clinical care. Based on unique symptoms, they can make a prediction of cirrhosis, hepatitis, and liver cancer as three major classes of liver diseases. The main goal here is to predict the class kinds from classes like cirrhosis, hepatitis, liver cancer, and "no diseases". This research compares the performance of the FT tree approach with Naïve Bayes and finds Naive Bayes to do much better.

S. E. Seker, Y. Unal, Z. Erdem, and H. Erdinc Kocer et al. [2014] proposed and applied some data mining techniques such as KNN, SVM, MLP, or decision trees over a unique dataset which was gathered over the course of a year from 16,380 analytical findings. Further research could be done based on this in order to bring down the number of analyses.

Onwodi Gregory [2015] proposed two real liver patient datasets that were investigated for building classification models to predict liver diagnosis. Classifications of datasets were made using eleven different data mining techniques, and the accuracy, precision, and recall of each classifier's output was compared to those of others. FT Tree method is considered better compared to other algorithms in terms of classification accuracy using the testing data. Based on the qualities, it comes out to be an accuracy of 78.0%, precision of 77.5%, sensitivity of 86.4%, and specificity of 38.2%, correspondingly.

Tapas Ranjan Baitharua, Subhendu Kumar Panib, et al. (2016) have proposed focusing on the aspect of medical diagnosis by learning patterns through the collected data of Liver disorders to develop intelligent medical decision support systems to help physicians. In this work, we examine the efficacy and correction rate of several classification algorithms (e.g., J.48, SVM, Random Forest, etc.) in the categorization of these disorders. This research presents a comparative investigation of the accuracy of data categorization utilizing data on liver disease in various settings. When compared to other classifiers, the Multilayer Perceptron has the best overall classification result, with an accuracy of 71.59%.

Prior studies have mostly focused on the examination of this Indian Liver Patient Records dataset, with little attention paid to its preparation. Thus, by recognizing preprocessing as an important step in data analysis, our study closes the gap. Additionally, several other algorithms are used in this study.

Attributes	Description
Age	A numeric value having range [4 -90] In the year
Gender	having two nominal value "male" or "female"
TB (Total Bilirubin)	A numeric value having range [0.4-75]
DB (Direct Bilirubin)	Numeric value having range [0.1-19.7]
Alkphos (Alkaline Phospotase)	A numeric value having range [63-2110]
Sgpt (Alamine Aminotransferase)	Numeric value having range [10-2000]
Sgot (Aspartate aminotransferase)	A numeric value having range [10-4929]
TP (Total Proteins)	A numeric value having range [2.7 - 9.6]
ALB (Albumin)	Numeric value having range [0.9-5.5]
Albumin and Globulin Ratio (A/G Ratio)	A numeric value having range [0.3 - 2.8]
Class	having the class value "1" represents Liver Disease present and "2" represent Liver
(Selector)	Disease not present

TABLE 1. DESCRIPTION OF LIVER PATIENT DISEASES DATASET ATTRIBUTES

IV. ILLUSTRATIONS AND PROPOSED TECHNIQUES

This research's primary goal is to effectively anticipate and assess liver illnesses using feature selection techniques and classification algorithms. Additionally, evaluate the outcomes of feature selection and non-selection methods on various classifiers to ascertain whether the approach yields a more accurately categorised result for liver disease diagnosis.

Below are brief descriptions of the different steps:

A. Database Selection: CSV files were used to construct the liver patient dataset. To apply various classifiers, the data was converted into the appropriate format. To get the dataset in a standard format, it is also vital to eliminate any duplicate entries, missing records, and extraneous fields.

B. Feature Selection: The process of choosing the important characteristics from the dataset that have a high correlation with the output to expedite model training, minimize complexity and dimensionality, facilitate easy interpretation, and enhance accuracy is known as feature selection. The pre-processing tools listed above also include feature selection packages.

C. Cross-Validation and Splitting of Data Sets A subset of the dataset is put aside for model testing in cross-validation, with the remaining data being used for model training. The issues of both overfitting and underfitting are circumvented. "Hold-out cross-validation (early stopping)" and "K-fold cross-validation" are two frequently used cross-validation approaches. To remove bias, a dataset is divided into two or three parts: training data, validation data, and test data. This process is known as data splitting. The prediction model is created by the ML algorithms using the training data, and it is then verified by the validation data. While the validation data is solely used to fine-tune the hyperparameters, the model gains knowledge from the training data. Once the hyperparameters have been determined, the test data is utilized to verify.

D. Classification algorithms: The process of building a model of class attributes from a dataset in order to accurately assign a class label to an unknown record is known as classification. Training data with important attributes is used by ML systems for classification. The sort of machine learning algorithm that is used depends on the kind of predictive and diagnostic model that must be developed. Three main categories may be used to group machine learning algorithms: reinforcement learning, unsupervised learning, and supervised learning. In our study, supervised learning is being used to train our dataset.

Supervised learning: Based on an intended outcome, these algorithms transfer an input to the intended output. Up until the model is built, these techniques allow for continuous training. Naïve Bayes (NB), Support Vector Machine (SVM), K-nearest neighbour (KNN), Decision Tree (DT), Random Forest (RF), Logistic Regression, and others are examples of these techniques. Prospective or retrospective, single- or multi-cantered, primary or secondary, can all be employed with the data. For doing the analysis, secondary data is information that the researcher manages from internet databases or from another researcher. Data that has already been brought up is referred to as retrospective.

Some of the algorithms we utilize are:

```
In [17]: import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         from matplotlib.colors import ListedColormap
         import warnings
         warnings.filterwarnings("ignore")
         sns.set()
         plt.style.use('ggplot')
         %matplotlib inline
```

V.PERFORMANCE ANALYSIS

416 of the 583 patient records have liver illness, whereas the remaining 167 do not. Table 2 displays the metadata for the dataset. The binary classification dataset is displayed in Figure 3 with class two values of '1' and '2,' where '1' indicates that patients have liver disease and '2' indicates that they do not.

```
df.describe()
```

	Age	Total_Bilirubin	Direct_Bilirubin	Alkaline_Phosphotase	Alamine_Aminotransferase	Aspartate_Aminotransferase	Total_Protiens	Albumin	Albumin_and_Globulin_Ratio
count	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	583.000000	579.000000
mean	44.746141	3.298799	1.486106	290.576329	80.713551	109.910806	6.483192	3.141852	0.947064
std	16.189833	6.209522	2.808498	242.937989	182.620356	288.918529	1.085451	0.795519	0.319592
min	4.000000	0.400000	0.100000	63.000000	10.000000	10.000000	2.700000	0.900000	0.300000
25%	33.000000	0.800000	0.200000	175.500000	23.000000	25.000000	5.800000	2.600000	0.700000
50%	45.000000	1.000000	0.300000	208.000000	35.000000	42.000000	6.600000	3.100000	0.930000
75%	58.000000	2.600000	1.300000	298.000000	60.500000	87.000000	7.200000	3.800000	1.100000
max	90.000000	75.000000	19.700000	2110.000000	2000.000000	4929.000000	9.600000	5.500000	2.800000

```
In [25]: df.Dataset.value_counts()
Out[25]: Dataset
         1    416
         0    167
         Name: count, dtype: int64
```

```
In [28]: plt.figure(figsize=(5,5))
         sns.countplot(y='Gender',data=df)
Out[28]: <Axes: xlabel='count', ylabel='Gender'>
```
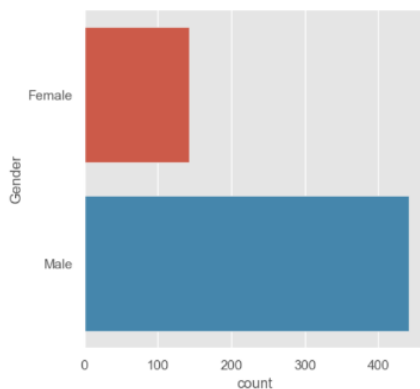

V. EXPERIMENTAL ENVIRONMENT

The data analysis of applying algorithms and finding accuracy is done using Python with version 1.3. The investigation starts by loading the dataset into Python and then modifying and preprocessing the data. Then different algorithms, namely Support Vector Machine (SVM), Naïve Bayes, K-Nearest Neighbours (K-NN), and Classification and Regression Trees (CART), are applied to the dataset.

1. SVC (Support Vector Classifier): SVC is a specific implementation of the Support Vector Machine algorithm that is designed specifically for classification tasks. In other words, SVC is an SVM used for classification. It seeks to find the hyperplane that best separates the data points into different classes.

2. Logistic Regression: Logistic regression is a supervised machine learning algorithm that accomplishes binary classification tasks by predicting the probability of an outcome, event, or observation. The model delivers a binary or dichotomous outcome limited to two possible outcomes: yes/no, 0/1, or true/false.

3. K-Nearest Neighbours (K-NN): K-NN is one of the most straightforward and efficient classification methods. This method predicts the test data point label with the superior class of its k most identical points of training data.

4. Classification and Regression Trees (CART): CART is a decision tree algorithm used for classification or regression, depending on the class label. If it is nominal, it classifies the dataset, or if it is numeric, it performs regression on the dataset using decision trees.

VII. EXPERIMENTAL RESULT ANALYSIS

After applying many algorithms to the liver disease data set, accuracy is seen. The confusion matrix predicts the behavioral structure of the supervised learning algorithms. This square matrix represents the actual and predicted class values. In the confusion matrix, the rows stand for the actual values, while the columns stand for the anticipated values.

```
In [72]: svc=SVC(C=0.01, gamma=0.0001,probability=True)
         svc.fit(x_train,y_train)
Out[72]:           SVC
         SVC(C=0.01, gamma=0.0001, probability=True)
```

```
In [73]: print(accuracy_score(y_train,svc.predict(x_train)))
         svc_acc=accuracy_score(y_test,svc.predict(x_test))
         print(svc_acc)
         print(confusion_matrix(y_test,svc.predict(x_test)))
         print(classification_report(y_test,svc.predict(x_test)))

         0.7117516629711752
         0.7168141592920354
         [[ 0 32]
          [ 0 81]]
                       precision    recall  f1-score   support

                    0       0.00      0.00      0.00        32
                    1       0.72      1.00      0.84        81

             accuracy                           0.72       113
            macro avg       0.36      0.50      0.42       113
         weighted avg       0.51      0.72      0.60       113
```

```
In [66]: from sklearn.neighbors import KNeighborsClassifier
         knn=KNeighborsClassifier()
         knn.fit(x_train,y_train)

Out[66]: ▸ KNeighborsClassifier
         KNeighborsClassifier()

In [67]: knn.predict(x_test)

Out[67]: array([1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0,
                1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0,
                0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1,
                0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1,
                1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1,
                1, 0, 1], dtype=int64)

In [68]: print(accuracy_score(y_train,lr.predict(x_train)))
         knn_acc=accuracy_score(y_test,knn.predict(x_test))
         print(knn_acc)
         print(confusion_matrix(y_test,knn.predict(x_test)))
         print(classification_report(y_test,knn.predict(x_test)))

         0.7117516629711752
         0.6637168141592921
         [[16 16]
          [22 59]]
                       precision    recall  f1-score   support

                    0       0.42      0.50      0.46        32
                    1       0.79      0.73      0.76        81

             accuracy                           0.66       113
            macro avg       0.60      0.61      0.61       113
         weighted avg       0.68      0.66      0.67       113
```

```
Out[63]: ▸ LogisticRegression
         LogisticRegression()

In [64]: y_pred=lr.predict(x_test)

In [65]: print(accuracy_score(y_train,lr.predict(x_train)))
         lr_acc=accuracy_score(y_test,lr.predict(x_test))
         print(lr_acc)
         print(confusion_matrix(y_test,lr.predict(x_test)))
         print(classification_report(y_test,lr.predict(x_test)))

         0.7117516629711752
         0.7699115044247787
         [[11 21]
          [ 5 76]]
                       precision    recall  f1-score   support

                    0       0.69      0.34      0.46        32
                    1       0.78      0.94      0.85        81

             accuracy                           0.77       113
            macro avg       0.74      0.64      0.66       113
         weighted avg       0.76      0.77      0.74       113
```

```
Out[74]: ▸ DecisionTreeClassifier
         DecisionTreeClassifier()

In [75]: print(accuracy_score(y_train,dtc.predict(x_train)))
         dtc_acc=accuracy_score(y_test,dtc.predict(x_test))
         print(dtc_acc)
         print(confusion_matrix(y_test,dtc.predict(x_test)))
         print(classification_report(y_test,dtc.predict(x_test)))

         1.0
         0.6283185840707964
         [[17 15]
          [27 54]]
                       precision    recall  f1-score   support

                    0       0.39      0.53      0.45        32
                    1       0.78      0.67      0.72        81

             accuracy                           0.63       113
            macro avg       0.58      0.60      0.58       113
         weighted avg       0.67      0.63      0.64       113
```

```
In [89]: rand_clf.fit(x_train,y_train)

Out[89]: ▸                    RandomForestClassifier
         RandomForestClassifier(criterion='entropy', max_depth=15, max_features=0.7
         5,
                                min_samples_leaf=7, min_samples_split=3,
                                n_estimators=130)

In [92]: print(accuracy_score(y_train,rand_clf.predict(x_train)))
         rand_clf_acc=accuracy_score(y_test,rand_clf.predict(x_test))
         print(rand_clf_acc)
         print(confusion_matrix(y_test,rand_clf.predict(x_test)))
         print(classification_report(y_test,rand_clf.predict(x_test)))

         0.8980044345898004
         0.7079646017699115
         [[13 19]
          [14 67]]
                       precision    recall  f1-score   support

                    0       0.48      0.41      0.44        32
                    1       0.78      0.83      0.80        81

             accuracy                           0.71       113
            macro avg       0.63      0.62      0.62       113
         weighted avg       0.69      0.71      0.70       113

In [99]: print(accuracy_score(y_train,gbc.predict(x_train)))
         gbc_acc=accuracy_score(y_test,gbc.predict(x_test))
         print(gbc_acc)
         print(confusion_matrix(y_test,gbc.predict(x_test)))
         print(classification_report(y_test,gbc.predict(x_test)))

         0.7117516629711752
         0.7168141592920354
         [[ 0 32]
          [ 0 81]]
                       precision    recall  f1-score   support

                    0       0.00      0.00      0.00        32
                    1       0.72      1.00      0.84        81

             accuracy                           0.72       113
            macro avg       0.36      0.50      0.42       113
         weighted avg       0.51      0.72      0.60       113
```

THE ACCURACY OF THE MODEL IS DETERMINED WITH THE HELP OF THE CONFUSION MATRIX.

```
Out[104...
```

	Model	Score
0	Logistic Regression	76.99
2	SVC	71.68
5	Gradient Boosting Classifier	71.68
4	Random Forest Classifier	70.80
1	KNN	66.37
3	Decision Tree Classifier	66.37

The results obtained from the experiment show that logistic regression gives the highest accuracy, while the other two algorithms give the same. The lowest accuracy is given by the decision tree classifier.

VIII. CONCLUSION

The Liver Patient data set allows the use of prediction and classification algorithms since it lessens the load for physicians. We would like to propose the usage of some machine learning methods in evaluating the general health of the liver of a patient. A chronic liver disease has lasted for at least six months. Thus, we will be using the percentage of people that have the ailment as both positive and negative

statistics. The result of the classifier processing the percentages of liver disease is in the form of the confusion matrix. Our proposed techniques can improve the classification performance drastically when the training data set is available. After that, the excellent and poor values will be identified with the help of a machine learning classifier. The outputs of the proposed classification model are, therefore, accurate in predicting the results.

The proposed liver disease prediction approach has made available the ideal route for the diagnosis of the said disease. The outcome of the study, after balancing the dataset, has varying top scores: Gradient Boosting Classifier is 71.68%, Random Forest Classifier is 70.8%, KNN is 66.37%, Decision Tree Classifier is 66.37%, and Logistic Regression is 76.99%.
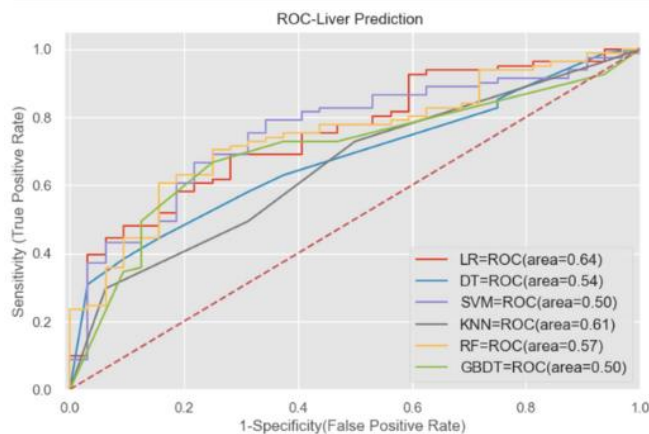
Along with that, increasing the instances can add different attributes, which are important for the prediction of liver disease-like triglycerides, urine copper, serum cholesterol, and serum glutamic-oxaloacetic transaminase, may be added to increase the chances of the prediction of liver disease.

VII. REFERENCES

[1] A.A. Mokdad, A.D. Lopez, S. Shahraz, R. Lozano, A.H. Mokdad, J. Stanaway, et al, Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. BMC Med 2014; 12:145.

[2] W. Nanyue, Y. Youhua, H. Dawei, X. Bin, L. Jia, L. Tongda X. Liyuan, S, Zengyu, C. Yanping and W. Jia. (2015), "Pulse Diagnosis Signals Analysis of Fatty Liver Disease and Cirrhosis Patients by Using Machine Learning," The Scientific World Journal, vol. 2015, Article ID 859192, pp. 1-9.

[3] A. Spann, A. Yasodhara, J. Kang, K. Watt, B. Wang, A. Goldenberg, and M. Bhat, "Applying machine learning in liver disease and transplantation: a comprehensive review," Hepatology, vol. 71, no. 3, pp. 1093–1105, 2020.

[4] S. Sontakke, J. Lohokare, and R. Dani, "Diagnosis of liver diseases using machine learning," in 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI). IEEE, 2017, pp. 129–133.

[5] J. C. Cohen, J. D. Horton, and H. H. Hobbs, "Human fatty liver disease: old questions and new insights," Science, vol. 332, no. 6037, pp. 1519– 1523, 2011.

[6] F. Himmah, R. Sigit, and T. Harsono, "Segmentation of liver using abdominal CT scan to detection liver disease area," in 2018Internationall Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC). IEEE, 2018, pp. 225–228.

[7] M. B. Priya, P. L. Juliet, and P. Tamilselvi, "Performance analysis of liver disease prediction using machine learning algorithms," International Research Journal of Engineering and Technology (IRJET), vol. 5, no. 1, pp. 206–211, 2018.

[8] T. R. Baitharu and S. K. Pani, "Analysis of data mining techniques for healthcare decision support system using liver disorder dataset," Procedia Computer Science, vol. 85, pp. 862–870, 2016.

[9] U. R. Acharya, S. V. Sree, R. Ribeiro, G. Krishnamurthi, R. T. Marinho, J. Sanches, and J. S. Suri, "Data mining framework for fatty liver disease classification in ultrasound: a hybrid feature extraction paradigm," Medical physics, vol. 39, no. 7Part1, pp. 4255–4264, 2012.

[10] N. Nahar and F. Ara, "Liver disease prediction by using different decision tree techniques," International Journal of Data Mining & Knowledge Management Process, vol. 8, no. 2, pp. 01–09, 2018.

[11] A. Naik and L. Samant, "Correlation review of classification algorithm using data mining tool: WEKA, rapid miner ,Tanagraa, 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 learning techniques," in 2022 2nd International Conference on Intelligent Technologies (CONIT). IEEE, 2022, pp. 1–6.

[15] M. K. Enduri, A. R. Sangi, S. Anamalamudi, R. C. B. Manikanta, K. Y. Reddy, P. L. Yeswanth, S. K. S. Reddy, and G. A. Karthikeya, "Comparative study on sentimental analysis using machine learning techniques," Mehran University Research Journal of Engineering and Technology, vol. 42, no. 1, pp. 207–215, 2023.

[16] M. Islam, C.-C. Wu, T. N. Poly, H.-C. Yang, Y.-C. J. Li et al., "Applications of machine learning in fatty live disease prediction," in Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth. IOS Press, 2018, pp. 166–170.

[17] S. Mohanty, P. K. Gantayat, S. Dash, B. P. Mishra, and S. C. Barik, "Liver disease prediction using machine learning algorithm," in Data Engineering and Intelligent Computing: Proceedings of ICICC 2020. Springer, 2021, pp. 589–596.

[18] C. Liang and L. Peng, "An automated diagnosis system of liver disease using artificial immune and genetic algorithms," JOURNAL OF MEDICAL SYSTEMS, vol. 37, no. 2, 2013.

[19] R. A. Khan, Y. Luo, and F.-X. Wu, "Machine learning based liver disease diagnosis: A systematic review," Neurocomputing, vol. 468, pp. 492–509, 2022.

[20] A. S. Abdalrada, O. H. Yahya, A. H. M. Alaidi, N. A. Hussein, H. T. Alrikabi, and T. A.-Q. Al-Quraishi, "A predictive model for liver disease progression based on logistic regression algorithm," Periodicals of Engineering and Natural Sciences (PEN), vol. 7, no. 3, pp. 1255– 1264, 2019. [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.

[29] G. Bonaccorso, Machine learning algorithms. Packt Publishing Ltd, 2017.

[30] I. Cohen, Y. Huang, J. Chen, J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," Noise reduction in speech processing, pp. 1–4, 2009.

[31] M. Ghosh, M. Raihan, M. Sarker, M. Raihan, L. Akter, A. K. Bairagi, S. S. Alshamrani, and M. Masud, "A comparative analysis of machine learning algorithms to predict liver disease." Intelligent Automation & Soft Computing, vol. 30, no. 3, 2021.

[32] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," International Journal of Computer Trends and Technology (IJCTT), vol. 48, no. 3, pp. 128–138, 2017.

[33] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: A survey and review," in Emerging technology in modeling and graphics. Springer, 2020, pp. 99–111.

[34] A. Arjmand, C. T. Angelis, A. T. Tzallas, M. G. Tsipouras, E. Glavas, R. Forlano, P. Manousou, and N. Giannakeas, "Deep learning in liver biopsies using convolutional neural networks," in 2019 42nd International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2019, pp. 496–499.

[35] L. A. Auxilia, "Accuracy prediction using machine learning techniques for Indian patient liver disease," in 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI). IEEE, 2018, pp. 45–50.