Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 13, No. 4, 2025, 467-479.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

An MLP-Based Deep Neural Network Incorporating SMOTE-Tomek Approach for Robust Prediction of Liver Disorders

Elahe Moradi*

Department of Electrical Engineering, YI.C, Islamic Azad University, Tehran, Iran.

Article Info

Article History:
Received 04 April 2025
Revised 14 June 2025
Accepted 10 July 2025

DOI:10.22044/jadm.2025.16006.2719

Keywords:
Deep Learning, MLP, SMOTE-Tomek, Data Balancing, Liver disease, Machine Learning.

**Corresponding author: Elahe.Moradi@iau.ac.ir (E. Moradi).*

Abstract

Liver disorders are among the most common diseases worldwide, and their timely diagnosis and prediction can significantly improve treatment outcomes. In recent years, the application of artificial intelligence, particularly machine learning and deep learning algorithms, has gained tremendous importance in the medical field, leading to reduced healthcare costs. This study utilized the ILPD dataset from the UCI Machine Learning Repository, which comprises 583 liver patient records with 11 features. A predictive framework based on a Multilayer Perceptron (MLP) was employed to predict liver disorders. To address class imbalance in the binary classification dataset, the Synthetic Minority Oversampling Technique (SMOTE)–Tomek approach was implemented to improve data balance. Robust scaling was applied to manage the presence of outlier values. Finally, the proposed method's performance was compared with three well-known machine learning algorithms. Five-fold cross-validation was employed across all classifiers to enhance evaluation robustness. All simulations were conducted using Python. The results indicate that the proposed method achieves superior performance, with an accuracy of 90.90%, surpassing state-of-the-art approaches.

1. Introduction

Recent advancements in artificial intelligence (AI) have enabled researchers to leverage techniques such as machine learning (ML) and deep learning (DL) to enhance disease diagnosis, ultimately improving treatment strategies and preventive care [1-3]. This approach has not only reduced healthcare costs but has also significantly contributed to improving the quality of medical services, expediting disease diagnosis, and facilitating more informed clinical decision-making [4, 5].

Liver disease is one of the most dangerous and fatal diseases, according to the World Health Organization (WHO) [6]. Liver diseases can develop due to various factors, including viral infections such as hepatitis, which cause liver inflammation, as well as obesity, excessive alcohol consumption, medication overuse, and inherited genetic risks [7, 8]. Liver disorders, including fatty liver disease, fibrosis, cirrhosis, and hepatitis, may disrupt normal physiological functions and, in severe cases, can lead to patient mortality [9].

Early detection of liver disease symptoms is particularly challenging because the liver continues to function normally until significant damage has occurred [10]. Therefore, timely diagnosis is crucial for preventing complications, controlling disease progression, and reducing healthcare costs [11]. In recent years, AI algorithms have gained considerable attention as powerful tools for disease diagnosis and prediction [3, 12-14]. Recent studies indicate that classification models in ML and DL can significantly assist physicians and healthcare professionals in early diagnosis and the timely administration of appropriate treatments [15, 16].

Section 2 reviews related work. Section 3 presents the dataset related to liver disorders and the proposed methodology. Section 4 evaluates the

proposed method and analyzes the results in comparison with three well-known machine learning algorithms. Finally, the conclusion summarizes the findings and presents the final conclusions.

2. Related Works

In recent years, a substantial number of scholars have focused on employing artificial intelligence algorithms to predict, diagnose, and classify patients with liver disease.

The authors of [17] worked with the Indian Liver Patient Dataset (ILPD) to classify liver patients by leveraging integrated projection-based statistical feature extraction with machine learning (ML) algorithms. They employed Support Vector Machine (SVM), Random Forest (RF), Multilayer Perceptron (MLP), Logistic Regression (LR), K-Nearest Neighbor (KNN), and the ensemble voting classifier. They achieved the highest accuracy of 88.10% using the Random Forest (RF) classifier. The authors of [18] proposed an enhanced preprocessing framework incorporating data balancing, feature scaling, and selection to improve testing accuracy in liver disease prediction using the ILPD dataset. The study evaluated ensemble methods including XGB, Bagging, RF, ET, GB, and Stacking and reported the highest accuracy of 86.06% for RF, underscoring the approach's efficacy. The authors of [19] examined supervised machine learning models for liver disease risk prediction. Employing SMOTE with 10-fold cross-validation, they evaluated and compared various methods. Their results revealed that the Voting classifier achieved the highest testing accuracy of 80.1%, underscoring its potential for effective early detection of liver disease.

The authors of [20] developed an early detection model for liver disorders from imbalanced Liver Function Test datasets. They combined the ILPD dataset from UCI and a primary dataset from Madhya Pradesh, India. Utilizing SVM and KNN algorithms with SMOTE for data balancing, they reported improved accuracy, specificity, and precision, demonstrating the system's potential to aid healthcare practitioners in early diagnosis. In [21], researchers analyzed registry data from Denmark spanning 1996–2014 to uncover comorbidities associated with alcoholic liver disease. They applied ML models, including SVM, RF, LightGBM, and Naive Bayes, achieving an AUC of 0.89 for detecting alcoholic liver cirrhosis. Their findings highlight the potential of high-dimensional statistical techniques in predicting disease progression. In [22], the authors proposed an intelligent model for early liver disease detection using machine learning techniques. Their system achieved an accuracy of 88.4% and a miss-rate of 11.6% by evaluating various ML algorithms. This study demonstrates the efficacy of automated diagnosis in reducing diagnostic costs and expediting treatment initiation.

The authors of [23] studied ensemble methods for early liver disease prediction by employing approaches such as AdaBoost (AB), LogitBoost, BeggRep, BeggJ48, and Random Forest. They analyzed clinical attributes including total bilirubin, direct bilirubin, age, sex, total protein, albumin, and globulin ratio. Their study demonstrated that LogitBoost achieved the highest testing accuracy of 71.53%, highlighting its comparative efficacy. The authors of [24] proposed a novel classifier by extending the XGBoost model with a genetic algorithm for predicting liver disease in Indian patients. They evaluated various classification models while incorporating feature selection and eliminating outliers using an isolation forest. Their experimental results revealed that the Random Forest classifier delivered the highest accuracy of 88.00%, underscoring its superior performance. The authors of [25] conducted a comparative study of machine learning and deep learning techniques for liver disease prediction using the ILPD dataset. They implemented models including MLP, SGD, RBM with logistic regression, SVM, and Random Forest. Their findings indicated that the deep learning-based MLP model achieved the highest accuracy of 72.00%, offering valuable insights into the efficacy of different predictive methods.

The authors of [26] analyzed liver function tests to predict liver disease using classifiers such as SVM, KNN, Hard Voting Classifier (HVC), and MLP. They evaluated models based on accuracy, precision, recall, specificity, and F-score. Their findings showed that HVC achieved the highest accuracy of 78.62%, making it the most effective model for diagnosing liver disease using patient data. The study in [27] focused on chronic liver disease prediction using machine learning models, including RF, Logistic Regression, DT, SVM, and KNN. The highest accuracy of 72.00% was achieved using the Random Forest classifier. Additionally, SHAP kernels were used to enhance model interpretability, improving trust in predictions and aiding in disease diagnosis. The authors of [28] developed the StackLD framework, a stacking-ensemble model for liver disease detection. They applied XGB, LGBM, DT, KNN, and RF on the ILPD dataset, which was balanced using SMOTE. The stacking approach outperformed other models, achieving an accuracy

of 86.22%. Their study highlighted the significance of Alkphos, SGOT, and SGPT in diagnosing liver disease.

The authors of [29] examined the performance of various classification algorithms, including Bagging, IBK, J48, JRip, MLP, and Naive Bayes, on multiple medical datasets from the UCI repository. Their study compared these methods across datasets such as Breast Cancer, Chronic Kidney Disease, Hepatitis, and ILPD, providing valuable insights into the relative testing accuracy of each classifier. The authors of [30] proposed a novel approach for predicting liver disease by enhancing classification performance and severity estimation using machine learning algorithms combined with GridSearchCV. Their method was evaluated on both the ILPD and HCV datasets. For the ILPD dataset, the Extra Tree classifier achieved a testing accuracy of 80%.

Table 1 presents a comprehensive comparison between state-of-the-art methods and the proposed technique on the ILPD dataset.

Based on the literature review, a significant research gap persists in effectively addressing class imbalance and outlier management within liver disorder prediction datasets while leveraging deep learning architectures to enhance prediction accuracy. This study introduces an innovative technique that employs an MLP-based deep neural network, incorporating the Synthetic Minority Oversampling Technique (SMOTE)-Tomek approach, a combination not previously explored on the ILPD. The SMOTE-Tomek approach strategically balances datasets by synthesizing minority-class instances while eliminating ambiguous boundary cases through Tomek Links, thereby mitigating noise-induced bias and enhancing the model's reliability and accuracy. This technique enhances data balance, reduces overfitting risks, and achieves superior performance metrics compared to existing state-of-the-art methods, particularly in medical fields like liver disease diagnosis.

3. Materials and Methods

3.1. ILPD Data

This research leverages the Indian Liver Patient Dataset (ILPD), accessible through the UC Irvine Machine Learning Repository, to advance the understanding of liver disease diagnostics. This dataset has been extensively utilized in contemporary studies aimed at enhancing the diagnosis and prediction of liver disorders [16, 17, 19]. The ILPD dataset comprises 11 features, with 10 attributes reflecting clinical symptoms and one attribute designated as the diagnostic outcome, serving as the target variable. Comprising 583 records, the dataset includes 416 instances of liver disorder patients and 167 instances of healthy (non-liver disorder) individuals. The target variable effectively delineates the presence or absence of liver dysfunction based on the sample's feature set. A comprehensive overview of the ILPD dataset's feature specifications is systematically presented in Table 2 [31].

3.2. Data Pre-processing

Data pre-processing represents the initial phase in developing any model based on data mining techniques. The statistical information of the ILPD dataset is presented in Table 3.

To enhance data quality, which ultimately improves the performance of the proposed AI methods, pre-processing was executed in four distinct stages as outlined below:

3.2.1. Missing Value Analysis

Missing values are common in real-world datasets and can adversely affect data analysis and the performance of machine learning (ML) models. To address this issue, various methods are available, such as deleting incomplete records, imputing with mean or median values, and employing advanced algorithms like KNN imputation [32]. Given the presence of missing values in the albumin to globulin (A/G) ratio feature of the ILPD dataset, the KNN imputation method with five neighbors (k equals five) was employed. This approach leverages the similarities among samples to estimate and replace missing values, thereby enhancing data quality and improving model performance.

3.2.2. *Data Normalization*

Model inputs are characterized by varying scales, which can hinder convergence, prolong training time, and increase the frequency of weight updates. Therefore, normalization is crucial to harmonize the data and prevent bias from features with excessively large values.

While z-score and min–max methods have been widely used in previous studies [20, 21], the robust scaling method was used due to the significant presence of outliers in the ILPD dataset. This approach relies on the median and quartiles instead of the mean to reduce the impact of outliers. It scales the data according to the interquartile range (IQR), defined as the difference between the 25th and 75th percentiles [33, 34]. The robust scaling equation is mathematically expressed as follows:

$$x_{robust} = \frac{x - Q_1}{Q_3 - Q_1} \quad (1)$$

Table 1. Comparison of the state-of-the-art methods and the proposed technique on the ILPD.

No.	Reference	Year	Classifier	Robust	SMOTE-Tomek	G-mean	Accuracy (%)
1	[17]	2023	RF, SVM, KNN, MLP, LR, Ensemble	×	×	×	88.10
2	[18]	2023	XGB, Bagging, RF, ET, GB, Stacking ensemble	✓	×	×	86.06
3	[19]	2023	Voting classifier	×	×	×	80.10
4	[21]	2020	SVM, RF, LGBM and NB	×	×	✓	89.00*
5	[24]	2020	RF, DT, KNN, LR, MLP, GB, XGB, AB, and LGBM	×	×	×	88.00
6	[25]	2022	MLP, RBM, SVM, SGD, and RF	×	×	×	72.00
7	[26]	2023	HVC, SVM, MLP, KNN				78.62
8	[27]	2024	RF, KNN, DT, and SVM	×	×	×	72.00
9	[28]	2024	XGB, LGBM, DT, KNN, RF, KNN	×	×	×	86.22
10	[30]	2023	ET, SVC, DT, GB, KNN, MLP, and BC	×	×	×	80.00
The proposed technique			MLP, RF, SVC, ET, and KNN	✓	✓	✓	**90.90**

Cross (×) sign indicates 'no', Tick (✓) sign indicates 'yes'

* AUC was used instead of accuracy.

Table 2. The ILPD data description.

No.	Feature	Description	Type	Range
1	Age	Patient age	Numeric	4-90
2	Gender	Patients gender (Male or Female)	Nominal	0/1
3	TB	Total Bilirubin	Numeric	0.4-75
4	DB	Direct bilirubin	Numeric	0.1-19.7
5	Sgpt	Alanine Aminotransferase	Numeric	10-2000
6	Sgot	Aspartate Aminotransferase	Numeric	10-4929
7	Alkphos	Alkaline phosphatase	Numeric	63-2110
8	ALB	Albumin	Numeric	0.9-5.5
9	TP	Total Protein	Numeric	2.7-9.6
10	A/G	Albumin and globulin ratio	Numeric	0.3-2.8
11	Selector Field	Selector which classifies the liver disorder	Nominal	1/2

Table 3. The ILPD data statistics.

Feature	Mean	Stdv	25%	50%	75%
Age	44.74	16.18	33.00	45.00	58.00
TB	3.29	6.20	0.80	1.00	2.60
DB	1.48	2.80	0.20	0.30	1.30
Sgpt	80.71	182.62	23.00	35.00	60.50
Sgot	109.91	288.91	25.00	42.00	87.00
Alkphos	290.57	242.93	175.50	208.00	298.00
ALB	3.14	0.79	2.60	3.10	3.80
TP	6.48	1.08	5.80	6.60	7.20
A/G	0.94	0.31	0.70	0.93	1.10

3.2.3. *Data Balancing*

The ILPD dataset exhibits a class imbalance, with a significantly higher number of samples diagnosed with liver disorder compared to healthy samples (a ratio of 2.491). This imbalance can negatively impact the performance of classification models. To mitigate this challenge, several techniques, including SMOTE, ROS, RUS, and Adasyn, have been applied to medical datasets [1, 3]. Previous research on liver diseases [19, 20, and 35] has commonly utilized the SMOTE method to balance the ILPD dataset. However, the SMOTE-Tomek method is employed in [36] for data balancing. The primary distinction between SMOTE-Tomek and SMOTE lies in the former's dual approach: it not only oversamples the minority class using SMOTE but also leverages the Tomek Links algorithm to eliminate majority class samples situated near decision boundaries. This process enhances the quality of the balanced dataset and minimizes class

overlap. According to [36], SMOTE-Tomek outperforms other methods, particularly in highly imbalanced medical datasets, owing to its capacity to reduce noise and improve class separability. This advantage has also been validated in achieving more precise predictions in medicine. Detailed information about the original ILPD dataset and its post-balancing state is provided in Table 4. Additionally, Figure 1 illustrates this data using a bar plot for improved visual comparison.

Table 4. Original and balanced ILPD.

ILPD	Minority		Majority		Sum
	No. Sample (%)		No. Sample (%)		
Original	167	28.65 %	416	50 %	583
Balanced	403	71.35 %	403	50 %	806

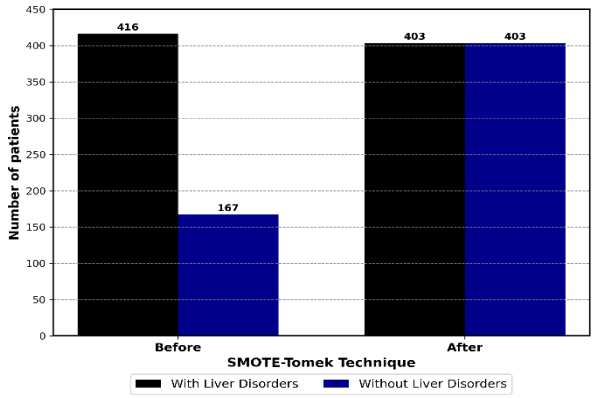

Figure 1. Distribution of the ILPD before and after applying the SMOTE-Tomek.

3.2.4. *Splitting ILPD*

In this subsection, random sampling was employed to allocate 80% of the data for training and 20% for testing the classifiers. For preprocessing the ILPD, prominent Python libraries such as Pandas, NumPy, Seaborn, Matplotlib, and Scikit-learn were utilized. Furthermore, to balance the ILPD, the SMOTE-Tomek method from the imbalanced-learn package was applied.

3.3. Proposed Hybrid AI Methods

This research proposes a hybrid AI method that synergistically combines advanced data pre-processing techniques with established machine learning (ML) and deep learning (DL) algorithms to improve classification performance on datasets characterized by class imbalance and outliers. Specifically, we employ the SMOTE-Tomek method to address class imbalance by simultaneously oversampling the minority class and under-sampling the majority class, thus creating a more balanced dataset. Additionally, Robust Scaling, which leverages the median and interquartile range to normalize features, is utilized to mitigate the adverse effects of outliers. These pre-processing techniques are applied to four AI classifiers as stated below.

3.3.1. *Robust SVM (RSVM)*

The Support Vector Machine (SVM), a robust classification technique widely acknowledged for its ability to delineate class boundaries through an optimal hyperplane [3], forms the foundation of the Robust SVM (RSVM). Traditional SVMs, however, are susceptible to distortions caused by outliers and imbalanced datasets, which can compromise their effectiveness. To overcome these limitations, RSVM integrates Robust scaling, normalizing features based on their median and interquartile range to diminish the influence of extreme values. Furthermore, the SMOTE-Tomek technique is incorporated to balance the ILPD. This dual pre-processing approach strengthens RSVM's capacity to construct a reliable decision boundary, thereby improving its generalization across diverse and challenging datasets.

3.3.2. *Robust Extra Tree (RET)*

The Extra Tree classifier, also known as Extremely Randomized Trees, is an ensemble technique that constructs multiple decision trees by employing randomized feature splits and utilizing the entire dataset for each tree, offering computational efficiency and potential performance gains [29]. To augment its resilience, the Robust Extra Tree (RET) incorporates Robust scaling, which normalizes feature values to minimize the disruptive impact of outliers on split decisions. Additionally, SMOTE-Tomek is applied to ensure a balanced class distribution, preventing the ensemble from disproportionately favoring the majority class. By synthesizing these pre-processing steps with the inherent strengths of the Extra Tree framework, RET achieves enhanced classification accuracy, particularly for minority class instances, making it well-suited for datasets with anomalies.

3.3.3. Robust KNN (RKNN)

The K-Nearest Neighbors (KNN) algorithm, a straightforward instance-based learning method that assigns class labels based on the majority vote of an instance's nearest neighbors [3], is adapted into the Robust KNN (RKNN) to address its sensitivity to feature scaling and class imbalance. In RKNN, Robust scaling is employed to standardize feature values, ensuring that distance computations remain consistent and reliable even in the presence of outliers. Concurrently, the SMOTE-Tomek technique balances the dataset,

enriching the neighbourhood of each instance to reflect both classes equitably. This pre-processing enhances RKNN's ability to accurately classify minority class samples, thereby overcoming the limitations of traditional KNN and improving its performance on complex datasets.

3.3.4. *Robust MLP-based Deep Neural Network (RMLP)*

The Multilayer Perceptron (MLP), a deep neural network architecture capable of capturing intricate patterns through multiple interconnected layers of neurons [3], serves as the basis for the Robust MLP-based Deep Neural Network (RMLP). While MLP excels in modeling complex relationships, its performance can be hindered by imbalanced datasets and outliers. To address these challenges, RMLP employs Robust scaling to normalize input features, reducing the impact of extreme values on the network's learning dynamics. Additionally, SMOTE-Tomek is utilized to equilibrate the training data, enabling the network to learn effectively from both majority and minority classes. This hybrid approach enhances RMLP's robustness and predictive accuracy, particularly for underrepresented samples, making it a powerful tool for classification tasks in the presence of data anomalies.

In this study, TensorFlow and Keras were employed to implement the Robust Multilayer Perceptron (RMLP) model, integrating dropout and L2 regularization techniques to enhance performance. Dropout functions by randomly deactivating a subset of neurons during training, which helps prevent overfitting by reducing inter-neuron dependencies. This method confirms that the model does not rely too heavily on specific neurons, promoting better generalization to unseen data [37]. L_2 regularization, commonly referred to as weight decay, incorporates a penalty term into the loss function that is proportional to the sum of the squared values of the model's weights. This technique encourages the model to maintain smaller weight values, thereby simplifying the model and reducing the risk of overfitting. By mitigating the presence of excessively large weights, L_2 regularization enhances the model's capacity to generalize effectively to unseen data. The combined application of dropout and L_2 regularization in the RMLP model significantly improved its predictive accuracy, demonstrating the effectiveness of these regularization methods in refining deep learning models for complex data scenarios [37].

3.4. Evaluation Metrics

This subsection presents the metrics employed to evaluate the performance of the algorithms proposed in Subsection 3.3. The evaluation criteria utilized in this study to assess the classifiers' performance encompass widely recognized metrics, including accuracy, precision, confusion matrix, recall, F1-score, and the area under the curve (AUC), as referenced in [38, 39, 40]. Additionally, given the imbalanced nature of the ILPD dataset, the G-mean metric has been incorporated alongside these conventional measures to provide a more comprehensive assessment of the classifiers' effectiveness.

The G-mean (Geometric Mean) serves as a critical evaluation metric for assessing classifier performance on imbalanced datasets, particularly within the domain of medical diagnostics. Defined as the square root of the product of sensitivity and specificity, G-mean offers a balanced indicator of a model's capability to accurately classify instances from both the minority and majority classes. This metric proves especially valuable in scenarios where misclassification of minority class instances, such as patients with a disease, carries significant consequences, offering a more reliable alternative to traditional accuracy [41]. Table 5 illustrates the confusion matrix.

Table 5. Confusion Matrix for Binary Classification

		Actual Class	
		Actual Positive	**Actual Negative**
Predicted Class	**Classified Positive**	**TP**	**FP**
	Classified Negative	**FN**	**TN**

In binary classification, the terms TP, TN, FP, and FN are defined as follows:

- True Positive (TP): The number of instances correctly identified as positive.
- True Negative (TN): The number of instances correctly identified as negative.
- False Positive (FP): The number of instances incorrectly identified as positive.
- False Negative (FN): The number of instances incorrectly identified as negative.

The formulas corresponding to the aforementioned metrics are provided below:

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (5)$$

$$Gmean = \sqrt{Specificity \times Recall} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (7)$$

4. Results and Discussions

During the pre-processing stage of the ILPD dataset, missing value handling and outlier management were initially applied. Subsequently, data normalization was performed using the Robust Scaling method, which is well-suited for datasets with outliers and commonly applied in medical applications. Considering the imbalanced nature of the ILPD dataset, the SMOTE-Tomek technique was employed for data balancing, increasing the number of samples from 583 to 806. The dataset was then split into training and testing sets in an 80:20 ratio, comprising 644 training samples and 162 testing samples, respectively. The hybrid AI classifiers introduced in Section 3.3 were subsequently applied to the dataset.

Hyperparameter tuning is a key element in the development of AI algorithms. The following details the tuning process for the proposed AI classifiers:

RSVM Classifier: The kernel type was identified as a pivotal hyperparameter. Its effect on accuracy was assessed by varying the kernel (Linear, Polynomial, Sigmoid, and radial basis function (rbf) kernel) while maintaining other hyperparameters at default values. The effect on accuracy is illustrated in Table 6 and Figure 2, from which the optimal kernel yielding the highest accuracy was determined to be the rbf kernel.

Table 6. Hyperparameter of the RML Classifiers.

Classifier	Hyperparameter		Accuracy
RSVM	Kernel Type	Optimal	
	Linear, poly, rbf, sigmoid	rbf	74.54 %
RET	Max_depth Range: [2-50] k-value	28	83.24 %

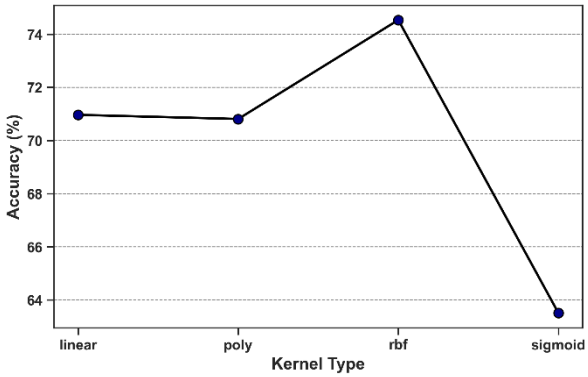

Figure 2. Effect of Kernel Type on Accuracy for the RSVM Classifier.

RET Classifier: For the RET classifier, maximum depth is a significant hyperparameter. By holding other hyperparameters constant, only the impact of variations in maximum depth on accuracy was evaluated and is illustrated in Table 6 and Figure 3. The optimal maximum depth was determined to be 28.

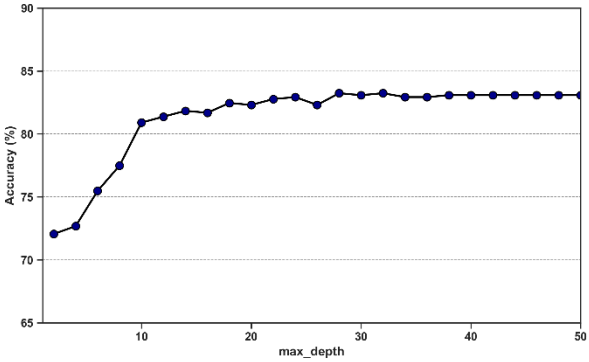

Figure 3. Effect of Maximum Depth on Accuracy for the RET Classifier.

RKNN Classifier: In the RKNN classifier, the value of k is an essential hyperparameter. With other parameters fixed, the effect of changing k on accuracy was examined and is shown in Table 6 and Figure 4, leading to the determination of the optimal k value, which was found to be 3.

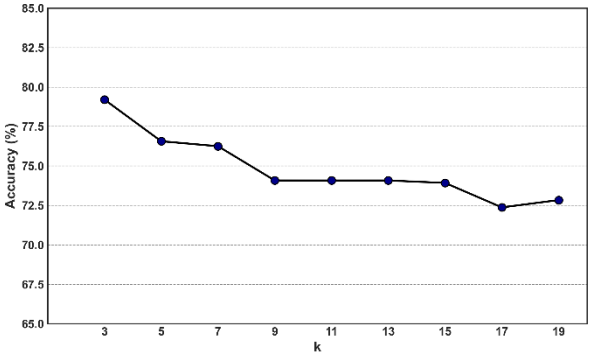

Figure 4. Effect of the k Value on Accuracy for the RKNN Classifier.

RMLP Classifier: For the proposed hybrid classifier RMLP, hyperparameter values along with changes in the optimizer were evaluated based

on accuracy and the loss function, using the binary cross-entropy loss function as the criterion. Figures 5 and 6 display the curves of accuracy and loss concerning different optimizers (Adam, SGD, RMSprop, Adadelta, Nadam, and Ftrl), from which Adam was identified as the optimal optimizer in both accuracy and loss curves. The hyperparameters of the proposed RMLP-based neural networks are presented in Table 7.

Table 7. Hyperparameters of the RMLP Classifier.

Hyperparameter	Value
Number of Hidden Layers	3
Learning rate	0.008
Activation Function (Hidden Layers)	ReLU
Activation Function (Output Layer)	Sigmoid
Optimizer	RMSprop
Dropout	0.1
Loss Function	Binary cross-entropy
Epochs	300
Batch Size	32

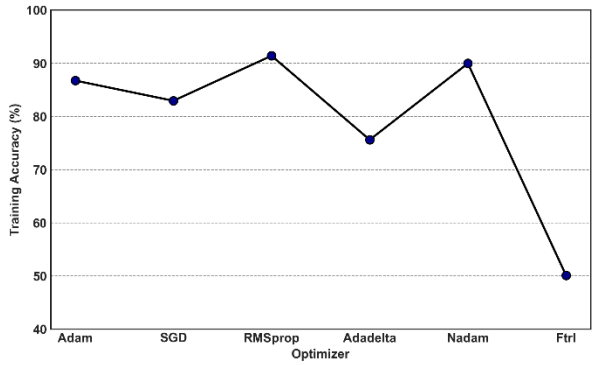

Figure 5. Accuracy Curve of the RMLP Classifier Across Different Optimizers.

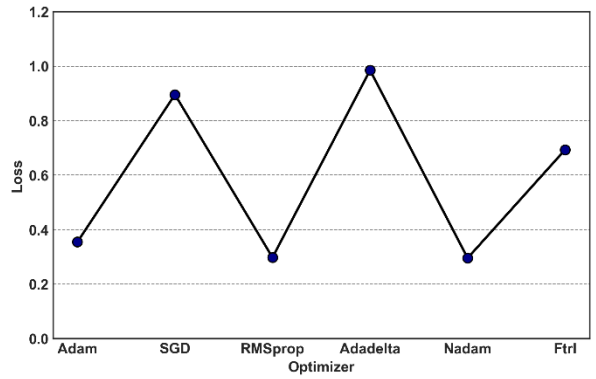

Figure 6. Loss Curve of the RMLP Classifier Across Different Optimizers.

The $RelU$ activation function, used within the RMLP classifier, is defined as:

$$RelU(x) = \max(0, x) \quad (8)$$

where, x is the input. If x is greater than 0, so $RelU(x)$ equals x and otherwise, it equals zero.

The binary cross-entropy loss function is formulated as:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (9)$$

where, y_i is the true label for the i-th sample, and $\hat{y}_i$ is the predicted probability of the positive class for the i-th sample. Also, N is the total number of samples in the dataset.

The output layer of the RMLP classifier employs the sigmoid activation function, expressed as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

here, x is the input to the function, and $\sigma(x)$ is the output of the sigmoid function.

Table 8 compares the accuracy, precision, recall, and F1-score metrics for the four proposed hybrid AI classifiers. A comprehensive comparison is illustrated in the bar plot presented in Figure 7.

Table 8. Comparative Performance Metrics for the Proposed AI Classifiers.

Proposed Classifier	Evaluation Metrics			
	Accuracy	Precision	Recall	F1-score
RSVM	74.54 %	68.93 %	90.06 %	78.03 %
RET	83.24 %	79.03 %	90.69 %	84.41 %
RKNN	79.20 %	73.66 %	91.61 %	81.54 %
RMLP	90.90%	88.07%	95.34 %	91.38 %

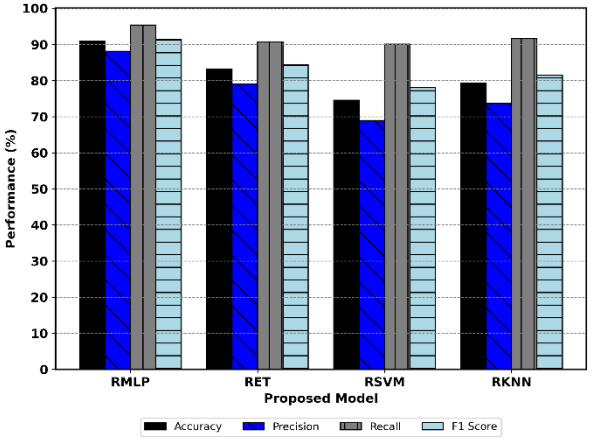

Figure 7. Comparing Accuracy, Precision, Recall, and F1-score Across the Proposed Hybrid Classifiers.

Furthermore, Table 9 presents additional evaluation metrics, including AUC, G-mean, p-value as a statistical significance indicator, and training time, providing a more comprehensive performance comparison across classifiers.

In addition, a t-test was conducted on the accuracy scores obtained through 5-fold cross-validation to statistically validate the classification results. The

resulting p-values, reported in Table 9, indicate that the proposed classifiers perform significantly better than the defined significance threshold ($\alpha = 0.05$). These analyses further strengthen the robustness of the findings and underscore the meaningful differences among the compared models.

Table 9. Comparison of AUC, G-mean, p-value, and training time for the proposed classifiers.

Criteria	Proposed Classifier			
	RSVM	RET	RKNN	RMLP
AUC (%)	81.50	92.79	85.33	96.78
G-mean (%)	72.71	82.83	78.00	90.65
P-value	0.00016	0.00002	0.00002	0.00003
Training Time	0.2576	0.6191	0.0029	16.5265

The corresponding linear curves for these metrics across the classifiers are shown in Figure 8.

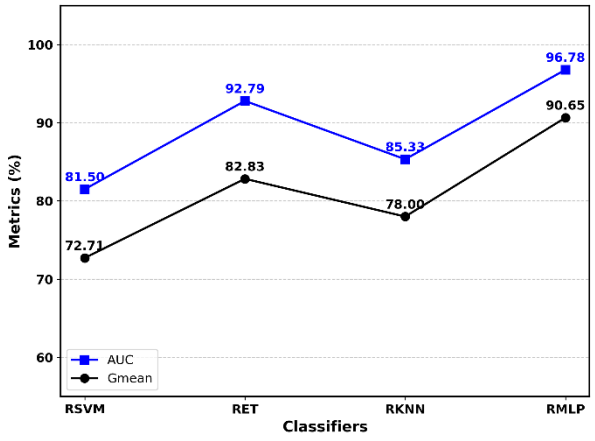

Figure 8. Comparing AUC and G-mean Metrics for the Proposed Hybrid Classifiers.

Furthermore, the performance metrics of accuracy, precision, recall, and F1-score for each classifier are individually plotted in Figures 9, 10, 11, and 12, respectively, for enhanced visualization.

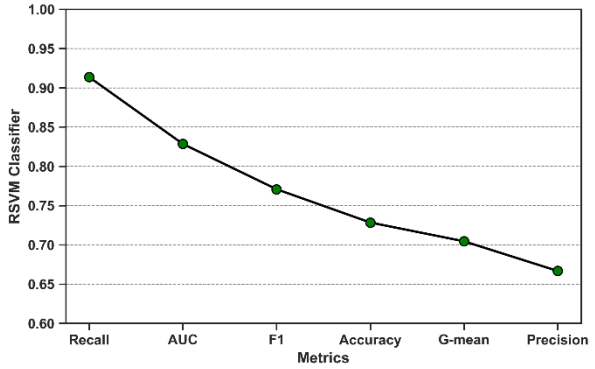

Figure 9. Performance Metrics of the RSVM Classifier.

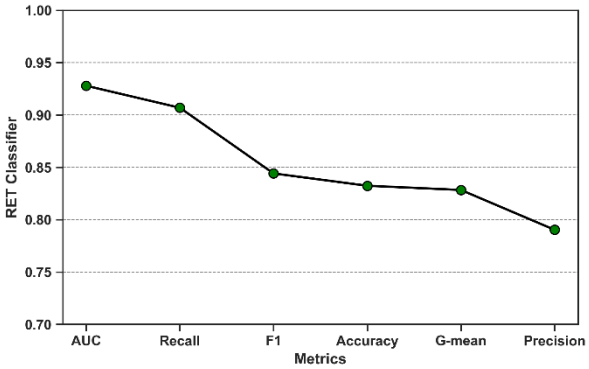

Figure 10. Performance Metrics of the RET Classifier.

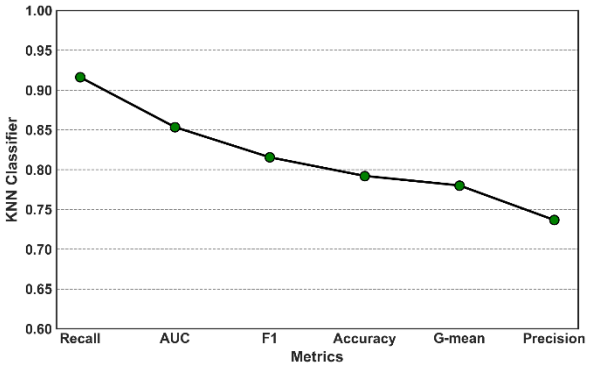

Figure 11. Performance Metrics of the RKNN Classifier.

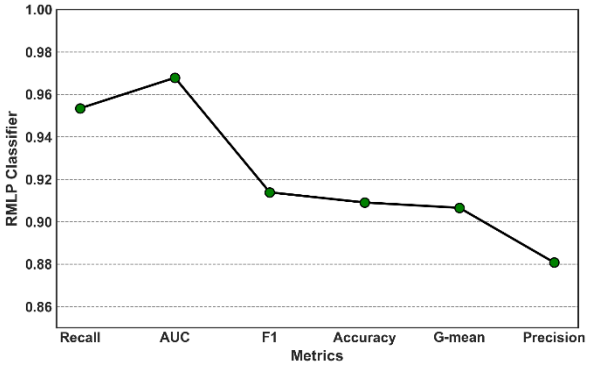

Figure 12. Performance Metrics of the RMLP Classifier.

The relative importance of the input features was evaluated using the ANOVA F-test, as illustrated in Figure 13. The analysis reveals that features such as DB, TB, SGPT, and SGOT exhibit the highest F-scores (135.46, 128.48, 122.66, and 118.37, respectively), indicating a strong statistical relationship with the target variable. These features are therefore considered to play a more significant role in classifying liver disorder status. Mid-ranked features such as Alkphos (86.09), A/G Ratio (27.17), ALB (17.21), and Age (11.67) also contribute to the classification task to varying degrees. In contrast, features such as Gender (1.12) and TP (0.22) demonstrated the lowest F-scores. An increased F-score reflects a greater dependency of the features on the target variable.

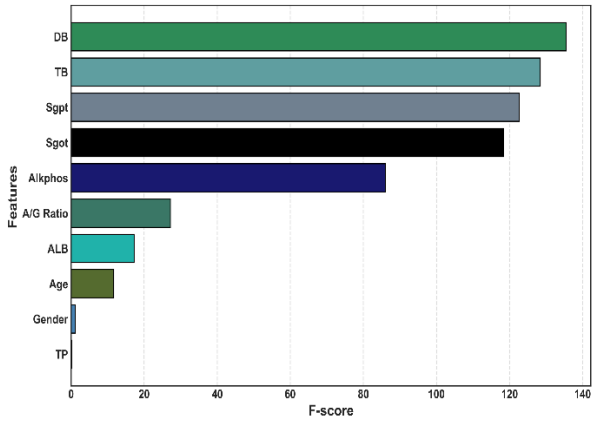

Figure 13. Feature importance based on ANOVA F-test scores.

To further evaluate the performance of the proposed hybrid classifiers, confusion matrices are presented in Figures 14, 15, 16, and 17 for the RSVM, RET, RKNN, and RMLP methods, respectively.

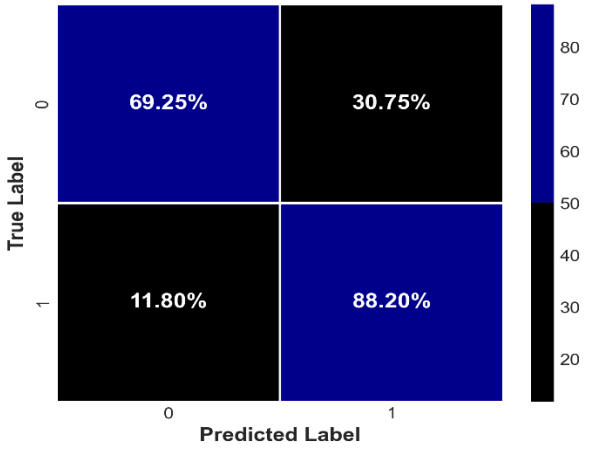

Figure 14. Confusion Matrix of the RSVM.

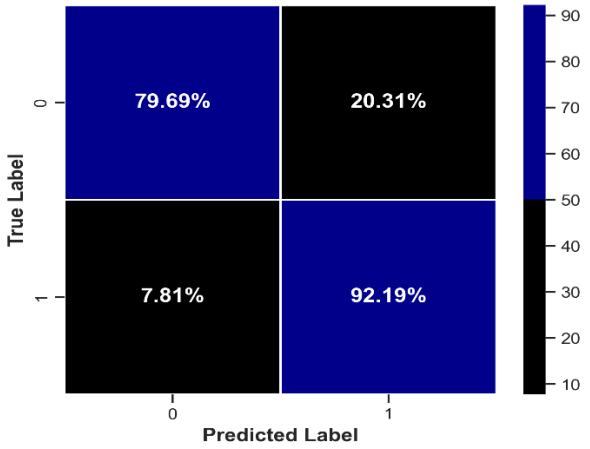

Figure 15. Confusion Matrix of the RET.

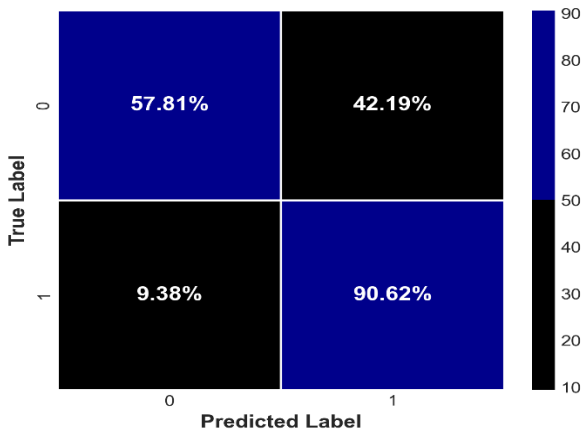

Figure 16. Confusion Matrix of the RKNN.

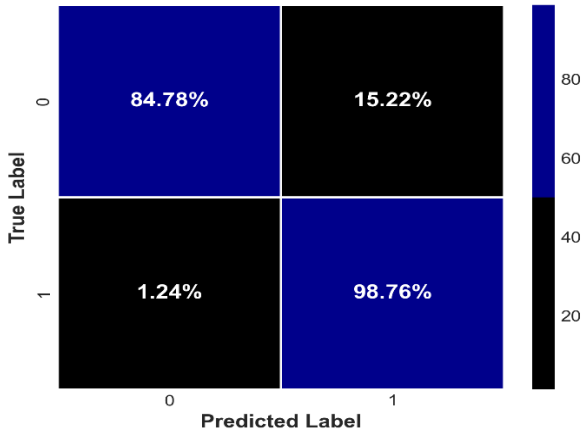

Figure 17. Confusion Matrix of the RMLP.

The RMLP-based model demonstrated superior performance across all evaluation metrics, including accuracy, precision, recall, F1-score, AUC, and G-mean, when compared to the baseline classifiers. Specifically, the RMLP model yielded an accuracy of 90.90 %, precision of 88.08 %, recall of 95.34 %, F1-score of 91.38 %, AUC of 96.78 %, and G-mean of 90.65%, indicating its robust classification capability. The model's behavior is further illustrated in Figures 18 and 19, which show the accuracy and loss curves across epochs and confirm its stable convergence on the ILPD dataset.

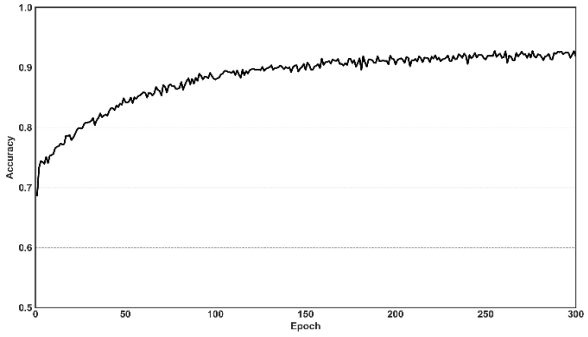

Figure 18. Accuracy Curve of the RMLP Classifier Over Epochs.

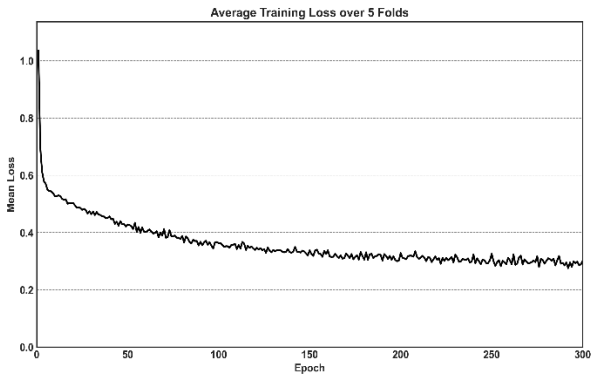

Figure 19. Loss Curve of the RMLP Classifier Over Epochs.

Although the results of this study are promising, two important limitations should be noted. First, the number of samples in the ILPD dataset is limited. While techniques such as SMOTE–Tomek balancing, dropout regularization, and cross-validation have been applied to reduce the risk of overfitting, it is expected that using larger datasets, especially in deep learning applications, would further improve model performance. Second, since the proposed model was evaluated solely on the ILPD dataset, using real-world clinical data in future studies could provide a valuable opportunity to enhance the generalizability of the results.

5. Conclusion

This research introduces an innovative hybrid framework for liver disorder prediction, combining a multilayer perceptron (MLP)-based deep neural network with advanced preprocessing techniques. Employing the SMOTE-Tomek method to mitigate class imbalance and robust scaling to handle outliers, the proposed Robust MLP (RMLP) classifier was assessed using the Indian Liver Patient Dataset (ILPD). A five-fold cross-validation strategy was employed to ensure robust evaluation, enhancing the reliability and generalizability of the results.

The model achieved superior performance, attaining an accuracy of 90.90%, precision of 88.07%, recall of 95.34%, F1-score of 91.38%, AUC of 96.78%, and G-mean of 90.65%, outperforming existing state-of-the-art methods. Statistical analysis further confirmed the significance of these results, with a p-value < 0.05, underscoring the model's robustness and reliability. These findings emphasize the potential of combining deep learning models with robust preprocessing techniques as an effective method to address prevalent challenges in medical datasets, particularly class imbalance and outlier sensitivity. The RMLP model's ability to provide consistent and precise predictions positions it as a promising tool for clinical decision-making, enabling early and accurate detection of liver conditions to enhance patient outcomes. Future research could explore more advanced deep learning architectures and apply metaheuristic optimization techniques for hyperparameter tuning to further enhance classification performance.

References

[1] E. Moradi, "Comparative Analysis of Tree-Based Machine Learning Algorithms on Thyroid Disease Prediction Using ROS Technique and Hyperparameter Optimization," *Journal of AI and Data Mining (JAIDM)*, vol. 12, no. 4, pp. 511-520, 2024.

[2] J. N. Kather, T. P. Schluechter, D. Nieratzki, V. Papanikolaou, J. Calderaro, and F. Marquardt, "Artificial intelligence in liver diseases: Improving diagnostics, prognostics and response prediction," *JHEP Reports*, vol. 4, no. 4, p. 100443, 2022.

[3] E. Moradi, "Diagnosis of Coronary Heart Disease via Robust Artificial Neural Network Classifier by Adaptive Synthetic Sampling Approach," *Iranian Journal of Electrical and Electronic Engineering*, vol. 20, no. 4, pp. 55-67, 2024.

[4] A. Z. Zakariaee, H. Sadr, and M. R. Yamaghani, "A New Hybrid Method to Detect Risk of Gastric Cancer using Machine Learning Techniques," *Journal of Artificial Intelligence and Data Mining (JAIDM)*, vol. 11, no. 4, pp. 505-515, 2023.

[5] J. Wolff, J. Pauling, A. Keck, and J. Baumbach, "The Economic Impact of Artificial Intelligence in Health Care: Systematic Review," *Journal of Medical Internet Research*, vol. 22, no. 2, 2020.

[6] G. Shaheamlung, H. Kaur, and M. Kaur, "A survey on machine learning techniques for the diagnosis of liver disease," in *Proc. 2020 Int. Conf. Intell. Eng. Manag. (ICIEM)*, London, UK, 2020, pp.17–19.

[7] K. Rezaee, J. Haddadnia, and M. R. Ghezelbash, "A novel algorithm for accurate diagnosis of hepatitis B and its severity," *International Journal of Hospital Research*, vol. 3, no. 1, pp. 1–10, Mar. 2014.

[8] A. Quadir MD, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease," *Biomedicines*, vol. 11, no. 2, Feb. 2023.

[9] P. Decharatanachart, R. Chaiteerakij, T. Tiyarattanachai, and S. Treeprasertsuk, "Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis,"*BMC Gastroenterol.*, vol. 21, no. 1, pp. 1–16, 2021.

[10] P. Kumar and R. S. Thakur, "An approach using fuzzy sets and boosting techniques to predict liver disease," *Comput. Mater. Contin.*, vol. 68, no. 3, pp. 3513–3529, 2021.

[11] A. G. Mabrouk, A. Hamdy, H. M. Abdelaal, A. G. Elkattan, M. M. Elshourbagy, and H. A. Y. Alansary, "Automatic classification algorithm for diffused liver diseases based on ultrasound images," *IEEE Access*, vol. 9, pp. 5760–5768, 2021.

[12] S. A. Bashiri Mosavi, O. Khalaf Beigi, and A. Mahjoubifard, "Designing a Visual Geometry Group-based Triad-Channel Convolutional Neural Network for COVID-19 Prediction," *Journal of Artificial Intelligence and Data Mining*, vol. 12, no. 3, pp. 423–434, 2024.

[13] Z. A. Varzaneh and S. Hosseini, "An Intelligent Fuzzy System for Diabetes Disease Detection using Harris Hawks Optimization," *Journal of Artificial Intelligence and Data Mining (JAIDM)*, vol. 11, no. 2, pp. 187–194, 2023.

[14] F. Fakouri, M. Nikpour, and A. S. Amiri, "Automatic Brain Tumor Detection in Brain MRI Images using Deep Learning Methods," *Journal of Artificial Intelligence and Data Mining (JAIDM)*, vol. 12, no. 1, pp. 27–35, 2024.

[15] R. A. Khan, Y. Luo, and F. X. Wu, "Machine learning based liver disease diagnosis: A systematic review," *Neurocomputing*, vol. 468, pp. 492–509, 2022.

[16] E. Dritsas and M. Trigka, "Supervised Machine learning models for liver disease risk prediction," *Computers*, vol. 12, no. 1, pp. 19, Jan. 2023.

[17] R. Amin, R. Yasmin, S. Ruhi, M. H. Rahman, and M. S. Reza, "Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms," *Informatics in Medicine Unlocked*, vol. 36, Dec. 2022.

[18] A. Quadir MD, S. Kulkarni, C. J. Joshua, T. Vaichole, S. Mohan, and C. Iwendi, "Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease," *Biomedicines*, vol. 11, no. 2, Feb. 2023.

[19] E. Dritsas and M. Trigka, "Supervised Machine learning models for liver disease risk prediction," *Computers*, vol. 12, no. 1, Jan. 2023.

[20] P. Kumar and R. S. Thakur, "Early Detection of the Liver Disorder from Imbalance Liver Function Test Datasets," *Int. J. Innov. Technol. Explor. Eng. (IJITEE)*, vol. 8, no. 4, pp. 179–186, Feb. 2019.

[21] D. Grissa, D. N. Rasmussen, A. Krag, S. Brunak, and L. J. Jensen, "Alcoholic liver disease: A registry view on comorbidities and disease prediction," *PLoS Computational Biology*, vol. 16, no. 9, Sep. 2020.

[22] T. M. Ghazal, A. U. Rehman, M. Saleem, M. Ahmad, S. Ahmad, and F. Mehmood, "Intelligent Model to Predict Early Liver Disease using Machine Learning Technique," in *Proc. 2022 Int. Conf. Business Analytics for Technology and Security (ICBATS)*, 2022, pp. 1–5.

[23] N. Nahar, F. Ara, Md. A. I. Neloy, V. Barua, M. S. Hossain, and K. Andersson, "A Comparative Analysis of the Ensemble Method for Liver Disease Prediction," 2*nd International Conference on Innovation in Engineering and Technology (ICIET)*, 2019, pp. 1–6.

[24] M. A. Kuzhippallil, C. Joseph, and A. Kannan, "Comparative analysis of machine learning techniques for Indian liver disease patients," *8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Mar. 2020.

[25] B. Singla, S. Taneja, R. Garg, and P. Nagrath, "Liver disease prediction using machine learning and deep learning: A comparative study," *Intell. Decis. Technol.*, vol. 16, pp. 71–84, 2022.

[26] V. Anthonysamy and S. K. K. Babu, "Multi perceptron neural network and voting classifier for liver disease dataset," *IEEE Access*, vol. 11, pp. 102149–102156, Jan. 2023.

[27] H. Raj, N. G, A. Kodipalli, and T. Rao, "Prediction of Chronic Liver Disease Using Machine Learning Algorithms and Interpretation with SHAP Kernels," *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems*, June. 2024.

[28] Md. S. H. Shaon, Md. F. Sultan, T. Karim, A. Cuzzocrea, and M. S. Akter, "An Advanced Liver Disease Detection Tool with a Stacking-Ensemble-based Machine Learning Approach," 2024 IEEE International Conference on Big Data (Big Data), Dec. 2024, pp. 6123–6131.

[29] B. V. Ramana and R. S. K. Boddu, "Performance comparison of classification algorithms on medical datasets," *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2019, pp. 0140–0145.

[30] I. M. Attiya, R. A. Abouelsoud, and A. S. Ismail, "A proposed approach for predicting liver disease," *Inf. Sci. Lett.*, vol. 12, no. 6, pp. 2447–2460, Jun. 2023.

[31] D. Dua and C. Graff, " 'UCI machine learning repository," University of California, School of Information and Computer Science Irvine, CA, USA, 2019. [Online]. Available: http://archive.ics.uci.edu/ml.

[32] S. Zhang, "Nearest neighbor selection for iteratively kNN imputation," *Journal of Systems and Software*, vol. 85, no. 11, pp. 2541–2552, Jun. 2012.

[33] E. Moradi, "A Data-Driven based Robust Multilayer Perceptron Approach for Fault Diagnosis of Power Transformers," *Proceedings of the 20th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP)*, Babol, Iran, Feb. 2024.

[34] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, "Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification," *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Aug. 2020, pp. 729–735.

[35] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: review of a decade of research, " *Artificial Intelligence Review*, vol. 57, no. 10, Sep. 2024.

[36] G. Yang, G. Wang, L. Wan, X. Wang, and Y. He, "Utilizing SMOTE-TomekLink and machine learning to construct a predictive model for elderly medical and daily care services demand," *Scientific Reports*, vol. 15, no. 1, Art. no. 8446, 2025.

[37] A. N. Safriandono, D. R. I. M. Setiadi, A. Dahlan, F. Z. Rahmanti, I. S. Wibisono, and A. A. Ojugo, "Analyzing quantum feature engineering and balancing strategies effect on liver disease classification, " *Journal of Future Artificial Intelligence and Technologies*, vol. 1, no. 1, pp. 51–63, Jun. 2024.

[38] I. Salehin and D.-K. Kang, "A Review on Dropout Regularization Approaches for Deep Neural Networks within the Scholarly Domain," *Electronics*, vol. 12, no. 14, p. 3106, Jul. 2023.

[39] A. A. Yaghoubi, P. Karimi, E. Moradi, and R. Gavagsaz-Ghoachani, "Implementing Engineering Education Based on Posing a Riddle in Field of Instrumentation and Artificial Intelligence,"9th International Conference on Control, Instrumentation and Automation (ICCIA) 2023, pp. 1–5.

[40] P. Rabiei and N. Ashrafi-Payaman, "Anomaly Detection in Dynamic Graph Using Machine Learning Algorithms," *Journal of Artificial Intelligence and Data Mining (JADIM)*, vol. 12, no. 3, pp. 359–367, 2024.

[41] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Mining*, vol. 16, no. 1, Feb. 2023.

یک شبکه عصبی عمیق مبتنی بر پرسپترون چندلایه (MLP) با استفاده از روش SMOTE-Tomek برای پیش‌بینی مقاوم اختلالات کبدی

الهه مرادی*

گروه مهندسی برق، واحد یادگار امام خمینی(ره) شهرری، دانشگاه آزاد اسلامی، تهران، ایران.

ارسال ۲۰۲۵/۰۴/۰۴؛ بازنگری ۲۰۲۵/۰۶/۱۴؛ پذیرش ۲۰۲۵/۰۷/۱۰

چکیده:

اختلالات کبدی از شایع‌ترین بیماری‌ها در سراسر جهان هستند و تشخیص و پیش‌بینی به‌موقع آن‌ها می‌تواند به‌طور قابل‌توجهی نتایج درمان را بهبود بخشد. در سال‌های اخیر، کاربرد هوش مصنوعی، به‌ویژه الگوریتم‌های یادگیری ماشین و یادگیری عمیق، اهمیت زیادی در حوزه پزشکی پیدا کرده و منجر به کاهش هزینه‌های مراقبت‌های بهداشتی شده است. این مطالعه از مجموعه‌داده ILPD موجود در مخزن یادگیری ماشین UCI استفاده کرده که شامل ۵۸۳ پرونده بیمار کبدی با ۱۱ ویژگی است. یک چارچوب پیش‌بینی مبتنی بر شبکه پرسپترون چندلایه (MLP) برای پیش‌بینی اختلالات کبدی به کار گرفته شد. برای رفع عدم تعادل کلاس‌ها در مجموعه‌داده طبقه‌بندی دودویی، از روش SMOTE–Tomek برای بهبود تعادل داده‌ها استفاده شد. مقیاس‌بندی مقاوم برای مدیریت مقادیر پرت به کار گرفته شد. در نهایت، عملکرد روش پیشنهادی با سه الگوریتم معروف یادگیری ماشین مقایسه شد. برای افزایش استحکام ارزیابی، Five-fold cross-validation در تمامی طبقه‌بندها به کار گرفته شد. تمامی شبیه‌سازی‌ها با استفاده از پایتون انجام شد. نتایج نشان می‌دهد که روش پیشنهادی عملکرد برتری داشته و با دقت ۹۰,۹۰٪، از روش‌های پیشرفته موجود پیشی گرفته است.

کلمات کلیدی: یادگیری عمیق، پرسپترون چندلایه، SMOTE-Tomek، تعادل داده‌ها، بیماری کبدی، یادگیری ماشین.