**Team leader**

Najila M

**Members**

Priyanka.S

Selvi K

Swethaa Shri J

**STATISTICAL MACHINE LEARNING APPROACHES TO LIVER DISEASE PREDICTION**

**LITERATURE SURVEY:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.NO** | **YEAR** | **AUTHOR** | **JOURNAL NAME** | **TITLE** | **METHODOLOGY** | **LIMITATIONS** |
| 1 | 2022 | A. Srivastava, V. V. Kumar, M. T. R and V. Vivek | IEEE | Automated Prediction of Liver Disease using Machine Learning (ML) Algorithms | Logistics Regression (LR), Naive Bayes Model (NB), K-Nearest Neighbor (KNN) | Data preprocessing is not done efficiently. |
| 2 | 2022 | Weidong Ji, Mingyue Xue, Yushan Zhang, Hua Yao, Yushan Wang | Frontiers | A Machine Learning Based Framework to Identify and Classify Non-alcoholic Fatty Liver Disease in a Large-Scale Population | Logistic Regression,Naïve Baiyes,Random Forest,XGBoost | Did not examine with an additional testing dataset for validation. |
| 3 | 2022 | Manas Minnoor  ,Veeky Baths | IEEE | Liver Disease Diagnosis Using Machine Learning | Logistic Regression, K-Nearest Neighbors (KNN), Extra Trees, LightGBM ,Multilayer Perceptron (MLP) neural network | Features used in the dataset is less which is not enough for the efficient prediction of liver disease. |
| 4 | 2021 | Fahad Mostafa,Easin Hasan,Morgan Williamson,Hafiz khan | MDPI | Statistical Machine Learning Approaches to Liver Disease Prediction | Support Vector Machine(SVM),Random Forest Classifier, Artificial Neural Network classifier | Data imbalance problem was detected in the class labels which may lead to inaccurate results. |
| 5 | 2021 | Saima Afrin, F. M. Javed Mehedi Shamrat, Tafsirul Islam Nibir, Mst. Fahmida Muntasim, Md. Shakil Moharram, M. M. Imran, Md Abdulla | iaes | Supervised machine learning based liver disease prediction approach with LASSO feature selection | LASSO feature selection, Decision tree, Random forest, K-Nearest Neighbor, Adaboost classifier, Logistic Regression, Support Vector Machine, Gradient Boosting | Dataset used in this paper contains only limited number of records. |
| 6 | 2021 | C. Geetha and A. Arunachalam | IEEE | Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms | Linear Regression, Support Vector Machine | Accuracy is less when compared with other models. |
| 7 | 2021 | Mylavarapu Kalyan Ram, Challapali Sujana, Rayudu Srinivas, G.S.N Murthy | Springer | A Fact-Based Liver Disease Prediction by Enforcing Machine Learning Algorithms | Multilayer Perceptron Classifier, K-Nearest Neighbor, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting,support Vector Machine | Data Preprocessing stage is not elaborated. |
| 8 | 2021 | Ritesh Choudhary, T. Gopalakrishnan, D. Ruby, A. Gayathri, Vishnu Srinivasa Murthy, Rishabh Shekhar | Wiley | An Efficient Model for Predicting Liver Disease Using Machine Learning | Support Vector Machine,Gradient Boosting, Naïve Baiyes, Logistic Regression, Random Forest | Naïve baiyes model’s Accuracy is less when compared with other models |
| 9 | 2021 | A. Durga Praveen, T. PanduRanga Vital, D. Jayaram, L. Venkata Satyanarayana | Springer | Intelligent Liver Disease Prediction (ILDP) System Using Machine Learning Models | KNN, SVM, RF, Naïve Bayes, and AdaBoost. | Models are trained with limited number of records. |
| 10 | 2021 | Shafiha R, Bahcivanci B, Gkoutos GV, Acharjee A. | Biomedicines | Machine Learning-Based Identification of Potentially Novel Non-Alcoholic Fatty Liver Disease Biomarkers | Random Forest | 1)Utilized a limited number of datasets.  2)Did not examine with an additional testing dataset for validation |
| 11 | 2021 | Shamima Akter, Hossain Uddin Shekhar, Sharif Akhteruzzaman | Advances in Bioscience and Biotechnology | Application of Biochemical Tests and Machine Learning Techniques to Diagnose and Evaluate Liver Disease | Random Forest, CART | Dataset used for training the model contains limited number of patient records. |
| 12 | 2021 | Krittika Dutta, Satish Chandra, Mahendra Kumar Gourisaria | Springer | Early-Stage Detection of Liver Disease Through Machine Learning Algorithms | Artificial Neural Network (ANN), Logistic Regression, K-Nearest Neighbor’s, SVC, Gaussian NB, Decision Tree, Random Forest, LR-SGD Classifier, Passive-Aggressive, Ada Boost | Evaluation metrics used for evaluating the models are not sufficient |
| 13 | 2021 | Md. Reshad Reza, Gahangir Hossain, Ayush Goyal, Sanju Tiwari, Anurag Tripathi, Anupama Bhan, Pritam Dash | Springer | Automatic Diabetes and Liver Disease Diagnosis and Prediction Through SVM and KNN Algorithms | Support Vector Machine, K-Nearest Neighbor | Manual exploration of patient data is difficult and patient data is not preprocessed efficiently. |
| 14 | 2021 | Liu YX, Liu X, Cen C, Li X, Liu JM, Ming ZY, Yu SF, Tang XF, Zhou L, Yu J, Huang KJ, Zheng SS. | Elsevier | Comparison and development of advanced machine learning tools to predict nonalcoholic fatty liver disease: An extended study | XG Boost | Limited number of records are used for training the model and did not examine with an additional testing dataset for validation |
| 15 | 2020 | A. Sivasangari; Baddigam Jaya Krishna Reddy; Annamareddy Kiran; P. Ajitha | IEEE | Diagnosis of Liver Disease using Machine Learning Models | Support Vector Machines (SVM), Decision Tree (DT) and Random Forest (RF) | Data preprocessing method used is not explained. |