Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 2975~2986
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.9412

Improved non-invasive diagnosis of hepatocellular carcinoma by optimized meta classifier with hybridized features

Babitha Thamby[1], Edwin Jayakaran Thomson Fredrik[2]
[1]Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India
[2]Department of Computer Technology, Karpagam Academy of Higher Education, Coimbatore, India

Article Info

Article history:

Received Oct 6, 2024
Revised May 7, 2025
Accepted May 27, 2025

Keywords:

Deep learning
Detection
Hepatocellular carcinoma
Hybrid
PIVKA-II
Stacking classifier

ABSTRACT

Hepatocellular carcinoma (HCC), the primary cancer of the liver, is life-threatening, with few or no symptoms, and detection in the early stage will help for successful treatment with surgery, and transplant for a better life quality. Here, we proposed two stacking classification models based on deep learning with differential hybrid feature selection for the early detection of HCC using novel non-invasive biomarker PIVKA-II. We showed how the variations in hybrid feature selection affect the performance of stacking classification and different supervised machine-learning algorithms on a metaclassifier. The base layers were support vector machine (SVM), gradient boosting (GB), and linear discriminant analysis (LDA). The meta classifier was a multilayer perceptron (MLP) with three different optimizers, stochastic gradient descent (SGD), adaptive moment estimation (ADAM), and Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS). Our first model outperformed the second with their hybrid features by improving accuracy by 1.5% and F1_score by 1% in both SGD and ADAM optimization, while MLP-LBFGS had a 1.4% increase in accuracy. The precision had hiked by 1.9%, 3.5%, and 1.7% in SGD, ADAM, and LBFGS, respectively, in Model-1. Matthew's correlation coefficient (MCC) for MLP-SGD increased from 0.82 to 0.85, MLP-ADAM from 0.81 to 0.85, and MLP-LBFGS from 0.75 to 78 for the first model.

Corresponding Author:

Babitha Thamby
Department of Computer Science, Karpagam Academy of Higher Education
Coimbatore, India
Email: babitha86@gmail.com

1. INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most prevalent, asymptomatic cancers around the globe. It is the most commonly seen liver cancer around the globe, having an elevated mortality rate [1]. More patients affected are above 60 years, especially males [2], [3]. It is the fourth most frequently found cancer and the second leading cause of death in Asia that is related to cancer [4], [5]. HCC diagnosis at the beginning stage can give the best remedies like ablation therapy, resection, and liver transplantation, thereby improving the quality of life and lifespan. Our study looked into the aspects of data mining for an earlier diagnosis of HCC for a better quality of life span and to reduce global mortality.

In the landscape of HCC, non-alcoholic fatty liver disease (NAFLD) based HCC is the prevalent type of HCC in Asia [6]. The stages of NAFLD-HCC are from non-alcoholic fatty liver (NAFL) to non-alcoholic steatohepatitis (NASH), to fibrosis, to cirrhosis, and then finally to HCC [7]. Although there are invasive and non-invasive procedures for HCC diagnosis, invasive procedures are risky [8]. In most people having a mean age above 60 years, invasive methods like biopsy show post-procedure bleeding and

complications [9]. So, the age group and physical fitness should be considered, as the difficulties during procedures affect physical, emotional, and mental stability to an extent. Another issue was that the liquid biopsy sometimes fails in the prognosis of HCC due to sampling error [10]. The early diagnosis of NAFLD-HCC using non-invasive methods is trending since those methods have a massive contribution to the diagnosis of the disease, in contrast to using invasive procedures. The non-invasive blood serum markers taken via blood test are strong enough to diagnose the disease, along with the ultrasonography findings like tumour size, portal vein thrombosis, and portal hypertension [11], [12]. Going deep into the recent aspects of HCC diagnosis, we understood that the disease can be diagnosed early by combining blood serum values with ultrasonography findings [13], [14].

Among blood serum markers, alpha-fetoprotein (AFP) was the prominent biomarker traditionally used for the HCC diagnosis, with other regular attributes [15]. However, the issue was that not all tumours develop a higher level of AFP [16], [17]. Therefore, it should not be used as one of the pre-eminent biomarkers for diagnosing HCC in some patients. So we need another first-line non-invasive blood biomarker. Our investigations revealed that the novel biomarker Des-gamma Carboxy Prothrombin (DCP), also known as PIVKA-II, can be used along with AFP and other detecting attributes to detect HCC in its early stage [18], [19]. From the statistics, it is evident that PIVKA-II is positive in HCC patients even if they are negative for AFP. Elevated values of PIVKA-II in some HCC patients are seen even if AFP is negative [20], [21].

Some of the existing works of HCC diagnosis using machine learning (ML) and their results are given. The study done by Ali *et al.* [22] used a hybrid concept of linear discriminant analysis (LDA) for reducing dimension, and a genetic algorithm (GA) for support vector machine (SVM) optimizer. For classification, they used SVM, and a hybridization of three models reached 90.3% accuracy, 96.07% specificity, and 82.25% sensitivity. In another deep-learning model of HBV-related HCC, the accuracy of the diagnosis was 76.3 [23]. Another deep learning recurrent neural network model of HBC-related HCC prediction showed a mean area under the receiver operating characteristic curve (AUROC) of 0.806 [24]. A different prediction model, k-nearest neighbor (K-NN), was chosen by Liu *et al.* [25] for detecting post-resection HCC recurrence and found an accuracy of 71% and precision of 70%. Another research using a few serum markers, PIVKA-II with AFP, increased sensitivity 67%, accuracy 90%, and 100% specificity [26]. They proved the efficiency of AFP and PIVKA-II in the diagnosis of HCC in its early stage.

The issue was that all those studies were focused on a particular category of patients, like Hepatitis B or Hepatitis C. Also, AFP was the major attribute for HCC diagnosis, even if some patients were negative for AFP. They did not address HCC occurrence in non-alcoholic fatty liver patients. The scope of a stacking classification is not mentioned anywhere, especially with a hybrid feature selection. Through our study, we showed that we can diagnose HCC, particularly in non-alcoholics infected with Hepatitis B or Hepatitis C, using a novel biomarker, PIVKA-II. We explored the potential of stacking classification by using a neural network as a meta-classifier with the help of hybrid selected features.

There are four major contributions from our study. The first contribution was to diagnose HCC, especially in non-alcoholic patients, using data mining algorithms with the help of novel non-invasive biomarker PIVKA-II in combination with AFP, with ultrasonography findings like tumor size. The second contribution was the improved accuracy and true positivity compared to the recent studies. The third contribution is the design of a novel algorithm based on a deep learning stacking classifier built on conventional ML algorithms, coupled with optimization and cross-validation. The final contribution is that different variations in feature selection can boost the performance of binary classifications, especially the hybrid approach of the embedded and filter method for feature selection. Age and gender details are also listed in the attribute set, as most patients were males older than 60. In this research, we created two different ML models with differential hybrid feature selection, followed by a common stacking classifier with three different optimizations for better prediction of HCC. Then, we compared the results of both models and chose the better one as the proposed model.

2. METHOD

The research applied feature selection by splitting the dataset into training and testing. Here we constructed two models. ‘Model-1’, the first model, uses hybrid feature selection using selectKBest and selectFromModel and applying a stacking classifier as a meta classifier. In the second model, ‘Model-2’, we used hybrid feature selection using selectKBest and recursive feature elimination (RFE), and applied a stacking classifier [27]. The workflow is given in Figure 1.

Multilayer perceptron (MLP) will be the meta classifier with three different optimizations. All the operations were performed in Anaconda Navigator 2.5.2, Jupyter Notebook 6.5.4, Windows 11 operating system, using Python as the language. The dataset has undergone bootstrapping to obtain multiple subsets of the data, thereby reducing the risk of overfitting.

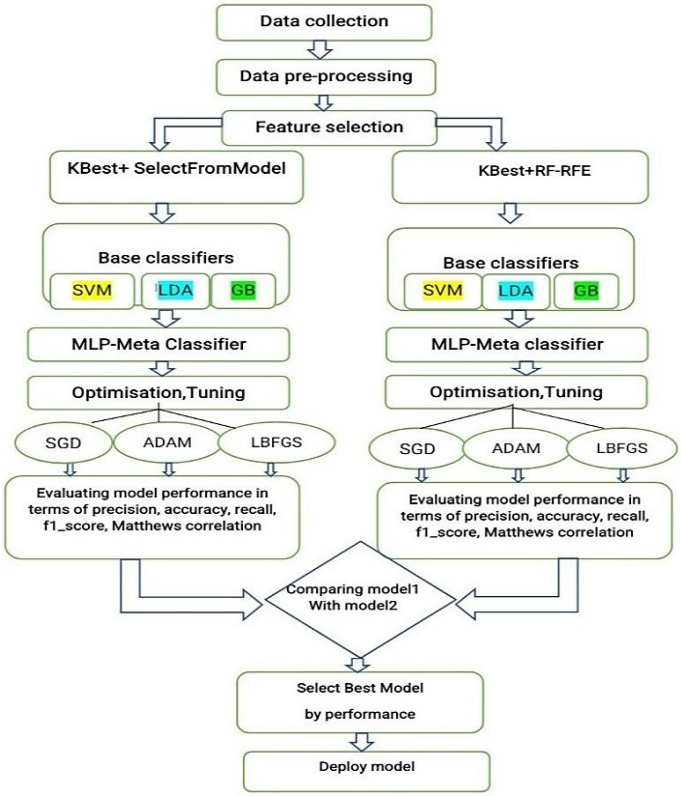

Figure 1. Overall flowchart of hybrid feature selection, stacking classification, and optimization

2.1. Dataset description

The primary data collection was done mainly from Kerala, South India. A cohort study on retrospective data was done with 364 non-alcoholic liver patients as participants, in which 233 were HCC victims and the others were HCC-negative. The data of the last three years' participants from Amala Institute of Medical Sciences and other clinics, Thrissur, Kerala, were collected. The data includes both HCC-positive as well as HCC-negative patients. The dataset contains blood biomarkers that are essential and inevitable for the detection of HCC [28]. Other pathological values, like tumor size and liver stiffness, are also taken into consideration. We gathered a dataset with 27 features as our biomarkers and one final 'CLASS' column as our target value. The HCC-positive patients were marked as '1' and others as '0'. The detailed biomarker list is given in Table 1.

Table 1. Biomarkers description in the dataset

Sl.No	Attribute	Description	Sl.No	Attribute	Description
1	AGE	Age of the participant	15	LYMPHO	Lymphocyte
2	GENDER	Sex	16	NEUTRO	Neutrophil
3	VIRUS	Hepatitis B, C	17	CREAT	Creatinine
4	TUMOUR	Tumors in the liver	18	TOT_BIL	Total bilirubin
5	PVT	Portal vein thrombosis	19	DIR_BIL	Direct bilirubin
6	PHTN	Portal hypertension	20	SGOT	Serum glutamic oxaloacetic transaminase
7	CIRRHOSIS	Liver cirrhosis	21	SGPT	Serum glutamate pyruvate transaminase
8	NASH	Non-alcoholic steatohepatitis	22	ALP	Alkaline phosphate
9	LIV_STIFF	Measurement of liver stiffness	23	A/G	Albumin-globulin ratio
10	PIVKA II	Protein induced in the absence of vitamin K Antagonist	24	NA	Sodium
11	Hb	Hemoglobin	25	K	Potassium
12	RBC	Red blood cells	26	INR	International normalized ratio
13	PLATELET	Platelet	27	HBA1c	The average blood sugar level for 3 months
14	AFP	AlphaFetoprotein	28	CLASS	HCC positive (1) or not (0)

2.2. Data pre-processing, cleaning, and missing value imputation

The dataset contains categorical and numerical data. Imputation of missing values was done using the concept of mean imputation. The research filled in the missing values of some variables by considering the

variable mean of cases that are not missing. Age and gender are the other biomarkers used in the dataset apart from the blood serum and pathology biomarkers. The categorical data takes male as '0' and female as '1' using a Label encoder module. The reason is that, as we said earlier, HCC affects more males [29] than females, especially those aged between 50 and 70. The statistics given in Figure 2 show that most patients were males above 60.

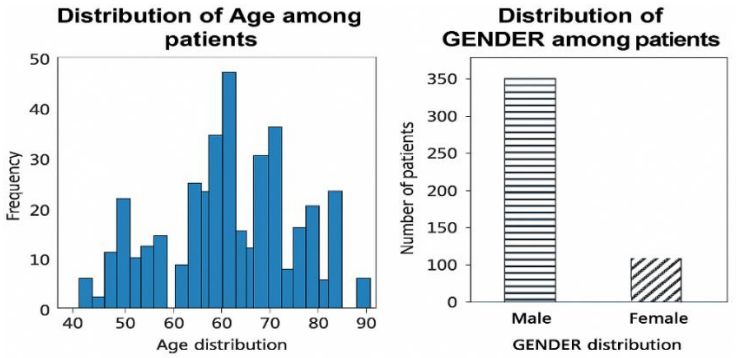

Figure 2. Frequency of age, gender among HCC-positive patients

2.3. Feature selection

To reduce the complexity and to improve the accuracy of results, we have reduced the number of features. For feature reduction, we made two parallel models. The first model (Model-1) used hybrid feature selection by combining the filter method and embedded method. That is, SelectKBest and SelectFromModel (SFM) use random forest (RF) as our estimator. It is described in section 2.3.1. The second model, Model-2, used another hybrid feature selection by combining the filter method and the wrapper method. That is, SelectKBest and RFE using RF as our estimator. It is described in section 2.3.2.

In the two models, we used the SelectKBest filter method [30] as a common method. It selects the top 'k' features and aids in focusing on the most relevant and robust features and reducing dimensionality. The k-best algorithm ranks the features based on a definite criterion and selects the top 'k' features. The score function was 'f_classif' to compute the ANOVA-F value between each feature and the target. Those features that are highly dependent on the target variable will be chosen. The following are the steps in the process:

- Step 1. The dataset is divided into a training and testing set to train and test with algorithms.
  X_train, X_test, y_train, y_test=train_test_split(X, y, test_size=0.2, random_state=42)
- Step 2. Score calculation
- Step 3. Feature ranking
- Step 4. Select top-ranked features

The main code snippet is given as (1) and (2):

```
selector_kbest=SelectKBest (f_classif, k=7)                    (1)
```

```
X_kbest=selector_kbest.fit_transform(X,y)                      (2)
```

The seven features were selected by assigning k=7 as shown in Figure 3. The two different hybrid feature selection methods using KBest are given in detail.

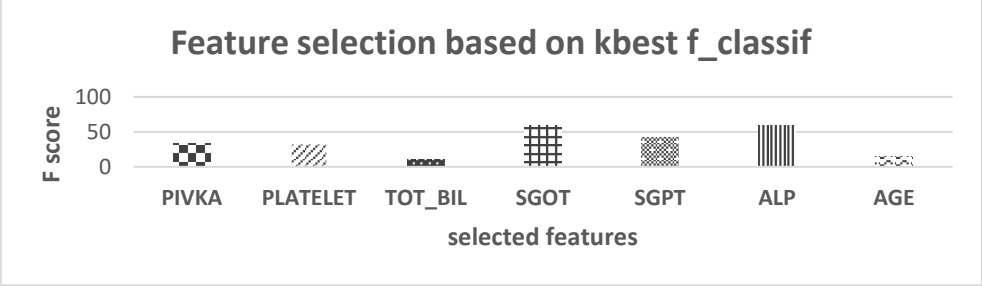

Figure 3. Features selected by KBest f_classif method using score value

2.3.1. Feature selection for Model-1 (SelectKBest+SelectFromModel)

In Model-1, we use the SelectKbest algorithm using the F-value for ranking features. The k value is the number of features to be selected using the f_classif score function. It is a simple and fast algorithm for feature ranking. But it may not capture the feature dependencies. So, we decided to integrate another algorithm and design a hybrid feature selection methodology, and integrated SFM [31], as it considers feature dependencies. It is a model-based feature selection, which can select the features mainly based on the important weights assigned to them using a learning estimator. The estimator RF will fit into the data so that the features will be selected based on the specific threshold value. Those features with weights above the threshold were selected, while others were discarded. The part of the code is given (3) to (5).

```
rf_model=RandomForestClassifier(n_estimators=100, max_features=7, random_state=42)   (3)

selector_rf=SelectFromModel(rf_model)                                                (4)

X_rf=selector_rf.fit_transform(X,y)                                                  (5)
```

The parameter ‘n_estimators’ is the number of decision trees we included in the RF (here given the value ‘100’). An increase in the value will improve the performance of the RF and the computational cost. Another parameter called ‘max_features’ refers to the maximum number of features we need to get as selected from the model. The selected 7 features were 'INR', 'TOT_BIL', 'PIVKA', 'HBA1c', 'AFP', 'ALP', and 'PLATELET' as shown in Figure 4.

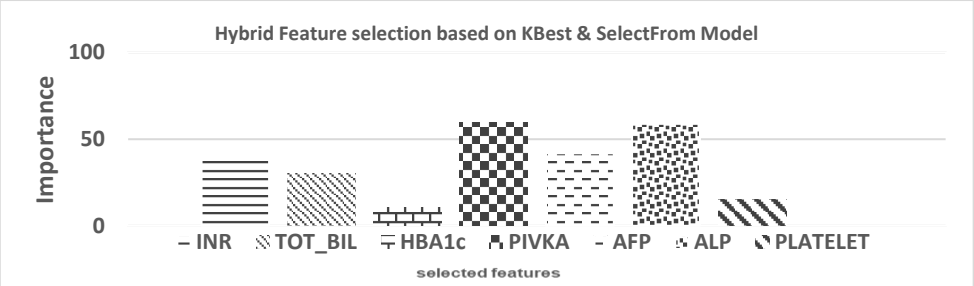

Figure 4. Features selected by KBest+select from model (SFM-RF) method in Model-1

2.3.2. Feature selection for Model-2 (SelectKBest+RFE)

Model-2 is built based on KBest and RFE hybrid feature selection. RFE [32] is a technique that removes features by recursion and builds models on the existing remaining features until the desired number of features is met. In (6), the research uses RF as our RFE estimator. The number of estimators denotes the number of trees in the forest, which is given as 500. ‘Max_features’ is the number of features under consideration; here it is ’7’ for the best split as in (7).

```
rf_model=RandomForestClassifier(n_estimators=500, max_features=10, random_state=42)   (6)

rfe_selector_rf=RFE (estimator=rf_model,n_features_to_select=7, step=1)              (7)

X_rfe_rf=rfe_selector_rf.fit_transform(X,y)                                          (8)
```

The features selected were 'PIVKA', 'PLATELET', 'NEUTRO', 'AFP', 'ALP', 'HBA1c', 'INR', 'SGPT', 'TOT_BIL', and 'SGOT' as in Figure 5.

Here we gave ‘10’ features for selection because we didn’t get PIVKA and AFP together when giving other values. Here we can see that PIVKA-II, our novel biomarker, was selected as the main feature in both the feature selection analysis, while reducing the entire 26 features into 7. Here, we can undoubtedly see that our novel biomarker is strong enough to aid in our study of diagnosis. After the feature reduction, we finally did the binary classification to know whether the patients have HCC or not.

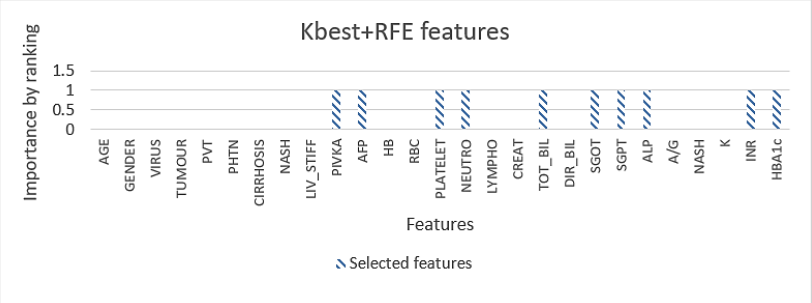

Figure 5. Features selected by the Kbest+RFE method in Model-2

2.4. Binary classification and prediction

In this paper, Model-1 and Model-2 used a stacking classifier concept to evaluate the classification performance. A stacking classifier is an ensemble learning method in which multiple models' predictions are combined after training using another model, called the meta-classifier. It aims to improve overall results by leveraging the strengths of diverse base models with a meta-classifier. Each layer consists of a set of models, and the predictions of those models can be fed into the meta-model for the next layer. This hierarchical approach will be able to capture the complex relationships in the data. A base model in stacking refers to the individual models that can be enough to form the initial layer of the ensemble. Such models will have different types or even the same type. The configuration will be different. Here in this study, we used SVM, gradient boost (GB), and LDA as the base layer prediction models. The idea behind the diverse base models is to get the different patterns or aspects in the data. Each base model can give specifically unique models of data in particular patterns, so that the research can combine them to predict a more robust and accurate overall model. The meta classifier above all layers will give an integrated output.

A validation set will be used after the creation of base layers. It will be a subset of our original dataset. It will be used to fine-tune and validate the performance of the above-said base models and the meta-model. The dataset is usually divided into three: i) training set, ii) validation set, and iii) test set. The layer of input grabs the input from the training dataset. It will contribute the same to the hidden layer. The number of nodes as input regulates the number of dataset features. Each input vector variable is dispensed to each node of the hidden layers. The hidden layer is the main computational part of the network, which uses the activation function. Weights are allotted to the edges of the hidden layer, which are multiplied by the values of the nodes. The output layer provides the output, which is already estimated.

The input node denotes the feature of the dataset. Every input node forwards the vector input value directly to the invisible hidden layer. The weighted sum is given by (9):

$$z=\sum_i w_i x_i+b_z \tag{9}$$

where X_i is the input feature, W_i is the corresponding weight, and b is the bias term.

The weighted sum 'z' is passed through an activation function to introduce non-linearity. We used activation 'ReLu' (max(0,z)) for the hidden layer. The output will be handed over to the output layer. Backpropagation is a method of fine-tuning the weights in a neural network. It is done by passing the error from the output back into the network. This improves the performance of the network and will be performed well by reducing the errors in the output. The function used in the output layer was 'sigmoid'. The loss function used was 'binary_crossentropy' for binary classification.
The binary cross-entropy loss function for MLP is (10):

$$L = -(\frac{1}{N}) \Sigma_{i=1}^{N} [y_i \, log(\hat{y}_i) + (1 - y_i) \, log(1 - \hat{y}_i)] \tag{10}$$

where yi the actual label, $\hat{y}_i$ is the predicted label, and N is the number of samples.

By backpropagation, the gradients of the loss function concerning each weight and bias were estimated. The error will pass back through layers: The weight and bias will undergo updating by going in the opposite direction of the gradient. So the loss will be reduced.

$$w = (w - \eta) * (\frac{\partial L}{\partial w}) \tag{11}$$

where w is the weight, η is the learning rate, and ƏL/Əw is the gradient of the loss function concerning the weight.

The loss was reduced by optimizing the model with the three optimizers. For MLP as a meta-classifier in the stacking method, we needed an optimization algorithm for training the neural network. Here we used three different popular optimization algorithms: stochastic gradient descent (SGD) [33], adaptive moment estimation (ADAM) [34], and Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [35]. Here is the code snippet of stacking an MLP classifier with SGD as optimizer.

```
mlp_sgd_classifier=MLPClassifier(random_state=42, solver='sgd', max_iter=1000)
```
(12)

The maximum iterations for ADAM were '700', whereas for L-BFGS it was '2000'. The learning_rate parameter for SGD for both models was given as ['constant', 'adaptive']. That means the rate of learning will be constant throughout the training, and it will be adjusted as the loss improves. We used LBFGS and ADAM instead of SGD to get other optimization results. The Grid search with five-fold cross-validation can be additionally used as a tuning tool.

```
stacking_classifier=StackingClassifier(estimators=[('svm',svm_classifier),
('lda',lda_classifier),('gb',gb_classifier)],final_estimator=mlp_sgd_classifier)
```
(13)

Hyperparameter tuning was done to find the best set of hyperparameters for the ML model to obtain optimal performance. Hyperparameters and optimal values for 3 optimizers are:

- Case1(SGD) - tuning with SGD optimizer for MLP in alpha value as 0.001 for hidden layer size (50,50) with an 'adaptive' learning rate. The tuning for Model-2 gave 0.0001 as alpha for hidden layer size (100) with a 'constant' learning rate.
- Case 2 (ADAM) - tuning with ADAM optimizer for MLP in Model-1 gave alpha value as 0.0001, hidden layer size as (50,50) and 'constant' learning rate. For Model-2, it was 0.0001 as alpha for hidden layer size (100) with a 'constant' learning rate.
- Case3(L-BFGS) - tuning with L-BFGS optimizer for MLP in Model-1 gave alpha value as 0.01, hidden layer size as (100), and 'constant' learning rate. For Model-2, it was 0.01 as alpha for hidden layer size (100) with a 'constant' learning rate.

The error rate was reduced to 0.26, improving the accuracy and robustness. The results are discussed in the following section.

3. RESULTS AND DISCUSSION

The models are evaluated concerning accuracy, precision, recall, and true positive rate. Also, Matthew's correlation coefficient (MCC), which is a statistical tool for model evaluation, is calculated. They were calculated using true positive (TP), true negative (TN), false positive (FP), and false negative (FN) terms.

$$\text{Accuracy}=(TP+TN)/(TP+FP+FN+TN) \quad (14)$$

$$\text{Precision}=TP/(TP+FP) \quad (15)$$

$$\text{Recall}=TP/(TP+FN) \quad (16)$$

$$F1 = 2 * (\frac{precision \times recall}{precision + recall}) \quad (17)$$

$$\text{MCC}=(TP*TN-FP*FN)/\text{np.sqrt}((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)) \quad (18)$$

The results obtained showed that among the two models proposed, Model-1 is showing more accuracy (MLP-SGD 93%, MLP-ADAM 93%, MLP-LBFGS 90.4%), precision for SGD, ADAM, and LBFGS were 93.9%, 93.9%, and 90.2%, respectively, with an improvement. The F1_score was 94.85 for SGD and ADAM, and 92.93 for LBFGS and F1_scores. The true positive rate increased by around 1% in MLP-ADAM. The recall rate was the same for SGD and LBFGS, except for a slight decline for ADAM. All other metrics in Model-1 performed better than in Model-2. A total of 95.8% of patients were truly identified as they are having HCC.

Model-2 obtained less accuracy, 91.8% for SGD and ADAM, and only 89% for LBFGS optimized MLP. The true positive rate was also increased for MLP-SGD by 1%, while all other methods remained the same. The performance metrics of Model-1 are described in Table 2. The performance metrics of Model-2 are given in Table 3.

Table 2. Performance comparison of three different MLP optimizations in Model-1

Feature selection	Base classifiers	Meta classifier	Accuracy	Precision	Recall	F1_score
KBest+RF-SFM	SVM, LDA, and GB	MLP-SGD	93.15	93.88	95.83	94.85
		MLP-ADAM	93.15	93.88	95.84	94.85
		MLP-LBFGS	90.41	90.2	95.83	92.93

Table 3. Performance comparison of three different MLP optimizations in Model-2

Feature selection	Base classifiers	Meta classifier	Accuracy	Precision	Recall	F1_score
KBest+RF-RFE	SVM, LDA, and GB	MLP-SGD	91.79	92	95.83	93.88
		MLP-ADAM	91.78	90.38	97.92	94
		MLP-LBFGS	89.04	88.46	95.83	91.99

Overall, the comparative study was done using both models. For a better performance evaluation, we calculated the MCC [36]. The MCC values for MLP-SGD of Model-1 and Model-2 were 0.85 and 0.82, respectively. The same for MLP-ADAM were 0.85 and 0.81, whereas for MLP-LBFGS, they were 0.78 and 0.75, respectively, as shown in Figure 6. It showed that Model-1 performs better than Model-2 again, by statistical evidence. We also tried the two models with the Kaggle HCC survival dataset [37] and found a 3% increase in accuracy, a 4% increase in F1_score, and a 10% decrease in False negative cases by Model-1. Performance metrics with 95% confidence interval for accuracy, precision, and MCC and are given in Table 4.

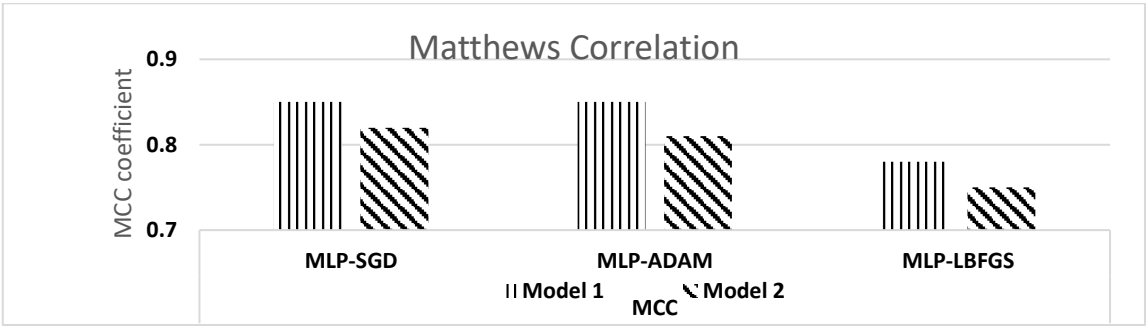

Figure 6. Performance evaluation of both models by the MCC

Table 4. Performance metrics with 95% confidence intervals

Model	Meta classifier	Accuracy (%)	Precision (%)	MCC
Model-1	MLP-SGD	92.92--93.48	93.42--93.96	0.845--0.855
Model-1	MLP-ADAM	93.02--93.28	93.71--93.99	0.845--0.855
Model-1	MLP-LBFGS	90.24--90.56	90.09--90.37	0.775--0.785
Model-2	MLP-SGD	91.55--92.01	91.78--92.18	0.815--0.825
Model-2	MLP-ADAM	91.64--91.92	90.12--90.44	0.806--0.814
Model-2	MLP-LBFGS	88.96--89.20	88.32--88.60	0.746--0.754

From the confidence interval, it is evident that Model-1 with MLP-ADAM shows the highest precision and ties with MLP-SGD for the highest MCC. MLP-LBFGS consistently underperforms across both models in all metrics. Model-1 outperforms Model-2 across all optimization algorithms, and MLP-ADAM provides the best balance of all three metrics.

For evaluating the performance ranking of the selected features from both models, we have done SHapley Additive exPlanations (SHAP) analysis and results are given in Figure 7. It was found that PIVKA and AFP are strong positive predictors. They produce higher values consistently, pushing the model toward predicting the positive class (meaning detecting HCC). The NEUTRO biomarker was highlighted with its positive prediction power in Model 2. PLATELET count has an inverse effect, as lower platelet levels contribute to a higher risk. So, it is convincing that low platelet count is a contributing factor to HCC. INR and HbA1c have nonlinear effects. Their high values increase prediction probability, low values decrease it. TOT_BIL and ALP contribute moderately important, varying per instance. SGOT and SGPT had minimal negative contributions and were not justified for the diagnosis.

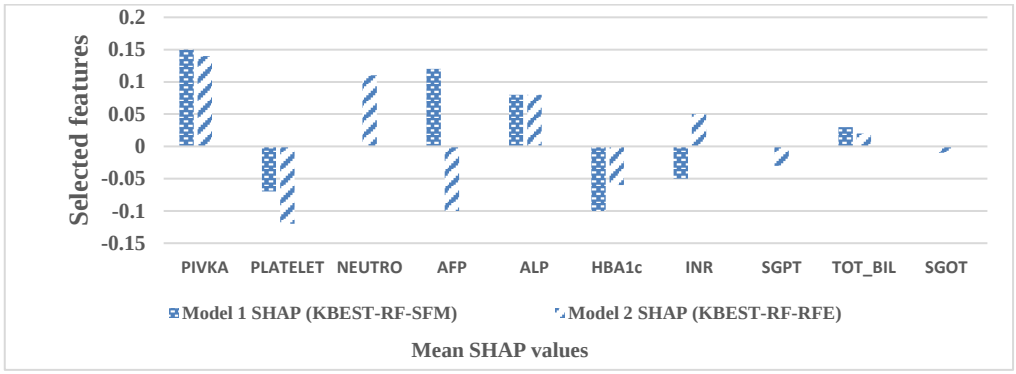

Figure 7. Feature ranking based on SHAP

For an evaluation in terms of error term, we calculated the error term, misclassification error rate (MER), as given (19):

$$MER = 1 - Accuracy \quad (19)$$

The MER for Model-1 (1-0.9315) and Model-2 (1-0.9179) were calculated as 0.0685 and 0.0821, respectively. Model-1 has a lower error rate (6.85%) vs. Model-2 (8.21%). Model-1 has higher precision and recall, meaning lower FPR and FNR, especially under MLP-ADAM and MLP-SGD. It maintains a better precision-recall trade-off (higher F1). Model-2 may be slightly overfitting or responding to noise in additional features.

To assess the statistical significance of Model-1 compared to Model-2, the Wilcoxon Signed-Rank Test was conducted by taking accuracy and F1_score across 10 folds of cross-validation, and the corresponding p-values were calculated as in Table 5.

Table 5. Significance of Model-1 by Wilcoxon Signed Rank Test

Metric	Optimizer	W Statistic	p-value	Significant (p<0.05)
Accuracy	SGD	0	0.001953	Yes
Accuracy	ADAM	0	0.001953	Yes
Accuracy	LBFGS	0	0.001953	Yes
F1_score	SGD	1	0.010862	Yes
F1_score	ADAM	17.5	0.375	No
F1_score	LBFGS	0	0.001953	Yes

For each optimization algorithm (SGD, ADAM, and LBFGS), and both evaluation metrics (accuracy and F1-score), the difference between Model-1 and Model-2 was all positive as Model-1 has higher values than Model-2. So, the W+ (sum of positive ranks) was 55.0, and W- (sum of negative ranks) was 0.0. The Wilcoxon test statistic is the smaller one (e.g.,: 0.0 for accuracy with SGD). The p-values were calculated, and they were well below 0.05, the standard threshold for statistical significance, except for the F1_score of ADAM. Since five p-values are below this threshold, the test rejects the null hypothesis that there is no difference between Model 1 and Model 2. The Wilcoxon for accuracy, Model-1 significantly outperforms Model-2 across all optimizers. For F1, the advantage of Model-1 is significant with SGD and LBFGS but not with ADAM (p>0.05).

4. CONCLUSION

Through the analysis of two models, we found that Model-1 outperforms Model-2 and the existing ML models. While applying the proposed 'Model-1' algorithm, the maximum accuracy was improved by 3%, and the F1_score by 4%, with an increase in true positivity. The method is helpful in clinical decision-making capability in healthcare diagnostic areas to identify non-alcoholic HCC from certain data in a primary stage, so that the patient can take remedial precautions and solutions for a quality lifestyle and life span improvement. When compared to standard deep learning architectures like CNNs, LSTMs, and Transformer-based models, the stacking approach is inherently better suited for tabular datasets, which are common in clinical settings. Furthermore, the model enhances interpretability through tools like SHAP, allowing for

transparent feature importance analysis, a high aspect in healthcare decision support systems where explainability is critical. In contrast, deep learning models often function as black boxes and demand significantly higher computational resources and data volume. Therefore, the proposed stacking model offers a more interpretable, resource-efficient, and equally accurate alternative to standard deep learning models.

The study also shows that HCC is a more widespread disease in human males who are between 50 and 70 years. The above proposed Model-1 shows its ability to improve accuracy, true positive rate, F1_score, and MCC, and the Wilcoxon test also proves the statistical relevance of Model-1. The hybrid feature selection using embedded and filter methods performs better with stacking classification than filter-wrapper features. The improved predictive power of stacking with diverse optimizers is also useful when dealing with complex datasets, so that we can improve the accuracy and robustness. The novel biomarker PIVKA-II is highlighted when used with the conventional biomarker AFP in HCC-positive patients. The study also shows that HCC is a more widespread disease in human males whose ages were between 50 and 70 years. The comparative SHAP analysis reveals both convergence and divergence in feature utilization between the two models and the strength of PIVKA and AFP in the early diagnosis. Such insights are valuable for improving model interpretability, stability, and performance in predictive biomedical modelling. The above proposed Model-1 shows its ability to improve accuracy, true positive rate, F1_score, and MCC. So, Model-1 can be deployed not only for HCC detection but also for the life-extending therapies and survival rate finding, as it performs well with the given data. One limitation of the current study is the absence of a standalone MLP model as a baseline. In this study, MLP was employed as the meta-classifier in the stacking ensemble to leverage the predictive power of base models, including SVM, LDA, and GB. Future work should incorporate this baseline to more clearly isolate the impact of stacking and to determine whether the complexity of the ensemble structure is justified when compared to a well-optimized individual model. The other limitation of the study was that the findings were based on a particular region of patients in South India. The sample size can be increased to get more accurate results. The extension of the study can be done with globally available data for extensive research on HCC occurrence. Also, the above-proposed algorithm can be useful in predicting HCC reoccurrence in patients after liver transplantation or ablation therapy. The survival rate prediction can also be done by doing appropriate feature engineering.

ACKNOWLEDGMENTS

We appreciate Karpagam Academy of Higher Education, Coimbatore, India, and Amala Institute of Medical Sciences, Thrissur, Kerala, India, for the support they have given to our research. No research grant or contract supported this work.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Babitha Thamby	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Edwin Jayakaran		✓			✓			✓		✓	✓	✓		
Thomson Fredrik														

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

Not applicable.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [Babitha Thamby]. The data, which contain information that could compromise the privacy of research participants, are not publicly available due to certain restrictions.

REFERENCES

[1] T. A. Addissouky *et al.*, "Latest advances in hepatocellular carcinoma management and prevention through advanced technologies," *Egyptian Liver Journal*, vol. 14, no. 1, pp. 1–18, Jan. 2024, doi: 10.1186/s43066-023-00306-3.

[2] T. C. F. Yip, H. L. Y. Chan, V. W. S. Wong, Y. K. Tse, K. L. Y. Lam, and G. L. H. Wong, "Impact of age and gender on risk of hepatocellular carcinoma after hepatitis B surface antigen seroclearance," *Journal of Hepatology*, vol. 67, no. 5, pp. 902–908, Nov. 2017, doi: 10.1016/j.jhep.2017.06.019.

[3] S. Chidambaranathan-Reghupaty, P. B. Fisher, and D. Sarkar, "Hepatocellular carcinoma (HCC): Epidemiology, etiology and molecular classification," in *Advances in Cancer Research*, vol. 149, 2021, pp. 1–61. doi: 10.1016/bs.acr.2020.10.001.

[4] A. Villanueva, "Hepatocellular carcinoma," *The New England Journal of Medicine*, vol. 380, no. 15, pp. 1450–1462, Apr. 2019, doi: 10.1056/NEJMra1713263.

[5] C.-h. Zhang, Y. Cheng, S. Zhang, J. Fan, and Q. Gao, "Changing epidemiology of hepatocellular carcinoma in Asia," *Liver International*, vol. 42, no. 9, pp. 2029–2041, Aug. 2022, doi: 10.1111/liv.15251.

[6] T. C. F. Yip, H. W. Lee, W. K. Chan, G. L. H. Wong, and V. W. S. Wong, "Asian perspective on NAFLD-associated HCC," *Journal of Hepatology*, vol. 76, no. 3, pp. 726–734, Mar. 2022, doi: 10.1016/j.jhep.2021.09.024.

[7] S. Singh, A. M. Allen, Z. Wang, L. J. Prokop, M. H. Murad, and R. Loomba, "Fibrosis Progression in Nonalcoholic Fatty Liver vs Nonalcoholic Steatohepatitis: A Systematic Review and Meta-analysis of Paired-Biopsy Studies," *Clinical Gastroenterology and Hepatology*, vol. 13, no. 4, pp. 643-654.e9, Apr. 2015, doi: 10.1016/j.cgh.2014.04.014.

[8] J. Neuberger *et al.*, "Guidelines on the use of liver biopsy in clinical practice from the British Society of Gastroenterology, the Royal College of Radiologists and the Royal College of Pathology," *Gut*, vol. 69, no. 8, pp. 1382–1403, Aug. 2020, doi: 10.1136/gutjnl-2020-321299.

[9] K. E. Maturen *et al.*, "Lack of tumor seeding of hepatocellular carcinoma after percutaneous needle biopsy using coaxial cutting needle technique," *American Journal of Roentgenology*, vol. 187, no. 5, pp. 1184–1187, Nov. 2006, doi: 10.2214/AJR.05.1347.

[10] C. Ayuso *et al.*, "Diagnosis and staging of hepatocellular carcinoma (HCC): current guidelines," *European Journal of Radiology*, vol. 101, pp. 72–81, Apr. 2018, doi: 10.1016/j.ejrad.2018.01.025.

[11] E. Shahini, G. Pasculli, A. G. Solimando, C. Tiribelli, R. Cozzolongo, and G. Giannelli, "Updating the Clinical Application of Blood Biomarkers and Their Algorithms in the Diagnosis and Surveillance of Hepatocellular Carcinoma: A Critical Review," *International Journal of Molecular Sciences*, vol. 24, no. 5, pp. 1–39, Feb. 2023, doi: 10.3390/ijms24054286.

[12] J. D. Debes, P. A. Romagnoli, J. Prieto, M. Arrese, A. Z. Mattos, and A. Boonstra, "Serum biomarkers for the prediction of hepatocellular carcinoma," *Cancers*, vol. 13, no. 7, pp. 1–13, Apr. 2021, doi: 10.3390/cancers13071681.

[13] D. Huang, J. Zhang, J. Xu, Q. Niu, and D. Zhou, "Utility of Alpha-Fetoprotein and Ultrasound in the Diagnosis and Prognosis of Patients with Hepatocellular Liver Cancer," *Journal of Multidisciplinary Healthcare*, vol. 17, pp. 1819–1826, Apr. 2024, doi: 10.2147/JMDH.S449276.

[14] A. H. Abduljabbar, "Diagnostic accuracy of ultrasound and alpha-fetoprotein measurement for hepatocellular carcinoma surveillance: a retrospective comparative study," *Egyptian Journal of Radiology and Nuclear Medicine*, vol. 54, no. 1, pp. 1–7, Feb. 2023, doi: 10.1186/s43055-023-00982-6.

[15] F. Özdemir and A. Baskiran, "The Importance of AFP in Liver Transplantation for HCC," *Journal of Gastrointestinal Cancer*, vol. 51, no. 4, pp. 1127–1132, Dec. 2020, doi: 10.1007/s12029-020-00486-w.

[16] P. Luo *et al.*, "Current Status and Perspective Biomarkers in AFP Negative HCC: Towards Screening for and Diagnosing Hepatocellular Carcinoma at an Earlier Stage," *Pathology and Oncology Research*, vol. 26, no. 2, pp. 599–603, Apr. 2020, doi: 10.1007/s12253-019-00585-5.

[17] T. Inoue and Y. Tanaka, "Novel biomarkers for the management of chronic hepatitis B," *Clinical and Molecular Hepatology*, vol. 26, no. 3, pp. 261–279, Jul. 2020, doi: 10.3350/cmh.2020.0032.

[18] D. Y. Kim *et al.*, "Utility of combining PIVKA-II and AFP in the surveillance and monitoring of hepatocellular carcinoma in the Asia-Pacific region," *Clinical and Molecular Hepatology*, vol. 29, no. 2, pp. 277–292, Apr. 2023, doi: 10.3350/cmh.2022.0212.

[19] S. Tian, Y. Chen, Y. Zhang, and X. Xu, "Clinical value of serum AFP and PIVKA-II for diagnosis, treatment and prognosis of hepatocellular carcinoma," *Journal of Clinical Laboratory Analysis*, vol. 37, no. 1, pp. 1–9, Jan. 2023, doi: 10.1002/jcla.24823.

[20] M. Kudo, "Urgent Global Need for PIVKA-II and AFP-L3 Measurements for Surveillance and Management of Hepatocellular Carcinoma," *Liver Cancer*, vol. 13, no. 2, pp. 113–118, 2024, doi: 10.1159/000537897.

[21] A. Latif *et al.*, "Validity of Prothrombin-induced Vitamin K antagonist versus Alpha-Fetoprotein (Tumor Markers) in Diagnosis of Hepatocellular Carcinoma, Using Computed Tomography scan as Gold Standard," *Pakistan Journal of Health Sciences*, pp. 158–162, Jul. 2024, doi: 10.54393/pjhs.v5i07.1804.

[22] L. Ali, I. Wajahat, N. A. Golilarz, F. Keshtkar, and S. A. C. Bukhari, "LDA–GA–SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2783–2792, Apr. 2021, doi: 10.1007/s00521-020-05157-2.

[23] J. Y. Nam, D. H. Sinn, J. Bae, E. S. Jang, J. W. Kim, and S. H. Jeong, "Deep learning model for prediction of hepatocellular carcinoma in patients with HBV-related cirrhosis on antiviral therapy," *JHEP Reports*, vol. 2, no. 6, pp. 1–7, Dec. 2020, doi: 10.1016/j.jhepr.2020.100175.

[24] G. N. Ioannou *et al.*, "Assessment of a Deep Learning Model to Predict Hepatocellular Carcinoma in Patients with Hepatitis C Cirrhosis," *JAMA Network Open*, vol. 3, no. 9, p. e2015626, Sep. 2020, doi: 10.1001/jamanetworkopen.2020.15626.
[25] C. Liu *et al.*, "A K-nearest Neighbor Model to Predict Early Recurrence of Hepatocellular Carcinoma After Resection," *Journal of Clinical and Translational Hepatology*, vol. 10, no. 4, pp. 600–607, Aug. 2022, doi: 10.14218/JCTH.2021.00348.
[26] A. Loglio *et al.*, "The combination of PIVKA-II and AFP improves the detection accuracy for HCC in HBV caucasian cirrhotics on long-term oral therapy," *Liver International*, vol. 40, no. 8, pp. 1987–1996, Aug. 2020, doi: 10.1111/liv.14475.
[27] D. Wu, X. Du, and F. Peng, "Multi-layer and multi-source features stacking ensemble learning for user profile," *Electric Power Systems Research*, vol. 229, p. 110128, Apr. 2024, doi: 10.1016/j.epsr.2024.110128.
[28] N. D. Parikh, N. Tayob, and A. G. Singal, "Blood-based biomarkers for hepatocellular carcinoma screening: Approaching the end of the ultrasound era?," *Journal of Hepatology*, vol. 78, no. 1, pp. 207–216, Jan. 2023, doi: 10.1016/j.jhep.2022.08.036.
[29] R. Nevola *et al.*, "Gender Differences in the Pathogenesis and Risk Factors of Hepatocellular Carcinoma," *Biology*, vol. 12, no. 7, pp. 1–25, Jul. 2023, doi: 10.3390/biology12070984.
[30] K. M. M. Uddin, A. Al Mamun, A. Chakrabarti, R. Mostafiz, and S. K. Dey, "An ensemble machine learning-based approach to predict cervical cancer using hybrid feature selection," *Neuroscience Informatics*, vol. 4, no. 3, pp. 1–15, Sep. 2024, doi: 10.1016/j.neuri.2024.100169.
[31] S. H. Choi, E. Kim, S. J. Heo, M. Y. Seol, Y. Chung, and H. I. Yoon, "Integrative prediction model for radiation pneumonitis incorporating genetic and clinical-pathological factors using machine learning," *Clinical and Translational Radiation Oncology*, vol. 48, pp. 1–8, Sep. 2024, doi: 10.1016/j.ctro.2024.100819.
[32] R. K. Sachdeva, K. D. Singh, P. Bathla, A. Jain, T. Choudhury, and K. Kotecha, "Empowering Hepatitis Diagnosis Using RFE Feature Selection," in *7th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2023 - Proceedings*, IEEE, Oct. 2023, pp. 1–5. doi: 10.1109/ISMSIT58785.2023.10304999.
[33] S. Shadkani, A. Abbaspour, S. Samadianfard, S. Hashemi, A. Mosavi, and S. S. Band, "Comparative study of multilayer perceptron-stochastic gradient descent and gradient boosted trees for predicting daily suspended sediment load: The case study of the Mississippi River, U.S.," *International Journal of Sediment Research*, vol. 36, no. 4, pp. 512–523, Aug. 2021, doi: 10.1016/j.ijsrc.2020.10.001.
[34] N. Calik, M. A. Belen, and P. Mahouti, "Deep learning base modified MLP model for precise scattering parameter prediction of capacitive feed antenna," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 33, no. 2, Mar. 2020, doi: 10.1002/jnm.2682.
[35] J. G. Park and S. Jo, "Approximate Bayesian MLP regularization for regression in the presence of noise," *Neural Networks*, vol. 83, pp. 75–85, Nov. 2016, doi: 10.1016/j.neunet.2016.07.010.
[36] D. Chicco, M. J. Warrens, and G. Jurman, "The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment," *IEEE Access*, vol. 9, pp. 78368–78381, 2021, doi: 10.1109/ACCESS.2021.3084050.
[37] Mrsantos, "HCC dataset," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/mrsantos/hcc-dataset. (Date accessed: Aug. 01, 2024).

BIOGRAPHIES OF AUTHORS

Babitha Thamby received a Master of Computer Applications from Calicut University, Kerala, India. She worked as an Assistant Professor for 6.5 years in Don Bosco College, Mannuthy, Kerala affiliated with Calicut University. Her area of interest is mainly data mining. Her study focuses on the contribution of data mining in healthcare. She has presented papers at national and international conferences and published papers in her research area. She can be contacted at email: babitha86@gmail.com.

Edwin Jayakaran Thomson Fredrik Professor in the Department of Computer Technology at Karpagam Academy of Higher Education, Coimbatore, holds both an MCA and a Ph.D. in Computer Science from Bharathiar University. With over two decades of teaching experience, he has mentored many Ph.D. students to completion. His research expertise lies in artificial intelligence and data mining, reflected in his four patents, numerous publications in esteemed journals, two authored books, and three contributions to the Springer Lecture Series. He can be contacted at email: thomson500@gmail.com.