**LITERATURE REVIEW**

Health care and medicine handles huge data on daily basis. This data comprises of information about the patients, diagnosis reports and medical images. It is important to utilize this information to decipher a decision support system. To achieve this, it is important to discover and extract the knowledge domain from the raw data. It is accomplished by knowledge discovery and data mining (KDD). The implementation of data mining techniques is widespread in biological domain. In recent years, liver disorders have excessively increased and liver diseases are becoming one of the most fatal diseases in several countries. In this study, liver patient datasets are investigated for building classification models in order to predict liver disease. Several feature model construction and comparative analysis are implemented for improving prediction accuracy of Indian liver patients. Different studies have been conducted for classification of liver disorders; they are discussed briefly. Classification algorithm is one of the greatest significant and applicable data mining techniques used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show good classification accuracy. In another study the UCI liver dataset was used for selection of sub features based on random forest classifier with multi-layer perceptron induced. Different approaches for artificial intelligence for the liver patient dataset, precise predictions of liver failure were applied. Identification of liver infection at preliminary stage is important to combat the frequency and severity deaths of patients in India. The patients must be screened based on initial symptoms for development of personalized therapy. In this study, an attempt is made for prediction of liver disease in patients using data mining techniques. Based on the review of literature, it was depicted that the past research studies have implemented different data mining techniques for classification of liver dataset. A hybrid model can be adapted to further increase the prediction accuracy of liver disease. It is followed by development of a graphical user interface would further aid the scientific community in early diagnosis of liver infection. It will provide a framework for end user application for generating promising treatment protocols.

Using machine learning algorithms to predict disease is made possible by increasing access to hidden attributes in medical data sets. Various kinds of data sets, such as blood panels with liver function tests, histologically stained slide images, and the presence of specific molecular markers in blood or tissue samples, have been used to train classifier algorithms to predict liver disease with good accuracy. The ML methods described in previous studies have been evaluated for accuracy by a combination of confusion matrix, receiver operating characteristic under area under curve, and k-fold cross-validation. Singh et al. designed software based on classification algorithms (including logistic regression, random forest, and naive Bayes) to predict the risk of liver disease from a data set with liver function test results. Vijayarani and Dhavanand found that SVM performed better over naive Bayes to predict cirrhosis, acute hepatitis, chronic hepatitis, and liver cancers from patient liver function test results. SVM with particle swarm optimization Livers 2021, 1 297 (PSO) predicted the most important features for liver disease detection with the highest accuracy over SVM, random forest, Bayesian network, and an MLP-neural network. SVM more accurately predicted drug-induced hepatotoxicity with reduced molecular descriptors than Bayesian and other previously used models. Phan and Chan et al. demonstrated that a convolutional neural network (CNN) model predicted liver cancer in subjects with hepatitis with an accuracy of 0.980. The ANN model has been used to predict liver cancer in patients with type 2 diabetes. Neural network ML methods can help differentiate between types of liver cancers when applied to imaging data sets. Neural network algorithms have even been trained to predict a patient’s survival after liver tumor removal using a data set containing images of processed and stained tissue from biopsies. ML methods can facilitate the diagnosis of many diseases in clinical settings if trained and tested thoroughly. More widespread application of these methods to varying data sets can further improve accuracy in current deep learning methods. This study aimed to

1. impute missing data using the MICE algorithm;
2. determine variable selection using eigen decomposition of a data matrix by PCA and to rank the important variables using the Gini index;
3. compare among several statistical learning methods the ability to predict binary classifications of liver disease;
4. use the synthetic minority oversampling technique (SMOTE) to oversample minority class to regulate overfitting;
5. obtain confusion matrices for comparing actual classes with predictive classes;
6. compare several ML approaches to assess a better performance of liver disease diagnosis;
7. evaluate receiver operating characteristic (ROC) curves for determining the diagnostic ability of binary classification of liver disease.

**References**

1. Wang, Y.; Li, Y.; Wang, X.; Gacesa, R.; Zhang, J.; Zhou, L.; Wang, B. Predicting Liver Disease Risk Using a Combination of Common Clinical Markers: A Screening Model from Routine Health Check-Up. Dis. Markers 2020, 2020, 8460883. [CrossRef]

2. Torkadi, P.P.; Apte, I.C.; Bhute, A.K. Biochemical evaluation of patients of alcoholic liver disease and non-alcoholic liver disease. Indian J. Clin. Biochem. 2014, 29, 79–83. [CrossRef]

3. Ceriotti, F.; Henny, J.; Queraltó, J.; Ziyu, S.; Özarda, Y.; Chen, B.; Boyd, J.C.; Panteghini, M. Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ-glutamyl transferase (GGT) in serum: Results from an IFCC multicenter study. Clin. Chem. Lab. Med. 2010, 48, 1593–1601. [CrossRef] [PubMed]

4. Chalasani, N.; Younossi, Z.; Lavine, J.E.; Charlton, M.; Cusi, K.; Rinella, M.; Harrison, S.A.; Brunt, E.M.; Sanyal, A.J. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018, 67, 328–357. [CrossRef] [PubMed]

5. Woreta, T.A.; Saleh, A.A. Evaluation of abnormal liver tests. Med Clin. 2014, 98, 1–16. [CrossRef]

6. Robles-Diaz, M.; Garcia-Cortes, M.; Medina-Caliz, I.; Gonzalez-Jimenez, A.; Gonzalez-Grande, R.; Navarro, J.M.; Castiella, A.; Zapata, E.M.; Romero-Gomez, M.; Blanco, S.; et al. The value of serum aspartate aminotransferase and gamma-glutamyl transpetidase as biomarkers in hepatotoxicity. Liver Int. 2015, 35, 2474–2482. [CrossRef] [PubMed]

7. Borroni, G.; Ceriani, R.; Cazzaniga, M.; Tommasini, M.; Roncalli, M.; Maltempo, C.; Felline, C.; Salerno, F. Comparison of simple tests for the non-invasive diagnosis of clinically silent cirrhosis in chronic hepatitis C. Aliment. Pharmacol. Ther. 2006, 24, 797–804. [CrossRef]

8. Asrani, S.K.; Devarbhavi, H.; Eaton, J.; Kamath, P.S. Burden of liver diseases in the world. J. Hepatol. 2019, 70, 151–171. [CrossRef]

9. Udell, J.A.; Wang, C.S.; Tinmouth, J.; FitzGerald, J.M.; Ayas, N.T.; Simel, D.L.; Schulzer, M.; Mak, E.; Yoshida, E.M. Does this patient with liver disease have cirrhosis? JAMA 2012, 307, 832–842. [CrossRef] [PubMed]

10. Munish, G.; Kaplan, H.C. Measurement for quality improvement: Using data to drive change. J. Perinatol. 2020, 40, 962–971