Volume (1) No (9), 2025
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Review Paper/Research Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

Artificial Intelligence Approaches for Diagnosis and Continuous Monitoring of Chronic Liver Disorders

Mr. Sachin Manekar
Assistant Professor
Department of Computer Science and Application
Mandsaur University, Mandsaur
Sachin.manekar@meu.edu.in

***Abstract*—Chronic liver disease, or CLD, is becoming a bigger problem in the world's health care system because it is hard to notice when it starts and is fatal in its later stages. It is difficult to detect the disease at an early stage since conventional diagnostic procedures including imaging, biopsies, and liver function tests are intrusive, expensive, or not widely available. This highlights the need for accurate, non-invasive, and data-driven approaches to support timely intervention and reduce clinical risks. Machine learning (ML) has shown promise in processing large, heterogeneous datasets to uncover hidden patterns and enhance predictive performance. Using the Liver Disease Patient Dataset from the UCI Repository, comprising 30,691 records with 11 attributes, this study applied extensive preprocessing, including missing value imputation, outlier removal with the IQR method, Min–Max normalization, and SMOTE for class balancing. Feature selection was employed to improve interpretability and efficiency. A model called Gradient Boosting (GB) was created and tested against many other methods, including SVM, Random Forest, MLP, and XGBoost. Surpassing baseline models, GB attained the best performance with a 98.60% ROC-AUC, 98.50% recall, precision, and F1-score. Based on these findings, ensemble approaches are reliable for predicting early CLD. Improving practicality will be the goal of future studies that investigate clinical validation and the integration of multimodal data.**

Keywords—Chronic Liver Disease (CLD), Machine Learning, Gradient Boosting, Healthcare Monitoring, Early Diagnosis, Clinical Decision Support.

I. INTRODUCTION

The human body's largest organ is the liver. All metabolic processes, including the transformation of dietary materials into usable compounds, their storage, and their eventual delivery to cells as required, are carried out by it [1]. Generation of bile, proteins, glucose storage and release, processing of haemoglobin, detoxification of blood, generation of immunological factors, and clearance of bilirubin are all essential tasks. Because of this, the liver is crucial to good health in general. But many people fail to prioritize their liver health. Liver problems, from moderate to severe, affect a large section of the world's population as a result of unhealthy lifestyle choices [2].

Chronic liver disease (CLD) includes hepatitis, cirrhosis, and non-alcoholic fatty liver disease (NAFLD), and it is a major public health concern since it contributes to the yearly deaths of millions of people [3][4]. Late diagnosis, complicated disease progression, and significant morbidity and mortality are clinical problems of CLD. To intervene quickly, it is essential to understand how liver disease develops. The American Liver Foundation states that inflammation is the usual starting point for liver disease, which can lead to the liver swelling up to its normal size. Fibrosis, scarring caused by inflammation, can develop as a result of this. Cirrhosis is a severe scarring stage that can develop from fibrosis if left untreated. It can lead to liver cancer or failure [5][6]. The liver's capacity to repair itself and carry out its vital tasks can be severely compromised by advanced fibrosis or cirrhosis.

Diagnostic procedures that have been around for a while include ultrasound, MRI, LFTs, and liver biopsies. Although imaging allows for a visual evaluation of liver morphology and the detection of anomalies, LFTs evaluate liver function by testing enzyme levels and other chemicals in the blood. Liver biopsies are invasive, expensive, and generally unavailable due to resource constraints; they include removing and analysing liver tissue under a microscope [7]. Early detection and ongoing monitoring of CLD are becoming increasingly important due to the limits of conventional diagnostics. Modern healthcare generates vast amounts of heterogeneous data, including lab results, imaging studies, genetic markers, and lifestyle information, which are challenging for clinicians to integrate manually[8]. Diagnostic precision, workflow, and individualized treatment plans can all be enhanced by data-driven methods, especially ML.

Machine learning is an AI subfield that focuses on improving the comprehension and analysis of massive datasets via the application of mathematical and statistical methods [9][10][11]. Predicting diseases, patient classification, and creating individualized treatment programs are all areas where it has found use in healthcare. When it comes to crunching numbers, finding trends, and predicting outcomes with pinpoint accuracy, ML models are unrivalled [12]. ML has the potential to be useful in medical diagnostics and risk assessment of chronic diseases, and successful cases have been reported in diabetes, heart disease, and cancer. Machine learning can be used with multimodal clinical information, including patient records, imaging, lab results, and genomic data, to provide a comprehensive view of liver disease [13][14]. Machine learning models trained on such datasets can aid in early diagnosis, predict disease progression, track patient response to treatment, and assess the risk of adverse consequences [15][16]. The approach enables healthcare professionals to identify individuals at risk, diagnose more effectively, prescribe interventions promptly, and ultimately improve patient outcomes. The use of ML in CLD management can be a great opportunity to address the classical constraints of diagnosis and help to provide personalized care.

A. Motivation and Contribution

The fact that chronic liver disease (CLD) frequently goes undiagnosed until it has progressed to a severe stage adds insult to injury: CLD is a growing world health concern. This, in turn, spurred this research. Timely intervention and subsequent patient outcomes, as well as a reduction in healthcare burdens, depend on early diagnosis. Liver biopsy and imaging are examples of traditional diagnostic procedures; nevertheless, they can be intrusive, costly, and time-consuming, which limits their accessibility. The growing availability of medical datasets, combined with advancements in machine learning, offers an opportunity to develop automated, accurate, and non-invasive predictive models. Early identification, continuous monitoring, and informed decision-making can all be facilitated by these technologies, which in turn can enhance patient care and prognosis. Recent studies have shown promising results in the fight against liver disease:

- The UCI Repository's Liver Disease Patient Dataset was used for this purpose; it has 30,691 records with 11 clinical variables, making it a diversified and large-scale dataset, ideal for predictive modelling.
- Data quality and model resilience were enhanced through extensive pre-processing, which included IQR outlier identification, Min-Max normalization, SMOTE for class balance, and missing value imputation.
- Feature selection to preserve the most important clinical variables enhanced computational performance, decreased dimensionality, and increased model interpretability.
- Improved prediction performance by using a Gradient Boosting (GB) model for accurate liver disease classification, which leverages its ensemble learning capabilities.
- A comprehensive evaluation of the reliability of the predictions was ensured by using the following metrics: F1-score, accuracy, precision, recall, and area under the curve (AUC-ROC).
- Developed a non-invasive data-driven prediction system to assist with chronic liver disease management care decision-making, monitoring, and early detection.

B. Justification and Novelty

The proposed study is based on the idea that, because of the limits of current diagnostic procedures, there is an urgent need to create efficient, non-invasive ways to diagnose and monitor chronic liver disease in its early stages. This study is novel because it combines extensive pre-processing methods, feature selection, and the Gradient Boosting model to process immense, skewed clinical data effectively. One advantage of this methodology compared to the traditional one is that it can maximize predictive performance and can also be interpreted and scaled to a typical clinical scenario. The resulting system provides a stable, evidence-based framework to enhance early diagnosis, ongoing monitoring, and support clinical decision-making.

C. Organization of the Paper

This paper is structured in the following way: Section II is a review of related studies. Section III outlines the recommended methodology, including data preprocessing, model implementation, and evaluation measures. In Section IV, the results of the experimental results and comparison with the existing methods are provided. Lastly, Section V provides a summary of the study, highlights the major findings, and outlines potential future research directions.

II. Literature Review

The current review shows recent progress in ML and DL in diagnosing and monitoring chronic liver disease, with a focus on predictive models, biomarkers, and ensemble-based methods on both clinical and imaging data.

Syed et al. (2025) trained various models—including CNN, SVM, LR, and KNN—using ILPD resources on Kaggle to improve liver disease diagnosis. Top performance was achieved by the CNN model, which achieved an accuracy rate of 96.21, precision rate of 74.76, recall rate of 92.77, and F1-score rate of 82.80. The paper identifies the promise of Deep Learning algorithms, specifically CNN, to be more predictive than Machine Learning algorithms [17].

Maurya et al. (2025) conducted research to improve liver disease diagnosis. They created a predictor model based on multiple clinical data sets, achieving an astonishing accuracy of 83% on the test set through hyperparameter optimisation of logistic regression. This method outperforms the conventional diagnostic approach for identifying liver-related conditions in a non-invasive manner, and, more importantly, a cost-effective and reliable solution has been achieved. Moreover, it is a good way to detect people at higher risk of progressive liver diseases, such as fibrosis or organ damage [18]

Hossain Shaon et al. (2024) developed StackLD using the stacking-ensemble machine learning approach. Seven highly effective models were quickly created after they harmonized with the ILPD dataset. The stacking method was found to be more accurate, sensitive, specific, and had a larger area under the curve (0.8622, 0.8933, 0.8369, and 0.9275, respectively). This is a useful method of separation of positive and negative classes in independent test procedures [19].

Kumar et al. (2024) created a StackLD architecture based on a machine learning stacking-ensemble method. They have utilized the dataset of Indian Liver Patient Dataset (ILPD) to correct the mismatches and have made seven powerful models. Results of 0.8622 for accuracy, 0.8933 for sensitivity, 0.8369 for specificity, and 0.9275 for the area under the curve all indicate the stacking method's superior performance. This method successfully separates positive and negative classes in independent test approaches, based on a framework of more than 1,600 patients, to forecast subtyping of liver diseases and prognosis using CNNs, RNNs, and LSTMs. The method enabled analysis of structural abnormalities, analysis of multimodal data sources and analysis of sequential clinical data. The findings were correct, and CNN with RNN performed the best at 97.8% and CNN with LSTM at 94.5%. The model's viability and accuracy in hepatology were enabled by the integration of multiple data sources [20]

P, S S and D (2023) exploring the potential for creating supervised ML systems to forecast liver failure. Instead, they utilised data visualisation and univariate and bivariate analysis as part of the pre-processing steps to gain insight into the dataset's properties. They compared their performance metrics of recall, F1 score, and accuracy after developing a multi-classification model using machine learning methods. By accurately predicting cases of liver failure, the results demonstrated the potential medical applications of machine learning algorithms. The target algorithm achieved an accuracy of 94.48, prompting researchers to develop more complex models [21].

Minnoor and Baths (2022) studied the performance of several ML models in detecting and grading liver disease. One example of a model was the LR model. Other models included the Extra Tree, LightGBM, and KNN. Findings also indicate that current approaches are intrusive and time-consuming, with the shortage of qualified experts exacerbating the situation. The 11 features utilised to train the models yielded a peak accuracy of 0.89 and an F1 score of 0.88; hence, the Extra Trees classifier showed superior accuracy and speed in predicting liver disease from blood enzyme levels. [22].

The Table I summarizes key studies on machine learning for chronic liver disease, highlighting methodologies, performance, limitations, and future work to guide improvements in diagnosis, monitoring, and clinical applicability.

TABLE I. REVIEW OF MACHINE LEARNING APPROACHES FOR CHRONIC LIVER DISEASE DIAGNOSIS AND MONITORING

Reference	Methodology	Dataset	Performance	Limitations	Future Work
Syed et al. (2025)	CNN, SVM, LR, KNN	ILPD, Kaggle	CNN: Accuracy 96.21%, Precision 74.76%, Recall 92.77%, F1-score 82.80	Small, imbalanced dataset; no external validation	Use larger, diverse datasets and integrate multimodal data (imaging + clinical)
Maurya et al. (2025)	Logistic Regression (hyperparameter tuned)	Clinical records	Accuracy 83%	Moderate accuracy; lacks multimodal data	Apply deep learning or ensemble methods and include imaging/sequential patient data.
Hossain Shaon et al. (2024)	Stacking ensemble (XGB, LGBM, DT, KNN, RF)	ILPD (SMOTE)	Accuracy 86.22%, Sensitivity 89.33%, Specificity 83.69%, AUC 92.75%	Synthetic balancing may not reflect real-world data	Validate models on real-world clinical datasets and larger cohorts
Kumar et al. (2024)	CNN + RNN + LSTM hybrid	Clinical + imaging + sequential data	CNN+RNN: 97.8%, CNN+LSTM: 94.5%	Limited demographic diversity; single-centre data	Test across diverse populations and develop multimodal prognosis models
P, S S and D (2023)	Multi-class ML classification	Pre-processed clinical dataset	Accuracy 94.48%, F1-score and recall comparable	Specific algorithms not clearly detailed; no deployment	Implement models in clinical workflows for practical validation
Minnoor and Baths (2022)	Logistic Regression, KNN, Extra Trees, LightGBM, MLP	Blood enzyme dataset (11 attributes)	Extra Trees: Accuracy 89%, F1-score 0.88	Small, non-diverse dataset; no imaging integration	Combine biomarkers with multimodal data to improve predictive accuracy

A. *Research Gap*

Existing studies on the diagnosis and monitoring of chronic liver disease demonstrate promising results using machine learning and deep learning techniques; however, critical gaps remain. Most approaches rely on limited, single-source datasets, restricting generalizability across populations. Challenges such as multimodal data integration, longitudinal patient monitoring, and real-world deployment are underexplored. Current research often uses either biochemical, imaging, or sequential clinical data, neglecting comprehensive integration. Furthermore, external validation, interpretability, and the implementation of ensemble or hybrid models in clinical workflows are insufficiently addressed. Building reliable, multi-modal, and therapeutically deployable prediction systems should be the goal of future research.

III. RESEARCH METHODOLOGY

The suggested approach for using ML for Chronic Liver Disease Diagnosis and Monitoring includes a well-organised pipeline to guarantee precise predictions, as shown in Figure 1. Retrieve the UCI repository's dataset of individuals with liver disease first. Among the 30,691 patient records with 11 variables included in the sample, 21,917 patients were found to have liver illness. Initial exploratory analysis used bar plots and correlation heatmaps to examine feature relationships and class imbalance. The preprocessing steps included filling in missing values, detecting outliers using the IQR approach, and normalizing the features to a range of 0 to 1 using the min-max methodology. Applying SMOTE helped level the playing field between the diseased and non-diseased groups, leading to a more representative sample. Feature selection is carried out to identify the most effective predictors. Next, the dataset was divided into two parts: the training set and the testing set. This would guarantee accurate results when evaluating the models and making predictions. The categorization data is then used to train a Gradient Boosting (GB) model. Metrics used to quantify the model's performance include AUC-ROC, F1-score, recall, accuracy, precision, and precision. The end product of this pipeline is a dependable system for predicting liver illness. This system can facilitate early diagnosis and continuous monitoring, ultimately enhancing patient outcomes and informing clinical decision-making.

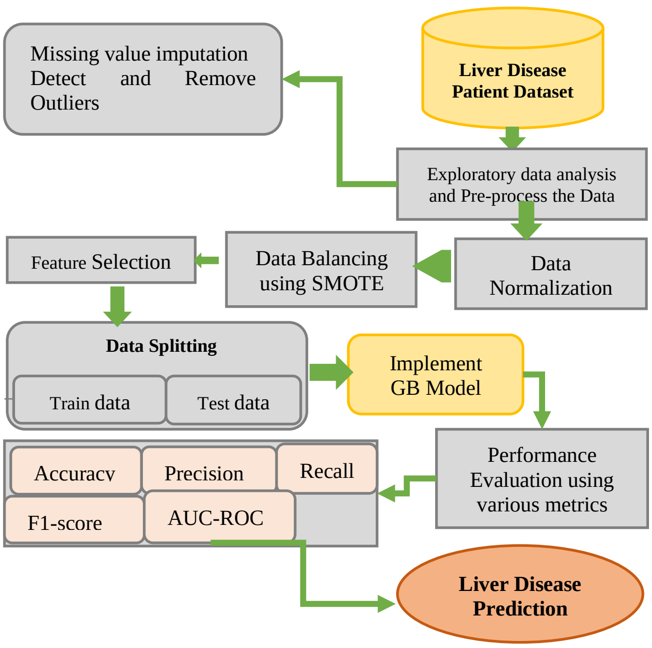

Fig. 1. Proposed Flowchart for Chronic Liver Disease Prediction

A. Data Collection

Data from patients with liver disease worldwide are housed in the Liver Disease Patient Dataset, which can be accessed openly at the UCI ML repository. Out of a total of 30,691 individuals recorded in this dataset, 21,917 were found to have liver disease, while the remaining 8,774 were healthy. Every record in the collection has eleven different properties. A predicate is the eleventh attribute, while a target attribute is the twelfth. There are two categories: attributes, five decimal attributes, and four integer attributes.

B. Exploratory Data Analysis

EDA was conducted to identify underlying patterns and relationships among variables in the liver disease dataset. It highlights the influence, along with clinical features, on the occurrence of liver disease, while also assessing the distributions of demographic and behavioural factors. The EDA for the dataset is presented in Figure 2 below.

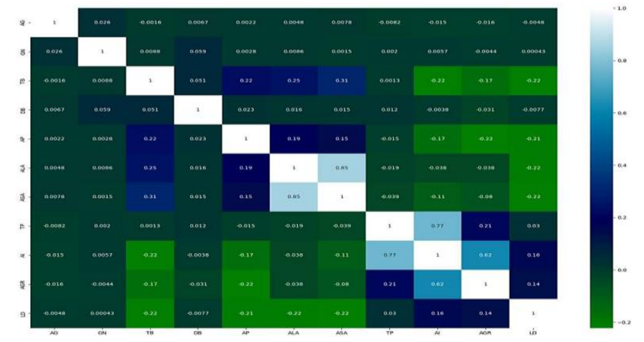

Fig. 2. Correlation Heatmap of Liver Disease Patient Features

Figure 2 shows the correlation matrix of clinical characteristics for the Liver Disease Patient Dataset. The direction and strength of feature-to-feature connections are graphically shown in the heatmap, where values range from -1 to 1. Positive correlations are shown in shades of green, while negative correlations appear in blue. Diagonal elements display perfect correlation (1.0) of features with themselves. Among the observed correlations, the strongest positive correlations are between TP (Total Protein) and A/G (Albumin and Globulin Ratio) (0.77), TP and Al (Albumin) (0.77), and Al and A/G (0.62). The strongest negative correlation is observed between LD (Liver Disease) and A/G (-0.25). This visualization aids in identifying potential interactions, multicollinearity, and informs feature selection for predictive model development.

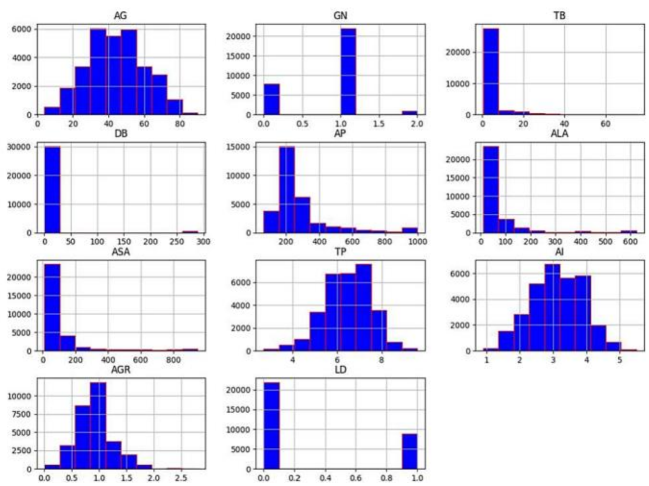

Fig. 3. Histogram of Dataset Attributes

The distribution of individual features in the Liver Disease Patient Dataset was shown in Figure 3, which is a set of histograms. Albumin (AL), albumin and globulin ratio (AGR), total proteins (TP), gender (GN), total bilirubin (TB), direct bilirubin (DB), alkaline phosphatase (AP), alanine aminotransferase (ALA), aspartate aminotransferase (ASA), total proteins (TP), and the liver disease class label (LD) are all displayed in separate subplots. The majority of characteristics, including TB, DB, AP, ALA, and ASA, have a high concentration of lower values and are right-skewed, suggesting that they do not follow a normal distribution. In contrast, Age, TP, AL, and AGR appear to be more normally distributed. The class label LD shows a clear imbalance, with a majority of instances belonging to one class, highlighting the need for techniques to address class imbalance during model training.

C. Data Pre-Processing

Preparing the data to construct a robust and dependable system is crucial before implementing ML approaches into the model. Different data preparation challenges were addressed in this study using various methods. The dataset was pre-processed using missing value imputation and Detection and Removal of Outliers. Preprocessing, the data transformation and normalization, is carried out. The following steps of pre-processing are as follows:

- **Missing value imputation:** Improving data representation, stability, and reducing bias are all goals of missing value imputation. It also computes the proportion of empty attributes by using isnull() to locate missing values.
- **Detect and Remove Outliers:** The method entails searching the dataset for data points that stand out significantly from the others. Prediction models are vulnerable to errors if they are not handled appropriately. To highlight any dataset outliers, used the IQR method and placed the feature cutoff at three points.

D. Data Normalization

The records were normalized using the min-max technique, which limits values to a range of 0 to 1. An increase in classifier efficiency and a decrease in the effect of outliers were the intended outcomes of this [23]. Or this normalization, consulted the following mathematical Equation (1):

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where X is the initial feature value, X' is its normalized value, X_{min}Is its lowest value, and X_{max}Is its highest.

E. Data Balancing using SMOTE

Machine learning model performance can be adversely affected by class imbalances in datasets; data balancing approaches are essential for rectifying this issue, especially in classification tasks where certain classes are underrepresented. The SMOTE method is effective because it enhances model stability and accuracy by generating new samples for the minority group, rather than simply replicating existing ones [24]. The distribution of patients in the Liver disease and no liver disease groups before and after SMOTE administration is shown in Figure 4. The dataset before SMOTE included 21,917 people with liver disease and 8,774 without liver disease. Following SMOTE, the distribution became almost balanced, with 13,050 and 13,161 patients in the two groups, which demonstrates how SMOTE is effective in reducing bias in the class and providing a more accurate model for training.

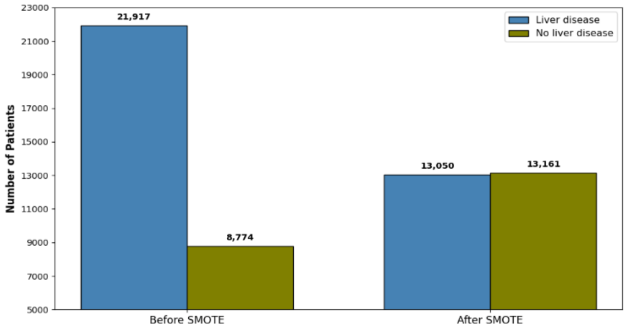

Fig. 4. Patient Distribution Before and After SMOTE

F. Feature Selection

The goal of variable or attribute selection, often called feature selection, is to build effective predictive models by picking a subset of important features. It is based on the assumption that datasets frequently have hidden or redundant features - redundant features do not add new information not already represented. Still, irrelevant features add little or no predictive information. In addition to enhancing model performance through complexity and overfitting reduction, feature selection also provides sufficient information about which features hold the most power, as well as their relationships with each other. Figure 5 shows a horizontal bar chart that summarizes the feature important scores of a Gradient Boosting model. The features are listed on the y-axis and the x-axis represents the importance score. AG, ALA, DB, and AP. The chart shows that the strongest feature is DB, followed by AP, and then ALA. The other features, such as AG and GN, have a small effect on model predictions. This visualization once again justifies the role of feature selection as a tool to illuminate important variables and make the model more intelligible and predictive.

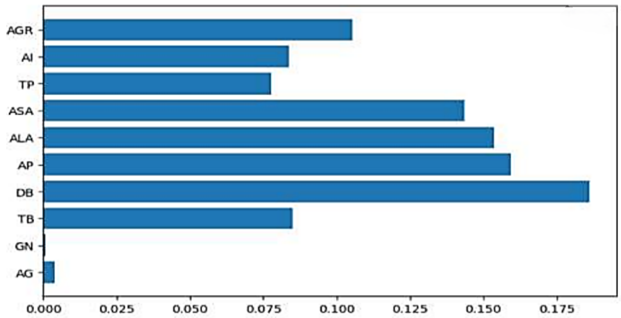

Fig. 5. Feature importance Plot

G. Data Splitting

Data was divided into training and testing samples to evaluate the model's performance and generalizability. To be more precise, the model was trained using 80% of the data in order to discover patterns and optimize its parameters, and its predicted accuracy on new data was evaluated objectively using the remaining 20%.

H. Proposed Gradient Boosting Model(GB)

The ensemble technique known as "Gradient Boosting" leverages the interdependence of base estimators. Using the errors of the previous iteration's base estimator as a learning tool, this algorithm trains a new one [25]. To merge weak learners at each stage to "boost" performance and produce a strong learner, this method is known as boosting. Three primary aspects are taken into account by GB to ensure optimal performance: the loss function, a weak learner, and an additive model to minimize the loss function in conjunction with the weak learner [26]. Step one is to write the formula as it appears in Equation (2).

$$F(x) = \sum_{m=0}^{M} f_m(x) \tag{2}$$

The optimization approach determines the sequential incremental functions, also called "boosts," where $f_0(x)$ is an initial estimate and $\{f_m(x)\}_1^M$. One of the most popular optimisation approaches for the loss function is steepest descent. As mentioned in Equation (3), it specifies the increment of $\{f_m(x)\}_1^M$.

$$g_m(x) = \left\{\left[\frac{\partial\emptyset(F(x))}{\partial F(x)}\right] F(x) = F_m - 1(x)\right\} \tag{3}$$

Where $F_{m-1} = \sum_{i=0}^{m-1} f_i(x)$, Then, the line search multiplier ρm is found in the same way as in Equations (4) and (5).

$$f_m(x) = -\rho_m g_m(x) \tag{4}$$

$$\rho_m = arg \min_{\rho} \emptyset[F(x) = F_m - 1(x) - \rho_m g_m(x) \tag{5}$$

Here, the negative gradient −gm(x) defines the **steepest descent direction**, ensuring that the algorithm minimizes the loss function, improves performance, and reduces the risk of overfitting.

I. Evaluation Metrics

The accuracy, precision, recall, and F1-score metrics were used to measure performance. These metrics were derived from the total number of correct and incorrect classifications. This information is inferred from the confusion matrix, which includes TP, TN, FP, and FN, which stand for true positives and true negatives, respectively. Further definitions of Accuracy, Precision, F1-Score, and Recall can be found below:

Accuracy: The trained model's accuracy as a percentage of the total occurrences in the dataset (input samples) is calculated using Equation (6)-

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \tag{6}$$

Precision: The ratio of a model's true positives to its overall number of true positives is a measure of its prediction accuracy. Accuracy shows. As Equation (7), shows how well the classifier predicts the positive classes-

$$Precision = \frac{\text{TP}}{\text{TP+FP}} \tag{7}$$

Recall: This statistic represents the proportion of positive events that were really predicted out of all the cases that were expected to be positive. It can be expressed mathematically as Equation (8)-

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

F1 score: It aids in maintaining a healthy equilibrium between recall and precision by combining the two concepts of the harmonic mean. Its range is [0, 1]. It is mathematically expressed as Equation (9)-

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

ROC-AUC Curve: The ROC curve is a popular tool for assessing the efficacy of classification models across a range of threshold values. The AUC is a probabilistic curve that measures the model's discriminatory strength. A better discriminator will have a higher AUC. To create a ROC curve, one needs to use Equations (10) and (11) to draw a straight line that connects the True Positive rate (TPR) and the False Positive rate (FPR). The Y-axis shows the TPR, and the X-axis shows the FPR.

$$FPR = \frac{FP}{TN+FP} \quad (10)$$

$$TPR = \frac{TP}{TP+FN} \quad (11)$$

A thorough understanding of the model's overall classification efficacy can be obtained by combining these evaluation metrics.

IV. Results and Discussion

This part describes the research design and the performance of the proposed Gradient Boosting model. The model was generated and tested on a high-performance computing platform with an Intel Core i9-10900 K processor, 3.70 GHz, 64 GB of DDR4 random-access memory, a 500 GB NVMe SSD, and 2 TB of disk memory, running Windows 11 Pro on a Jupyter notebook. The system provides effective data processing and computing power, supporting the development and testing of robust models. As it is observed in Table II, the performance of the model can be summarized as follows: The model reached an accuracy of 98.80, precision of 98.50, recall of 98.50, F1-score of 98.50, and ROC-AUC of 98.60, which is relatively high in its ability to classify patient data. Those findings indicate the model's effectiveness in detecting actual positive cases and reducing false positives, thereby demonstrating its power and overall efficacy in forecasting chronic liver disease.

TABLE II. Experiment Results of Proposed Models for Chronic Liver Disease Prediction on Liver Disease Patient Dataset

Performance matrix	Gradient Boosting
Accuracy	98.80
Precision	98.50
Recall	98.50
F1-Score	98.50
ROC-AUC	98.60

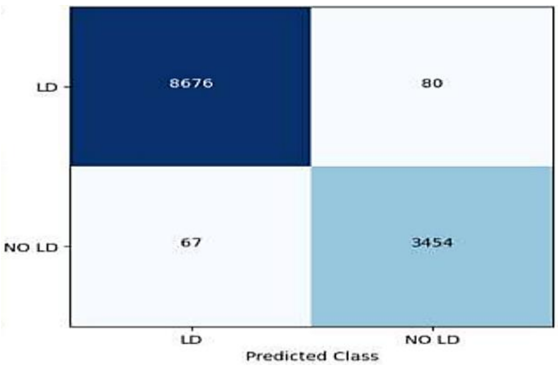

Fig. 6. Confusion Matrix of the Gradient Boost Model for Liver Disease Prediction

Figure 6 shows the GB model's confusion matrix for LD prediction; the model works as expected. The matrix indicates that of all cases 8676 were rightfully found to have LD (true positives), 3454 were rightly found to not have LD (true negatives). This indicates that the accuracy of classifying positive and negative cases is high. With only 80 false negatives (LD cases wrongly named as not having the disease) and 67 false positives (LD cases wrongly branded as having the disease), the model proved to be highly accurate. The GB model appears to be highly accurate and trustworthy in predicting liver illness, based on the minimal number of false positives and false negatives.

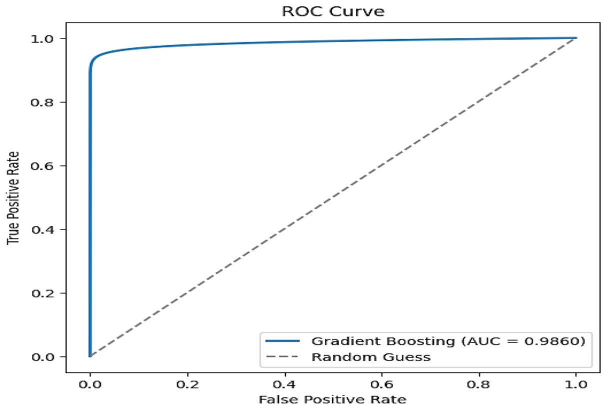

Fig. 7. ROC Curve of the Gradient Boost Model for Liver Disease Prediction

Figure 7 shows a ROC curve depicting a GB model used for the prediction of liver disease. See how various threshold values affect the TPR and FPR in this graph. The solid blue curve of the model remains far higher than the diagonal dashed line, which stands for a random guess. Differentiating between individuals with and without liver illness is certainly a strong suit of the model, as evidenced by an AUC of 0.9860, which is quite close to 1. A high area under the curve (AUC) indicates that this predictive model is robust and accurate.

A. *Comparative Analysis*

In this section, the comparison of various ML models in chronic liver disease prediction is given, which is summarized in Table III. Of all the models compared, SVM achieved an accuracy of 71.35%, while RF performed better at 87%. The MLP has shown a significant improvement in its accuracy to 96.43%. XGBoost (XGB) then improved predictive precision to 92.07%. The GB model outperformed the competition, demonstrating its superior performance and reliability for patient data classification with an impressive accuracy rate of 98.80%. Table III explicitly shows the incremental accuracy improvement with models and the merit of ensemble-based approaches, particularly Gradient Boosting, in achieving effective and accurate chronic liver disease prediction.

TABLE III. Accuracy Comparison of different Predictive models of Chronic Liver Disease prediction using the Liver Disease Patient Dataset

Models	Accuracy
SVM[27]	71.35
RF[28]	87
MLP[29]	96.43
XGBoost[30]	92.07
GB	98.80

The primary advantage of the proposed Gradient Boosting (GB) model is that it is highly predictive and capable of diagnosing chronic liver disease. The GB model can be used effectively to learn the complex non-linear patterns and interactions of the dataset by sequentially incorporating several weak learners, which is two times superior to conventional machine learning algorithms. It also demonstrates a good level of accuracy and recall, reducing false positives and false negatives, which is crucial in clinical decision-making. Additionally, the ability to work with asymmetrical and heterogeneous data, along with the model's flexibility, allows it to be applied in the medical field on a large scale.

V. CONCLUSION AND FUTURE STUDY

Machine learning holds significant promise for improving CLD diagnosis and surveillance. Using extensive preprocessing methods, feature elimination, and data balancing of the UCI Liver Disease Patient Dataset, a GB model was constructed that achieved high performance, with 98.80% accuracy, 98.50% precision, recall, and F1-score, and an AUC of 98.60. These results suggest that ensemble-based models can generate more complex data patterns than traditional models, such as SVM, RF, and XGBoost. The suggested solution will overcome the drawbacks of invasive and expensive diagnostic methods, providing a scalable, non-invasive, and effective framework to assist in clinical decision-making, early intervention, and continuous patient monitoring. However, there are still several challenges. The majority of existing datasets lack diversity and consist of single-source clinical data, limiting their ability to generalise to the population. Moreover, the actual application and clinical interpretability of such models are not studied.

Future research should consider using multimodal data, including imaging data, biochemical data, genomic data and sequential patient data to increase predictive ability. The implementation of explainable AI will also increase usability and trust among healthcare experts. Last but not least, clinical trials and the ability to integrate them into hospital decision-support systems are the key to transforming predictive models into practical applications that enhance patient outcomes and advance precision medicine.

REFERENCES

[1] W. Książek, M. Abdar, U. R. Acharya, and P. Pławiak, "A novel machine learning approach for early detection of hepatocellular carcinoma patients," *Cogn. Syst. Res.*, vol. 54, pp. 116–127, May 2019, doi: 10.1016/j.cogsys.2018.12.001.

[2] M. A. Uddin, A. Stranieri, I. Gondal, and V. Balasubramanian, "Rapid health data repository allocation using predictive machine learning," *Health Informatics J.*, 2020, doi: 10.1177/1460458220957486.

[3] D. A. Ofori, P. Anjarwalla, and L. Mwaura, "Asuhan Keperawatan Pada Pasien Dengan Gagal Ginjal Kronis Yang Di Rawat Di Rumah Sakit," *Molecules*, 2020.

[4] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 7, pp. 8459–8486, Jul. 2023, doi: 10.1007/s12652-021-03612-z.

[5] S. Pandya, "Predictive Modelling for Cancer Detection Based on Machine Learning Algorithms and AI in the Healthcare Sector," *TIJER – Int. Res. J.*, vol. 11, no. 12, 2024.

[6] M. A. Mostafiz, "Machine Learning for Early Cancer Detection and Classification : AI-Based Medical Imaging Analysis in Healthcare," *Int. J. Curr. Eng. Technol.*, vol. 15, no. 3, pp. 251–260, 2025, doi: https://doi.org/10.14741/ijcet/v.15.3.7.

[7] S. Tokala *et al.*, "Liver Disease Prediction and Classification using Machine Learning Techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 2, 2023, doi: 10.14569/IJACSA.2023.0140299.

[8] R. Dattangire, D. Biradar, L. Dewangan, and A. Joon, "Unlocking Healthcare Fraud Detection Using Innovations of Machine Learning Strategies," in *International Conference on Data Science and Big Data Analysis*, Springer Nature Singapore, 2025, pp. 233–249. doi: 10.1007/978-981-97-9855-1_16.

[9] R. Dattangire, R. Vaidya, D. Biradar, and A. Joon, "Exploring the Tangible Impact of Artificial Intelligence and Machine Learning: Bridging the Gap between Hype and Reality," in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, IEEE, 2024, pp. 1–6. doi: 10.1109/ACET61898.2024.10730334.

[10] R. Q. Majumder, "Machine Learning for Predictive Analytics : Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, 2025.

[11] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[12] G. Shaheamlung, H. Kaur, and M. Kaur, "A Survey on machine learning techniques for the diagnosis of liver disease," in *Proceedings of International Conference on Intelligent Engineering and Management, ICIEM 2020*, 2020. doi: 10.1109/ICIEM48762.2020.9160097.

[13] S. A. Pahune, "How does AI help in Rural Development in Healthcare Domain: A Short Survey," *IJRASET*, vol. 11, no. 6, pp. 4184–4191, 2023.

[14] S. A. Pahune, "A Brief Overview of How AI Enables Healthcare Sector Rural Development," no. January, 2024, doi: 10.13140/RG.2.2.16675.63525.

[15] A. Balasubramanian, "Intelligent Health Monitoring: Leveraging Machine Learning and Wearables for Chronic Disease Management and Prevention," *Int. J. Innov. Res. Eng. Multidiscip. Phys. Sci.*, vol. 7, no. 6, pp. 1–13, 2019, doi: 10.5281/zenodo.14535443.

[16] S. Pandya, "Integrating Smart IoT and AI-Enhanced Systems for Predictive Diagnostics Disease in Healthcare," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 10, no. 6, pp. 2093–2105, Dec. 2024, doi: 10.32628/CSEIT2410612406.

[17] A. A. Syed, A. B. Shaik, S. Konatam, M. V. P. Kumar, and N. Gupta, "CNN-Based Analytical Approach for Liver Disease Classification and Prediction," in *2025 3rd International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, IEEE, Apr. 2025, pp. 682–685. doi: 10.1109/InCACCT65424.2025.11011283.

[18] S. Maurya, A. Khatoon, A. Hassan, A. K. Badhan, M. Rakhra, and T. Sarkar, "Advanced Diagnostic Framework for Liver Disease Identification Utilizing Integrated Ensemble Machine Learning Models," in *2025 International Conference on Networks and Cryptology (NETCRYPT)*, 2025, pp. 1064–1068. doi: 10.1109/NETCRYPT65877.2025.11102555.

[19] M. S. Hossain Shaon, M. F. Sultan, T. Karim, A. Cuzzocrea, and M. S. Akter, "An Advanced Liver Disease Detection Tool with a Stacking-Ensemble-based Machine Learning Approach," in *2024 IEEE International Conference on Big Data (BigData)*, 2024, pp. 6123–6131. doi: 10.1109/BigData62323.2024.10825884.

[20] M. Kumar, N. A. Shelke, J. Singh, K. N. Sharma, R. Sharma, and R. Kumar, "A Multimodal Deep Learning Approach for Advancing Liver Disease Diagnosis and Prognosis Prediction," in *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, 2024, pp. 838–843. doi: 10.1109/ICICAT62666.2024.10923423.

[21] A. V P, S. N. S S, and R. D, "Cirrhosis Prediction in Chronic Liver Disease Patients Using Machine Learning Techniques," in *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, 2023, pp. 437–441. doi: 10.1109/ICPCSN58827.2023.00077.

[22] M. Minnoor and V. Baths, "Liver Disease Diagnosis Using Machine Learning," in *2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*, 2022, pp. 41–47. doi: 10.1109/AIC55036.2022.9848916.

[23] M. Abdar, "A Survey and Comparison of the Performance of IBM SPSS Modeller and Rapid Miner Software for Predicting Liver Disease by Using Various Data Mining Algorithms," *Cumhur. Sci. J.*, vol. 36, no. 3, 2015.

[24] N. Nishida and M. Kudo, "Artificial intelligence models for the diagnosis and management of liver diseases," *Ultrasonography*, vol. 42, no. 1, pp. 10–19, Jan. 2023, doi: 10.14366/usg.22110.

[25] D. Sharma, N. Gotlieb, M. E. Farkouh, K. Patel, W. Xu, and M. Bhat, “Machine Learning Approach to Classify Cardiovascular Disease in Patients With Nonalcoholic Fatty Liver Disease in the UK Biobank Cohort,” *J. Am. Heart Assoc.*, 2022, doi: 10.1161/JAHA.121.022576.

[26] S. M. Ganie and P. K. Dutta Pramanik, “A comparative analysis of boosting algorithms for chronic liver disease prediction,” *Healthc. Anal.*, vol. 5, p. 100313, 2024, doi: https://doi.org/10.1016/j.health.2024.100313.

[27] A. Gulia, R. Vohra, and P. Rani, “Liver Patient Classification Using Intelligent Techniques,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5110–5115, 2014.

[28] P. S. Harshini, K. Naresh, S. R. Pamulapati, and A. Lavanya, “Diagnosis of Liver Diseases Using Machine Learning Algorithms and their Prediction Using Logistic Regression and ANN,” in *2023 3rd International Conference on Intelligent Technologies, CONIT 2023*, 2023. doi: 10.1109/CONIT59222.2023.10205819.

[29] M. E. Haque *et al.*, “StackLiverNet: A Novel Stacked Ensemble Model for Accurate and Interpretable Liver Disease Detection,” *ArXiv*, 2025.

[30] S. Noor, S. A. AlQahtani, and S. Khan, “XGBoost-Liver: An Intelligent Integrated Features Approach for Classifying Liver Diseases Using Ensemble XGBoost Training Model,” *Comput. Mater. Contin.*, vol. 83, no. 1, pp. 1435–1450, 2025, doi: 10.32604/cmc.2025.061700.

