

A NOVEL METHOD FOR DETECTION OF LIVER DISEASE USING HYBRID MACHINE LEARNING ALGORITHM

¹Cherakulam Vidya, ²K. Lakshman Kumar, ³Sakamuri Srinivasa Rao,
⁴Shaik Baba Fariddin

¹Assistant professor, Department of Electronics and Communication Engineering, Sree Venkateswara College of Engineering, North Rajupalem, Nellore Dist., Andhra Pradesh, India

²Assistant Professor, Department of Computer Science & Engineering, Sri Venkatesa Perumal College of Engineering and Technology, Puttur, Andhra Pradesh, India

³Assistant Professor, Department of Computer Science & Engineering, Chalapathi Institute of Technology, A.R Nagar, Mothadaka, Guntur, Andhra Pradesh, India

⁴Assistant Professor, Department of Electronics and Communication Engineering, NRI Institute of Technology, Guntur, Andhra Pradesh, India

ABSTRACT: In Human beings, Liver is the most primary part of the body that performs many functions including the production of Bile, excretion of bile and bilirubin, metabolism of proteins and carbohydrates, activation of Enzymes, Storing glycogen, vitamins, and minerals, plasma proteins synthesis and clotting factors. The liver easily gets affected due to intake of alcohol, pain killer tablets, food habits, and includes plenty of wired practices. Currently, the liver related diseases are identified by analyzing liver function blood test reports and scan reports. There are several traditional methods to diagnose liver diseases, but they are expensive. Early prediction of liver disease would benefit all individuals who suffer with liver diseases for early treatment. As technology is growing in health care, machine learning significantly affects health care for predicting conditions at early stages. Therefore, by using hybrid machine learning algorithm (Support Vector Machine + Decision Tree) predicts liver disease at early stage. This method shows better results interms of accuracy, precision and detection rate.

Keywords: Liver, Support Vector Machine (SVM), Decision Tree (DT), Bilirubin, Enzymes

I. INTRODUCTION

The liver is the largest organ of the body and it is essential for digesting food and releasing the toxic element of the body. The viruses and alcohol use lead the liver towards liver damage and lead a human to a life-threatening condition. There are many types of liver diseases whereas hepatitis, cirrhosis, liver tumors, liver cancer, and many more. Among them liver diseases and cirrhosis as the main cause of death [1].

A total of 264,193 deaths as a result of liver disease were reported in India in 2018, according to the latest World Health Organization data [2]. There are about 23.00 deaths per 100,000 people based on age-adjusted death rates for the population. With a weight of approximately 1.36 kg, the liver is the largest organ in the body. It has four lobes of differing sizes and shapes, and is dark reddish-brown in color. The liver is located right behind the diaphragm beneath the abdominal cavity. The hepatic artery and the portal vein are two major arteries that transport blood to the liver. Its primary function is to eliminate poisonous and damaging compounds from the bloodstream before they are distributed to other regions of the body. WHO officials have identified liver disease as one of the most serious and deadly diseases[3].

Hepatitis infection, fatty liver, cirrhosis, liver fibrosis, high alcohol intake, drug exposure, and genetic anomalies can all cause liver disease[4]. A liver transplant is the only treatment option left if the liver has completely failed, and there is no way to recover it. Timely identification of liver illnesses can aid in therapy and speedy recovery. The phases of liver disease are: healthy, fibrosis, cirrhosis, and the last stage is cancer. Detecting liver disease in its early stages can be difficult, even after there is significant damage to liver tissue. This would lead to failure to provide

proper treatment and drugs. An early diagnosis of the disease is crucial to preventing this and saving the patient's life. Internal bleeding, dry mouth, constipation, and stomach pain are a few signs of liver disease that can affect the digestive system[5]. Some other signs include brain and nervous system anomalies such as loss of memory, numbness, and fainting, as well as skin concerns such as yellow skin, spider veins, and feet redness. Visiting a doctor regularly, getting vaccinated, drinking less soda and alcohol, exercising regularly, and keeping your weight in check can prevent liver diseases. The advancement of artificial intelligence has led to the development of numerous machine learning algorithms which enhance the accuracy and effectiveness of diagnosing and prognosticating liver disease[6].

The prevention of liver failure is possible by diagnosing and treating liver diseases at an early stage. There are four stages of liver disease, among which the first stage is marked by inflammation, which may or may not show any symptoms in patients. Prolonged inflammation replaces the healthy liver tissue with the scar tissue, which causes the disease to enter in the second stage, i.e., fibrosis, which is also mostly asymptomatic. Severe scarring on the liver causes cirrhosis, which is the third stage[7]. In this stage, the patient starts to experience symptoms like abdominal pain, fatigue, weakness, jaundice, etc.

When there is a drastic deterioration in liver function, then it is End-Stage Liver Disease (ESLD). In this, the patient shows severe complications but can be treated without transplanting the liver. In the fourth stage, unhealthy cells start to develop and expand, which causes Liver cancer. These conditions pointed to the need for methods that provide early prediction of liver diseases so that the

effect can be mitigated, and damage can be controlled by providing appropriate treatment at an early stage[8]. The commonly used diagnosis and tests for liver diseases or Hepatic diseases include a liver blood test, Complete Blood Count (CBC), Abdominal and Pelvic CT, Abdominal Ultrasound, Elastography, ERCP (Endoscopic Retrograde Cholangiopancreatography), Lactate Dehydrogenase (LDH) Isoenzymes Test, Lactate Dehydrogenase (LDH) Test, Liver Biopsy, Liver Function Tests (LFT), Magnetic Resonance Cholangiopancreatography (MRCP), MRI of the body (Chest, Abdomen, Pelvis), and Needle Biopsy[9].

In many automatic medical diagnostic tools, classification approaches are particularly common. Due to the fact that liver diseases do not manifest until the organ is partially damaged, it is difficult to detect early. The presence of enzymes in the blood can be used to identify liver disease[10]. Furthermore, mobile devices are increasingly being utilized to track the health of humans. In this case, it is also necessary to use automatic classification algorithms. Mobile and online technologies capable of automatically identifying liver illnesses can be used to reduce patient wait times with liver specialists such as endocrinologists.

II. LITERATURE SURVEY

N. Afreen, R. Patel, M. Ahmed and M. Sameer, et.al [11] implemented gradient boosting based machine learning classifier to achieve the results. CatBoost and LightGBM model are employed for prediction and classification of liver disease with feature selection approach. Preprocessing is performed on the original dataset to remove deviated values using isolation forest and to get relevant features for better results. Model performance is calculated in respect of precision, accuracy, recall and f1-score. CatBoost

resulted in highest accuracy of 86.8% and LightGBM achieves 82.6% accuracy with feature selection on Indian Liver Patient Dataset.

N. Giannakeas *et al.*, [12] introduces an automated method for measuring steatosis in liver biopsies, using both machine learning and classical image processing techniques. Clustering is employed for tissue specimen detection, while an iterative morphological procedure is used for steatosis revealing. The method has been evaluated in a set of 20 liver biopsy images and the obtained results present ~1% mean percentage error.

A. Brankovic, A. Zamani and A. Abbosh, et.al [13] propose an electromagnetic system, including an antenna operating across the band 0.4-1 GHz as a data acquisition device and a supervised Machine Learning (ML) framework to learn an inferring model for Fatty Liver Disease (FLD) directly from collected data. This paper reports the system configuration, ML problem setup and the obtained results, which show an accuracy of more than 97% for the simulated torso model. Computational methods reduce the operator dependability and hence improve diagnostic reliability. Most of these methods are based on automated analysis of Computed Tomography (CT), Ultrasound (US), and Magnetic Resonance (MRI) images. Besides the high costs and harmful radiation involved in the conventional imaging tools, the outcome of the automated imaging strongly depends on the image quality.

V. Singh, M. K. Gourisaria and H. Das, et.al [14] diagnosis of liver disease needs high accuracy and precise results for predicting whether a person is suffering from liver disease or not. Major disastrous repercussions can be the result of minor errors in the diagnosis of liver diseases.

The major goal of this paper is for the detection of liver disease at right time and helping the doctors and combating the increasing number of cases. In this paper, we implemented various machine learning techniques like logistic regression, KNN, XG-Boost, SVM, Gaussian NB, Random forest, Decision tree, Gradient Boosting, CatBoost, AdaBoost, and LightGBM on selected features from the dataset for predicting liver disease and it was found that Random Forest performed best among all the technique and gained high accuracy and performed outstandingly in all metric evaluations.

K. Hamid, A. Asif, W. Abbasi, D. Sabih and F. -u. -A. A. Minhas, et.al [15] a novel approach for detection of liver abnormalities in an automated manner using ultrasound images. For this purpose, we have implemented a machine learning model that can, not only generate labels (normal and abnormal) for a given ultrasound image but, it can also detect when its prediction is likely to be incorrect. The proposed model abstains from generating the label of a test example if it is not confident about its prediction. Such behavior is commonly practiced by medical doctors who, when given insufficient information or a difficult case, can choose to carry out further clinical or diagnostic tests before generating a diagnosis. We have proposed a novel stochastic gradient descent based solver for the learning with abstention paradigm and use it to make a practical, state of the art method for liver disease classification.

S. Panigrahi, R. Deo and E. A. Liechty, et.al [16] paper focuses on machine learning-based intelligent model development using liver functionality and physiological parameters for Hepatic Steatosis (Non-alcoholic Fatty Liver) screening. Gender-specific models were

developed separately. Customized data processing techniques were incorporated. Publicly available, population data (NHANES-III) was used. The maximum sensitivity provided by the models were approximately 72% and 71% for male and female, respectively. Maximum specificities obtained by the models were 74% and 75% for male and female, respectively.

X. Li, X. Chen and Z. Yuan, et.al [17] Classification and Regression Tree (CART) as a weak classifier of the AdaBoost framework to propose a Classification and Regression Tree-Adaptive Boosting (CART-AdaBoost) model. Moreover, the authors trained and verified the model basing on the Indian Liver Patient Dataset (ILPD) of UCI. The results showed that the model's accuracy was 83.06%, and its precision was 84.31%. Besides, F1-score could reach 80.75%, and the recall metric was 77.48%.

K. Srivastava, G. Malhotra, M. Chauhan and S. Jain, et.al [18] proposes a design for early detection and classification of liver cancer which consists of discrete wavelet transform image fusion technique, extraction of Speeding up robust features, feature selection using Cuckoo meta-heuristic approach and machine learning algorithms from contrast enhanced CT and MRI images. The proposed framework is fully automated which requires no user interaction. To overcome this problem, in this paper, detection of liver cancer is done using fused images of CT scan and Magnetic Resonance Imaging (MRI) that plays a crucial role in the choice of different strategies for liver disease and also for treatment monitoring.

D. S. Reddy, R. Bharath and P. Rajalakshmi, et.al [19] propose a novel CAD framework using convolution neural

networks and transfer learning (pre-trained VGG-16 model). Performance analysis shows that the proposed framework offers an FLD classification accuracy of 90.6% in classifying normal and fatty liver images. However, this quantification or diagnostic accuracy depends on the expertise and skill of the radiologist. With the advent of Health 4.0 and the Computer Aided Diagnosis (CAD) techniques, the accuracy in detection of FLD using the ultrasound by the sonographers and clinicians can be improved. Along with an accurate diagnosis, the CAD techniques will help radiologists to diagnose more patients in less time.

T. A. Joarder, B. Ahmed and A. S. Sattar, et.al [20] proposed a method to organize online search log data to detect liver cancer early. If we have online search log data containing symptoms searched on search engine and the time of the searches of each symptom, they can be easily organized by our method of dataset creation and detect liver cancer. We have estimated the performance of Support Vector Machine (SVM) based on different kernel scale values. The best classification accuracy for $\gamma = 0.03$ is 94.30%. We have also compared the performance of different machine learning techniques. Among them, Random Forest has the highest classification accuracy of 97.50%.

A. Kalsoom, A. Moin, M. Maqsood, I. Mehmood and S. Rho, et.al [21] propose an unsupervised machine learning technique combined with a supervised mechanism that accurately performs liver tumor segmentation. We perform clustering on our collected dataset and extract LBP features as well as HOG features from these clusters. Furthermore, we perform classification which is based on these extracted features using KNN. Furthermore, we have compared our

results with two classifiers namely SVM and Ensemble to achieve a better understanding. Our proposed technique outperformed existing techniques and showed encouraging results when compared to other methods.

W. Cao, P. Yang, Z. Ming, S. Cai and J. Zhang, et.al [22] proposed Improved Fuzziness - Random Vector Functional Link network (IF-RVFL) is a semi-supervised learning algorithm using the self-training strategy, which can make full use of a large number of unlabeled samples to improve the performance of the model. At the same time, the SMOTE technique enables the IFRVFL to effectively solve the class imbalanced problem. The effectiveness of the proposed IF-RVFL has been verified on a real-life liver disease data set. Extensive experimental results show that the IF-RVFL algorithm can achieve better generalization ability than the RVFL, F-RVFL, and their variants. IF-RVFL also provides a new technique with great potential for other disease detection.

III. A FRAMEWORK FOR DETECTION OF LIVER DISEASE USING HYBRID MACHINE LEARNING ALGORITHM

In this section, a framework of a novel method for detection of liver disease using hybrid machine learning algorithm is observed in Figure.1.

Input data is given and then the data will be pre-processed. After pre-processing, attributes are selected. The data is classified as training data and testing data. The training data as well as testing data is given to the machine learning classifier. Then the disease is predicted.

A real-world data generally contains noises, missing values, and maybe in an

unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model. Attribute Selection Measures is a heuristic approach to select the best splitting criterion that separates a given data partition, D , of class-labeled training tuples into individual classes. Splitting criterion is called the best when after splitting, each partition will be pure. Attributes are the items of data that are used in machine learning. Attributes are also referred as variables, fields, or predictors. In predictive models, attributes are the predictors that affect a given outcome.

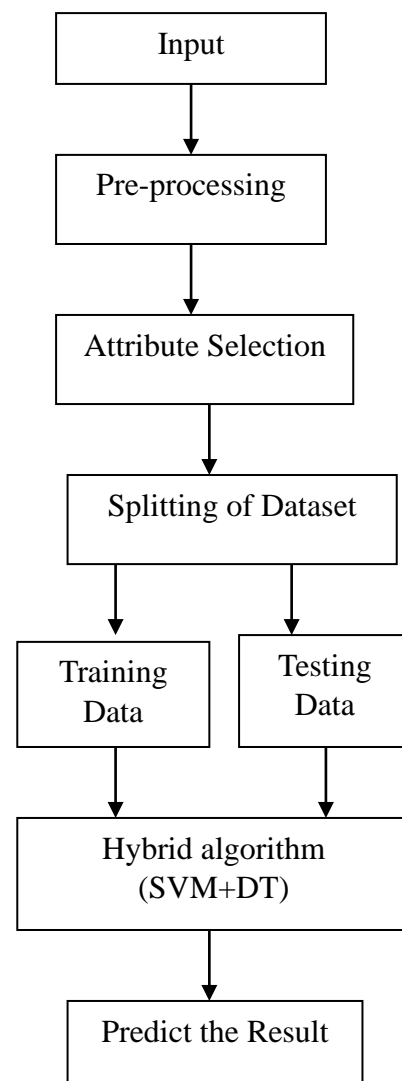


Figure1. Framework of A Novel Method For Detection Of Liver Disease Using Hybrid Machine Learning Algorithm

Data splitting is a crucial process in machine learning, involving the partitioning of a dataset into different subsets, such as training, validation, and test sets. This is essential for training models, tuning parameters, and ultimately assessing their performance. Training data is the data you use to train an algorithm or machine learning model to predict the outcome you design your model to predict. If you are using supervised learning or some hybrid that includes that approach, your data will be enriched with data labeling or annotation. Finally, the test data set is a data set used to provide an unbiased evaluation of a final model fit on the training data set. If the data in the test data set has never been used in training (for example in cross-validation), the test data set is also called a holdout data set. Support Vector Machine (SVM) is a powerful machine learning algorithm used for linear or nonlinear classification, regression, and even outlier detection tasks. SVMs can be used for a variety of tasks, such as text classification, image classification, spam detection, handwriting identification, gene expression analysis, face detection, and anomaly detection. SVMs are adaptable and efficient in a variety of applications because they can manage high-dimensional data and nonlinear relationships. A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes. It is the possible outcomes of a scenario. A prediction is a statement that should tell something about the future.

The results that can expect the hypothesis is correct.

IV. RESULT ANALYSIS

In this section result analysis for a novel method for detection of liver disease using hybrid machine learning algorithm is observed.

Table.1: Performance Comparison

Parameters	Random Forest (RF)	Hybrid Method
Accuracy	89.6	92.7
Precision	91.8	96.6
Detection Rate	90.3	95.2

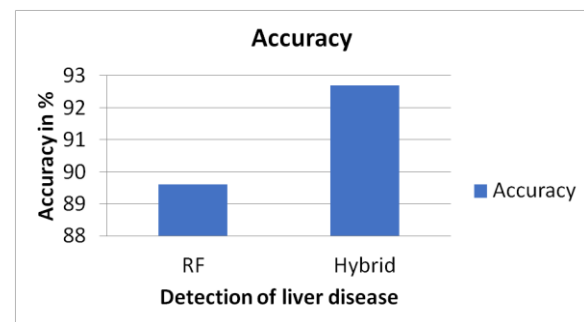


Figure 2. Accuracy Comparison Graph

In Figure.2 accuracy comparison graph between Random Forest and Hybrid method is observed in novel method for detection of liver disease using hybrid machine learning algorithm. Hybrid method shows high accuracy.

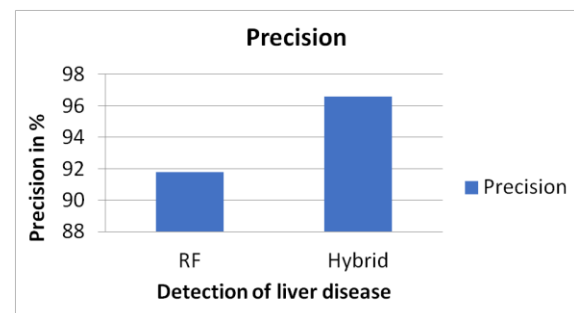


Figure 3. Precision Comparison Graph

Hybrid method shows high precision in Figure.3 for a novel method for detection of liver disease using hybrid machine learning algorithm.

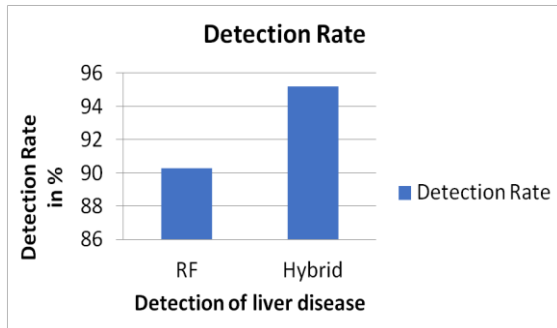


Figure 4. Detection Rate Comparison Graph

In Figure.4 detection rate comparison graph between Random Forest and Hybrid method is observed in novel method for detection of liver disease using hybrid machine learning algorithm.

V. CONCLUSION

In this section a novel method for detection of liver disease using hybrid machine learning algorithm is concluded. Chronic Liver Disease is the leading cause of global death that impacts the massive quantity of humans around the world. This disease is caused by an assortment of elements that harm the liver. Therefore, by using this hybrid machine learning algorithm liver disease is detected accurately. Hence, this method achieves better results in terms of accuracy, precision and detection rate.

VI. REFERENCES

[1] G. Shaheamlung, H. Kaur and M. Kaur, "A Survey on machine learning techniques for the diagnosis of liver disease," *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, London, UK, 2020, pp. 337-341, doi: 10.1109/ICIEM48762.2020.9160097.

[2] M. A. Kuzhippallil, C. Joseph and A. Kannan, "Comparative Analysis of Machine Learning Techniques for Indian Liver Disease Patients," *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2020, pp. 778-782, doi: 10.1109/ICACCS48705.2020.9074368.

[3] 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)*, Dhaka, Bangladesh, 2020, pp. 111-116, doi: 10.1109/ICAICT51780.2020.9333528.

[4] 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)*, Coimbatore, India, 2021, pp. 1-4, doi: 10.1109/ICCCI50826.2021.9402463.

[5] 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)*, Palladam, India, 2020, pp. 627-630, doi: 10.1109/I-SMAC49090.2020.9243375.

[6] P. Hamsagayathri and S. Vigneshwaran, "Symptoms Based Disease Prediction Using Machine Learning Techniques," *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Tirunelveli, India, 2021, pp. 747-752, doi: 10.1109/ICICV50876.2021.9388603.

[7] S. Gupta, U. Namdev, V. Gupta, V. Chheda and K. Bhowmick, "Data-driven

- Preprocessing Techniques for Early Diagnosis of Diabetes, Heart and Liver Diseases," *2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Erode, India, 2021, pp. 1-8, doi: 10.1109/ICECCT52121.2021.9616835.
- [8] N. Nahar, F. Ara, M. A. I. Neloy, V. Barua, M. S. Hossain and K. Andersson, "A Comparative Analysis of the Ensemble Method for Liver Disease Prediction," *2019 2nd International Conference on Innovation in Engineering and Technology (ICIET)*, Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICIET48527.2019.9290507.
- [9] A. L. Kostka. J.E. and S. V. Jinny, "Automatic Extraction of Lesions and Hepatic Structures in Liver using Segmentation Techniques," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, Coimbatore, India, 2020, pp. 114-117, doi: 10.1109/ICIRCA48905.2020.9182888.
- [10] H. Li, L. Yang, T. He, Y. Xiao, Z. Liang and X. Wu, "B-mode Ultrasound Texture Recognition Algorithm of Liver Based on Random Forests," *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, Chengdu, China, 2020, pp. 614-618, doi: 10.1109/CISP-BMEI51763.2020.9263593.
- [11] N. Afreen, R. Patel, M. Ahmed and M. Sameer, "A Novel Machine Learning Approach Using Boosting Algorithm for Liver Disease Classification," *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2021, pp. 1-5, doi: 10.1109/ISCON52037.2021.9702488.
- [12] N. Giannakeas *et al.*, "Measuring Steatosis in Liver Biopsies Using Machine Learning and Morphological Imaging," *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, Thessaloniki, Greece, 2017, pp. 40-44, doi: 10.1109/CBMS.2017.98.
- [13] A. Brankovic, A. Zamani and A. Abbosh, "Electromagnetic Based Fatty Liver Detection Using Machine Learning," *2019 13th European Conference on Antennas and Propagation (EuCAP)*, Krakow, Poland, 2019, pp. 1-3.
- [14] 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)*, Kuala Lumpur, Malaysia, 2021, pp. 1-7, doi: 10.1109/GUCON50781.2021.9573803.
- [15] K. Hamid, A. Asif, W. Abbasi, D. Sabih and F. -u. -A. A. Minhas, "Machine Learning with Abstention for Automated Liver Disease Diagnosis," *2017 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan, 2017, pp. 356-361, doi: 10.1109/FIT.2017.00070.
- [16] S. Panigrahi, R. Deo and E. A. Liechty, "A New Machine Learning-Based Complementary Approach for Screening of NAFLD (Hepatic Steatosis)," *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Mexico, 2021, pp. 2343-2346, doi: 10.1109/EMBC46164.2021.9629507.
- [17] X. Li, X. Chen and Z. Yuan, "Applicable model of liver disease detection based on the improved CART-AdaBoost algorithm," *2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, Dalian, China, 2021, pp. 1177-1181, doi: 10.1109/ICAICA52286.2021.9498046.

- [18] K. Srivastava, G. Malhotra, M. Chauhan and S. Jain, "Design of Novel Hybrid Model for Detection of Liver Cancer," *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, Greater Noida, India, 2020, pp. 623-628, doi: 10.1109/GUCON48875.2020.9231143.
- [19] D. S. Reddy, R. Bharath and P. Rajalakshmi, "A Novel Computer-Aided Diagnosis Framework Using Deep Learning for Classification of Fatty Liver Disease in Ultrasound Imaging," *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*, Ostrava, Czech Republic, 2018, pp. 1-5, doi: 10.1109/HealthCom.2018.8531118.
- [20] T. A. Joarder, B. Ahmed and A. S. Sattar, "Detecting Liver Cancer from Online Search Logs," *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, Khulna, Bangladesh, 2019, pp. 1-5, doi: 10.1109/EICT48899.2019.9068842.
- [21] A. Kalsoom, A. Moin, M. Maqsood, I. Mehmood and S. Rho, "An Efficient Liver Tumor Detection using Machine Learning," *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2020, pp. 706-711, doi: 10.1109/CSCI51800.2020.00130.
- [22] W. Cao, P. Yang, Z. Ming, S. Cai and J. Zhang, "An Improved Fuzziness based Random Vector Functional Link Network for Liver Disease Detection," *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, Baltimore, MD, USA, 2020, pp. 42-48, doi: 10.1109/BigDataSecurity-HPSC-IDS49724.2020.00019.