EPH–International Journal of Science & Engineering
Published by: E-Pub Solutions Private Limited
ISSN (Online): **2454-2016**
Volume 12 Issue 04 November 2026
DOI:10.67772/t0xrxy05

ADAPTIVE WEIGHTED MULTI-MODEL PREDICTIVE FRAMEWORK FOR LIVER HEALTH ASSESSMENT

Paras Soni[1], Dr Sumit Kumar Sar[2]

[1]M. Tech Research Scholar, Bhilai Institute of Technology, Durg, Chhattisgarh, Email ID: soni.paras2015@gmail.com, ORCID ID: 0009000711938722
[2]Assistant Professor, Bhilai Institute of Technology Durg, Chhattisgarh, Email ID: sumit.sar@bitdurg.ac.in, ORCID ID: 0009-0007-7776-4622

Article History:
Article Type: **Research**
Received Date: **23/06/2026**
Revised Date: **11/10/2026**
Accepted Date: **05/11/2026**
Published Date: **26/11/2026**

Keywords: Adaptive ensemble learning, healthcare analytics, Indian Liver Patient Dataset, liver disease prediction, machine learning, multi-metric evaluation

ABSTRACT
This paper presents an Adaptive Weighted Multi-Model Predictive Framework for assessment of liver health condition by utilizing demographic and biochemical data of patients. For this study, Indian Liver Patient Dataset has been used as primary source that describes a dataset of 12500 records. The workflow pipeline involves handling of missing value using imputation techniques, exploratory data analysis, feature scaling, noise removal (handling outliers), assessment and handling of class imbalance. Using Grid Search CV and Shuffle Split Cross Validation a set of 8 machine learning classifiers—Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Naïve Bayes, K-Nearest Neighbors, Extra Trees, and XGBoost are tuned. The performance evaluation of each model is done by assessing and evaluating the set of classification metrics— Accuracy, Precision, Recall, F2-Score, and ROC-AUC. The adaptive weights are calculated from Accuracy, Precision, Recall, ROC-AUC, and F2-Score, normalized to give overall sum as one. The derived adaptive weights are applied to output probability of trained classifiers. The designed ensemble achieves 99.7984% Accuracy, 99.6422% Precision, 99.6422% Recall, 99.7509% ROC-AUC, and 99.6422% F2-Score. Extra Trees remain stronger with respect to individual ROC-AUC and Precision, which indicates that the contribution of proposed framework is interpreted as balanced multi-metric aggregation rather than individual superiority of every metric. The result shows that the potential of adaptive weighted ensemble learning for research-oriented liver health assessment systems, however clinical validation remains necessary before real-world use.

Introduction

Liver disease comprises of a wide range of disorders that affect liver health, which is a vital organ that is directly involved in several biochemical processes, detoxification, protein synthesis and metabolism. Several liver-based conditions may progress without prominent early symptoms, so timely assessment is critical healthcare requirement.

Conventional assessments usually involve medical procedures such as liver function tests, imaging techniques such as magnetic resonance imaging, ultrasounds and computed tomography and in selected cases liver biopsy. These methods remain important in clinical practice, but the project identifies challenges in detecting risk during early or asymptomatic stages. This conventional assessment can be complemented with a predictive system developed on routinely available biomarkers and demographic variables providing additional

A predictive system based on routinely available demographic and biochemical variables can complement conventional assessment by providing an additional data-driven indication of liver health. Several Machine Learning algorithms have become increasingly relevant to medical prediction since they can model nonlinear relationships and complex patterns that are difficult to capture using single statistical rules. The Indian Liver Patient Dataset (ILPD) provides a compact set of demographic and biochemical features such as age, gender, bilirubin levels, liver enzymes, total proteins, albumin, and albumin-to-globulin ratio. These attributes make the dataset suitable for comprehensive evaluation of diverse classifiers and proposed adaptive weighted ensembles.

One of the important issues in medical classification is that no single algorithm essentially performs best on every evaluation metric. For example, a classifier may have high Recall but weaker Precision, while another classifier may give high accuracy but lower Recall. Using conventional equal weights or fixed weight in ensemble does not explicitly handle these differences. The proposed framework therefore uses performance metric driven adaptive weights method so that models with better overall performance score has greater weightage and stronger overall influence on final probability and overall prediction.

The most important contributions of this study are: (1) development of multi-model liver health prediction pipeline; (2) preprocessing of biochemical and demographic features on basis of observations obtained using exploratory data analysis; (3) using GridSearchCV and ShuffleSplit for tuning of several hyperparameter; (4) Using Accuracy, Precision, Recall, F2-Score and ROC-AUC for multi-metric performance evaluation; (5) deriving the adaptive weights on basis of five selected performance measures; and (6) aggregation of weighted probability for final liver health classification.

Related work

Liver disease prediction methods has progressed from conventional classification methods to advanced machine learning, ensemble-based approach and deep learning-based classification methods.

Singaravelu et al. [1] performed the classification of liver patients using the Indian Liver Patient Dataset (ILPD) and developed different classification algorithms utilizing WEKA establishing a comparative method to identify suitable machine learning algorithms for predicting liver health

Singh et al. [2] utilized feature selection and several classification algorithms on ILPD and proposed a liver disease prediction system which is software and comparing classifier performance before and after feature selection. Singla et al. [3] extended the comparative direction by evaluating machine learning and deep learning approaches, including Multi-Layer Perceptron, Stochastic Gradient Descent, Restricted Boltzmann Machine with Logistic Regression, Support Vector Machines, and Random Forest and identified the Multi-Layer Perceptron as the best-performing approach among the evaluated models.

Hendi et al. [4] proposed an adaptive framework to explore deep learning techniques for liver disease subtypes and prediction, discussing the application of neural architectures to complex medical data. Similarly, Dashti et al. [5] proposed a self-predictive liver failure diagnosis approach utilizing backpropagation and multilayer neural networks demonstrating usefulness of neural-

network-based approaches for learning complex clinical feature patterns and helping automated predictions.

Ensemble methods can improve overall predictive performance by combining multiple classifiers that utilizes the strengths of individual classifiers. Ganie et al. [6] incorporated preprocessing, feature selection and hyperparameter tuning and evaluated nine algorithms and three ensemble approaches on dataset of liver patient comprising of 30,691 samples and 11 features with Gradient Boosting being the best overall performance in their experiments. Alizargar et al. [7] used NHANES data and combined Random Forest, XGBoost and Support Vector Machine along with contribution of gender-related information to investigate ensemble machine learning for diagnosis of non-alcoholic fatty liver disease.

Charu et al. [8] evaluated a Super Learner framework for the non-invasive diagnosis of liver fibrosis against established clinical risk scores and presented the potential of ensemble-based prediction to identify liver fibrosis and benchmarked conventional methods of clinical risk assessment methods.

Mohamed et al. [9] used ILPD dataset and presented a two-level ensemble stacking model to predict liver disease. The framework presented combined multiple base classifiers along with feature selection to improve predictive performance. Islam et al. [10] applied Tree-Structured Parzen Estimator-based hyperparameter optimization for classifiers such as Decision Tree, Random Forest, Extra Trees, LightGBM, and AdaBoost on ILPD and found that optimized Extra Trees model achieved 95.8% accuracy, thereby highlighting the importance of handling class imbalance in dataset along with systematic hyperparameter optimization.

Ganji et al. [11] proposed a framework that used feature selection and machine learning algorithms to develop an enhanced early-diagnosis framework which was evaluated on ILPD dataset. Their optimized LightGBM model achieved an 82.12% accuracy compared with its initial accuracy of 76.21%. Rani et al. [12] further applied hybrid SMOTE-ENN-based balancing to address the issues caused by class imbalance during model development.

The existing studies explored feature selection, hyperparameter optimization, deep learning, and conventional ensemble techniques, however scope for investigating multi-model performance-driven adaptive weighted approach remains. The present work addresses it by developing a framework that evaluates eight machine learning classifiers and calculates overall performance score for each trained model using multiple metrics for evaluation which are than normalized to derive adaptive ensemble weights, higher weights are assigned to strong performers which allows more significant contribution to final prediction from classifiers showing stronger performance.

Dataset And Data Preparation

A. Dataset Description

Indian Liver Patient Dataset (ILPD) has been used as its primary source. Since the original ILPD contains 583 records, a similar dataset from Kaggle was also merged. To make the final data more practical and realistic, controlled noise was introduced using the developed simulator. The resulting working dataset consists of 12500 observations, 10 attributes, and one target variable.

Table 3.1 summarizes the feature present in the dataset. The target feature Liver Disease has two labels 0 for Diseased and 1 for Healthy.

Table 3.1: Features used for Liver Health Assessment

Feature	Role	Type / description
Age	Feature	Integer; patient age
Gender	Feature	Binary categorical
Total Bilirubin (TB)	Feature	Continuous biochemical marker
Direct Bilirubin (DB)	Feature	Continuous biochemical marker
Alkaline Phosphatase (ALP)	Feature	Enzyme measurement
ALT / SGPT	Feature	Liver enzyme measurement
AST / SGOT	Feature	Liver enzyme measurement

Total Proteins (TP)	Feature	Continuous biochemical marker
Albumin (ALB)	Feature	Continuous biochemical marker
A/G Ratio	Feature	Albumin-to-globulin ratio
Liver Disease	Target	Binary class

B. Preprocessing

Data preprocessing is one of the most essential stages in the proposed framework pipeline. The preprocessing sequence includes removal of duplicates, imputation of missing-values, encoding of categorical features, feature scaling, detection of outliers, class balancing of dataset and collinearity handling. These operations are essential to improve the predictive performance of proposed framework.

Simple imputation that replaces missing values with mode in case of categorical variables and with median in case of continuous numerical variables is used. Label encoder is used to encode the categorical feature, Gender. Standard Scalar is used to standardize so that difference in numerical magnitudes do not dominate the algorithms that are sensitive to the scale.

Z-score based outlier detection is applied to detect and remove the outliers, where observation having absolute Z-score greater than 3 are identified and removed. Descriptive statistics, histograms, correlation heatmaps and potential class imbalance are checked during exploration data analysis. Feature-importance analysis focuses on meaningful biochemical variables, particularly bilirubin and liver-enzyme measurements. Class imbalance was observed in original imbalanced ILPD and hence data balancing using SMOTE is included in the pre-processing stage of pipeline

The models are evaluated after the common preprocessing and balancing stage.

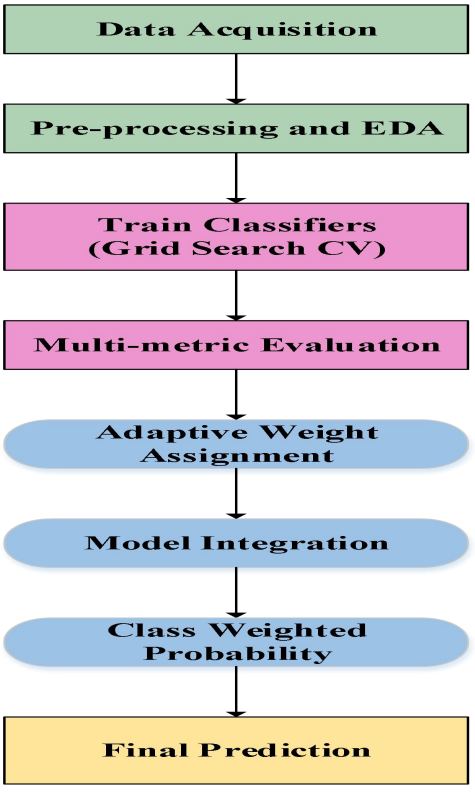


Figure 3. 1. Workflow of the proposed adaptive weighted multi-model framework.

Proposed methodology

Step by step pipeline in proposed methodology begins with data preparation and ends with model prediction. The objective is liver health assessment, which is critical task therefore instead of relying on one metric and maximizing one score, the framework considers set of metrics such as Accuracy, Precision, Recall, F2-Score and ROC-AUC which is used to calculate adaptive weights. F1-Score is

also calculated and is retained and used as the GridSearchCV optimization criterion.

A. Model Selection

Logistic Regression provides a linear probabilistic baseline that establishes the performance using a relatively simple decision boundary while Decision Tree is based on interpretable rule-based structure. KNN performs distance-based classification, Naïve Bayes provides a probabilistic classification approach based on conditional independence assumptions. Ensemble learners such as random Forest and Extra Trees are also part of ensembled multi-model framework

Sequential boosting strategies such as Gradient Boosting and XGBoost capable of modelling nonlinear interactions among biochemical variables are also used. The idea behind using diversity is that the adaptive ensemble must constitute models that can provide complementary predictive behavior.

B. Hyperparameter Optimization

GridSearchCV combined with ShuffleSplit cross-validation is used to perform hyperparameter tuning. Three Shufflesplit cross-validation with test size of 20% is specified. F1-Score is used as optimization criteria since it balances precision and Recall as Accuracy alone is not suitable especially in case of dataset having class imbalance

Number of estimators, maximum depth and criterion are tuned parameters in case of Random Forest and Extra Trees Classifiers while learning rate, number of estimators and maximum depth are tuned for Gradient Boosting and XGBoost. For KNN, number of neighbors, weighing strategy and distance metric are tuned while regularization parameters and parameter C and solver are tuned for Logistic regression. The estimator producing best F1-Score is retained for downstream evaluation.

C. Performance Metrics

Accuracy, Recall, Precision, F2-Score and ROC-AUC are the metrics used to evaluate normalized score based on which the models are ranked and weights are assigned. While F1-Score is used to select the optimized models during GridCV Search based hyperparameter tuning. Accuracy is the ratio of correctly classified observation, while Precision is the measure of correctly predicted positive samples out of positive predicted samples. Recall measures proportion of correctly identified positive samples from total positives. F1-Score provides a harmonic balance between Recall and Precision while F2-Score gives greater emphasis to Recall. ROC-AUC measures the ability of classifier to discriminate between two classes across decision thresholds.

Using multiple metrics in case of medical prediction is highly relevant because an individual model with high accuracy may still miss significant fraction of diseased cases. Hence, using multi-metric score is much better way to deal in such cases rather than relying on one single measure to evaluate ensemble contribution.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.3)$$

$$\text{F1} - \text{Score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (3.4)$$

$$\text{F2} - \text{Score} = \frac{5\times\text{Precision}\times\text{Recall}}{(4\times\text{Precision})+\text{Recall}} \quad (3.5)$$

D. Adaptive Weight Calculation

The core contribution of the proposed framework is the adaptive weighing mechanism. For an individual model i, the average performance score over five measures Accuracy, Precision, Recall, ROC-AUC and F2-Score is calculated, where Acc_i, $Prec_i$, Rec_i, AUC_i, and $F2_i$ denote the metric value of model i. which is mathematically presented in equation 3.6

$$S_i = \frac{Acc_i+Prec_i+Rec_i+AUC_i+F2_i}{5} \quad (3.6)$$

The adaptive weights are then calculated by normalizing the scores using equation 3.7

$$w_i = \frac{S_i}{\sum_j S_j} \quad (3.7)$$

Using normalization ensures that the sum of all model weights is exactly one. Hence, a classifier with stronger and performance across all selected metrics have comparatively higher contribution towards the final prediction, while weaker classifier remains part of ensemble, but its influences are less proportional.

Table 4.1: Average Performance Score and adaptive weights for all models

Model	Average Performance Score	Assigned Weight
Extra Trees	99.51%	0.1402
Random Forest	99.39%	0.1400
XG Boost	99.32%	0.1399
Gradient Boosting	99.25%	0.1398
Decision Tree	98.41%	0.1386
KNN	97.77%	0.1377
Naive Bayes	67.65%	0.0953
Logistic Regression	48.67%	0.0686

The evaluated average performance score and the adaptive weights are calculated and presented in table 4.1. Clearly on basis of average performance score, we can say that higher weighted classifiers have stronger influence on prediction probability.

E. Weighted Ensemble Integration

The probability of liver disease is generated for every optimized classifier for an input patient. Using the calculated adaptive weights, the ensemble combines these probabilities as per the equation 3.8

$$P_{ensemble}(x) = \sum_i w_i P_i(x) \quad (3.8)$$

where $P_i(x)$ is the probability generated by model i for patient x. To obtain final binary class, a threshold of 0.5 is applied. The project describes the same mechanism as a weighted aggregation of the probability outputs of all individual models.

F. Model Saving and Deployment

Once the training and evaluation is done, the final models, scalers, encoders and adaptive weights configurations are saved using joblib library that allows the learned components to be reused for predicting new observations, without repeating the entire training process. The architecture is modular and scalable and identifies that deployment is feasible through application or via API. The concept of deployment is considered to support clinical decisions rather than replacing a professional diagnosis systems and process. Any future implementation can take patients biochemical inputs and demographics and obtain prediction using the trained saved model, calculate the weighted ensemble probability score and provides the predicted class and confidence information.

Results

The eight individual classifiers and trained and their average performance scores are calculated and based on average performance scores the weights are assigned which are presented in figure 5.1. The result of the experiment shows substantial differences among the eight classifiers as shown in table 5.1. With respect to accuracy the Extra-trees classifiers seem to be the best performing model, while Random Forest, Gradient Boosting and XG-Boost also provide satisfactory values across reported metrics, however Naïve Bayes and Logistic Regression show substantially weaker performances. The proposed adaptive weighted model achieves the strongest overall Accuracy among the reported models while also showing strong performance on other metrics.

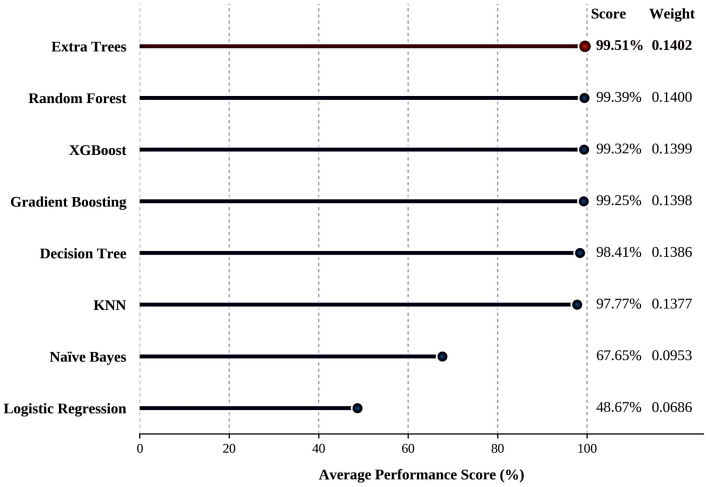


Fig 5.1: Performance Score and assigned weights

A. Evaluation Metric Analysis

The proposed model achieves an accuracy of 0.9980 exceeding the accuracy of other individual models used. Since the proposed model combines diverse classifiers and is performance driven weighted approach. The improvement supports the intention of reducing dependence on an individual model. In terms of Precision, Extra Trees Classifiers gives the best performance with precision of 0.9972 while proposed model records 0.9964 slightly lower than Extra Trees but better than other classifiers while the recall of proposed model is 0.9964 which is better than all other individual models reported in table 5.1, this is the most relevant metric because recall gives the measure of model's ability to identify actual positive cases correctly.

Table 5.1: Comparative performance of evaluated models

Model	Accuracy	Precision	Recall	AUC	F2-Score
Random Forest	0.9954	0.9944	0.9897	0.9995	0.9906
Gradient Boosting	0.9943	0.9916	0.9886	0.9986	0.9892
Decision Tree	0.9892	0.9804	0.9822	0.9871	0.9818
KNN	0.9832	0.973	0.9684	0.9945	0.9693
Naïve Bayes	0.5604	0.3923	0.9558	0.7316	0.7425
Logistic Regression	0.7208	0.5474	0.1903	0.7562	0.2187

Extra Trees	0.9964	0.9972	0.9904	0.9995	0.9917
XGBoost	0.9948	0.9928	0.9893	0.9993	0.99
Proposed Model	0.998	0.9964	0.9964	0.9975	0.9964

The evaluated metric of Naïve Bayes clearly shows the importance of using multi-metric approach. As it has relatively high recall value of 0.9558, but its precision is only 0.3932 and AUC is 0.7316, clearly considering only Recall for evaluation would be misleading. The adaptive score approach shrinks the influence of imbalanced performance profiles.

ROC-AUC values of Extra Trees, Random Foresr, Gradient Boost and XG-Boost exceeds 0.998 clearly showing strong discrimination between the classes. The proposed model records and ROC-AUC value of 0.9975 which is strong but relatively less strong than some of the individual classifiers.

F2-Score emphasize stronger Recall more compared to F1-Score. The highest individual F2-Score is seen in case of Extra Trees approximately 0.9917, while the proposed model achieves 0.9964. Clearly the obtained result shows that the weighted ensemble provides a strong balance when sensitivity has greater significance.

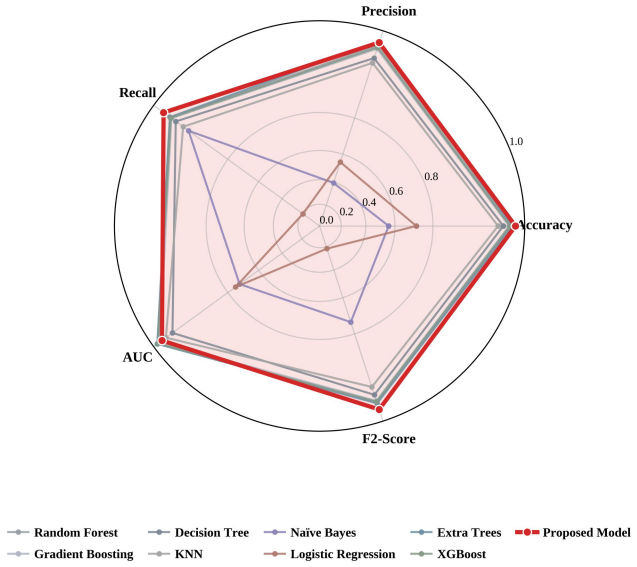


Figure 5.2: Performance comparison of proposed models with individual models across the evaluation metrics.

Figure 5.2 clearly compares the performance of proposed model with individually trained and optimized classifiers using the five metrics discussed.

B. Why the Adaptive Ensemble Performs Better

The results clearly indicate that model's performance is not identical across all metrics. For example, Extra Trees is strongest on several individual metrics, Random Forest have very high ROC-AUC while boosting classifiers like XGBoost and Gradient Boost also give strong performances. At the same time, Logistic Regression and Naïve Bayes shows weak results for some of the metrics. Using an Equi-weight ensemble would consider all models equal, however in the proposed work weights are assigned as per measured competence.

The individual classifier contributes an estimate of probability and adaptive weights calculates how strongly the estimate affects the final confidence score, creating a contiguous aggregation which is compared with the classification threshold of 0.5. Thus, it can be concluded that proposed approach also retains additional information.

The proposed framework us robust ensemble rather than simply an individual metric maximizing

classifier, the primary benefit of which is a systematic integration of heterogeneous models that uses an explicit and reusable weighting mechanism.

C. Computational Considerations

Computational complexity of the proposed methodology is the main trade-off as wight classifiers are trained and hyperparameters are tuned using cross-validation before constructing the final ensemble, which needs greater computation compared to deployment of an individual optimized classifier. The work explicitly discovers this increased computational cost as limitation for the model proposed,

Since the trained models and components used in preprocessing and encoding are serialized using joblib library the main computations are concentrated in model development phase and during periodic retraining. The saved components and models can be reutilized for future deployments in case latency of prediction is more critical than the training costs.

D. Generalizability and Limitations

Interpretation of results should be done within the scope of experimental pipeline and dataset available. The primary source of data that is ILPD is relatively sparse so it has been supplemented using similar another source and generated noise using controlled simulator, therefore the reported accuracy of 99.80% must not be interpreted as universal accuracy.

The limitations related to dataset bias, class imbalance, feature availability, selection of models, generalization and population differences have been identified. The model trained on given demographics may show different behavior on data from another region. Therefore, external validation becomes on individual clinical dataset becomes extremely necessary before deploying for real-world clinical use.

Another limitation of the proposed framework is that the adaptive score uses arithmetic means of five metrics that are easy to reproduce and transparent, but future research could identify metrics that are clinically motivated, differently calibrated probabilities, or optimization of the weighting rule on a validation set. Explainability techniques such as LIME or SHAP can also be considered as future prospect.

Discussions

The most important discovery in this study is that the adaptive weighted ensemble integration provides an overall stronger classification result than any single evaluated model. The proposed framework achieves an accuracy of 99.8% while the highest individual accuracy of 99.64% was recorded for Extra Trees. The ensemble also produces very strong performances for other metrics such as Precision of 99.64%, Recall of 99.64%, ROC-AUC of 99.76%, and 99.64% F2-Score.

The proposed model does not essentially maximize every individual metric. For instance, Extra Trees shows slightly higher Precision and ROC-AUC. The proposed model tries to achieve an overall balanced performance across five criteria which makes this distinction very important while presenting the contribution. The novelty of approach is the adaptive multi-metric weight mechanism, not the claim that every metric is better than every individual classifier trained.

The results clearly indicate that model diversity is very essential as the behavior of different models differs on same biomedical features. Several learners are retained and performance-based weights are assigned; such framework reduces the dependency on the inductive bias of a single classifier.

With respect to healthcare analytics perspective, this framework can be best positioned as decision-support component and can produce an additional risk classification on basis of demographic and biochemical attributes, however the final treatment and diagnosis must remain dependent on medical expertise.

Conclusion And Future Work

An adaptive weighted multi-model predictive framework has been presented in this paper for assessment of Liver Health. ILPD dataset has been used as primary source and applies a structured pipeline that applies preprocessing, exploratory analysis, feature scaling, hyperparameter tuning, multi-metric evaluation, adaptive weighting, and weighted probability aggregation.

Evaluation of eight individual machine learning algorithms—Logistic Regression, Decision Tree,

Random Forest, Gradient Boosting, Naïve Bayes, KNN, Extra Trees, and XGBoost—were done. The proposed adaptive ensemble shows an Accuracy of 99.80%, Precision and Recall of 99.64%, ROC-AUC of 99.76 and F2-Score of 99.64%. From the results it is evident that proposed multi-model approach achieves highest accuracy among the evaluated classifiers while maintaining a balanced predictive performance and strong sensitivity.

The major methodological contribution is the use of normalized performance-based weights rather than assigning fixed or equal contributions, the framework assigns greater influence to models that perform consistently well on five metrics: Accuracy, Precision, Recall, ROC-AUC, and F2-Score. The integration of probability-based confidence score from all constituent models to contribute before final classification provides additional information.

Future work should focus on external validation using multi-center dataset and independent hospital or integration of additional features. LIME and SHAP based explanations, and hybrid deep learning architectures. Future integration with electronic health records, real-time clinical workflows, deployment on web applications or API and expanding prediction towards liver conditions such as fatty liver, hepatitis, fibrosis etc. can be done.

Implementation

The framework implemented is sequential and reproducible. The dataset is loaded and during pre-processing it is analyzed for structure, missing values, duplicates, encoding of categorical variables and distribution of class. The preprocessing stage is applied prior to model training to ensure that all classifiers receive same representation of patient data. The preprocessing components are stored together along with trained classifiers so that the same transformation procedure can be applied when a new record is evaluated.

The pipeline of experiment consists of eight individual classifiers: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Naïve Bayes, K-Nearest Neighbours, Extra Trees, and XGBoost. To achieve best possible configuration of hyperparameter GridSearchCV with ShuffleSplit is applied. Here we are using three rounds of ShuffleSplit with 20% split as test size and remaining for training with F1-Score as the search objective.

Once the optimized classifiers are obtained the selected estimators are evaluated using five metrics: Accuracy, Precision, Recall, F2-Score, and ROC-AUC. This prevents a model from receiving the largest ensemble contribution simply because it performs well on one measure while performing poorly on another. The average score S_i is calculated for every model and normalized so that the complete set of weights sums to one.

In the prediction stage probability outputs rather than hard class labels are used for prediction. For a patient x, each optimized classifier generates $P_i(x)$, representing its estimated probability for the target class. These probabilities are multiplied by their corresponding adaptive weights and summed. The resulting ensemble probability is converted into a binary decision by comparing with the threshold of 0.5. Since the trained models, scaler, encoder, and weights are serialized using Joblib, the prediction stage can be repeated without retraining the models. For reproducibility, preprocessing configuration, hyperparameter grids, train-test splitting procedure, cross-validation settings, metric definitions, and saved model artifacts must be retained.

.

References

[1] S. Muthuselvan, S. Rajapraksh, K. Somasundaram, and K. Karthik, “Classification of liver patient dataset using machine learning algorithms,” International Journal of Engineering & Technology, vol. 7, no. 3.34, pp. 323–326, 2018. https://doi.org/10.14419/ijet.v7i3.34.19217.

[2] J. Singh, S. Bagga, and R. Kaur, “Software-based prediction of liver disease with feature selection and classification techniques,” Procedia Computer Science, vol. 167, pp. 1970–1980, 2020. https://doi.org/10.1016/j.procs.2020.03.226.

[3] B. Singla, S. Taneja, R. Garg, and P. Nagrath, “Liver disease prediction using machine learning and deep learning: A comparative study,” Intelligent Decision Technologies, vol. 16, no. 1, 2022. https://doi.org/10.3233/IDT-210065.

[4] A. M. Hendi, M. A. Hossain, N. A. Majrashi, S. Limkar, B. M. Elamin, and M. Rahman,

"Adaptive method for exploring deep learning techniques for subtyping and prediction of liver disease," Applied Sciences, vol. 14, no. 4, Art. no. 1488, 2024. https://doi.org/10.3390/app14041488.
[5] F. Dashti, A. Ghaffari, A. Seyfollahi, and B. Arasteh, "A self-predictive diagnosis system of liver failure based on multilayer neural networks," Multimedia Tools and Applications, vol. 83, pp. 83769–83788, 2024. https://doi.org/10.1007/s11042-024-18945-y.
[6] S. M. Ganie, P. K. D. Pramanik, and Z. Zhao, "Improved liver disease prediction from clinical data through an evaluation of ensemble learning approaches," BMC Medical Informatics and Decision Making, vol. 24, Art. no. 160, 2024. https://doi.org/10.1186/s12911-024-02550-y.
[7] A. Alizargar, Y.-L. Chang, M. Alkhaleefah, and T.-H. Tan, "Precision non-alcoholic fatty liver disease (NAFLD) diagnosis: Leveraging ensemble machine learning and gender insights for cost-effective detection," Bioengineering, vol. 11, no. 6, Art. no. 600, 2024. https://doi.org/10.3390/bioengineering11060600.
[8] V. Charu, J. W. Liang, A. Mannalithara, A. Kwong, L. Tian, and W. R. Kim, "Benchmarking clinical risk prediction algorithms with ensemble machine learning for the noninvasive diagnosis of liver fibrosis in NAFLD," Hepatology, vol. 80, no. 5, pp. 1184–1195, 2024. https://doi.org/10.1097/HEP.0000000000000908.
[9] M. H. Mohamed, B. H. Ali, A. I. Taloba, A. O. Aseeri, M. A. Elaziz, S. H. Ali El-Sappagh, and N. El-Rashidy, "Toward an accurate liver disease prediction based on two-level ensemble stacking model," IEEE Access, vol. 12, pp. 180210–180237, 2024. https://doi.org/10.1109/ACCESS.2024.3459429.
[10] R. Islam, A. Sultana, and M. N. Tuhin, "A comparative analysis of machine learning algorithms with tree-structured parzen estimator for liver disease prediction," Healthcare Analytics, vol. 6, Art. no. 100358, 2024. https://doi.org/10.1016/j.health.2024.100358.
[11] A. Ganji, D. Usha, and P. S. Rajakumar, "Enhanced early diagnosis of liver diseases using feature selection and machine learning techniques on the Indian Liver Patient Dataset," Scalable Computing: Practice and Experience, vol. 26, no. 3, 2025. https://doi.org/10.12694/scpe.v26i3.4254.
[12] R. Rani, G. Jaiswal, Nancy, Lipika, S. Bhushan, F. Ullah, P. Singh, and M. Diwakar, "Enhancing liver disease diagnosis with hybrid SMOTE-ENN balanced machine learning models—an empirical analysis of Indian patient liver disease datasets," Frontiers in Medicine, vol. 12, Art. no. 1502749, 2025. https://doi.org/10.3389/fmed.2025.1502749.