

STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION

LITERATURE SURVEY

1) Liver Disease Prediction Using Machine Learning Algorithms

The liver disease is categorized by using feature selection and fuzzy K-means classification methods. To classify the liver disease, Euclidean distance method was used to calculate the distance between each data and assuming k value from 0 to 3. Also, Decision tree, Adaptive Neuro-Fuzzy Inference System (ANFIS) was also used for prediction. Higher accuracy of 98% is obtained from ANFIS.

Advantages: By Comparing the parameter values, the ANFIS classification technique is more effective than the other and obtains the highest accuracy.

Disadvantages: Here the model is trained with a smaller number of parameters.

Algorithms used: K-means clustering, Decision tree, ANFIS.

2) Prognosis of Liver Disease: Using Machine Learning Algorithms

Various parameters are studied from Liver Function Test (LFT) and they are analysed and used for the prediction of the liver disease. The data is transformed into scaled values so that it can fit with the minimum range. Algorithms like SVM, Logistic Regression, Decision Tree and Linear Discriminant analysis were also implemented and the highest accuracy of 95.8% is obtained with Logistic Regression.

Advantages: Techniques like redundancy elimination, Integration helped in better and fast prediction

Disadvantages: Only a very few parameters were considered which is not sufficient for accurate predictions.

Algorithms used: Decision Tree, Linear Discriminant, SVM Fine Gaussian, Logistic Regression.

3) Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms

Initially after pre-processing the dataset, feature selection is performed which extracts the most sensitive and high impact feature from the dataset. The popular SVM classifier is used with maximised geometric margin and minimised error classification and accuracy of 75.04% is obtained. Also, Logistic Regression method is also used for prediction which acquired the accuracy of 73.23%.

Advantages: Feature selection extracts important feature that helps in efficient model training.

Disadvantages: The accuracy of the model is low.

Algorithms used: Logistic Regression and SVM

4) Diagnosing for Liver Disease Prediction in Patients Using Combined Machine Learning Models

Three models such as KNN, Decision tree and Artificial Neural Network (ANN) were used for prediction. Also, a combined model is proposed by combining the above-mentioned models. Higher accuracy of about 96% is obtained by the combined model.

Advantages: The combined model makes the prediction easier and accurate.

Disadvantages: More number of parameters can be considered for prediction.

Algorithms used: KNN, Decision Tree, Artificial Neural Network (ANN)

5) Performance Analysis of Machine Learning Algorithms for Prediction of Liver Disease

Most common algorithms like Random Forest, Boosting technique like XGBoost, Logistic Regression, Gaussian Naïve Bayes, KNN, Support Vector Machine, Gradient Boosting, CatBoost, AdaBoost, LightGBM, and Decision Tree were used for training and testing of the dataset and results of the model were considered.

Highest accuracy of 88% is obtained by Random Forest Classifier.

Advantages: The model predicted the result with very a smaller number of time and the performance of different algorithms were analysed.

Disadvantages: Though different models were considered; the accuracy is low.

Algorithms used: Random Forest, Boosting technique like XGBoost, Logistic Regression, Gaussian Naïve Bayes, KNN, Support Vector Machine, Gradient Boosting, CatBoost, AdaBoost, LightGBM, and Decision Tree.

6) Prediction of Liver Disease using Gradient Boost Machine Learning

Techniques with Feature Scaling

A feature reduction technique is offered in this study that uses recursive feature elimination and machine learning boosting methods. With fewer characteristics, Logistic regression and Multi-Layer Perceptron achieved higher prediction accuracy when basic machine learning models were applied to the dataset. On the dataset, boosting algorithms like CatBoost, LGBM Classifier, XGBoost, and Gradient Boost were used. The effects of feature reduction on the gradient boosting machine learning methods were examined.

Advantages:

Logistic Regression and Multi-Layer Perceptron performed efficiently compared to the other basic machine learning models. Recursive Feature Elimination technique effectively optimized the prediction accuracy of the Gradient Boosting algorithms to 94%.

Disadvantages:

The prediction accuracy of the fundamental models has not been increased by feature selection.

7) Diagnosis of Liver Disease using Machine Learning Models

In order to forecast liver illness more accurately, precisely, and consistently, the approaches of Support Vector Machines (SVM), Decision Trees (DT), and Random Forests (RF) are proposed in this study.

Advantage:

Better accuracy precision and reliability

Disadvantage:

The model couldn't diagnose the presence of liver disease in ppl with subtle symptoms

8) Prediction of Liver Disorders using Machine Learning Algorithms: A Comparative Study

In this study, we examine four different machine learning (ML) techniques for classifying the Indian Liver Patient Dataset, including Logistic Regression, Decision Tree, Random Forest, and Extra Trees (ILPD). To eliminate irrelevant features from the dataset, feature selection based on Pearson Correlation Coefficient is used.

Advantages:

In the data preprocessing step, to overcome the issue of imbalanced class distribution, an oversampling technique is used

Disadvantages:

The obtained results are comparatively inadequate. Robust scaling worked for a specific range only

9) Prediction of Liver Disease using Classification Algorithms

Support vector machines, K-Nearest Neighbor, and logistic regression are the methods utilised for this task. This classification technique is compared using the accuracy score and confusion matrix.

Advantages:

Comparing various types of algorithm has been done based on classification accuracy by using confusion Matrix.

Disadvantages:

Even though KNN having high accuracy, we can use only logistic regression for predicting liver disease as it has the highest sensitivity.

10) Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection

Using an extraction, loading, transformation, and analysis (ELTA) approach to predict liver disease, the goal of this study is to choose relevant features for accurate diagnosis. As a result, the ELTA technique is used to analyse various data mining models, including random forest, Multi-Layer Perceptron (MLP) neural networks, Bayesian networks, Support Vector Machine (SVM), and Particle Swarm Optimization (PSO)-SVM.

Advantages:

The proposed model demonstrated better performance in terms of accuracy, f-measure, precision, sensitivity, specificity, AUC, and FPR criteria.

Disadvantages:

Deep neural networks based on the dataset for liver disease and other datasets can be used to increase accuracy of detection by choosing more useful and

accurate features in the use of smart algorithms in the diagnosis and prediction of diseases, in particular liver disease.

REFERENCES:

- [1] R. Kalaiselvi, K. Meena and V. Vanitha, "Liver Disease Prediction Using Machine Learning Algorithms," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-6, doi: 10.1109/ICAECA52838.2021.9675756.
- [2] V. J. Gogi and V. M.N., "Prognosis of Liver Disease: Using Machine Learning Algorithms," 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 2018, pp. 875-879, doi: 10.1109/ICRIEECE44171.2018.9008482.
- [3] C. Geetha and A. Arunachalam, "Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9402463.
- [4] C. Anuradha, D. Swapna, B. Thati, V. N. Sree and S. P. Praveen, "Diagnosing for Liver Disease Prediction in Patients Using Combined Machine Learning Models," 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), 2022, pp. 889-896, doi: 10.1109/ICSSIT53264.2022.9716312.
- [5] V. Singh, M. K. Gourisaria and H. Das, "Performance Analysis of Machine Learning Algorithms for Prediction of Liver Disease," 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), 2021, pp. 1-7, doi: 10.1109/GUCON50781.2021.9573803.
- [6] G. Shobana and K. Umamaheswari, "Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 2021, pp. 1223-1229, doi: 10.1109/ICCMC51019.2021.9418333.

- [7] A. Sivasangari, B. J. Krishna Reddy, A. Kiran and P. Ajitha, "Diagnosis of Liver Disease using Machine Learning Models," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 627-630, doi: 10.1109/I-SMAC49090.2020.9243375.
- [8] M. F. Rabbi, S. M. Mahedy Hasan, A. I. Champa, M. AsifZaman and M. K. Hasan, "Prediction of Liver Disorders using Machine Learning Algorithms: A Comparative Study," 2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT), 2020, pp. 111-116, doi: 10.1109/ICAICT51780.2020.9333528.
- [9] 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), 2018, pp. 1-3, doi: 10.1109/CCAA.2018.8777655.
- [10] Hassannataj Joloudari, Javad & Saadatfar, Hamid & Dehzangi, Iman (Abdollah) & Shamshirband, Shahaboddin. (2019). Computer aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection. Informatics in Medicine Unlocked. 17. 100255. 10.1016/j.imu.2019.100255.