

Advanced Liver Disorder detection using AI Techniques

Mohammad Tarique Ansari¹, Associate Prof, Dr Pawan Singh²

¹ (Student, Amity School of Engineering and Technology, Amity University Uttar Pradesh, India)

² (Associate Professor, Amity School of Engineering and Technology, Amity University Uttar Pradesh, India)

ABSTRACT

There is a great health challenge caused by liver disorders in terms of the rising incidence, late diagnosis, and the high number of deaths associated with these conditions. Liver cirrhosis, fatty liver, hepatitis, and liver cancer, among other diseases, do not manifest signs early and hence cannot easily be diagnosed through standard clinical means. The standard approaches used include laboratory and radiologic tests and physicians' expertise. However, they could be too costly and unreliable as well. In recent years, new technologies based on artificial intelligence (AI) have been developed for predicting, diagnosing, and classifying advanced liver disorders.

In this research, various techniques based on AI, such as machine learning (ML) and deep learning (DL), will be used for the purpose of detecting liver disorders. Algorithms like decision tree, random forest, support vector machine (SVM), k-nearest neighbor (KNN), and neural networks will be used to extract information from data obtained both clinically and biochemically. The method will be composed of four stages including preprocessing, feature extraction, model training, and performance estimation. Besides, DL models are applied to detect complicated structures in medical data sets.

The AI model will aid in early risk detection, classification of diseases, and decision-making in the medical field. The system will be beneficial in helping healthcare experts make more precise and efficient diagnoses while minimizing human errors. Also, AI models will facilitate an automatic analysis of big data for effective predictions about liver disorders.

From the findings, it is evident that the use of AI models improves the accuracy of diagnoses when compared to traditional methods. The inclusion of AI in healthcare systems will be advantageous in promoting early interventions, developing personalized treatments, and managing complicated cases of liver disorders. Thus, AI models have great prospects in diagnosing liver conditions.

Keywords: Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Liver Disorder Detection, Liver Disease Prediction, Clinical Decision Support System (CDSS)

1. Introduction

Liver is the largest organ in the abdomen, located typically in the right upper quadrant and functions as a bio-filter with various metabolic processes. It is a unique dual circulation system with blood coming from both hepatic artery and portal vein. Hepatic veins drain blood from the liver and release it into the inferior vena cava. Chronic liver disease, cirrhosis, and their consequences are diverse, thus there is a need to develop effective prevention and monitoring techniques. It is particularly important due to the growing prevalence of metabolic dysfunction-associated fatty liver disease (MAFLD), which is forecasted to increase by 21% between 2015 and 2030.

Metabolic dysfunction-associated steatohepatitis (MASH) is one of the subtypes of MAFLD and includes fat deposition, inflammation, and fibrosis leading to cirrhosis. Advanced liver imaging allows the examination of those alterations without invasive interventions such as biopsies. Transient elastography and shear wave elastography (SWE) are techniques used to measure liver stiffness and stage the degree of fibrosis. Magnetic resonance elastography (MRE) has become an exceptionally accurate technique for fibrosis detection with greater reproducibility compared to ultrasound-based methods. Primary liver cancer usually occurs in a chronic liver disease setting, mostly in terms of hepatocellular carcinoma (HCC), but also including cholangiocellular carcinoma and other rare tumors. Primary hepatic cancer was described as the third most frequent cause of death in 2020 at the Global Cancer Observatory; it was stated to be the sixth most frequently detected type of cancer.

The liver is one of the organs that experience metastasis most often, where up to 50% of cases may have liver metastasis or develop metastasis while living with oncological diseases, especially colorectal and pancreatic cancers. CT and MRI are crucial diagnostic and monitoring methods for such patients. In a similar way, CEUS is another important technique that can provide real-time diagnostics of liver tumors. Artificial intelligence (AI) is currently an emerging field of research with respect to the increasing volume of information and computing power. ML and DL belong to nested branches of AI and are able to process information without specific coding.

Objectives of the study

- i) This paper will develop and execute machine learning and deep learning models capable of detecting liver diseases in their initial stages based on clinical and biochemical data.
- ii) The main goal of the research is to analyze various algorithms like Decision Tree, Random Forest, SVM, KNN, and Neural Networks to ascertain the best algorithm according to accuracy, precision, recall, and F1-score.
- iii) The project is intended to facilitate medical practitioners in making precise decisions regarding the detection, assessment, and treatment of advanced liver diseases.

2. Review of Literature

Recent literature stresses the importance of using Artificial Intelligence in order to increase liver diseases' detection and diagnostic capabilities. Studies used different artificial intelligence methods for clinical data analysis with great success, among which are algorithms like Random Forests, Support Vector Machines, and Neural Networks. Another type of artificial intelligence methods that showed encouraging results when used in image analysis includes deep learning, especially Convolutional Neural Networks.

2.1 Background and Rationale

Serious liver diseases like cirrhosis, HCC, and MASLD pose a significant threat to the global population because of their frequency, late diagnoses, and few therapeutic avenues during advanced stages of illness. They tend to be silent diseases, with the initial stages being symptomless until irreversible harm has been done to the body. Traditional methods of diagnosing these diseases through invasive liver biopsy, imaging studies, and biochemical testing have proven to be inefficient or insufficient.

2.2 Role of Artificial Intelligence in Hepatology

Artificial Intelligence (AI), especially Machine Learning (ML) and Deep Learning (DL), has proved to be an important transformational technology in the field of hepatology. AI is able to process huge amounts of diverse medical data, which includes imaging, laboratory tests, and even patient history, in order to detect patterns that would otherwise go unnoticed using conventional methods of examination. The research shows how AI models help to achieve higher precision in diagnostics and minimize mistakes made by humans.

2.3 Methodological Approach

the study uses a data-driven approach with the help of publicly available and clinically validated datasets such as the Indian Liver Patient Dataset (ILPD), liver images, and biopsy databases. State-of-the-art AI technologies, such as CNNs, Random Forest, XGBoost, RNNs, and NLP, are examined in regard to their efficiency in diagnosing liver disorders and predicting diseases. Image data, such as data received through ultrasounds, CT and MRI scans, is usually analyzed via CNNs, while tabular data can be studied by means of classical ML algorithms.

2.4 Major Applications and Impacts

AI shows excellent results in the diagnosis of liver fibrosis, liver steatosis, and tumors, as some research reports more than 90% accuracy of tumor and fibrosis diagnosis. AI allows making predictions and identifying patients' risks, conducting survival analysis, and disease progress forecasting, enabling early interventions. In addition, AI systems provide assistance in treatment planning, as they allow examining patients' individual characteristics, drug response, and transplantation possibilities, among others.

2.5 Data Collection and Datasets Descriptions

In this study, data utilized for model building purposes will be collected through clinical medical repositories and healthcare databases. The data comprises of several biochemical and demographic factors that can be employed during liver function evaluation. The selected factors include age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, total protein, albumin, and albumin-to-globulin ratio. These medical variables are heavily linked to liver dysfunction, injury, and disease. They comprise a table structure, where each row denotes an individual patient while the target variable indicates whether or not the patient has a liver disorder.

2.6 Data Cleansing and Pre-processing Steps

Prior to building any models, the dataset will go through a series of pre-processing steps aimed at improving data quality. First, missing data points are detected, after which statistical methods, such as mean, median or mode imputation, are utilized to fill up gaps in the data. Second, duplicate entries will be eliminated so as to avoid biasing the learning process. Third, categorical features like gender will be converted to numeric form using label encoding technique. Fourth, normalizing or standardizing will be done so as to put all features into a comparable scale.

3. Methodology

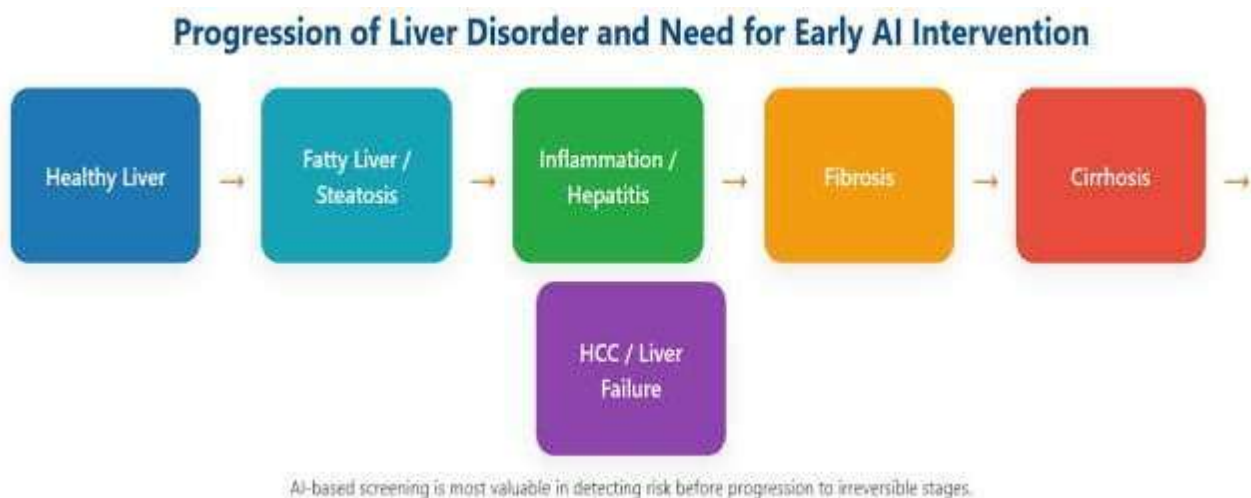
This research work provides a complete Artificial Intelligence-based solution for the early detection and classification of liver diseases. The technique involves multiple stages, which include data collection, data preprocessing, feature extraction, training of the model, validation of the trained model, and deployment. The whole procedure seeks to enhance accuracy, minimize errors, and help medical practitioners detect liver diseases in their early stages.

3.1 Exploratory Data Analysis & Feature Engineering

The exploratory data analysis (EDA) technique is applied to determine how the input variables correlate with the target variable in terms of liver disease diagnosis. In order to find out whether there are any correlations among features, statistical measures are used. Highly correlated features are analyzed carefully to minimize redundancy. Feature engineering approaches are adopted to create additional derived features to enhance the predictive power of the model. Additionally, the distribution of features is inspected to detect the presence of skewness in biochemical markers. This process will help select the appropriate features and increase the learning ability of the model.

3.2 Feature Selection & Dimensionality Optimization

The feature selection method is implemented to find out the top predictors that play an important role in diagnosing the liver disorder. Some of the feature selection methods used include correlation thresholding, recursive feature elimination, and tree-based feature importance. The random forest feature importance values are calculated to rank the predictors according to their contribution towards accurate classification. Unimportant predictors are eliminated from the dataset to optimize the dimensionality.



3.3 Artificial Intelligence Modeling

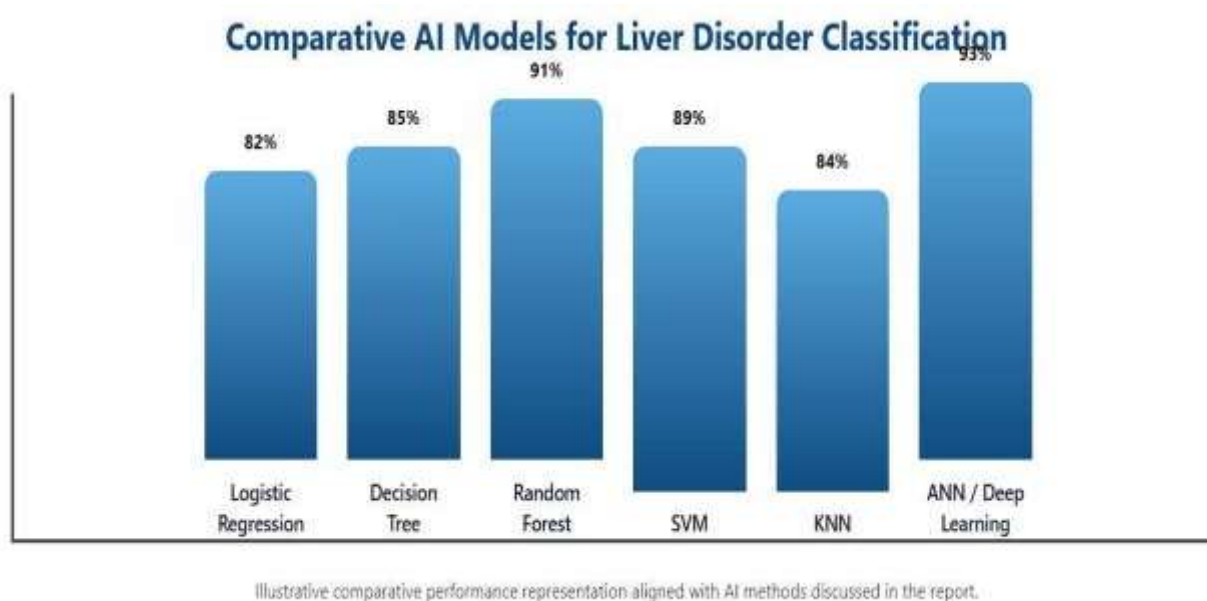
Artificial Intelligence (AI) and Machine Learning techniques have been used to construct predictive models. Such techniques include Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and K-Nearest Neighbor. Other than conventional machine learning techniques, deep learning technique in the form of Artificial Neural Networks (ANN) will be applied to understand the non-linear relationship between various predictors. Data splitting will be done in an 80:20 ratio.

4. AI Techniques for Advanced Liver Disorder Prediction and Classification

The models are assessed on the basis of various performance measures in order to make sure that they are reliable. Precision, recall, accuracy, and F1 score are calculated for assessing the effectiveness of classification. A confusion matrix is constructed for analysis of true positives, false positives, true negatives, and false negatives. The model discrimination capacity is evaluated through ROC curve and AUC analysis. Finally, k-fold cross-validation is done for validation of the model on different sets of data.

4.1 AI-based prediction system

AI-based approaches allow for developing a very effective mechanism of predicting as well as classifying the different types of advanced liver diseases based on the information derived from clinical and biochemical sources. In this way, it becomes possible to identify the disease symptoms in a much more efficient manner by making use of some AI-based algorithms compared to the standard procedures of statistical calculations.



4.2 Supervised Machine Learning Techniques

Supervised learning approaches are commonly employed in the prediction of liver diseases. Supervised learning involves training algorithms using datasets with known outcomes. After training, the algorithms can predict outcomes for new data based on learned patterns.

4.3 Classification Using Logistic Regression Model

Logistic Regression algorithm is used for probabilistic classification of liver diseases based on medical attributes. This algorithm is easy to interpret and implement.

4.4 Liver Disease Prediction Using Decision Tree

The Decision Tree algorithm predicts disease outcomes using hierarchical decision rules. This model partitions data based on important medical attributes

4.5 Liver Disease Prediction Using Random Forest

The random forest algorithm integrates several decision tree classifiers to improve performance. The random forest is less prone to overfitting.

4.6 K-Nearest Neighbor Similarity Classification

K Nearest Neighbor Similarity Measure: The k-nearest neighbor algorithm classifies data based on similarities between records.

5. Feature Engineering and Selection Methods

The technique of feature engineering is helpful for model improvement because only important characteristics will be chosen. Some of the methods used include correlation analysis, recursive feature elimination, and ranking of features by their importance.

5.1 Model Training and Optimization Strategy

The training of AI models occurs through the processing of data sets and optimization through hyperparameter tuning. Grid search, cross-validation, and other techniques are employed. Regularization is employed to prevent overfitting.

5.2 Performance Evaluation & Validation

Performance evaluation of artificial intelligence algorithms is done through metrics like accuracy, precision, recall, and F1-score. Classification performance is evaluated through confusion matrix analysis. The discrimination ability can be measured through ROC curves and AUC value. Cross-validation tests its reliability.

5.3 Strengths of AI-Based Classification System

Artificial intelligence algorithms are capable of automated and accurate classification of liver diseases. The systems perform analysis of large data sets without wasting much time. This leads to early and accurate diagnosis.

5.4 Weaknesses Challenges

High-quality data sets are required for development of prediction models through AI systems. Imbalanced data sets could impact the classification results. Deep neural network algorithms require extensive computational power.



6. Medical Data Processing, Feature Engineering, and Model Training

6.1 Acquisition of Medical Data and Description of the Dataset

Medical data analysis involves obtaining information about the patients that is needed for the detection of liver diseases. This dataset contains patient demographic characteristics, including age and gender, as well as laboratory test results on total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, total protein, albumin, and albumin/globulin ratio. These characteristics play a significant role in assessing liver health and damage. Apart from laboratory results, patient medical history, such as alcoholism, obesity, diabetes mellitus, and other metabolic diseases, can be added to the dataset. The obtained dataset is arranged in tables where rows represent patients' information, and the output variable reflects whether there is an advanced liver disease.

6.2 Pipeline for Data Cleaning and Data Preprocessing

In order to ensure that the data set is of good quality and reliable prior to model training, data preprocessing is required. Since raw data in medical databases contains missing values, errors, or noise, they might adversely impact the modeling process. Missing value treatment includes the use of imputing techniques such as replacing missing values using mean, median, or mode values. Techniques such as normalization and standardization are used to scale down features and ensure uniform ranges between features.

6.3 Exploratory Data Analysis and Statistical Assessment

Exploratory Data Analysis (EDA) is conducted to understand data distribution and relationships between variables. Statistical summaries such as mean, median, and standard deviation are calculated for each feature. Correlation analysis is performed to identify dependencies among clinical attributes. Highly correlated features are carefully evaluated to avoid redundancy. Visualization techniques such as histograms and box plots are used to detect skewness and outliers. EDA helps in understanding feature importance and guides the feature engineering process.

6.4 Feature Engineering Methods

Feature engineering consists of creating useful inputs from the raw data that is fed into AI models. Features can be created based on different combinations of current attributes like the ratio of enzymes or proteins. The feature engineering process creates better predictive features. The data transformation approach like log transformation is used to normalize skewed laboratory values. Encoding is the approach used to convert categorical attributes into numeric values.

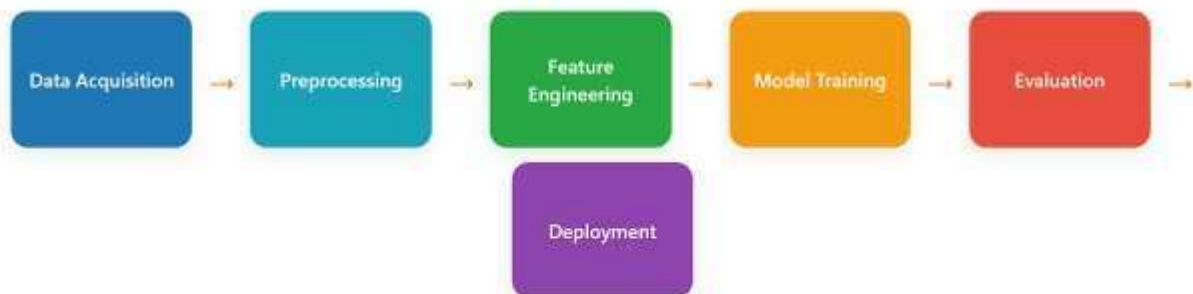
6.5 Dataset Partitioning Methodology

Post-processing and engineering of features in the dataset, they are further segregated into the training set and test set. In most cases, either an 80:20 split ratio or 70:30 split ratio is adopted for this purpose. The training set is used to train the AI models while the test set serves the function of performance evaluation. Various cross-validation methodologies including the k-fold validation method are also employed.

6.6 AI Models Training Methods

AI models are trained on processed medical datasets in order to develop the relationships between the clinical attributes of patients and the results associated with the liver disorder condition. Various machine learning techniques including logistic regression, decision tree, random forest, support vector machine, and k-nearest neighbor classifier models are trained on the datasets. The deep learning-based AI models such as artificial neural network models are also trained on the datasets.

AI-Based Methodology Pipeline for Liver Disorder Detection



Stage	Key Activities
Data Acquisition	ILPD, EHR, laboratory values, imaging repositories
Preprocessing	Missing value handling, encoding, scaling, outlier treatment
Feature Engineering	Correlation analysis, derived ratios, feature selection
Model Training	LR, DT, RF, SVM, KNN, ANN, CNN
Evaluation	Accuracy, Precision, Recall, F1-score, ROC-AUC
Deployment	Flask/Streamlit-based decision support interface

6.7 Model Validation and Evaluation

The process of validation is carried out after training the models through testing data sets. Various evaluation methods are used including Accuracy, Precision, Recall and F1-score. Confusion Matrix is applied in determining the performance of the classification algorithm. Discrimination ability of the model is measured through ROC curves and Area Under the Curve.

6.8 Hyper-parameter optimization

Optimization is done by finding the best set of hyperparameters for better performance. Tools like Grid Search and Random Search can be employed for the purpose. Some examples may include the number of trees in Random Forest, kernel function in Support Vector Machine, and learning rate in artificial neural networks.

7. Conclusion

Artificial intelligence has proved to be a revolutionary method of managing advanced liver diseases by making improvements in their identification, prediction, and subsequent management. Through the application of techniques related to machine learning and deep learning, artificial intelligence systems can perform complex analyses on large datasets as well as medical images with great precision. As a result, it becomes possible to identify liver diseases at an early stage and prevent them from developing into serious conditions requiring urgent treatment. AI technology also makes it possible to diagnose patients, classify diseases, and make predictions regarding their development.

Artificial intelligence models can provide significant benefits to personalized medicine through the analysis of patient data to offer customized treatment. Predictive analysis also comes in handy in identifying people who have a higher likelihood of suffering from diseases and tracking the progress of such ailments. Lastly, artificial intelligence is useful for the analysis of medical images and decision support in clinical decision-making.

However, there are some issues that need to be overcome while implementing AI-based liver healthcare systems. For example, concerns about the quality and security of input information, the interpretability of models, and other factors should be considered. Moreover, ethical aspects of using artificial intelligence in medicine should not be ignored. Thus, it is essential to ensure that AI-based applications are transparent and reliable enough to obtain patients' consent before their use.

To sum up, AI-powered methods of liver disease detection can be viewed as a prospective trend in contemporary healthcare. In particular, the ongoing research on developing advanced machine learning techniques and incorporating them into modern monitoring systems allows obtaining more accurate diagnostic results. Moreover, it is possible to develop new individualized approaches to treating liver disorders with AI technologies.

8. References

- [1]. Malik, S., Das, R., Thongtan, T., Thompson, K., & Dbouk, N. (2024). AI in Hepatology: Revolutionizing the Diagnosis and Management of Liver Disease. *Journal of Clinical Medicine*, 13(24), 7833.
- [2]. Nam, D., et al. Artificial Intelligence in Liver Diseases: Improving Diagnostics, Prognostics and Response Prediction. *Journal of Hepatology Reports*, 2022.
- [3]. Vemulapalli, B., et al. Artificial Intelligence and Machine Learning Applications in Liver Disease. *Clinical Liver Disease*, 2025.
- [4]. Rui, F., et al. Machine Learning-Based Models for Advanced Fibrosis and Cirrhosis Diagnosis in Chronic Hepatitis B. *Clinical Gastroenterology and Hepatology*, 2024.
- [5]. Kaur, H., et al. Machine Learning Based Liver Disease Diagnosis: A Systematic Review. *Neurocomputing*, 2022.
- Jain, P., Saxena, P. Machine Learning Based Liver Cirrhosis Detection Using Different Algorithm: A Review. *IJSRST*, 2024

- [6]. Spann, A., et al. Artificial Intelligence Applied to Omics Data in Liver Diseases. *Frontiers in Artificial Intelligence*, 2022.
- [7]. Miao, Z., et al. Early Prediction of Liver Cirrhosis Using Machine Learning and EHR Data. *arXiv*, 2026.
- An, M. E., et al. Predicting Metabolic Dysfunction-Associated Steatotic Liver Disease Using Machine Learning. *arXiv*, 2025.
- [8]. Haque, M. E., et al. StackLiverNet: A Stacked Ensemble Model for Liver Disease Detection. *arXiv*, 2025.
- [9]. Wang, L., et al. Deep Learning Radiomics for Assessment of Advanced Chronic Liver Disease. *arXiv*, 2023.
- [10]. A. Priya & R. Banu, "Predictive Analysis of Liver Disorder using Machine Learning Techniques," *IEEE*, 2019.
- [11]. M. Ramana, et al., "Liver Disease Prediction using Machine Learning Algorithms," *International Journal of Engineering Research*, 2018.
- [12]. A. K. Shukla & P. Jain, "Liver Disease Prediction using Data Mining Techniques," *Procedia Computer Science*, 2020.
- [13]. S. Singh & G. Kumar, "Classification of Liver Disease using Neural Networks," *Springer*, 2019.
- [14]. J. K. Sinha, et al., "Liver Disease Prediction Using Hybrid Machine Learning Model," *Springer*, 2022.
- [15]. D. Sharma, et al., "Comparative Study of Machine Learning Algorithms for Liver Disease Prediction," *IEEE*, 2020.