Journal of Information Systems Engineering and Management
2025, 10(13s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Diagnostic Predictive Approaches for Liver Disease Detection using Stacked Ensemble Model with Data Augmentation

Mahendran. S [1*], Venkatasekhar. D [2], Shanmugasundaram. G [3]
[1*] *Research Scholar, Department of Information Technology, Annamalai University, Chidambaram, India*
[2] *Professor, Department of Information Technology, Annamalai University, Chidambaram, India*
[3] *Professor, Department of Computer Science and Engineering, Chennai Institute of Technology, Chennai, India*
**Corresponding author, E-mail: smpdy963@gmail.com*

ARTICLE INFO	ABSTRACT

'Received: 28 Nov 2024

Revised: 10 Jan 2025

Accepted: 31 Jan 2025

The global medical fraternity is challenged with evolving a perfect prediction model that diagnoses a liver ailment at the right time and calls for an immediate medical intervention for its critical need. Potential threat to life of liver disease chronically contracted stresses for necessity for its root cause and time-bound medical remedy. By making use of a dataset of Indian liver patients from the pool of data maintained at the national level, this work is entailed with the introduction of an architecture of innovative nature that desegregates Stacked Ensemble Model, feature engineering for foretelling liver ailments. The core contribution to this work is the harness of feature engineering utilizing SHapely Additive exPlanations (SHAP), encompassing the state-of-the-art techniques that are not delved into by data-driven machine learning approaches that are in vogue now as a vaticinator of liver ailment. The labyrinthine design of Stacked Ensemble Model facilitates timely prediction of liver disease, exploiting the maximum of learning techniques that ramp up the detection activity and augment accurate diagnosing. The methodology exploits the potencies of varied base learners like Logistic Regression, Multi-layer Perceptron, Support Vector Machine, Decision Tree, K-nearest neighbor and Extra trees classifier. These diverse views are homogenized as input to the meta-learner, the Random Forest, to build a sturdy and trustworthy predictive model. The commissioning of Stacked Ensemble Model with feature engineering produced an accuracy of a cent percent short of 3. This methodology is aimed at fine tuning the prediction of liver ailments, ensuring smooth, effective and timely intervention, and assuring of an effective management.

Keywords: Liver disease, Stacked Ensemble Model, SHapely Additive exPlanations, base learners, meta-learner.

I. INTRODUCTION

Liver, the vital organ is indispensable to wholesome of human being. It plays an important role for secretion of bile, detoxification and blood clotting by means of protein synthesis. The integrity of hepatocyte is prone to damage through infection of microbes like bacteria, virus or fungus and is susceptible to a slew of liver diseases. In addition to this, consumption of alcohol is a substantial risk factor for liver malfunctions [1,2].

The hardship involved in diagnosing early-stage symptoms of liver illness is aggravated by the people who are usually not outspoken by nature, albeit having affected liver, may satisfy they are fit as a fiddle [3]. The emergence of machine learning approaches is a boon to prediction and decision making in almost all fields, particularly in healthcare [4]. An optimum efficiency is not yielded when a single high accuracy machine learning classifier is employed on voluminous datasets. Thus, the classifiers are put together with the help of ensemble learning techniques.

To attain a great deal of accuracy in prediction, the ensemble model can be entwined with SHapley Additive ExPlanations (SHAP). SHAP is designed with a holistic platform for ascertaining feature importance in model prediction by engaging a new class of feature, catering a unique, theoretically sound solution. Thus, the tough tasks of prediction are eased for better accuracy and actionable insights [5,6].

The objective of this work is entailed with introduction of architecture of innovative nature that desegregates Stacked Ensemble Model, feature engineering for foretelling liver ailments. The core contribution to this work is the harness of feature engineering utilizing SHapely Additive exPlanations (SHAP), encompassing the state-of-the-art techniques that are not delved into by data-driven machine learning approaches that are in vogue now as a vaticinator of liver ailment. The labyrinthine design of Stacked Ensemble Model facilitates timely prediction of liver disease, exploiting maximum of learning techniques that ramp up the detection activity and augment accurate diagnosis.

The sections hereafter to organize are as follows. Section II deals with the reviews of Related Work, taking summary of research on liver disease prediction before employing single classifiers, ensemble models, and SHAP-based feature engineering techniques. Section III deals with the Methods, elaborating the Stacked Ensemble Model set up and unification of SHAP for feature fabrication. Section IV includes Results and Discussion, portraying significance of SHAP-based feature and the model’s paramount performance in finding root cause of liver disease. Section V ends up with Conclusion, stating precisely about findings and putting forward further deep search into model interpretability and benefits of detecting the disease at primary level.

II. RELATED WORKS

Individual Machine Learning based methods employed a single classifier [7], chosen as per dataset characteristics, patient feature distribution, and the classifier's potential to have control over skewed boundaries and meagre data. For the classification of ILPD patients [8] for diagnosing liver disease, feature selection techniques adopted different algorithms. Their outcomes underscore the Logistic Regression secured maximum accuracy of 74.36% with feature selection. The researcher [9] put into use twelve classification algorithms for liver disease prediction, including Logistic regression, Naive Bayes, K-nearest neighbor, Random Forest, Multilayer Perceptron, Decision Tree, Support Vector Machine, Gradient Boosting, Bayesian, XGBoost, AdaBoost and Bagging. Although a variety of algorithms were adopted, the models fell short in efficiency in respect of measuring the liver disease.

The drawbacks of individual methods are overcome by ensemble techniques congregating various classifiers to attain an assured classification model. The ensemble methods comprise bagging, boosting, and voting schemes, operate within distinct domains and show multiple performances under the influence of the aggregation method, dataset distribution, non-linearity, and class imbalances.

An empirical model for diagnose of diabetes mellitus [10] with employment of a stacked ensemble approach emphasizes the method's effectiveness in handling complex medical data, leading to significant improvements in prediction accuracy. The research [11] constructed a hybrid ensemble classifier model using supervised classifiers aiming for improved precision and reduced time entanglement. Employing a stacking ensemble of give machine learning algorithms and 58 variable predictors, a tailored machine learning model was evolved towards estimation of the enhancement of Alpha-1 Antitrypsin Deficiency-associated Liver Disease (AATD-LD). This approach attained [12] high AUROC scores, comprising 91.20% for liver transplants and 75.9% for deaths pertained to liver. The employment of permutation methods emphasized the superior performance of the ensemble model by evaluation of feature importance.

The authors [13] put forward this by proposing a hybrid model that combines a stacked ensemble strategy with Kernel SHapley Additive exPlanations. It yielded high accuracy and F1 scores on both public and private datasets, with SHAP exposing that accelerations and abnormal short-term variability are pivotal to influencing fetal states. The approach establishes robust performance and increased interpretability for fetal monitoring. An empirical model for diagnosing diabetes in females [14] harnessing stacked ensemble model engaging Shapley additive explanation was evolved to augment the prediction accuracy of 92.91% to enable early detection and intervention.

This write-up accentuates the hardships embedded with feature selection methods and interleaved training, ending up with overfitting and suboptimal performance. This work manifests all of the constraints by drawing a modern and powerful model, the stacked ensemble model, which is put together with feature engineering employing shapely additive explanations and advanced data pre-processing techniques to improve performance.

III. METHODS

This research work puts into use a stacked ensemble classifier, which binds tier-0 classifiers and feeds their outputs as input for a tier-1 classifier. This tier-1 classifier is trained to give final predictions, augmenting the classifier's overall performance.

A. Stacked Ensemble Model Architecture

This work sums up a Stacked Ensemble Model with two tiers: tier-0 and tier-1. At tier-0, foundational models are subjected to execution on the dataset, with their results performing as features in a subsequent dataset. Subsequently, the predictions from these models, obtained using the validation set, are fed into the meta-model at tier-1. The tier-1 classifier, a machine learning algorithm, puts into use the outcomes of tier-0 classifiers to attain the final classification outcome. This technique enhances prediction accuracy by making most of the advantages of many classifiers, harvesting a more resilient outcome. The Stacked Ensemble model process is outlined below:

Tier-0 Classifiers

- Classifiers: Represented as $c_1(x)$, $c_2(x)$, ..., $c_k(x)$, with t representing the class label for the data point x.
- Predictions: Tier-0 classifiers weighted outputs are entered into the tier-1 classifier as input.

Teir-1 Classifier

- Function (x) = ($c_1(x)$, $c_2(x)$, ..., $c_k(x)$) is the combination of predictions from tier-0 classifiers to give the ultimate prediction.
- Functions f and g: Aggregation and final prediction functions, respectively.
- Weight Learning: During training, tier-0 classifier weights are optimized to attain maximum efficiency.

Training

- Loss Function: (y, C(x)), with y represents the true label of x.
- Optimization Algorithm: Weights are estimated using Newton's method during the learning process.

Prediction

- During prediction, the process starts with applying the tier-0 classifiers, followed by the utilization of the learned weights by the tier-1 classifier.

Tier-0 Output (Z)

- Matrix $Z = [z_1, z_2, \dots, z_n]$, where z_i represent the output vector for the i^{th} instance, representing class probabilities.

Tier-1 Input (X)

- Matrix $X = [x_1, x_2, \dots, x_n]$, where x_i represent the input vector for the i^{th} instance, composed of the output probabilities of the tier-0 classifiers for each class.

Tier-1 Output (y)

- Matrix $y = [y_1, y_2, \dots, y_n]$, where y_i represent the output vector for the i^{th} instance, containing predicted probabilities for each class.

Objective Function

- For each instance i, the predicted class label $\hat{y}_i = \mathrm{argmax}_j(y_{i,j})$, where $\hat{y}_i$ represents the estimated class label for the i^{th}instance. This is achieved by choosing the class with the maximum predicted probability $y_{i,j}$ for the i^{th} instance and the j^{th} class.

The Stacked Ensemble Model pseudo code, as specified in Algorithm 1, narrates the implementation. The model in this work follows two-tier architecture to augment Liver disease prediction accuracy.

Algorithm 1 Stacked Ensemble Model pseudocode

```
Input:
  - Training set (Xtrain, ytrain)
  - Testing set (Xtest)
  - Base models (M1, M2, ..., Mk)
  - Meta-model (M_meta)
  - Evaluation metric (e.g., accuracy, F1-score)
Training Phase:
  For each base model Mi:
   Mi ← Train (Xtrain, ytrain)
  For each base model Mi:
   Pi ← Predict(Mi, Xtrain)
M_meta ← Train(P, ytrain)
  where P = [P1, P2, ..., Pk]
Prediction Phase:
  For each base model Mi:
P_test,i ← Predict(Mi, Xtest)
Ensemble_Predictions ← Predict(M_meta, P_test)
Output:
  Return Ensemble_Predictions
```

B. Tier-0 and Tier-1 Model

Algorithm 1 engages a versatile technique, which yields flexibility in grabbing classifiers subject to certain aggregation criteria. These criteria entail the management of nonlinear borders, the handling of imbalanced classes, the management of sparse data, the handling of categorical features, and the mitigation of overfitting. A robust ensemble that fulfils these requirements is generated by making use of six baseline classifiers that are used every now and then. These classifiers contain Logistic Regression, Support Vector Machine, Multi-layer Perceptron, Decision Tree, k-nearest neighbor, and Extra Tree classifier.

Each baseline model satisfies a unique function: Support Vector Machine is well-versed for high-dimensional datasets and imbalanced classes; Logistic Regression gives room for large datasets with probabilistic outputs; Decision Tree adept at managing both categorical and numerical features, Multi-layer Perceptron registers complex patterns and nonlinearities; K-nearest neighbour (KNN) surpasses in delineating nonlinear decision boundaries; and Extra Tree Classifier is chosen for its resilience to overfitting, ability to navigate nonlinear decision boundaries, and versatility in processing both categorical and numerical features.

At tier-1, the combination of these baseline models ensures that categorical features are managed properly, decreases the risk of overfitting, sets right the non-linearity of decision boundaries, and efficiently manifests classes of imbalance within the ensemble. Harnessing Random Forest as the meta-learner permits to grab advantage of its resistance to over-fitting, its capability to deal with non-linearity, its prowess with categorical data, and its resistance to imbalanced classes.

As in Fig. 1, the two-tier Stacked Ensemble Model grabs the advantages of a large number of classifiers at tier-0, ensued by the Meta classifier at tier-1 of the model. This architecture augments the accuracy and resilience of predictions while decreasing the likelihood of overfitting at the same time.

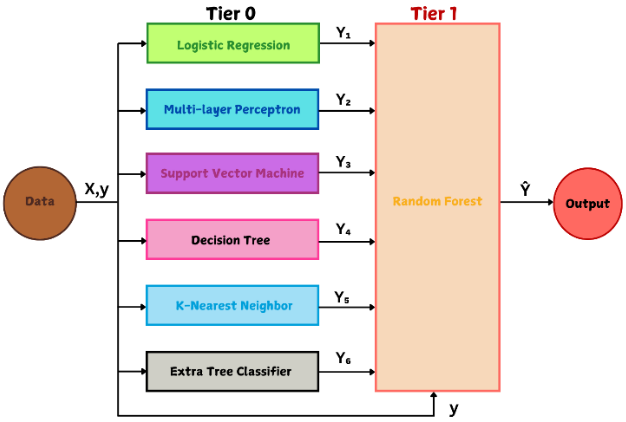

Fig. 1 Architecture of Stacked Ensemble Model

C. Feature Engineering using SHapley Additive exPlanations

Feature engineering [15] is a prominent processing part of machine learning model indulged in up taking representative features from the data that is supplied so as to augment the ability of models for precise predictions. Being a pivotal part of feature engineering, feature creation is meant that the process is merging of original features with an idea of tracking features that are of largely predictive nature [16]. The SHapley Additive exPlanations (SHAP) [17] perspective was made use of in an inquisitive way for supply of information to bring about new features in this work.

SHAP is employed with significant features identified as alkaline phosphatase concentration, direct bilirubin levels, and total bilirubin levels for diagnosis of liver ailments. The output of the model is affected by the correlation of these features from SHAP. The features take in the SHAP values so as to understand that the value affects the diagnosis of liver malady or not. Greater the SHAP value for alkaline phosphatase concentration shows a higher prognostic value for liver ailment. Conversely, a negative SHAP value tanks the variability in liver diagnosis in a great deal.

Further, the SHAP relative scores show that the features are tasked altogether to have an effect on predictions. SHAP illustrates with an example that diagnosis of liver disease is ascertained with a combination of high bilirubin and low albumin. The apprehensions are about the interactions among the features. It augments the obviousness and the ability to understand the model, which makes decision-making practical in a clinical environment. Algorithm 2 furnished as under is the SHAP algorithm for feature estimation and generation.

Algorithm 2 Feature estimation and generation using SHAP

Step 1: Initialize Inputs

- Input: Dataset X with N samples and M features.
- Trained model: f (Random Forest).

Step 2: Train the Model and Calculate SHAP Values

1. Train the Model:

- Train the predictive model f using dataset X.

$$f(x) = \hat{y}$$

2. SHAP Value Calculation:

- Use a SHAP explainer to compute SHAP values for each feature i in the dataset.
- For each sample x_j the SHAP value for feature i is computed as:

$$\phi_i^{(j)} = \sum_{S \subseteq \{1,2,..,M\}\setminus\{i\}} \frac{|S|!\,(M-|S|-1)!}{M!}[f(S\cup\{i\}) - f(S)]$$

Where S denotes a subset of features, and $\phi_i^{(j)}$ denotes the Shapley value for feature i for the j-th data sample.

Step 3: Analyze Feature Importance

- Rank the features based on their average SHAP values across all samples j:

$$\phi_i = \frac{1}{N}\sum_{j=1}^{N}\phi_{\mathrm{i}}^{(j)}$$

A higher SHAP value ϕ_i indicates that feature i is more significant for the model's prediction.

Step 4: Create Synthetic Features

1. Identify Feature Interactions:
   - Use SHAP interaction values to identify combinations of features that work together in predicting the target.
   - For two features i and j, the interaction value is:
     $SHAP\ \ Interaction_{i,j} = \phi_{i,j} - \phi_i - \phi_j$
2. Generate Synthetic Features:
   - Interaction Features: If the SHAP interaction between features i and j is significant, create a new synthetic feature:
     $Synthetic\ Feature_{i,j} = X_i \times X_j$
   - Nonlinear Transformation: For features showing nonlinearity in SHAP dependence plots, apply transformations:
     $Synthetic\ Feature_i = log(X_i + 1)$

Apart from playing up the momentous of individual features, this approach unwinds how their interactions impact predictions. By producing synthetic features from prominent interactions and nonlinearities, SHAP steps up the model's accuracy and exegesis. Desegregating SHAP into the Stacked Ensemble Model besides escalates the transparency of the classification process, ending up with more well-founded and impressive clinical decision-making, and consolidating its role as an indispensable tool for liver diagnosis and its therapy.

IV. RESULTS AND DISCUSSION

A search is made, in this section, on an elaborate analysis of the outputs through different factors comprising of precision, accuracy, recall, F1-score, and AUC score. In addition, SHAP (SHapley Additive exPlanations) criteria are utilized to prefatorily evaluate model interpretation, furnishing significant insights into the obscurant causes leading to liver ailment. This work underscores the potential of the model by blending the interpretability and performance measures to ascertain the efficiency in the decision-making process for the diagnosis of liver disease well in advance.

An inclusive comparison of tier-0 classifiers pre-feature engineering is furnished in Table 1, which is fundamentally on essential performance indicators such as accuracy, precision, recall, F1-score, and AUC score. It also portrays the efficiencies and bottlenecks of the six classifiers, thereby underscoring their overall prediction ability. Extra tree, logistic regression, and support vector machine classifiers obtained accuracy with a value of 73%. K-nearest neighbour attained an accuracy score of 74%. Multilayer view and decision tree classifiers yielded accuracy with a value of 70%.

Table 1 Tier-0 classifiers performance pre-feature engineering

Classifier	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**AUC Score**
Logistic Regression	73	0.76	0.73	0.73	0.60
Multi-layer Perceptron	70	0.70	0.70	0.70	0.62
Support Vector machine	73	0.67	0.73	0.63	0.51
Decision Tree	70	0.71	0.70	0.70	0.64
K-nearest neighbour	74	0.73	0.74	0.74	0.66
Extra Tree classifier	73	0.71	0.73	0.71	0.62

A correlation among Tier-0 classifiers post-feature engineering is made centered on influencing factors of accuracy, precision, recall, F1-score, and AUC score to unearth their foretelling vigor as shown in Table 2. Amongst, multi-layer perceptron tops all classifiers in performance, securing an accuracy of 95%, while the decision tree classifier proved to be the lowest accuracy of 90%. The remaining classifiers of Logistic Regression, Support Vector Machine, Extra Tree classifier, and K-nearest neighbour exhibited accuracies less than 95% and greater than 90% as 94%, 94%, 93%, and 92% respectively.

Table 2 Tier-0 classifiers performance post- feature engineering

Classifier	**Accuracy(%)**	**Precision**	**Recall**	**F1-Score**	**AUC Score**
Logistic Regression	94	0.94	0.94	0.94	0.93
Multi-layer Perceptron	95	0.95	0.95	0.95	0.94
Support Vector machine	94	0.94	0.94	0.94	0.90
Decision Tree	90	0.91	0.90	0.90	0.90
K-nearest neighbour	92	0.92	0.92	0.92	0.87
Extra Tree classifier	93	0.93	0.93	0.93	0.89

The particulars of pre- and post-feature engineering performance of the Proposed Models are furnished in Table 3 and Table 4 respectively. A remarkable accuracy of 97.14% was obtained by the model of post-feature engineering. It established a sizeable enhancement of 14.29% off the 82.85% accuracy shown by pre-feature engineering together with refinements in other appraisal metrics.

Table 3 Pre-feature Engineering performance of the Model

Proposed Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**AUC Score**
Stacked Ensemble Model	83	0.83	0.83	0.82	0.74

Table 4 Post-feature Engineering performance of the Model

Proposed Model	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**	**AUC Score**
Stacked Ensemble Model	97	0.97	0.97	0.97	0.94

A comparison in terms of accuracy is made between tier-0 classifier pre-feature engineering and tier-0 classifier post-feature engineering as shown in Fig. 2. An attainment of improved accuracy post-feature engineering was evident by each and every classifier. Logistic Regression witnessed an accuracy gain of 21% (from 73% to 94%), Multi-layer Perceptron obtained an improvement of 25% (from 70% to 95%), Support Vector Machine received 21% (from 73% to 94%), Decision Tree scaled up by 20% (from 70% to 90%), K-nearest Neighbor gained 18% (from 74% to 92%), and Extra Tree Classifier enhanced by 20% (from 73% to 93%).

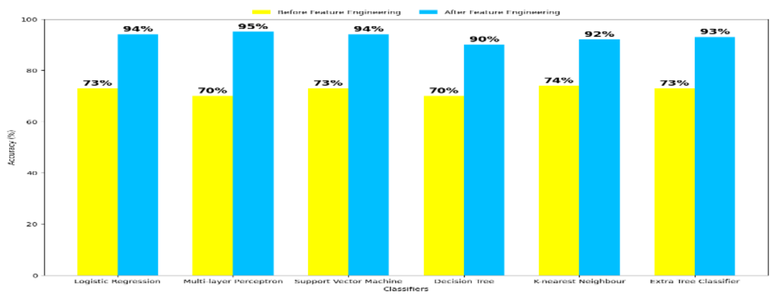

Fig. 2 Accuracy comparison of Tier-0 classifiers Pre- and Post-feature engineering

With a view to appraise the classification models and to reflect their outcome, the AUC criterion is mandated. The correlation between sensitivity and specificity's complement is denoted by the ROC curve, and the area below this curve is quantified by the AUC. Significance is observed that when AUC values increase, there exists an enhancement of model performance. Fig. 3 elucidates the ROC curves for the Stacked Ensemble Model post-feature engineering.

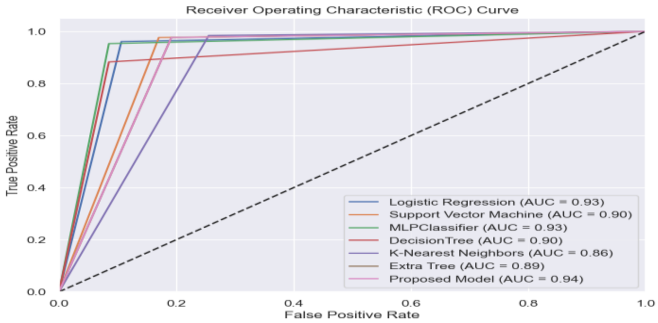

Fig. 3 Analyzing the ROC Curve of the Stacked Ensemble Model post-feature engineering

Fig. 4 is portrayed with a bar graph featuring every feature's average influence on the intended model. The vertices are the characteristics in downward influence as the x axis and the mean absolute SHAP values as the y axis showing the average effect that each feature holds on the diagnosis suggested by the model. The bar diagram interprets that the inherited feature, said to be SHAP feature, possess the most consequential effect on the diagnosis done by the model, suggesting the crucial role that it has in affecting the values that are produced by the model. Conversely, the 'gender' variable's least impact shows that there is a minuscule influence on the findings that the model develops. Each feature's relative importance containing the most impact on the performance of the model is represented by this.

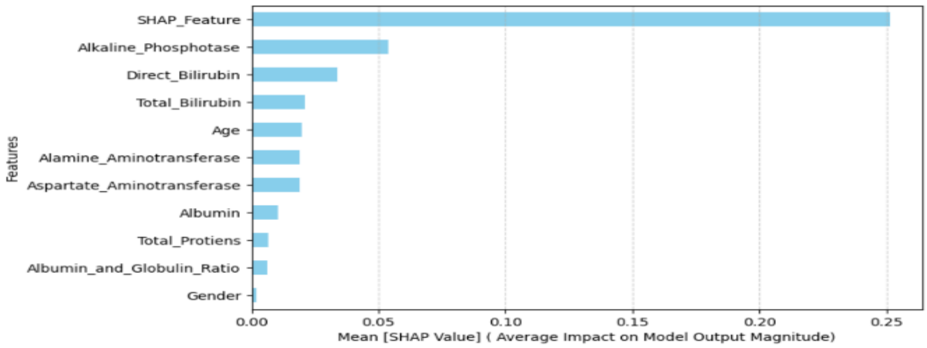

Fig. 4 Impact of features to the proposed model

A comparison between the performance of the Stacked Ensemble Model and other current works is shown in Table 5 that employs the machine learning models for forecasting liver-related malady. The output shows that the Stacked Ensemble Model's accuracy is phenomenal. With an accuracy of 97.14%, the Stacked Ensemble Model has one-upmanship to several other models evaluated in contemporary research. The results signify that the stacked ensemble model phenomenally augments performance by unifying the outputs of the tier-0 classifiers and feeding them as input value for the tier-1 classifier. Demonstration of this model proves to be efficient for forecasting hepatic complications at fledgling stage.

Table 5 Stacked Ensemble Model post-feature engineering performance assessment in comparison to contemporary research

Reference	Algorithm	Accuracy(%)	Reference	Algorithm	Accuracy(%)
[14]	**KNN+SV+RF**	**88**	[21]	Linear regression	86.90
[18]	Linear Regression	72.89		Gaussian Naïve Bayes	82.60
	Decision Tree	81.32		**Random Forest**	**92.69**
	Multilayer Perceptron	60.24		ANN	79
	Random Forest	**86.14**		CNN	73
	ANN	75.61	[22]	Logistic	55.40

	CNN	65.52		K-NN	67.90
[19]	SMO	71.36		**Random Forest**	**88.10**
	Naïve Bayes	55.90		SVM	67.90
	IBk	67.41		MLP	83.53
	J48	70.67		Ensemble	82.09
	Random Forest	71.87	Proposed Model	Logistic Regression	94.08
	Logistic Regression	**74.36**		MLP	95.18
[20]	Extra tree classifier	89.91		SVM	94.35
	Random Forest	85.93		Decision Tree	90.30
	Stacking	**93.15**		KNN	91.63
	Bagging	85.21		Extra tree classifier	93.16
	XGBoost	85.81		**Stacked Ensemble Model**	**97.14**
	Gradient Boosting	80.41			

On comparison of the recent results with the Stacked Ensemble Model towards foreseeing hepatic illness and suggesting beforehand for real-world healthcare applications, the latter succeeded in accomplishing a precise accuracy of 97.14%. Incorporation of SHAP values escalated desirable discernment into the key factors evoking predictions. But there exists a restraint that this work is deficient to fathomed analysis pertinent to the interpretability of the model and the specific influence of each feature. Besides, the research paves the way for an advanced development in the field of prediction of hepatic disease, highlighting its capability for early diagnosis, medical intervention, and strategizing for personalized treatment.

V. CONCLUSION

The change in lifestyle leads to cozy life involving sedentary works, lack of manual dexterity in daily chores, environmental pollutants et al., and deleterious practices like intoxication due to over-consumption of alcohol, narcotic contrabands like stuff, drugs, and kind, irregular food habits, becoming glutton to junk and fast foods et al. are the influencing factors that predispose to hepatic disease, which is perpetually proliferated globally in the hustle and bustle world. As a result, timely diagnosis for liver ailments is mandated to secure lives. In this work, for the prognostication of hepatic conditions, employment of stacked ensemble model is undertaken, manifesting a prominent diversion from customary moves that, in general, rely on individual classifiers or fledgling ensemble models short of feature engineering. The efficiency of the model is ascribed to its capability to conglomerate the outputs of tier-0 classifiers—Logistic Regression, Support Vector Machine, Multi-layer Perceptron, Decision Tree, K-nearest Neighbours, and Extra Trees—fed as an input to the tier-1 classifier—Random Forest, ending up with an augmented predictive performance.

The perception of significant contribution for attainment of awesome outputs of the machine learning model by harnessing Shapley Additive Explanations (SHAP) is evident from this work. It is proved in the findings of this research that the achievement of 97% accuracy by the Stacked Ensemble Model in tandem with SHAP overtopped the other machine learning models. The amalgamation of SHAP into the proposed model has led to delve into deep discernment of the individual features and characteristics, which shows a right direction to the medical fraternity in a pursuit of well-in-advance detection and treatment of hepatic patients.

In spite of having harvested overarching results, it is the need of the hour to give preference to enhancing the interpretability of the proposed model by means of a thorough, elaborate examination of the influence that feature has on predictions. The findings of such an analysis might bring about deep discernment into the metabolic processes and a more widespread comprehension of the decision-making processes that are the factors for driving hepatic ailments.

It is to be honestly admitted that this work has its own shortcomings. But this work has a cutting edge on pushing up the field of diagnosis of hepatic disease. The potential benefits earned out of harnessing the Stacked Ensemble Model along with SHAP feature engineering are indispensable to the medical fraternity in terms of precision in prediction, cost effective medication, and medical intervention for early detection. For any research scholars in the years to come, pursuing research on future applications as that of medical diagnostics work as this, where accuracy and interpretability are critical, this approach could greatly benefit them.

REFRENCES

[1] Cheemerla S, Balakrishnan M. Global Epidemiology of Chronic Liver Disease. Clin Liver Dis (Hoboken). 2021;17(5):365–70. doi: 10.1002/cld.1061.

[2] Seto WK, Mandell MS. Chronic liver disease: Global perspectives and future challenges to delivering quality health care. PLoS One. 2021 Jan 4;16(1):e0243607. doi: 10.1371/journal.pone.0243607.

[3] Y L, G M, AS H, SM T, Åberg F Ruuskanen M et al. Early prediction of incident liver disease using conventional risk factors and gut-microbiome-augmented gradient boosting. Cell Metab 2022 May 3;34(5):719-730 e4 doi: 2022;10. doi: 10.1016/j.cmet.2022.03.002.

[4] Zhao J, Wang P, Pan Y. Predicting liver disorder based on machine learning models. J Eng 2022, 978–984. 2022; doi: 10.1049/tje2.12184.

[5] Feretzakis G, Sakagianni A, Anastasiou A, Kapogianni I, Bazakidou E, Koufopoulos P, et al. Integrating Shapley Values into Machine Learning Techniques for Enhanced Predictions of Hospital Admissions. Applied Sciences. 2024 Jul 7;14(13):5925. , doi: 10.3390/app14135925.

[6] Luo H, Xiang C, Zeng L, others. SHAP based predictive modeling for 1 year all-cause readmission risk in elderly heart failure patients: feature selection and model interpretation. Sci Rep 14. 2024;17728. doi: 10.1038/s41598-024-67844-7.

[7] Alnuaimi AFAH, Albaldawi THK. “An overview of machine learning classification techniques, ” BIO Web of Conferences. 2024 Jan;97:133. doi: 10.1051/bioconf/20249700133.

[8] Singh J, Bagga S, Kaur R. “Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques, ” Procedia Computer Science. 2020 Jan;167:1970–80. doi: 10.1016/j.procs.2020.03.226.

[9] Ram MK, Sujana C, Srinivas R, Murthy GSN. A Fact-Based Liver Disease Prediction by Enforcing Machine Learning Algorithms. Advances in intelligent systems and computing. 2021;10.:567–86. doi: 10.1007/978-981-33-6862-0_45.

[10] R S, M S, Hasan MK, Saeed RA, Alsuhibany SA, Abdel-Khalek S. “An Empirical Model to Predict the Diabetic Positive Using Stacked Ensemble Approach, ” Frontiers in Public Health. 2022 Jan;9. doi: 10.3389/fpubh.2021.792124.

[11] Choubey R, Gautam P. Supervised ensemble classifier algorithm for prediction of liver disease, lung cancer and brain stroke. Int J Health Sci (Qassim). 2022;10.:9581–92. doi: 10.53730/ijhs.v6ns4.11241.

[12] Meng L, Treem W, Heap GA, Chen J. A stacking ensemble machine learning model to predict alpha-1 antitrypsin deficiency-associated liver disease clinical outcomes based on UK Biobank data. Sci Rep. 2022 Oct;12(1). doi: 10.1038/s41598-022-21389-9.

[13] Feng J, others. “A hybrid stacked ensemble and Kernel SHAP-based model for intelligent cardiotocography classification and interpretability, ” BMC Medical Informatics and Decision Making. 2023 Nov;23(1). doi: 10.1186/s12911-023-02378-y.

[14] Oliullah K, Rasel MH, Islam MdM, Islam MdR, Wadud MdAH, Whaiduzzaman Md. A stacked ensemble machine learning approach for the prediction of diabetes. J Diabetes Metab Disord. 2023 Nov 22;23(1):603–17. doi: 10.1007/s40200-023-01321-2.

[15] Zheng A, Casari A. Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. O’Reilly Media, Inc.; 2018.

[16] Verdonck T, Baesens B, Óskarsdóttir M, vanden Broucke S. Special issue on feature engineering editorial. Mach Learn. 2024 Jul 6;113(7):3917–28. doi: 10.1007/s10994-021-06042-2.

[17] Lundberg SM, Lee SI. “A Unified Approach to Interpreting Model Predictions. ” arXiv org, May. 2017;22. doi: 10.48550/arXiv.1705.07874.

[18] Hossen MdS, Haque I, Sarkar PR, Islam MdA, Fahim WA, Khatun T. Examining The Risk Factors of Liver Disease: A Machine Learning Approach. In: 7th International Conference on Communication and Electronics Systems (ICCES). IEEE; 2022. p. 1249–57. doi: 10.1109/ICCES54183.2022.9835732.

[19] Singh J, Bagga S, Kaur R. Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques. Procedia Comput Sci. 2020 Jan;167:1970–80. doi: 10.1016/j.procs.2020.03.226.

[20] MD AQ, Kulkarni S, Joshua CJ, Vaichole T, Mohan S, Iwendi C. Enhanced Preprocessing Approach Using Ensemble Machine Learning Algorithms for Detecting Liver Disease. Biomedicines. 2023 Feb;11(2):581. doi: 10.3390/biomedicines11020581.

[21] Bhusnurmath RA, Betageri S. Performance Comparison of Machine Learning and Deep Learning Algorithms for Liver Disease Detection. In: Advances in Computing and Information, ERCICA 2023. 2024. p. 337–47. doi: 10.1007/978-981-99-7622-5_23.

[22] Amin R, Yasmin R, Ruhi S, Rahman MH, Reza MS. Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms. Inform Med Unlocked. 2023;36:101155. doi: 10.1016/j.imu.2022.101155.