ISSN: 0067-2904

A Comparative Study of Probabilistic and Ensemble Learning for Liver Disease Diagnosis

Israa Mohammed Hassoon*
Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, Iraq

Received: 22/10/2023 Accepted: 10/6/2024 Published: 30/4/2025

Abstract

Early diagnosis of liver disease is extremely challenging because it lacks recognizable symptoms. When liver disorders are identified early, patients can start treatment before it's too late, perhaps saving their lives. It is imperative to propose a preprogramming diagnosis model to avoid misdiagnoses or delayed diagnoses. This article aims to give a comparative analysis of probabilistic and ensemble learning, both of which have demonstrated efficacy in resolving real-world problems. There are four methods used in total: two probabilistic and two ensemble learning. A substantial liver dataset is employed, containing the records of 30691 individuals, 21917 of whom have liver disease and 8774 of whom do not. First, enough features are discovered by applying ten patient attributes. After preprocessing, 30% of the patient data is used for testing, and 70% is used for training. Then, Naïve Bayes, logistic regression, random forest, and extreme gradient boosting receive them. The parameters (such as gamma number, number of estimators, max-iter, max-leaf nodes, max-depth, etc.) for each of the four algorithms are established. Specificity, sensitivity, accuracy, and f1-score are the four quantitative evaluation parameters used to evaluate the performance of each model. The results obtained from the four models are compared with previous research and with each other. More lives would be saved as a result of this work's decreased rate of wrong diagnoses. The outcomes show that using and fine-tuning hyperparameters optimizes the model's performance. By combining the output of multiple weak models using ensemble methods, increased accuracy is achieved. By reaching a high accuracy of 100%, ensemble algorithms fared better than probabilistic approaches, which had an accuracy of 91.64%.

Keywords: Probabilistic Learning, Ensemble Learning, Liver Disease, Naïve Bayes, Logistic Regression, Extreme Gradient Boosting, Random Forest.

دراسة مقارنة للتعلم الاحتمالي والجماعي لتشخيص أمراض الكبد

اسراء محمد حسون
قسم الرياضيات، كلية العلوم، الجامعة المستنصرية، بغداد، العراق

*Email: isrmo9@uomustansiriyah.edu.iq

الخلاصة

يعد التشخيص المبكر لأمراض الكبد أمرًا صعبًا للغاية لأنه يفتقر إلى الأعراض المميزة. عندما يتم تحديد اضطرابات الكبد مبكرًا، يمكن للمرضى بدء العلاج قبل فوات الأوان، وربما ينقذون حياتهم. من الضروري اقتراح نموذج تشخيص مسبق البرمجة لتجنب التشخيص الخاطئ أو التشخيص المتأخر. تهدف هذه المقالة إلى تقديم تحليل مقارن للتعلم الاحتمالي والتعلم الجماعي، وكلاهما أثبت فعاليته في حل مشاكل العالم الحقيقي. هناك أربع طرق مستعملة في المجموع: اثنتان للتعلم الاحتمالي واثنتان للتعلم الجماعي. تم استعمال مجموعة كبيرة من بيانات الكبد، والتي تحتوي على سجلات 30691 فردًا، 21917 منهم مصابون بأمراض الكبد و8774 منهم لا يعانون منها. أولاً، يتم اكتشاف الميزات الكافية من خلال تطبيق عشر سمات للمريض. بعد المعالجة المسبقة، يتم استخدام 30% من بيانات المريض للاختبار، و70% للتدريب. بعد ذلك، يتم استقبالهم من قبل Naïve Bayes والانحدار اللوجستي والغابة العشوائية وتعزيز التدرج الشديد. تم إنشاء المعلمات (مثل رقم جاما، وعدد المقدرات، والحد الأقصى، والعقد ذات الحد الأقصى، والعمق الأقصى، وما إلى ذلك) لكل من الخوارزميات الأربعة. الخصوصية والحساسية والدقة ودرجة f1 هي معلمات التقييم الكمي الأربعة المستعملة لتقييم أداء كل نموذج. وتمت مقارنة النتائج التي تم الحصول عليها من النماذج الأربعة مع البحوث السابقة ومع بعضها البعض. سيتم إنقاذ المزيد من الأرواح نتيجة لانخفاض معدل التشخيص الخاطئ لهذا العمل. تظهر النتائج أن استعمال المعلمات الفائقة وضبطها يؤدي إلى تحسين أداء النموذج. ومن خلال الجمع بين مخرجات النماذج الضعيفة المتعددة باستعمال طرق التجميع، يتم تحقيق دقة متزايدة. ومن خلال الوصول إلى دقة عالية تبلغ 100%، حققت خوارزميات المجموعة أداءً أفضل من الأساليب الاحتمالية، التي بلغت دقتها 91.64%.

1. Introduction

Disease diagnosis is the process of finding the disease that most closely matches a person's symptoms. The hardest problem to diagnose is that some symptoms and indicators are difficult to interpret, and identifying the disease is essential to treating any illness [1].

The liver is an essential and delicate organ located in the abdomen that is involved in numerous major body processes, including digestion, detoxification, metabolism, and filtration. Any difficulty carrying out these tasks results in grave issues and may even be fatal [2].

The three main symptoms of liver disease are excessive weight gain or loss, weakness, and abdominal pain. Because liver disease quadruples a person's risk of dying, it is known as the "silent killer." Certain behaviors and diseases, such as diabetes, hepatitis, obesity, and alcoholism, are linked to it [3]. Liver diseases cause the death of 70% of people around the world. A medical professional's experience and diagnostic procedures are used to diagnose a disease. But occasionally, an inaccurate diagnosis can result in the patient receiving an improper course of care. Finding effective methods for making accurate diagnoses and avoiding costly tests is therefore essential [4].

Machine learning is a branch of study that can use historical training data to predict sickness. In the medical domain, machine learning algorithms are one of the latest technologies that can manage a lot of unobserved difficulties and complex, large-scale, nonlinear data. It provides assurance for enhancing disease prediction and decision-making integrity. Several machine learning methods have been developed by scientists to efficiently identify a broad variety of circumstances. A model that anticipates illnesses and their remedies can be produced by machine learning algorithms [5].

The data-generating distribution need not be straightforward in many applications of machine learning models in the biomedical domain. Improved performance can be obtained by

adjusting the internal settings after becoming familiar with the model. Thus, applying machine learning (ML) models as an accurate, non-invasive, and highly reliable way to forecast liver disease is beneficial [6].

Predictions based on the fundamental principles of probability and statistics are produced by a family of machine learning algorithms known as probabilistic ML methods. In theory, probabilistic machine learning is competent. It is dependent on thoroughness and could reveal the degree of guarantee associated with any ML model. It offers crucial information for decision-making processes and is an excellent method for handling uncertainty in performance analysis and risk assessment. Stated differently, probabilistic machine learning algorithms are quantitative modeling techniques that estimate the probability of future events by utilizing the impacts of random activities [7, 8].

Performance and reliability cannot be increased by a single model, especially for complex issues. The accuracy of a model can be increased by combining many models. Combining two or more models is the foundation of ensemble machine learning algorithms, which aim to reduce model prediction dispersion, achieve high performance, and produce accurate predictions. To get the lowest potential error, the ensemble approach's model predictions should all be uncorrelated. Better accuracy in ensemble learning can be achieved by matching component classifier numbers to category labels [9, 10].

2. Related Works

In this section, a concise review of some related works is provided.

In [8], modern machine-learning algorithms to predict and diagnose the grades of ascites in the liver are presented using four machine-learning algorithms. The dataset was acquired from the research institute for gastroenterology and liver diseases, Shahid Beheshti University of Medical Sciences. The dataset was gathered from 492 patients. 20 attributes are selected for predicting ascites. First, the dataset is preprocessed, normalized, and missing data is replaced with the average of the columns. After that, four machine learning methods are applied: KNN, RF, SVM, and ANN. The KNN algorithm showed the highest accuracy of 94%. Their work had two limitations: first, an inadequate number of dataset samples, particularly in the case of patients who did not have ascites; and second, not all attributes were used in their work and may be considered in the evaluation of cirrhosis patients. Due to these limitations, these models may need more validation.

In [11], a prediction model based on hybrid ensemble methods (LightGBM, KNN, and random forest) is introduced. The output is fed to the voting classifier in order to obtain the final result. Three main diseases are predicted in their work: brain stroke, liver cancer, and lung cancer. The dataset is collected from the UCI repository for liver and lung cancer diseases. The dataset contained 11 features and 410 instants for liver disease, as well as 16 features and 309 instants for lung cancer. For brain stroke, their dataset contained 8 features and was obtained from https://www.kaggle.com. The preprocessing steps eliminated its imbalance, outliers, and null values. The voting classifier is utilized as an extra classifier, whereas LightGBM, KNN, and random forest are utilized as the main classifiers. The outcomes showed perfect accuracy of 98% and 90.32% for brain stroke and lung cancer, respectively, and good accuracy of 81.20% for liver disease.

In [12], an approach is proposed for the detection of liver disease based on ensemble machine learning algorithms. The dataset contained 583 records of Indians' patients. A number of preprocessing steps are performed, such as data balancing, imputation, encoding, and filling all null values. Ten features are selected based on various methods, such as feature importance,

correlation matrix, and univariate selection. Six ensemble algorithms are used: xgboost, bagging, gradient boosting, extra tree, stacking, and random forest algorithms. Their approach achieved the best outcomes compared to the previous works for the same model; the extra tree classifier, which has been utilized for liver disease detection for the first time, overran all the other studies. Random forest and extra tree classifiers achieved high performance of 86.06% and 91.82%, respectively.

In [13], an MLP-based deep learning system is developed to detect cirrhosis liver disease. The dataset was collected from https://www.kaggle.com; it contained 418 data rows and 19 important features. Their model included demographic data on the patients and blood test data, which represent the input. Two hidden layers for their model are used. Hyperparameter evaluation studies are performed utilizing GridSearchCV to define the number of neurons and epochs in the hidden layers. The ReLU activation function is utilized in the input layer. The sigmoid activation function is utilized in the output layer. Their model is compared with KNN, DT, LR, RF, SVM, and NB. The model is evaluated in terms of F1-score, recall, precision, and accuracy. The results showed that the MLP model was the best compared to other models in terms of accuracy (80.48%). Their model can be employed in real-world applications to improve healthcare and present an accurate and effective diagnosis system. On the other hand, there is one limitation to their work: the small size of the database.

In [14], a model to diagnose hepatitis B disease is proposed based on supervised machine learning. The model used 14 attributes gathered from Arba Minch General Hospital to classify factors relevant to hepatitis B disease, such as chronic and acute hepatitis B disease factors. The dataset contained 50032 patients, which were split into 80% for training and 20% for testing. Four models—J48, Bayes Net, PART, and REP Tree—are trained. The WEKA tool and the Asp.Net programming language are used for implementation. The results showed that the J48 model had the best accuracy (85.58% for the training and 82.7% for the testing). All models are evaluated using accuracy, precision, recall, and ROC area.

In [15], a system to predict liver disease risk is proposed based on supervised machine learning models. The dataset was gathered from the Indian Liver Patients' dataset, which contained 579 participants. First, the dataset is balanced using the synthetic minority sampling method, then trained on: probabilistic classifiers (naive Bayes, logistic regression, SVM), decision-tree-based models (random tree, reduced error pruning tree, and J48), ensemble ML algorithms (random forest, bagging, rotation forest, voting, adaboostm1, and stacking), KNN, and multilayer perceptron. The models are evaluated using accuracy, precision, recall, and f1-score. The results showed that the voting method achieved the best accuracy of 80.10% among other methods.

In [16], a hepatitis C virus prediction framework is proposed using machine learning. The dataset contains 859 instances and 12 attributes that were collected from the National Liver Institute, Menoufiya University, and Egypt. Four models are utilized: logistic regression, random forest, naïve bayes, and K-nearest neighbor. The authors proposed two frameworks: the first one used all attributes (12), and the second one used some attributes based on sequential forward selection that selected the best attributes. The results showed that RF achieved the best performance without or with sequential forward selection; the RF classifier achieved 94.06% and 94.29% accuracy, respectively. The hyperparameter values of the RF are adjusted to: n_estimators = 200, max_depth = 9, max_features=auto, criterion = gini, and the accuracy is improved to 94.88%. The model achieved excellent results, but it had limitations. 1) small HCV dataset; 2) the model used only 11 attributes; more attributes are needed to obtain more information that may be beneficial in predicting new infected instances of HCV; 3) the sample selected for their work was particularly Egyptian, which worked in a high-risk environment, not for patients in general.

Many research studies used reasonably small sample sizes. For the purpose of precisely identifying and treating liver diseases, more data is required. This work addresses prior research

gaps by introducing an ensemble machine learning model-based approach to liver disease diagnostics based on big data. The following list outlines the unique contributions of this work that go beyond earlier research:

1- To evaluate a case study with a large number of patients. Significant data on symptoms of liver disease has been obtained and then transformed into a suitable form for ML system coding.

2- To provide ML models that combine the fundamental classifiers in order to enhance their initial performance.

3- To design and execute influential probabilistic/ensemble ML algorithms, and then the best one for diagnosing liver disease has been chosen.

The output from this work would reduce incorrect diagnoses, which would preserve more lives. Two types of ML algorithms are used in this work: probabilistic and ensemble ML algorithms.

3. Materials and Methods

The presented comparative analysis system's architecture is illustrated in Figure. 1.

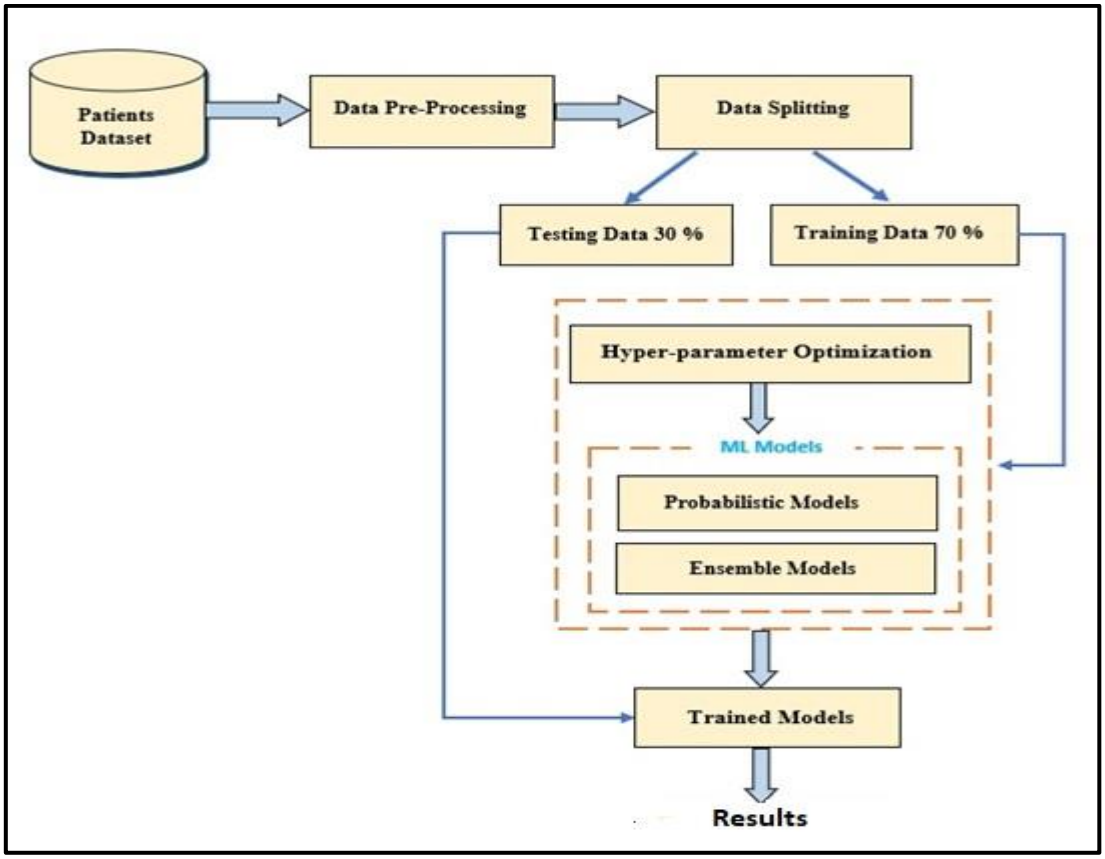

Figure1: The Presented Work Architecture

3.1 Data Set

In the proposed work, the required dataset is imported from [17]. The dataset contains 30691 people's records; 8774 of them don't have liver disease, and 21917 have liver disease. This dataset includes 11 columns, 10 features, and 1 output. The output serves as a category label, dividing the diagnosis results into two groups: liver disease patients and those without. The output will later change to a numerical value of 0 or 1. For more details, see Table 1.

Table 1: General Dataset details

Dataset	No. of patients	No. of patients/have disease	No. of patients/don't have disease	No. of attributes	Classes
Patients' Dataset [17]	30691	21917	8774	10	2

3.2 Data Preprocessing

Useful methods and techniques such as encoding data, handling null values, eliminating duplicate values, and correlation analysis are utilized to obtain an efficient dataset [18]. This work addresses missing or null values using the mean imputation method. Each variable's observed value mean is calculated, and the mean is used to impute the variable's missing values. The benefit of this method is that it yields fair estimates and reliable statistical results [19]. One crucial machine-learning preprocessing technique is data encoding. It describes the process of transforming textual or category data into numerical format so that algorithms can process it as input. Because machine learning algorithms often operate on numerical data rather than text or category variables, encoding is necessary. So, gender is mapped to a numerical type of 0 or 1 [20]. Duplicate data is eliminated in order to enhance the quality of the dataset. It becomes vital to combine several data representations and get rid of redundant information in order to give users access to correct and consistent data [21]. The liver dataset is unbalanced; it consists of 21917 patients with liver disease and 8774 who do not, so balancing the data is necessary. An efficient oversampling method is employed; oversampling is a data augmentation technique that is employed. By raising the quantity of samples in the minority class, it equalizes the distribution of classes [22]. The relationship between dataset attributes is analyzed using data correlation analysis [23]. The output is encoded as a numerical type: 1 when a patient has a liver disease, 0 when he does not.

3.3 Features Extraction

This work utilized data that was obtained from [17]. Ten features that are predictive factors represent the initial feature vector of the patient data. A two-case class variable outcome (yes or no) is the eleventh variable. It is translated to a numerical value and is regarded as a dependent variable. Table 2 illustrates the ten features.

Table 2: List of features utilized in this work.

No.	Feature Name	Range
1	Age	years > 12
2	Gender	Male or Female
3	Total Bilirubin	0 - 1.2 mg/dL
4	Direct Bilirubin	< 0.3
5	Alkphos Alkaline Phosphatase	44 to147 U/L
6	Sgpt Alamine Aminotransferase	4 to 36 U/L
7	Sgot Aspartate Aminotransferase	8 to 33 U/L
8	Total Proteins	6.0 to 8.3 g/dL
9	ALB Albumin	3.4 to 5.4 g/dL
10	Ratio of Globulin and A/G Ratio Albumin	1.10 to 1.80

3.4 The Proposed Models

In this work, two types of machine learning algorithms are used: probabilistic and ensemble ML algorithms.

- Probabilistic machine learning algorithms

Probabilistic ML algorithms are algorithms that make predictions based on the primary rules of probability. These algorithms identify uncertain correlations between variables and use the following: 1) an observation as an input; 2) an objective function to determine the cost of misprediction; and 3) a prediction function to make an intelligible prediction [22].

Two probabilistic ML algorithms are used: Naïve Bayes [23], which is simple to use, can outperform the most complex classification algorithms, is useful with large datasets, and can classify the new case faster because it does not wait for test data to learn. The second probabilistic ML algorithm is logistic regression, which is a supervised learning algorithm used in binary classification problems [24].

One of the most important advantages of probabilistic ML algorithms is that they provide a good understanding of the uncertainty related to predictions. The fundamental reasons why probabilistic ML algorithms are highly popular nowadays are that they give protection against overfitting and permit fully coherent inferences over complicated datasets [25].

- Ensemble machine learning algorithms

Ensemble ML algorithms are algorithms that use multi-learning models; they get the best prediction by employing the strengths of multiple algorithms [4]. Two types of ensemble ML algorithms are used in this work: the random forest algorithm and the extreme gradient boosting algorithm (XGBoost).

Random forest is an ensemble learning algorithm that harnesses the bagging technique's advantages to construct multiple decision trees based on boot-strapped samples. It is simple, fast, and obtains high performance [7]. The Extreme Gradient-Boosting Algorithm is an ensemble algorithm that uses the gradient-boosting framework. The extreme gradient boosting algorithm loss function has a regularization term and is based on subsampling to protect the model against overfitting [25].

The important advantages of utilizing ensemble algorithms are: its efficiency to solve the overfitting problem in decision tree approaches; minimal attribute engineering requirements; its ability to handle null values; its speed compared to most ML algorithms; and its ability to handle large datasets [26].

3.5 Evaluation of Models

The most realistic estimations that estimate the test execution of various ML algorithms are used to assess the performance of the proposed models. Several evaluation metrics exist, including accuracy, recall, precision, and f1-score. [15], which is based on an error matrix, is used to gauge how well machine learning algorithms work. A prediction table that shows the number of accurate and inaccurate predictions for each class is called an error (or confusion) matrix. It is utilized in binary classification to provide an overview of a classification model's performance [27].

High levels of recall, precision, and accuracy are desirable for a categorization technique. False-positives should be preferred over false-negatives in healthcare prediction models [28].

Four terms are used in this work to evaluate the four machine learning techniques (ensemble and probabilistic): accuracy, recall, precision, and f1-score. See equations (1–5).

$$AC = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\mathrm{SN} = \frac{TP}{TP+FN} \quad (2)$$

$$\mathrm{PR} = \frac{TP}{TP+FP} \quad (3)$$

$$\mathrm{F1-Score} = 2 * \frac{SN*PR}{SN+PR} \quad (4)$$

$$\mathrm{SP} = \frac{TN}{TN+FP} \quad (5)$$

4. Experimental Results

The proposed work is implemented using the processor Intel (R) Core (TM) i5-2430M CPU @ 2.40GHz, RAM 4GB, an x64-based processor, Windows 10, and Anaconda 3.5.2.0 64-bit, which is an open-source platform used to manage Python. Of the 30691 patient records in the sample, 21917 have liver disease, and 8774 do not. The preprocessed dataset is divided into 70% training and 30% testing, with the same dataset being given to each model.

When training a machine learning model, values called hyperparameters should be selected as settings; they could manage the model's learning process from the data. The process of fine-tuning or modifying a machine learning model's hyperparameter settings after initial selection is known as hyperparameter tuning. Hyperparameter tweaking can help the model perform better by identifying more accurate or ideal values that are appropriate for the given set of data. Selecting the right hyperparameters will aid in improving the model's accuracy, speed of learning, ability to adapt to new data, and ability to prevent overfitting and underfitting.

Table 3 presents further details regarding the hyperparameters that yielded optimal accuracy for the machine learning models.

Table 3: Models' hyperparameters

Models	Hyperparameters
Naïve Bayes	alpha=0.1 useKernelEstimator: False useSupervisedDiscretization: True
Logistic Regression	Max-iter=1000 Penalty=12 Tol=0.0001
Random Forest	n-estimators= 100 max-leaf nodes=9 max-depth=9
Extreme Gradient Boosting	n-estimators= 100 gamma=0 random-state=none

The scikit-learn hyperparameter tuning tool is utilized in this work. After splitting the dataset into training and testing sets, the hyperparameters and their search ranges are defined, and then these hyperparameters are given to the grid search algorithm, which examines the

hyperparameter search space and attempts to find the optimal values that maximize accuracy. The outcomes are assessed in terms of accuracy, recall, precision, and f1-score. The assessment of the suggested models is shown in Table 4.

Table 4: The proposed models' evaluation in terms of accuracy, recall, precision, and f1-score

Algorithms	Sensitivity	Precision	F1-Score	Specificity	Accuracy (%)
Naïve Bayes	0.9611	0.9660	0.9620	0.9610	96.41%
Logistic Regression	0.8681	0.8466	0.8411	0.8686	86.87%
Random Forest	1.0	1.0	1.0	1.0	100%
Extreme Gradient Boosting	1.0	1.0	1.0	1.0	100%

Table 4 shows how ensemble classifiers like random forest and extreme gradient boosting outperformed probabilistic classifiers like naïve bayes and logistic regression in terms of accuracy. While Naïve bayes and logistic regression attain 96.41% and 86.87% accuracy respectively, the most accurate models are random forest and extreme gradient boosting, which have 100% of accuracy. Figure. (2) Shows a comparison graph of various models with respect to sensitivity, precision, f1-score, and specificity

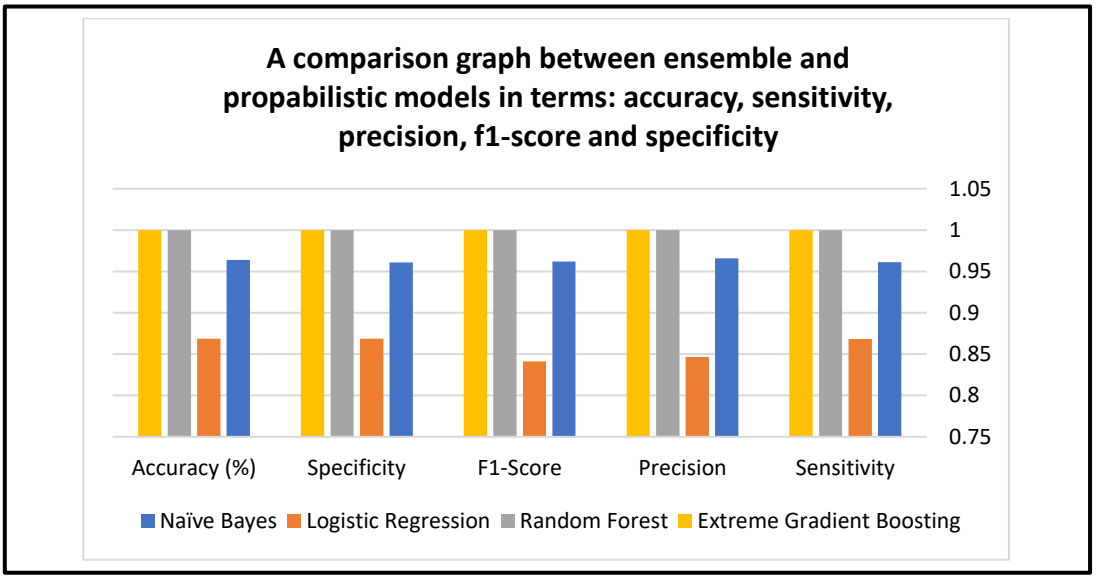

Figure 2: a comparison graph of ml models in terms of accuracy, sensitivity, precision, f1-score, and specificity

The two models—gradient boosting and random—performed much better overall than the other models. As demonstrated by the comparison results, ensemble algorithms are the most effective for liver disease diagnosis. Table 5 compares the proposed models with previous studies that utilized the same features.

Table 5: A comparison of the proposed models with previous studies that utilized the same attributes

Ref.	No. of Attributes	No. of Models	Best Models	Accuracy
[12]	10	6	Extra tree classifier	91.82%
[15]	10	14	Ensemble algorithm (Voting)	80.10%
The proposed work	10	4	Ensemble algorithms	100%

We can acquire the accuracy of the previous studies from Table 4. The extra-tree classifier [12] showed an accuracy of 91.82%. Voting [15] scored 80.10% for accuracy. Both algorithms were based on small databases and employed the same number of attributes. Small datasets might not have the required sample size to train the models, which means they may not include enough data to generate accurate prediction models. Furthermore, the limited performance of statistical inference on small datasets may limit the accuracy of the analysis. The presented ensemble algorithms yielded an accuracy of 100%. When evaluating the quality of a classifier for liver disease prediction, the presented model's relatively high sensitivity—as shown in Table 3—reflects its ability to diagnose liver disease accurately.

Also, as shown in Table 6 below, the suggested approach outperforms previous models that employed various attributes:

Table 6: A comparison of the proposed models with the previous studies, which used different attributes

Ref.	No. of Attributes	No. of Models	Best Models	Accuracy
[8]	20	4	KNN	94%
[13]	19	1	MLP-based deep learning	80.48%
[14]	14	4	J48 model	82.7%
[17]	12	4	Random forest	94.29%
The proposed work	10	4	Ensemble algorithms	100%,

KNN attained 94% accuracy. The accuracy of deep learning based on MLP was 80.48%. The accuracy of the J48 model was shown to be 82.7%. 94.29% accuracy was demonstrated by a random forest. Based on the algorithms' outputs, the presented ensemble methods (RF and XGBoost) generated results that were satisfactory. The algorithms are based on various attribute counts and database sizes. Small data bases were employed in the research [8, 13, and 17], but large data bases were used in the study [14]. This paper found that the provided algorithms performed better than the other algorithms based on a variety of statistical performance indicators.

Better and more accurate machine-learning models are produced when there is enough data. More trustworthy outcomes arise from handling missing and outlier values correctly. The performance of the model is optimized by using hyperparameters and tuning them. Improved accuracy is obtained by merging the output of several weak models through the use of ensemble methods.

There are a number of limitations with this work that need to be addressed:

First, the presented work has the drawback of being susceptible to outliers because every classifier is required to correct the mistakes made by the previous models. Each estimator bases its accuracy on previous predictions, so streamlining the process is difficult. Second, the data is obtained only from a significant database [17]. The samples would be more realistic if they were from many medical centers, hospitals, and different localities, so the outcomes would be more general. Third, understanding the patient's journey through the disease is one of the most essential aspects of liver disease research; thus, it's crucial to investigate all the factors influencing the rapid progression of the disease.

5. Conclusion

Global liver disease infection and dissemination have grown to be serious health risks that require greater focus. Owing to the challenge of identifying liver disease, it is imperative that liver disease be identified early and correctly. Robust and advantageous probabilistic and ensemble machine learning techniques are presented to address this issue. Following the selection of the penitent dataset, preprocessing is carried out, including balancing, resampling, and resolving missing variables. 30% of the dataset is designated for testing, while the remaining 70% is for training. Following that, various ensemble learning and probabilistic algorithms are used, including random forest, logistic regression, the extreme gradient boosting method, and naïve bayes. Every algorithm's hyperparameters, including the gamma number, number of estimators, max-iter, max-leaf nodes, max-depth, etc., are found. The outcomes demonstrate that the ensemble models (random forest and severe gradient boosting) are 100% accurate. The accuracy of probabilistic models, such as Naïve Bayes and logistic regression, is 96.41% and 86.87%, respectively. Assessing sensitivity, precision, f1-score, and specificity for the ensemble models yields excellent results. Based on the same or different attributes as prior works, this work performs better than those works. Future research can make use of more sophisticated datasets, analysis techniques like region of practical equivalency and maximum density intervals, and more recent classifiers like case-based reasoning.

Acknowledgments

The author would like to thank Mustansiriyah University (www.uomustansiriyah.edu.iq) in Baghdad, Iraq, for its support in the present work.

References

[1] M. A. Kuzhippallil, C. Joseph, and A. Kannan, “Comparative analysis of machine learning techniques for Indian liver disease patients,” in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*. India, 2020, pp. 778–782, Doi: 10.1109/ICACCS48705.2020.9074368.

[2] A. Gaber, H. A. Youness, A. Hamdy, H. M. Abdelaal, and A. M. Hassan, “Automatic Classification of Fatty Liver Disease Based on Supervised Learning and Genetic Algorithm,” *Appl. Sci.*, vol. 12, no. 1, p. 521, Jan. 2022, Doi: 10.3390/app12010521.

[3] M. Ringehan, J. A. McKeating, and U. Protzer, “Viral hepatitis and liver cancer,” *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, vol. 372, no. 1732, p. 20160274, Oct. 2017, Doi: 10.1098/rstb.2016.0274.

[4] Y. Zhang, J. Liu, and W. Shen, “A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications,” *Appl. Sci.*, vol. 12, no. 17, p. 8654, Aug. 2022, Doi: 10.3390/app12178654.

[5] N. H. Ali, M. E. Abdulmunem, A. E. Ali," Learning Evolution: A Survey,” *Iraqi J. Sci*, 2021, Vol. 62, No. 12, pp: 4978-4987, Feb. 2021, DOI: 10.24996/ijs.2021.62.12.34.

[6] L. Göcs and Z. C. Johanyák, "Feature Selection with Weighted Ensemble Ranking for Improved Classification Performance on the CSE-CIC-IDS2018 Dataset," *Computers*, vol. 12, no. 8, p. 147, Jul. 2023, Doi: 10.3390/computers12080147.

[7] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, “Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics,” *Electronics*, vol. 10, no. 5, p.593, Mar. 2021, Doi: 10.3390/electronics10050593.

[8] B. Hatami, F. Asadi, A. Bayani, M. R. Zali, and K. Kavousi, "Machine learning-based system for prediction of ascites grades in patients with liver cirrhosis using laboratory and clinical data: design and implementation study," *Clin. Chem. Lab. Med*., vol. 60, no. 12, pp. 1946-1954, May 2022, Doi: 10.1515/cclm-2022-0454.

[9] P. Mahajan, S. Uddin, Farshid Hajati, and Mohammad Ali Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare,* vol. 11, no. 12, p.1808, Jun. 2023, Doi: 10.3390/healthcare11121808.

[10] D. Ramesh and Y. S. Katheria, "Ensemble method based predictive model for analyzing disease datasets: a predictive analysis approach," *Health and Technol.*, vol. 9, pp. 533–545, Feb. 2019, Doi: https://doi.org/10.1007/s12553-019-00299-3.

[11] M.d. Abdul Quadir, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease," *Biomedicines,* vol. 11, no. 2, p. 581, Feb. 2023, Doi: https://doi.org/10.3390/biomedicines11020581.

[12] A. Utku," Deep Learning Based Cirrhosis Detection," *Operational Research in Engineering Sciences: Theory and Applications*, vol. 6, no. 1, pp. 95-114, 2023, DOI: https://doi.org/10.31181/oresta/060105.

[13] A.O. Assegie," Classification Model for Hepatitis B Disease Using Supervised Machine Learning Technique," *Computer Engineering and Intelligent Systems*, vol.13, no.3, pp.1-9, 2022, DOI: 10.7176/CEIS/13-3-01.

[14] I.M. Hassoon, S. A. Qassir, M. Riyadh," PDCNN: FRAMEWORK for Potato Diseases Classification Based on Feed Foreword Neural Network," *Baghdad Sci. J*., vol. 18, No.2, pp. 1012-1019, http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).1012.

[15] H. M. Farghaly, M. Y. Shams, T. A. El-Hafeez," Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt," *Knowl. Inf. Syst*., vol. 65, no.6, pp.2595–2617, Jun. 2023, https://doi.org/10.1007/s10115-023-01851-4.

[16] G.-W. Ji et al., "Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study," *EBioMedicine*, vol. 50, pp. 156–165, 2019, Doi: 10.1016/j.ebiom.2019.10.057.

[17] https://www.kaggle.com/datasets/abhi8923shriv/liver-disease-patient-dataset.

[18] H. Huang et al., "A 7 gene signature identifies the risk of developing cirrhosis in patients with chronic hepatitis C," *Hepatology*, vol. 46, no. 2, pp. 297–306, Aug. 2007, Doi: https://doi.org/10.1002/hep.21695.

[19] X. Dong et al., "An Improved Method of Handling Missing Values in the Analysis of Sample Entropy for Continuous Monitoring of Physiological Signals," *Entropy*, vol. 21, no. 3, p. 274, Mar. 2019, Doi: 10.3390/e21030274.

[20] N. R. Njeri, " Data Preparation for Machine Learning Modelling," *IJCTR* , vol. 11, no.06, pp. 231-235, 2022, DOI:10.7753/IJCATR1106.1008.

[21] D. G. Cuautle et al.," Synthetic Minority Oversampling Technique for Optimizing Classification Tasks in Botnet and Intrusion-Detection-System Datasets," *Appl. Sci.*, vol. 10, no. 794, pp.1-19, 2020, doi:10.3390/app10030794.

[22] J. H. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine," *Ann. Statist.* , vol. 29, no.5, pp. 1189–1232, 2001. DOI: 10.1214/aos/1013203451.

[23] S. M. Abd-Elsalam, M. M. Ezz, S. Gamalel-Din, G. Esmat, A. Salama, and M. ElHefnawi, "Early diagnosis of esophageal varices using Boosted-Naïve Bayes Tree: A multicenter cross-sectional study on chronic hepatitis C patients," *Inform. Med. Unlocked*, vol. 20, no.3, p. 100421, Jan. 2020, Doi: https://doi.org/10.1016/j.imu.2020.100421.

[24] A. G, B. Ganesh, A. Ganesh, C. Srinivas, Dhanraj, and K. Mensinkal, "Logistic regression technique for prediction of cardiovascular disease," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 127–130, Jun. 2022, Doi: https://doi.org/10.1016/j.gltp.2022.04.008.

[25] K.G. Dinesh, K. Arumugaraj, K.D. Santhosh and V. Mareeswari, "Prediction of cardiovascular disease using machine learning algorithms," 2018 *International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, India, 2018, pp. 1-7, Doi: 10.1109/ICCTCT.2018.8550857.

[26] Singh, V.; Gourisaria, M.K.; Das, H. Performance Analysis of Machine Learning Algorithms for Prediction of Liver Disease. *In Proceedings of the 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON*), Kuala Lumpur, Malaysia, 24–26 September 2021; pp. 1–7.

[27] D. Božić, B. Runje, D. Lisjak, and D. Kolar, "Metrics related to confusion matrix as tools for conformity assessment decisions," *Appl. Sci.*, vol. 13, no. 14, p. 8187, Jul. 2023, Doi: 10.3390/app13148187. [Online]. Available: https://doi.org/10.3390/app13148187.

[28] G. S. Handelman et al., "Peering into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods," *AJR. Am. J. Roentgenol.*, vol. 212, no. 1, pp. 38–43, Jan. 2019, Doi: 10.2214/AJR.18.20224