International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I5.102 ISSN: 1308-5581 Vol 14, Issue 05 2022

ANALYSIS ON DIAGNOSTIC METHODOLOGIES TO PREDICT LIVER DISEASE-A COMPREHENSIVE STUDY

S.Kowsalya,
Research Scholar (PhD), Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore-641008.
skowsalya@skasc.ac.in
Dr.S.Saraswathi
Associate Professor, Department of Computer Applications, Sri Krishna Arts and Science College, Coimbatore-641008.
saraswathys@skasc.ac.in

Abstract

The Medical data mining (MDM) is a fundamental component to extract the vital sensitive information that results in automated illness diagnosis and effective prediction. The MDM entails the utilization of data mining algorithms and methodologies for enhancing the analysis strategies of medical records. In recent years, liver illnesses have increased dramatically, to the point that liver ailments are now lethal in numerous nations. The humans suffering from liver disease has increased dramatically in recent years as a result of excessive alcohol use, unpureed air, and drug use, increased level of chemicals in food items, the proposed effort of contaminating food and pickling food. The proposed effort of Automatic prediction will be aided by the medical expert system to design the design making rubrics. The Early identification of liver disease is now achievable due to improvements in the Machine Learning (ML) technologies and successfully proven algorithms. This is more advantageous in the healthcare industry to overcome the dynamic challenges emerging every day. The effective function of liver is necessary for every human to survive because it aids in the removal of toxins from the body. As a result, early detection and treatment are crucial for sickness diagnosis and recovery. The study focuses on ML approaches are utilized, including supervised, unsupervised, semi-supervised, and reinforcement learning to detect liver illness. The objective of this research is to undertake a complete and comparative study of all machine learning algorithms currently being used in the medical industry for diagnosing and forecasting liver disease.

Keywords: Machine Learning, Liver disease

I. INTRODUCTION

The Medical data mining (MDM) is the process of extracting and analyzing hidden medical patterns to create predictive models that will be used to enhance the accuracy of diagnosis in the dynamic condition. In the present scenario applications of data mining methodologies are applied over the dataset and analysis studies have been conducted on liver illnesses, to control the disease is rapid expansion of the disease, though this and become one of the leading causes of death in several countries [1].

The Liver Infection (LD) is an inflammatory condition that affects the liver and impairs its function. It is caused by microorganisms or poisonous chemicals. In 2005, the World Health Organization (WHO) anticipated that 7.6 million people will die of cancer in the next decade, followed by another 84 million [1].The Liver cancer is the sixth most common kind of cancer globally and the third biggest cause of death and disability, proved by many research effort. It is effort undeniable that advancements in technology and increasing access to the internet have made it easier for people to detect the infections and advocate for patients with special needs of disorders [4]. The term "bioinformatics" refers to the use of computer and analytic tools to the collection and interpretation of biological data. The demand of using analytical tools increase day by day to ensure the accuracy [2-8]. It is very complicated to recognize liver illness in its early stages, even when liver tissue has been substantially damaged, so the medical specialists frequently fail to diagnose the condition at the earliest stage [14]. This may cause incorrect medication and therapy, It is the challenging aspect to predictive the infection in advance and vital to preserve the patient's life [15].

The Researchers list out the Characteristics that are possible to predict using, technology will assist in resolving health-related issues. One of the dynamic computing technology support to developing tools to ease labor or solve issues is termed as artificial intelligence. Machines are programmed to assume like people and trained systems, in which machine will act intelligently to solve problems like an expert. As with hepatitis, it may assist physicians who are not liver specialists in diagnosing and reporting the liver infections [16].

The Significant challenges placed on humans with liver infections are not immediately discovered during the first period. The Early identification of liver problems improves the patient's survival Possibilities. Indians are at an increased risk of developing liver failure [23]. India is expected to overtake the United States as the World Capital in this aspect by 2025. Due to a deskbound lifestyle, increasing alcohol use and smoking, India's ubiquitous incidence of infection inside the liver leads to around 100 different types of liver infections.

Additionally, it would be beneficial in the medical industry to develop a computer that would aid in illness Prediction. This method may assist physicians in making the best treatment decisions. Additionally, liver specialists such as endocrinologists will reduce the patient backlog, which will be supported by the Automated Categorization Methods for Liver Disorders section. Classification techniques are commonly used in medical diagnosis and illness prediction [24].

II. LITERATURE SURVEY FOR LIVER DISEASE DETECTION

S. Ambesange, A. V. et al. [1] developed the machine-learning model by, utilizing the Indian Liver Patient Dataset (ILPD) available at UCI.edu, which is based on Indian patients and the Random Forest (RF) method is used to forecast illness using various preprocessing Facts the data set is analyzed univariately and bivariately for skewness, outliers and imbalance. The data is then balanced using appropriate algorithms for different sampling strategies.

I. Arshad et al. [2] suggest employing the data extraction methods to identify and forecast the existence of the Liver Disease. The authors achieved maximum accuracy of 97.3913% by using the SMO algorithm. In comparison, the authors obtained the lowest accuracy, 70.7246

percent, using the Naive Bayes approach. The SMO algorithm required the most time (2.36 seconds), while the J48 algorithm required the least time (0 seconds).

Geetha, C., and Arunachalam, A. [5] devised and assessed approaches for identifying liver illness in patients using Machine Learning techniques. SVM, or LR, is one of the two primary ML algorithms. All models were used in the prediction study, and their performance was evaluated. The chance of developing the LD was predicted with 96% accuracy.

Using Machine Learning and image processing approaches, Giannakeas et al. [6] provide an automated method for assessing Non-Alcohol LD. The approach outperforms manual fat assessment in the liver.

K. Hamid et al. [8] provide a unique method for categorizing liver ultrasound pictures into normal and pathological categories. This technique is innovative in that it employs learning with the abstinence model for categorization. Our suggested technique can automatically identify the instances in which it lacks confidence in providing correct forecasts and identifying outliers.The researchers present a unique stochastic gradient-based solution for the paradigm of Learning with Abstention.

G. Pennazza et al. [17] The purpose of this research is to determine if this sort of analysis is capable of discriminating between liver cirrhosis (LC) and chronic hepatitis (CH), as well as defining the LC stage. The research will examine the discriminative and classificatory features of the e-nose in patients with chronic liver disease (CLD), as well as the ability of the e-nose to discriminate between distinct Child-Pugh (CPC) classes in LC patients.

R. T. Ribeiro et al. [22] offer a categorization and stage technique for CLD based on the disease's natural history. The system, dubbed the clinical-based classifier, involves a pipeline of binary classification steps similar to those used in clinical practice.

A. Sivasangari et al. [25] test several ML techniques for predicting the liver illness. The Liver illness is notoriously difficult to identify due to its modest signs. Prediction of LD followed the stage of data preparation, in which data were acquired from public databases and preprocessed for -1 value substitution. The Data segmentation into the complete data collection and its separation into training and research.

Thirunavukkarasu k et al. [29] employed various classification methods to predict liver illness in this study effort, including LR, SVM, and KNN. The classification accuracy of all these methods was determined using the confusion matrix. The experiment indicates that LR and KNN have the best accuracy, while LR has the highest sensitivity.

According to K. Wahab et al. [31], information in the biomedical area is concealed in an unstructured manner inside publications and on the World Wide Web. To make use of this semantic information, a web is created. Due to the mismatch in words, it isn't easy to transform the web into a semantic web. As a result of this reasoning, it is challenging are compelled to construct ontology. This effort proposes a basic biomedical ontology for chronic Liver issues, using the Unified Medical Language System as a data source to aid in the finding of medical information. The ontology's connections can be used to represent knowledge by processing data via a protégé toolkit interface. The ontology was built and then examined by domain experts to ascertain the ontology's quality.

Zhao et al. [33] proposed an ensemble empirical mode decomposition technique for correctly diagnosing faults. It was used to decompose vibration signals into a succession of intrinsic mode functions using a variety of physical significance feature extraction approaches. The

correlation coefficient analysis approach to get the entropy values for the intrinsic mode function to accomplish fault pattern detection. The benefit is that multi-scale fuzzy entropy ensures that attenuated fault frequency signals retain their relative consistency. The drawback is that it is critical to increasing rolling bearing dependability accurately over time. The Rolling bears are a critical component of rotating equipment must be operated safely, effectively, and reliably. The rolling bearing's defect rate must be high without decreasing.

III. COMPARATIVE ANALYSIS OF SURVEY

Table 1 Feature extraction for liver disorders

Reference Paper	Methodology	Advantages	Disadvantages
Acharya et al. (2016) [3]	A feature extraction method to differentiate between original fatty liver disorder and cirrhosis ultrasound images by utilizing curvelet transform.	Different non-invasive computer-aided methods were utilized to detect fatty liver disorders.	If liver disorder had been diagnosed at a later stage, then it leads to death ultimately.
Patidar et al. (2017) [18]	Alcoholic index Utilizing tunable Q-wavelet transform to detect small changes between alcoholic and actual parameters precisely.	It provides a better correlation coefficient than other methods.	Overconsumption of alcoholism would lead to liver disorder.
Rajathi and Jiji (2019) [21]	A chronic liver Disorder extraction and classification utilizing hybrid whale optimization method to extract features of liver patients.	Inflammatory It provides best-extracted features for liver patients	The condition would lead to incomplete or entire liver dysfunction.
Yu et al. (2019) [32]	A pathogenesis method to detect the deficiency of copper in liver disorders.	It is imaginable to prevent deficiency of copper.	Lesser serum copper was linked with low hepatic for partial patients as the availability of zinc component affects the synthesis of metallothionein.

Tomeno et al. (2015) [28]	Major depressive disorders to detect the clinical features of non-alcoholic fatty LD.	It reduces the risk of mortality in liver patients.	Appropriate pharmacological treatment for the non-alcoholic fatty LD has not been established still.

Table 2: Ensemble Based feature Extraction

Artical	Methodology	Advantages	Disadvantages
Shang et al. (2017) [26]	New feature extraction methods depend on ensemble empirical mode decomposition and sample entropy to reduce the influence of noise.	It resolves high-dimension and signals attenuation estimation problems easily.	If the feature extraction utilized the existing partial discharge method, then recognition results were affected by the attenuation of signals.
Mollaee and Moattar (2016) [11]	A new feature extraction method for microarray ensemble data classification reduces the computational cost.	It provides better extraction performance in a supervised way.	The fitness function calculation of particle swarm optimization provides lower computational complexity than discriminative independent component analysis.
Piao et al. (2019) [19]	Asymmetrical uncertainty-based feature extraction on identifying modification of customer class types.	It is very flexible to store consumed data by loading profiles and customer extraction. quantified class	Customers must protect information within the given period from the datasets.
Rai and Upadhyay (2018) [20]	An integrated bearing prognostic to enhance performance degradation assessment.	A divergence method had been utilized to differentiate defective posterior probability distribution from strong ones.	An extremely unpredictable behavior causes severe degradation to train Gaussian mixture models.

Zhao et al. (2017) [33]	An ensemble empirical mode decomposition to diagnose fault very accurately.	Multi-scale fuzzy entropy provides relative consistency of attenuated fault frequency signals.	It is significant to enhance the reliability of rolling bearings precisely within a period.

Table 3: Classification of Liver Disorders

Article	Methodology	Advantages	Disadvantages
Vijayarani and Dhayanand (2015) [30]	The LD prediction to predict disease from a huge number of medical databases by using two classification algorithms such as SVM and NB	Only less training data was required to calculate mean and variance for classification.	It is required to calculate a certain dataset of liver disorder
Sindhuja and Priyadarsini (2016) [27]	A classification technique in data mining to analyze the disorder for LD in human	It holds continuous and discrete values very easily.	The usability of insignificant branches could not be reduced.
Nahar and Ara (2018) [13]	LD prediction predicts the early stage by utilizing different decision tree methods.	It provides better accuracy to diagnose LD disorder	It is essential to secure a liver disorder dataset within a specific period.
Maaitah et al. (2017) [12]	An intelligent method for the liver disorder based on a fuzzy neural system to identify a LD for specific patients.	The computational time was high.	It requires a certain time to cure the liver disorder of a particular patient.
Nagaraj and Sridhar (2015) [14]	An identification of the liver to detect infection at the initial stage	SVM method performs better accuracy.	Only appropriate features for LD disorder must be selected.

In Table 1, 2, 3 are comparative evaluation on various authors' views.

Table 4: Accuracy Comparison Table

Author	Method	Accuracy
Geetha, C., & Arunachalam, A. (2021) [5]	SVM	75.04 %

Hartatik, H. et al. (2020) [7]	NB	72.5 %
	KNN	63.19 %
Kumar, P., & Thakur, R. S. (2019) [9]	Fuzzy Adaptive KNN	91.12 %
Kumar, S., & Katyal, S. (2018) [10]	K-MEANS	69.74 %
	RANDOM FOREST	72.18 %
	C5.0 with Adaptive Boosting	75.19 %

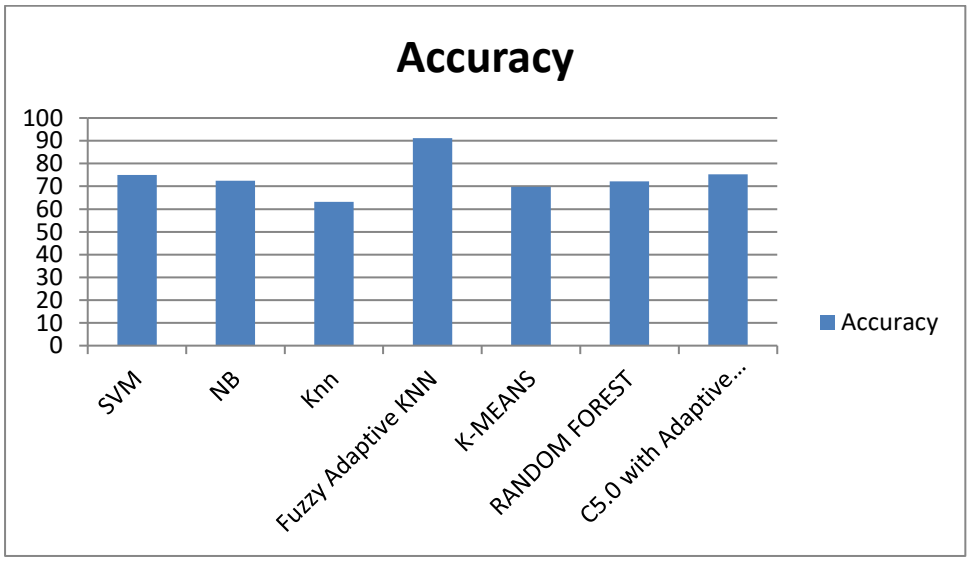

Figure 1: Accuracy Comparison Chart

Figure 1 represents the accuracy comparison chart for various algorithms like SVM, NB, RF, KNN, K-Means, Fuzzy Adaptive KNN algorithms.

IV. CONCLUSION

The research effort summarizes the previous research on identifying and diagnosing Liver Disease using various Machine Learning techniques. This review and research discovered and observed that some Machine Learning methods, such as decision trees, SVMs, and fuzzy adaptive KNN, are more accurate in detecting and predicting liver illness. Each algorithm performs differently in various settings, most crucially, the dataset and feature selection are critical for obtaining superior prediction outcomes. The article provides an analysis on literature survey and reported the applications on different kinds of ML approaches. Each ML approach produces desirable and undesirable results based on the datasets with appropriate feature selection. This study discovered that accuracy and performance might be enhanced by combining or hybridizing ML methods. The challenging aspects in the future to get higher performance than the present approach.

V. REFERENCES

[1] Ambesange, S., A, V., Uppin, R., Patil, S., & Patil, V. (2020). Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques. 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM). doi:10.1109/ccem50674.2020.00030

[2] Arshad, I., Dutta, C., Choudhury, T., & Thakral, A. (2018). Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques. 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE). doi:10.1109/icacce.2018.8441721

[3] Acharya, U. R., Raghavendra, U., Fujita, H., Hagiwara, Y., Koh, J. E., Hong, T. J., Sudarshan, V. K., Vijayananthan, A., Yeong, C. H., Gudigar, A. et al. (2016), 'Automated characterization of fatty liver disease and cirrhosis using curvelet transform and entropy features extracted from ultrasound images', Computers in biology and medicine 79, 250–258.

[4] Bahramirad, S., Mustapha, A., & Eshraghi, M. (2013). Classification of liver disease diagnosis: A comparative study. 2013 Second International Conference on Informatics & Applications (ICIA). doi:10.1109/icoia.2013.6650227

[5] Geetha, C., & Arunachalam, A. (2021). Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms. 2021 International Conference on Computer Communication and Informatics (ICCCI). doi:10.1109/iccci50826.2021.94024

[6] Giannakeas, N., Tsipouras, M. G., Tzallas, A. T., Vavva, M. G., Tsimplakidou, M., Karvounis, E. C., … Manousou, P. (2017). Measuring Steatosis in Liver Biopsies Using Machine Learning and Morphological Imaging. 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). doi:10.1109/cbms.2017.98

[7] Hartatik, H., Tamam, M. B., & Setyanto, A. (2020). Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms. 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS). doi:10.1109/icoris50180.2020.9326

[8] Hamid, K., Asif, A., Abbasi, W., Sabih, D., & Minhas, F.-A. A. (2017). Machine Learning with Abstention for Automated Liver Disease Diagnosis. 2017 International Conference on Frontiers of Information Technology (FIT). doi:10.1109/fit.2017.00070

[9] Kumar, P., & Thakur, R. S. (2019). Diagnosis of Liver Disorder Using Fuzzy Adaptive and Neighbor Weighted K-NN Method for LFT Imbalanced Data. 2019 International Conference on Smart Structures and Systems (ICSSS). doi:10.1109/icsss.2019.8882861

[10] Kumar, S., & Katyal, S. (2018). Effective Analysis and Diagnosis of Liver Disorder by Data Mining. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). doi:10.1109/icirca.2018.8596817

[11] Mollaee, M. and Moattar, M. H. (2016), 'A novel feature extraction approach based on ensemble feature selection and modified discriminant independent component analysis for microarray data classification', Biocybernetics and Biomedical Engineering 36(3), 521–529.

[12] Maaitah, M. K. S., Abiyev, R. and Bush, I. J. (2017), 'Intelligent classification of liver disorder using fuzzy neural system', International Journal of Advanced Computer Science and Applications 8(12), 25–31.

[13] Nahar, N. and Ara, F. (2018), 'Liver disease prediction by using different decision tree techniques', Int. J. Data Min. Knowl. Manag. Process 8(2), 01–09.

[14] Nagaraj, K. and Sridhar, A. (2015), 'Neurosvm: a graphical user interface for identification of liver patients', arXiv preprint arXiv:1502.05534 .

[15] Netzer, M., Millonig, G., Pfeifer, B., Kusonmano, K., Praun, S., Villinger, J., … Baumgartner, C. (2009). Stacked Feature Selection in Liver Disease Using IMR-MS Analysis. 2009 20th International Workshop on Database and Expert Systems Application. doi:10.1109/dexa.2009.20

[16] Olteanu, V. A., Mitrica, D. E., Balan, G., & Stefanescu, G. (2017). Technical and clinical biases of transient elastography in liver diseases. 2017 E-Health and Bioengineering Conference (EHB). doi:10.1109/ehb.2017.7995466

[17] Pennazza, G., Santonico, M., Vernile, C., Antonelli Incalzi, R., De Vincentis, A., & Picardi, A. (2015). Breathprinting of liver diseases. 2015 XVIII AISEM Annual Conference. doi:10.1109/aisem.2015.7066849

[18] Patidar, S., Pachori, R. B., Upadhyay, A. and Acharya, U. R. (2017), 'An integrated alcoholic index using tunable-q wavelet transform based features extracted from eeg signals for diagnosis of alcoholism', Applied Soft Computing 50, 71–78.

[19] Piao, M., Piao, Y. and Lee, J. Y. (2019), 'Symmetrical uncertainty-based feature subset generation and ensemble learning for electricity customer classification', Symmetry 11(4), 498.

[20] Rai, A. and Upadhyay, S. (2018), 'An integrated approach to bearing prognostics based on eemd-multi feature extraction, gaussian mixture models and jensen-r´enyi divergence', Applied Soft Computing 71, 36–50.

[21] Rajathi, G. I. and Jiji, G. W. (2019), 'Chronic liver disease classification using hybrid whale optimization with simulated annealing and ensemble classifier', Symmetry 11(1), 33.

[22] Ribeiro, R. T., Marinho, R. T., & Sanches, J. M. (2013). Classification and Staging of Chronic Liver Disease From Multimodal Data. IEEE Transactions on Biomedical Engineering, 60(5), 1336–1344. doi:10.1109/tbme.2012.2235438

[23] Sasso, M., Miette, V., & Sandrin, L. (2009). Novel Controlled Attenuation Parameter for the evaluation of fatty liver disease. 2009 IEEE International Ultrasonics Symposium. doi:10.1109/ultsym.2009.5441681

[24] Shaheamlung, G., Kaur, H., & Kaur, M. (2020). A Survey on machine learning techniques for the diagnosis of liver disease. 2020 International Conference on Intelligent Engineering and Management (ICIEM). doi:10.1109/iciem48762.2020.9160

[25] Sivasangari, A., Krishna Reddy, B. J., Kiran, A., & Ajitha, P. (2020). Diagnosis of Liver Disease using Machine Learning Models. 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). doi:10.1109/i-smac49090.2020.9243375

[26] Shang, H., Lo, K. L. and Li, F. (2017), 'Partial discharge feature extraction based on ensemble empirical mode decomposition and sample entropy', Entropy 19(9), 439.

[27] Sindhuja, D. and Priyadarsini, R. J. (2016), 'A survey on classification techniques in data mining for analyzing liver disease disorder', International Journal of Computer Science and Mobile Computing 5(5), 483–488.

[28] Tomeno, W., Kawashima, K., Yoneda, M., Saito, S., Ogawa, Y., Honda, Y., Kessoku, T., Imajo, K., Mawatari, H., Fujita, K. et al. (2015), 'Non-alcoholic fatty liver disease comorbid with major depressive disorder: The pathological features and poor therapeutic efficacy', Journal of gastroenterology and hepatology 30(6), 1009–1014.

[29] Thirunavukkarasu, k., Singh, A. S., Irfan, M., & Chowdhury, A. (2018). Prediction of Liver Disease using Classification Algorithms. 2018 4th International Conference on Computing Communication and Automation (ICCCA). doi:10.1109/ccaa.2018.8777655

[30] Vijayarani, S. and Dhayanand, S. (2015), 'Liver disease prediction using svm and na¨ıve bayes algorithms', International Journal of Science, Engineering and Technology Research (IJSETR) 4(4), 816–820.

[31] Wahab, K., Qamar, U., Arif, K. S., & Ali, U. (2019). Building a Biomedical Ontology for Chronic Liver Disease. 2019 International Conference on Computer, Information and Telecommunication Systems (CITS). doi:10.1109/cits.2019.8862104

[32] Yu, L., Liou, I. W., Biggins, S. W., Yeh, M., Jalikis, F., Chan, L.-N. and Burkhead, J. (2019), 'Copper deficiency in liver diseases: a case series and pathophysiological considerations', Hepatology communications 3(8), 1159–1165.

[33] Zhao, H., Sun, M., Deng, W. and Yang, X. (2017), 'A new feature extraction method based on eemd and multi-scale fuzzy entropy for motor bearing', Entropy 19(1), 14. 124