AIMS Bioengineering

AIMS Bioengineering, 12(1): 50–68.
DOI: 10.3934/bioeng.2025003
Received: 23 December 2024
Revised: 12 January 2025
Accepted: 21 January 2025
Published: 06 February 2025

https://www.aimspress.com/journal/Bioengineering

Research article

Chronic liver disease detection using ranking and projection-based feature optimization with deep learning

Sumaiya Noor[1], Salman A. AlQahtani[2] and Salman Khan[2,*]

[1] Business and Management Sciences Department, Purdue University, West Lafayette, IN, USA
[2] New Emerging Technologies and 5G Network and Beyond Research Chair, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

* **Correspondence:** Email: snawab@ksu.edu.sa.

Abstract: The liver is a vital gland responsible for various essential functions such as digestion, metabolism, detoxification, and immunity. Liver diseases caused by infections, injuries, or genetic factors are dangerous and require prompt diagnosis and treatment to improve survival rates. Early detection of liver conditions is crucial, and recent advancements in machine learning (ML) have proven highly effective in predicting diseases like chronic obstructive pulmonary disease (COPD), hypertension, and diabetes. Additionally, the rise of deep learning has begun transforming liver research, offering powerful tools to aid doctors in diagnosis and treatment. This study presents a novel and efficient learning method to identify liver patients accurately. The approach integrates multiple ranking and projection techniques for features, utilizing deep learning to detect early signs of liver disease. Additionally, Shapley Additive exPlanations (SHAP) are applied to perform global interpretation analysis, helping to select optimal features by assessing their contributions to the overall model. Our experimental results demonstrate that this proposed model outperforms traditional machine learning algorithms, achieving superior accuracy. Cross-validation and various testing methods confirm that the deep neural network (DNN) we developed surpasses other classifiers, reaching an accuracy rate of 90.12%. This paper explores how machine learning can be integrated into healthcare, particularly for predicting liver disease. Our findings show that the proposed model can potentially improve diagnostic accuracy and support timely medical intervention, ultimately enhancing patient outcomes.

Keywords: machine learning; deep neural network; SHAP; liver disease; classification; SMOTE

1. Introduction

The liver is a vital organ in the human body that maintains basic physiological activities necessary for health. The liver is essential for digestion, metabolism, detoxification, and immunity, highlighting its critical role in maintaining bodily equilibrium. Despite being one of the most diverse and intricate organs, the liver is susceptible to infections, traumas, and genetic disorders, jeopardizing human life and longevity [1,2]. These disorders will adversely affect the liver and lead to severe consequences; thus, preventative actions and cures for general liver issues are essential. However, persistent injury to this organ may disrupt its activities and overall health despite its significant ability for renewal and adaptation. Consequently, understanding the factors that lead to liver function deterioration, inflammation, and jaundice, alongside pursuing timely diagnosis and prevention, are two essential steps toward enhancing the quality of life and mitigating mortality risks associated with liver diseases (LD).

The healthcare industry has seen significant transformations in its delivery systems and organizations as a result of data processing, primarily via the use of machine learning (ML) and artificial intelligence (AI) [3]. These innovations have transformed data collection, storage, and analysis methods, allowing clinical practitioners to integrate cutting-edge decision-making models into traditional healthcare procedures. This integration has significantly enhanced inspection and informed decision-making in the healthcare industry. It is important to note that, because of the progressive advancement of AI and ML applications, early illness prediction is feasible for conditions such as diabetes, hypertension, COPD, and cardiovascular diseases [4]. These technologies use big data to provide significant insights into a specific diagnosis, aiding illness prognosis management and treatment. Although closely linked with technology, we must recognize that preventative efforts promoting healthy behavior have enhanced the positive transformation this innovation brings. Integrating AI and ML into healthcare procedures is already underway, with potential advancements in precision medicine, medication discovery, and diagnostic tools. The following sections illustrate that these technological advancements include enhancements in healthcare delivery, including improved diagnostics, more focused medication, and overall population health outcomes.

Researchers have proposed many machine learning techniques in the literature for accurately and quickly classifying liver patients. In 2016, Babu et al. [5] introduced a K-means clustering methodology for identifying liver illness, using several classification models to assess its effectiveness. Their study evaluated the efficacy of the Naive Bayes (NB), C4.5 decision tree, and k-nearest neighbors (KNN) classifiers. The results indicated that NBC obtained an accuracy of 56%, KNN reached 64%, and C4.5 attained 69%. This approach illustrates the differential efficacy of several classifiers in conjunction with k-means clustering for identifying liver patients. Likewise, Gan et al. [6] investigated several categorization methods for predicting liver illness. They suggested that the AdaC-TANBN approach had an accuracy of 69.03%. The integrated strategy, which merges AdaBoost with a modified tree-augmented Naive Bayes (TANBN) model, outperformed other classification techniques evaluated in their research. The result highlights the efficacy of integrating boosting methods with probabilistic models to improve prediction accuracy in liver disease categorization. Subsequently, Amare et al. [7] proposed an innovative predictor for identifying LD via sophisticated machine-learning techniques. They created a classification model that combines the complement Naive Bayes (CNB) technique with other classifiers to improve predictive accuracy, comparing it to traditional Naive Bayes classifiers and other models. Their results demonstrated that the CNB approach achieved an accuracy of 71.36%, exceeding the performance of many classifiers analyzed in the literature. This method underscores the efficacy of the CNB technique in improving the accuracy of liver disease categorization. Sreejith and colleagues [8] evaluated classification performance using the ILPD, Pima Indian Diabetes (PID), and Thoracic Surgery (TSD) datasets. They used chaotic multi-verse optimization (CMVO) for feature selection and the synthetic minority oversampling technique (SMOTE), achieving an accuracy of 82.62% on the ILPD dataset with enhanced methodologies. P. Kumar et al. improved liver disease classification accuracy with the ILPD dataset. They used 10-fold cross-validation and utilized variable neighbor-weighted fuzzy KNN, neighbor-weighted K-NN (NWKNN), and fuzzy-neighbor-weighted K-NN classifiers. To address dataset imbalance, they used Tomek connections and redundancy-based under-sampling (TR-RUS), attaining accuracies of 72.31% with NWKNN, 76.61% with fuzzy NWKNN, and 87.71% with their developed variable-NWFKNN approach, which surpassed existing classifiers. Kuzhippallil et al. [9] enhanced the categorization of chronic liver disease by several data preparation and feature selection techniques. Their methodologies improved classification efficacy, achieving a maximum accuracy of 88%. Ruhul et al. [10] recently presented a sophisticated feature extraction and classification technique for predicting LD with the ILPD dataset. Their study integrated projection-based statistical methods to enhance feature selection with relevant machine learning approaches, including support vector machine, logistic regression, and random forest. The suggested technique improves the accuracy of liver disease diagnosis to 88.10%, facilitating early identification and treatment alternatives.

This research introduces a precise and robust learning approach to successfully differentiate liver disease. We first use the random oversampling strategy to address overfitting and class imbalance during model building. Subsequently, we employ two feature methodologies: ranking and statistical projection-based strategies. The Pearson correlation coefficient (PCC), gain ratio (GR), and ReliefF algorithm (RFA) are utilized to rank and evaluate each feature's contribution to the target class (liver illness). We use principal component analysis (PCA) and linear discriminant analysis (LDA) for statistical projection-based techniques. A hybrid feature vector is generated by combining these five feature vectors. We use a feature selection method, Shapley additive explanations (SHAP) [11], to improve computational efficiency, identifying the most significant characteristics by analyzing their impact on model predictions. Diverse hypothesis learners, such as logistic regression (LR), k-nearest neighbor (KNN), random forest (RF), support vector machine (SVM), and Naive Bayes (NB), are used to assess prediction rates. A thorough performance evaluation using measures such as Matthew's correlation coefficient, accuracy, sensitivity, specificity, and area under the curve (AUC) illustrates the enhanced efficacy of our proposed DNN model in predicting liver disease. This methodology offers significant insights into predicting early liver disease, improving diagnostic accuracy, and enabling prompt medical treatments.

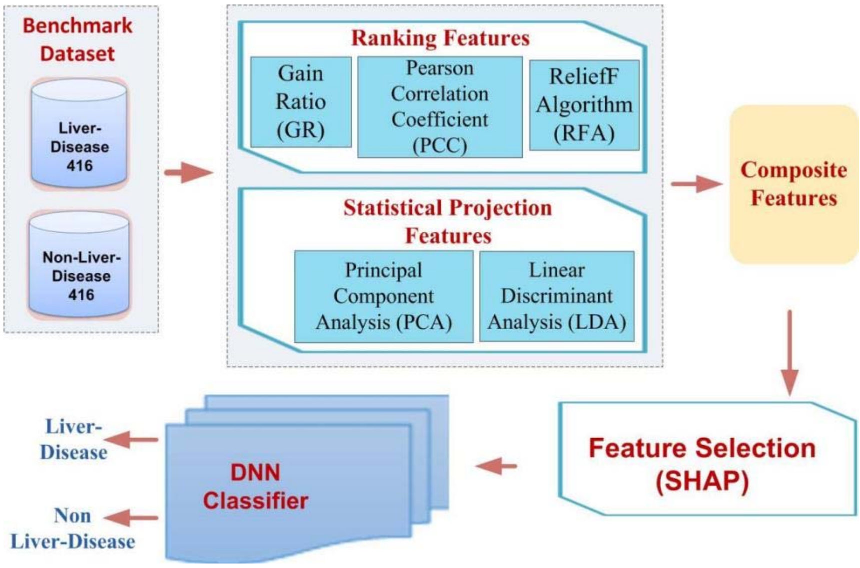

Figure 1. Architecture of the DNN-Liver.

2. Proposed model framework

This section delivers a detailed, illustrated representation of data transfers via the prototype and architecture of a suggested predictor model. Figure 1 illustrates essential design components to be entirely explored in subsequent sections.

2.1. Benchmark dataset

The dataset under study, the ILPD dataset [10], comprises 583 observations and encompasses 10 distinct features. A comprehensive summary of the dataset ILPD is presented in Table 1. The dataset under consideration comprises gender distribution revealing 439 male participants, constituting approximately 75.30%, and 144 female participants, accounting for 24.70% of the total cohort. The dataset's primary target class revolves around diagnosing liver disease, distinguishing participants based on whether they have been diagnosed with this condition.

Table 1. Description of features in the ILPD dataset.

Feature	Description
Gender	Participant's gender
Albumin (ALB)	Albumin level
Direct bilirubin (DB)	Direct bilirubin level
Age (years)	Age of participants (4–90 years)
Total bilirubin (TB)	Total bilirubin level
Aspartate aminotransferase (SGOT)	Aspartate aminotransferase level
Total protein (TP)	Total protein level
Alkaline phosphatase (ALP)	Alkaline phosphatase level
Alanine aminotransferase (SGPT)	Alanine aminotransferase level
Albumin and globulin ratio (AGR)	Albumin and globulin ratio

The long-term liver disease risk is addressed as a classification challenge with *Liver-Disease* (LD) or *Non-Liver-Disease* (non-LD) classes. The total dataset comprises 416 positive cases of liver disease and 167 controlled cases. Considering the significance of balanced datasets in machine learning analysis, we applied a data balancing technique, i.e., random oversampling, resulting in a dataset of 832 cases, evenly distributed between instances of liver disease and non-disease cases (i.e., LD = 416, non-LD = 416). Finally, Table 2 outlines the balanced dataset statistical properties of the features, highlighting essential measurements, i.e., standard deviation, mean, minimum, and maximum, for each variable. This analysis provides a complete analysis of the dataset's numerical distributions. Table 2 summarizes the statistical properties of features, including min, max, mean, and standard deviation, providing numerical distribution analysis.

Table 2. Numerical features for the statistical property features.

Features	Min	Mean	Max	± Stdv
TB	0.4	2.65	75	± 5.32
SGOT	10	88.78	4929	± 245.07
Age	4	43.55	90	± 16.28
ALP	63	267.26	2110	± 212.62
SGPT	10	66.78	2000	± 155.16
AGR	0.3	0.98	2.8	± 0.30
DB	0.1	1.16	19.7	± 2.42
ALB	0.9	3.19	5.5	± 0.76
TP	2.7	6.5	9.6	± 1.02

2.2. Feature formulation technique

ML algorithms often face the obscenity of dimensionality, which arises when there are many data points but relatively few meaningful features or when the feature space contains irrelevant information. To address this challenge, we first implement three ranking strategies [12,13], i.e., Pearson correlation coefficient (PCC), gain ratio (GR), and ReliefF algorithm (RFA), and statistical projection-based strategies such as principal component analysis (PCA) and linear discriminant analysis (LDA) to examine the involvement of each feature to the target class (i.e., liver disease). Subsequently, we use Shapley additive explanations (SHAP) to determine how much each feature matters. These techniques help simplify the dataset by focusing only on the most important parts, making it easier to analyze and reducing the computing power needed. These methods are crucial to finding the most valuable features when predicting liver disease, leading to more accurate and understandable diagnostic and prediction models [14,15].

2.2.1. Feature importance evaluation

In this section, three ranking methods, i.e., PCC, GR, and RFA, are nominated to calculate the involvement of a feature (i.e., LD class). Table 3 provides an evaluation of feature contributions to predicting the target class.

Table 3. Feature importance analysis for liver disease prediction.

Feature	PCC	GR	RFA
SGPT	0.2141	0.0701	0.2613
TB	0.2874	0.1373	0.2848
Gender	0.0857	0.0065	0.025
SGOT	0.2017	0.1005	0.2393
AGR	0.2046	0.0822	0.2599
DB	0.3205	0.1421	0.2895
ALB	0.1836	0.0408	0.2883
Age	0.1596	0.0372	0.2625
ALP	0.246	0.0867	0.1936
TP	0.0443	0.00	0.2161

First, using PCC analysis, some significant relationships were detected, including highly significant ones, for example, 0.88, indicating a relationship between TB and DB. The percentage for this association was comparatively low, at 0.32. DB remained the most frequent feature with 32 DDR, followed by task-interdependent, task-supportive, and task-contrast with 14 DDR, 13 DDR, 11 DDR, and 5 DDR, respectively. Second, the GR method used in the "screening variable selection" highlighted DB, TB, and SGOT as prominent ones. Third, using the Gini impurity ReliefF algorithm suggested that gender had the least priority, whereas DB, ALP, and TB were the least essential features. These methods all emphasize the importance of all features in screening for liver diseases.

2.2.2. Principle components analysis (PCA)

PCA reduces data dimensions using covariance matrices [16] with minimum loss of discriminative features. Let us consider a feature vector S with the dimension of i*j, where i represents the number of extracted features and j represents the number of samples. Let k represent the number of desired features. The value of k must be less than i. Let us consider the following input feature vector S using Eq. (1):

$$\left[s_1, s_2, s_3, \ldots s_n\right] \quad (1)$$

Through the PCA algorithm, we implemented the following steps to reduce the dimensionality of the feature vector:

Step 1: Compute the mean

$$\bar{S} = \frac{1}{i}\sum_{n=1}^{i} s_n \quad (2)$$

Step 2: Subtract the mean

$$S_n^{'} = s_n - \bar{s} \qquad \text{n} = (1, 2, 3 \ldots) \quad (3)$$

Step 3: Compute the covariance

$$C_n = S_n^{'}.(S_n^{'})^T BB^T \quad (4)$$

Where $B = \left\{S_1^{'},\ S_2^{'},\ \ldots,\ S_p^{'}\right\}$ $(i \ * \ j)$ and T= transpose of a matrix.

Step 4: Calculate the eigenvalues in ascending order. The first eigenvalue should be greater than the second and fourth.

$$\{\gamma_1 > \gamma_2 > \gamma_3 \ldots \gamma_n\} \quad (5)$$

Step 5: Compute the corresponding eigenvectors:

$$c_n : \{\sigma_1, \sigma_2, \sigma_3 \ldots \sigma_n\} \quad (6)$$

Last step: Select k eigenvectors corresponding to the largest eigenvalues to reconstruct a new set of feature vectors in the space. This selection process helps avoid dimensionality problems, keeping the most essential data. It also improves the computational speed and interpretability of various other tasks being performed in the analytical methods of the dataset, which is the main aim of PCA in the case of data preprocessing and data analysis.

$$PC_i = \sum_{j=1}^{s} C_n * x_j \quad (7)$$

2.2.3. Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is a supervised dimensionality reduction technique that projects data onto a lower-dimensional space while maximizing class separability. The main goal of LDA is to find a transformation matrix W that maximizes the ratio of between-class variance to within-class variance, ensuring that different classes are well-separated in the projected space. Mathematically, the objective function of LDA is to maximize the following criterion.

$$j(W) = {[W^T S_b W]}\Big/{[W^T S_w W]} \tag{8}$$

Here, S_b is the between-class scatter matrix, S_w is the within-class scatter matrix, and W is the transformation matrix to be optimized.

2.3. Hybrid feature

The methodology proposed in this paper consists of five feature extraction mechanisms described in section 2.2. To develop a hybrid feature vector, we combine the five feature vectors as given in Eq. (9). Mathematically, the hybrid feature vector is also called the hybrid feature vector. Table 4 shows the overview of features contributed by each technique and the total number of features created by the hybrid vector.

$$j(W) \cup PC_i \cup RFA \cup GR \cup PCC \tag{9}$$

Table 4. Dimension of feature vectors.

Method	Number of features (dimensionality)
Principal component analysis (PCA)	10
Linear discriminant analysis (LDA)	10
Gain ratio (GR)	10
Pearson correlation coefficient (PCC)	10
ReliefF algorithm (RFA)	10
Hybrid features	50

2.4. SHAP methods

Feature selection is vital in the machine learning pipeline to identify the most relevant features to improve model performance, reduce overfitting, and lower computational costs [17]. This paper employs Shapley additive explanations (SHAP) for feature selection, which leverages cooperative game theory to assign importance to input features based on their contribution to predictions. SHAP quantifies each feature's impact, providing insights into its influence on model outcomes, as detailed in Eq. (10). Understanding the structure and size of the dataset is equally essential, as it informs preprocessing steps like data splitting, regularization, and feature selection. Properly shaping data ensures efficient processing and supports informed decision-making throughout the data science process.

$$SHAP_i(x) = \phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|(|N| - |S| - 1)}{|N|} \left[f(S \cup \{i\}) - f(S) \right] \tag{10}$$

Where ϕ_i, represents the SHAP value for the feature i, N is the set of all features, S is a subset of features excluding i, $f(S)$ is the model's prediction given features in S, and $f(S \cup \{i\})$ is the model's prediction given features in S plus feature i.

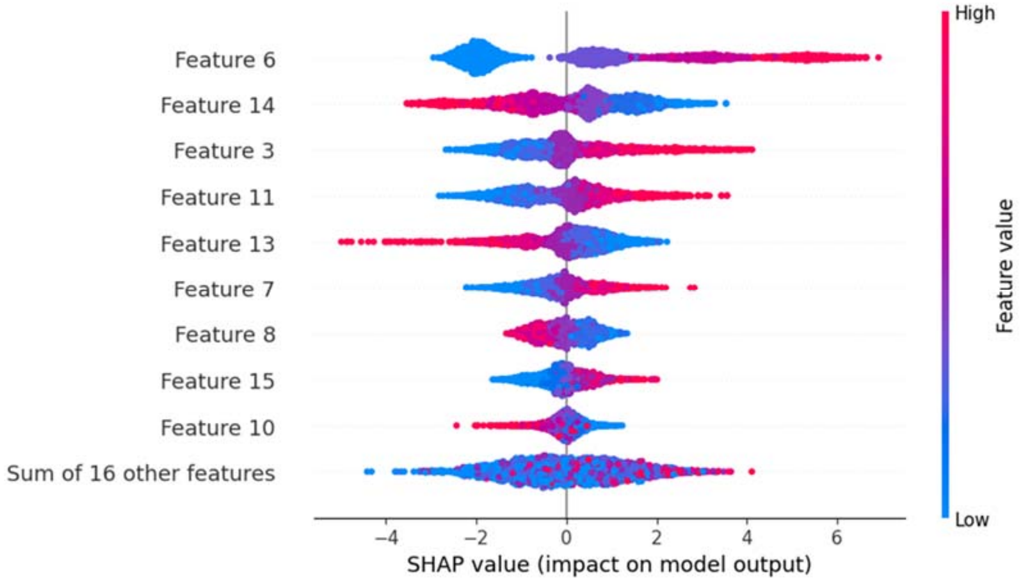

Figure 2. Vital features selected from hybrid features using SHAP.

In this paper, we selected the vital features shown in Figure 2. Each row represents a feature, with dots indicating their corresponding SHAP values. Blue dots signify low values, while red dots denote high values. Positive SHAP values indicate a higher likelihood of liver disease, whereas negative values suggest a lower risk. Testing different feature groups revealed that the top 25 features significantly improved the proposed method's performance.

2.5. Deep neural network

Deep neural networks (DNNs) mimic the human brain's functioning, consisting of an input layer for receiving data, an output layer for producing results, and multiple hidden layers that process information through interconnected tiers [18]. These layers transform input data, enabling the model to identify intricate patterns and representations. The deep structure of DNNs makes them highly effective in tasks such as image recognition and natural language processing, as each layer extracts increasingly complex features. However, while additional hidden layers improve the model's ability to learn intricate patterns, they also raise computational costs, increase the risk of overfitting, and complicate the model. DNNs excel in feature extraction from unstructured or unlabeled data and have proven valuable in biotechnology, natural language processing, image analysis, audio processing, and engineering [19,20].

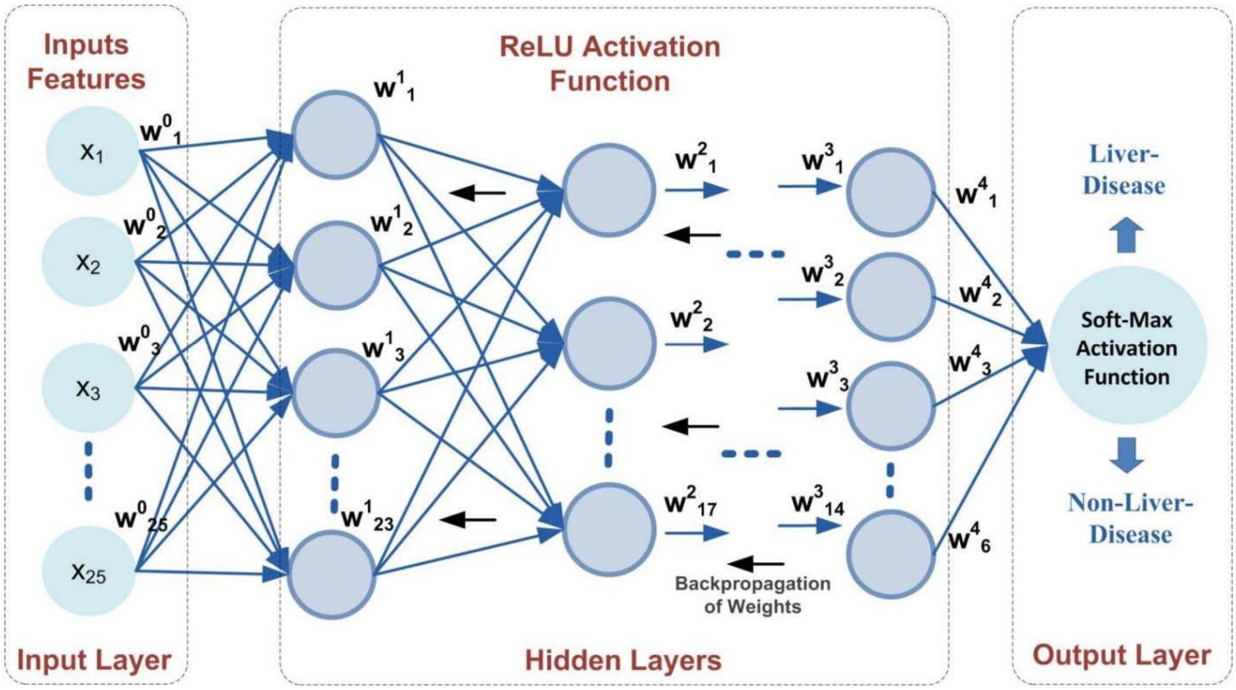

Figure 3. DNN configuration topology.

2.5.1. Model training

This study focuses on the early prediction of liver illness using benchmark data and a deep neural network model with four hidden layers, as illustrated in Figure 3. The DNN architecture comprises multiple neurons that process the input feature vector to produce output, as detailed in Eqs (11) and (12). The neurons' weights in each layer are initialized using the Xavier function to ensure equal variance across layers [14,15]. A backpropagation method adjusts the weight matrices to minimize the discrepancies between predicted and target classes. The ReLU activation function is applied in the input and hidden layers to capture nonlinearity and patterns in the data, while the logistic sigmoid activation function is used in the output layer. This mechanism determines whether a neuron is activated or remains inactive based on its output. Additionally, the Softmax activation function in the output layer converts the generated values into probabilities between 0 and 1, representing the likelihood of the data belonging to a specific class.

$$y_a = f(B_a + \sum_{b=1}^{m} x_b\, w_b^a) \tag{11}$$

$$f(i) = \frac{e^i}{1+e^i} \tag{12}$$

3. Performance evaluation

To thoroughly evaluate the performance of the proposed models, five essential evaluation metrics were employed, i.e., sensitivity (SN), specificity (SP), overall accuracy (ACC), Matthew's correlation coefficient (MCC), and the area under the curve (AUC) [21]. These critical indicators provide insights into the predictor's effectiveness from multiple perspectives. The ACC represents the proportion of correct predictions relative to the total number of samples. The SN measures the proportion of actual positives correctly identified, while SP reflects the proportion of actual negatives accurately detected. The AUC assesses the model's classification ability across multiple classes, with the receiver operating characteristic (ROC) curve plotting the true positive rate against the false positive rate. The MCC evaluates the classifier's performance by accounting for all four confusion matrix elements. The corresponding calculation formulas are provided below.

$$\mathrm{ACC} = \frac{TP+FN}{TN+TP+FN+FP} \tag{13}$$

$$\mathrm{SP} = \frac{TN}{TN+FP} \tag{14}$$

$$\mathrm{SN} = \frac{TP}{TP+FN} \tag{15}$$

$$\mathrm{MCC} = \frac{(TP\ X\ TN)-(FP\ X\ FN)}{\sqrt{[(TP+FP)\ X\ (TP+\ FN)\ X\ (TN+FP)\ X\ (TN+FN)\ ]}} \tag{16}$$

Here, TP and FN represent the true positive and negative samples (i.e., liver and non-liver diseases). Similarly, TN and FP are errors as the model inaccurately predicted the samples to be true or false.

4. Results and analysis

This section provides a detailed evaluation and discussion of the proposed model's effectiveness. The experiments were conducted using machine learning tests implemented with Python and TensorFlow, both open-source tools that facilitated efficient setup and testing of the prediction models. The tests were executed on a system equipped with an 11th Gen Intel Core i7-1165G7 processor (2.80 GHz), 16 GB of RAM, running Windows 11 Home on a 64-bit architecture.

4.1. Hyperparameters optimization

This section aims to identify the optimal configuration for hyperparameters in DNN topology. To analyze the performance of the DNN on different hyperparameters, we used a grid search technique that applies different combinations of parameters. We considered only those hyperparameters that primarily affect the DNN model's performance. These parameters include activation function, learning rate, and the number of iterations [15,18,20,22].

Table 5. Optimum hyperparameters values.

Parameters list	Optimize values
Hidden layers	4
Updater	ADAGRAD function
Hidden layers neurons	23-17-14-6
Seed	12345L
Epoch	10
Momentum	0.9
LR	0.01
Dropout	0.25
Regularization l2	0.001
Activation functions	ReLU & Softmax
Optimizer	SGD Method
Weight initialization function	XAVIER function

First, experiments were conducted to evaluate the impact of activation functions and learning rates on the model's performance. As shown in Table 6, the DNN classifier achieved its highest accuracy of 90.12% while using the ReLU activation function and a learning rate of 0.01. Furthermore, a gradual reduction in the learning rate improved the model's performance; however, further decreases beyond this point did not result in substantial accuracy gains. Thus, a learning rate of 0.01 and the ReLU activation function were identified as optimal parameters for achieving high accuracy in the DNN model.

Table 6. Performance comparison using a grid search technique.

LR	ReLU	Tanh	Sigmoid
0.01	90.12	88.10	89.44
0.02	89.91	88.09	89.23
0.03	89.45	87.71	89.01
0.04	89.16	87.58	88.80
0.05	88.49	87.38	88.17
0.06	87.82	87.19	87.54
0.07	87.55	86.99	86.91
0.08	87.09	86.80	86.28
0.09	87.05	86.60	85.65
0.10	87.03	86.41	85.02

Next, we conducted numerous experiments to evaluate the performance of the DNN model by varying the number of training epochs. Figure 4 illustrates the error loss observed when using the ReLU activation function. The results indicate a consistent decrease in the error rate with increased training epochs. For example, the error loss began at 0.892 during the initial epoch and progressively declined to 0.002 by the 50th epoch. These findings suggest that 50 epochs are optimal, as the error rates stabilize beyond this point. The optimal configuration determined from this analysis is detailed in Table 5.

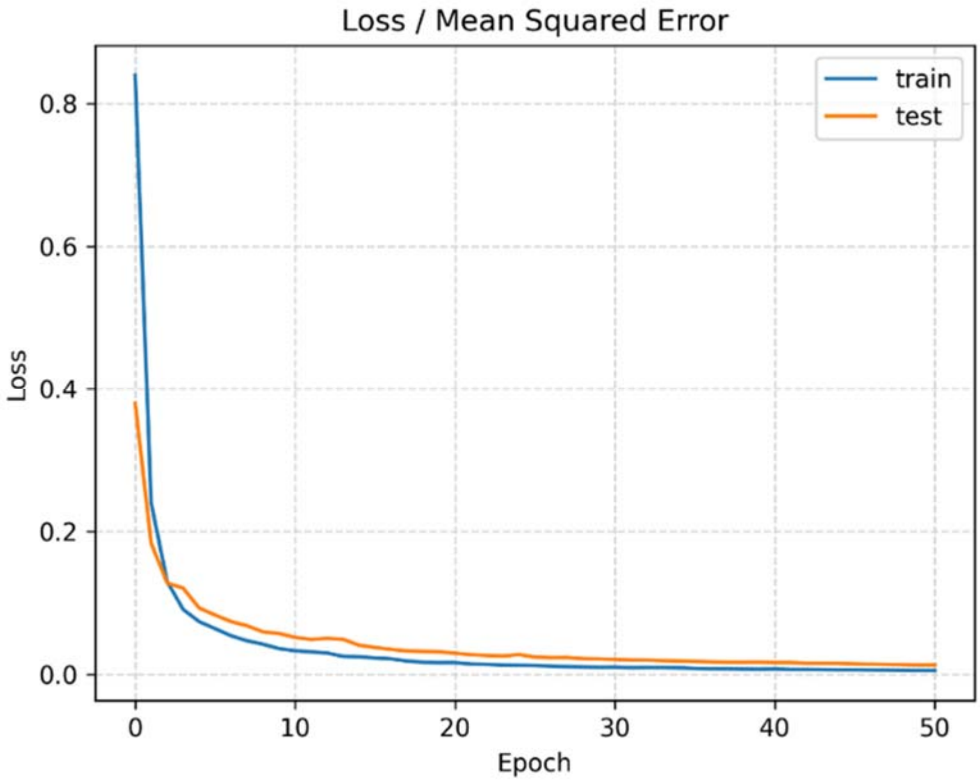

Figure 4. Number of training epochs versus error loss.

4.2. Performance analysis using sequence formulation techniques

This section examines the prediction results of the deep neural network (DNN) model using individual and hybrid feature extraction methods. Machine learning techniques such as k-fold cross-validation (CV) have been widely used in the literature to evaluate model performance. A 10-fold cross-validation approach was adopted in this study, offering an optimal balance between computational efficiency and comprehensive performance evaluation while minimizing the influence of data variability. Table 7 presents the DNN model's predicted outcomes on the balanced dataset using various feature vectors.

Table 7. Different feature extraction methods on the balance dataset.

Methods	ACC (%)	SP (%)	SN (%)	MCC
PCA	72.37	70.61	74.13	0.481
LDA	84.88	81.22	88.54	0.703
GR	77.85	76.38	79.32	0.561
PCC	82.38	79.12	85.64	0.662
RFA	81.32	78.18	84.46	0.665
Hybrid feature (before SHAP)	89.36	88.21	90.44	0.797
Hybrid feature (after SHAP)	90.12	88.17	92.05	0.825

Table 7 demonstrates that the DNN model attained superior performance with a hybrid feature vector compared to the individual feature extraction technique. The DNN model achieved an average accuracy of 89.36%, sensitivity of 90.44%, specificity of 88.21%, and an MCC of 0.797 utilizing a hybrid feature vector. In order to further enhance the performance of the proposed DNN model, we apply the feature selection method, i.e., SHAP. SHAP-selected features attained an accuracy of 90.12%, sensitivity of 92.05%, specificity of 88.17%, and a Matthew's correlation coefficient of 0.825. The result demonstrates the effectiveness of feature selection in enhancing prediction accuracy and target result identification in deep neural networks (DNNs).

4.3. Performance comparison with other classifiers

In this section, we compare the performance of various classification methods based on accuracy (ACC), sensitivity (SN), specificity (SP), and Matthew's correlation coefficient (MCC). The classifiers being assessed are random forest (RF) [23], support vector machine (SVM) [24], and k-nearest neighbor (KNN) [25]. RF is a popular ensemble learning method for classification and prediction, building multiple decision trees from random samples. SVM is widely used in life sciences for linear and nonlinear classification, as it finds the best boundary to separate classes [26]. KNN, commonly used in image processing, is a distance-based method that classifies by comparing examples. Table 8 shows the overall performance comparison of the different classifiers on the benchmark dataset.

Table 8. Comparison of the DNN model with other commonly used machine learning algorithms.

Method	ACC (%)	SN (%)	SP (%)	MCC
Proposed DNN	90.12	92.05	88.17	0.825
SVM	89.66	87.75	91.67	0.794
RF	87.86	84.98	90.91	0.759
KNN	85.96	84.98	86.96	0.719

The proposed DNN model achieved the highest accuracy of 90.12%, with a sensitivity of 92.05% and specificity of 88.17%, resulting in an MCC of 0.825. Comparatively, among the traditional machine learning models, the SVM demonstrated an accuracy of 89.66%, a sensitivity of 87.75%, a specificity of 91.67%, and an MCC of 0.794. Similarly, the RF and KNN models achieved lower accuracy, at 87.86% and 85.96%, respectively. Overall, the proposed DNN demonstrates the most balanced and robust performance, improving average accuracy by 2.29%. The superior performance of the DNN model compared to traditional machine learning algorithms can be attributed to its use of multi-stack processing layers. These layers enable DNN to effectively handle complex datasets with high nonlinearity, a task that traditional ML algorithms struggle with due to their reliance on single-stack processing layers, which are insufficient for such intricate data. These findings confirm that the DNN model outperforms conventional learning methods in predicting liver and non-liver diseases.

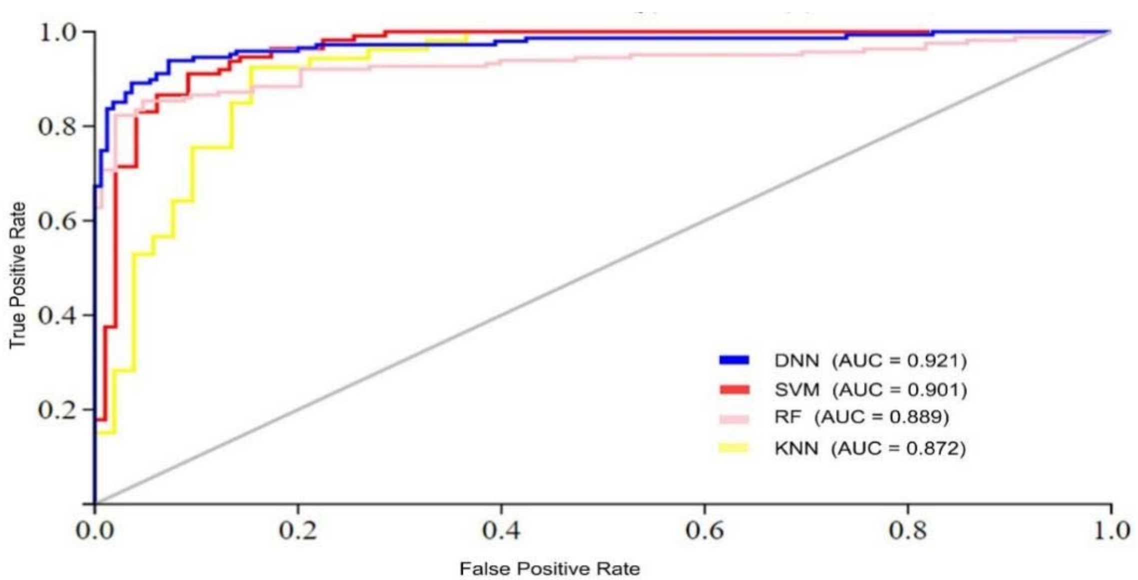

Figure 5. Area under the curve (AUC) performance comparison of machine learning algorithms.

Furthermore, we also compared the performance of the DNN model in terms of AUC (area under the ROC curve) [27] with traditional learning algorithms, as shown in Figure 5. The figure shows that the DNN model achieved the highest AUC value compared to the other algorithms. Specifically, the DNN model had an AUC of 0.921, while the SVM, RF, and KNN algorithms had AUC values of 0.901, 0.889, and 0.872, respectively. The main reason for the DNN model's superior performance is its multi-layer structure (hidden layers) with backpropagation, which effectively optimizes weights. This allows the DNN to handle complex and highly nonlinear datasets better than other models, which use a single-layer processing approach and struggle with such data. Additionally, the confusion matrix analysis in Figure 6 shows that our proposed DNN model achieves balanced recognition of both positive and negative. This strongly confirms the model's accuracy and reliability in distinguishing between positive and negative samples, making it particularly suitable for predictive tasks in the specific problem domain

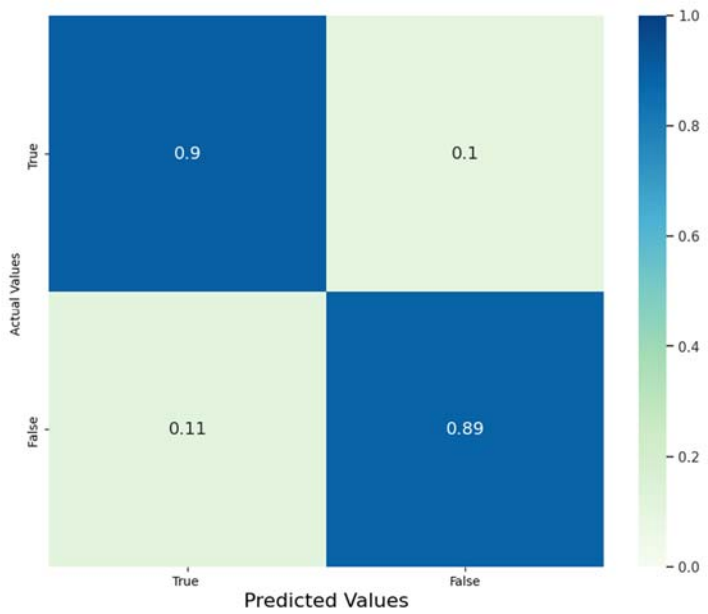

Figure 6. Evaluating DNN model performance through visualization with confusion matrix analysis.

4.4. Performance comparison with the existing predictors

In this section, we compare the performance of the proposed model with the existing state-of-the-art models, as mentioned in [10,28–30]. Table 9 shows the comparative findings. Moreover, we provide the proposed models from published research works based on the same dataset. Table 9 shows that the proposed model provided the best prediction accuracy among four previously published predictors. For example, K. Gupta et al. [29] achieved an accuracy level of 63%, whereas the recently published predictor, R. Amin et al. [10], achieved an accuracy of 88.10%. Similarly, D. Elias et al. [30] achieved 80.10%. These results confirm that the proposed model performed better than the existing predictors, with an average accuracy improvement of 13.93%.

Table 9. Comparison of the proposed model performance with the existing models.

Method	Classifiers	ACC (%)
K. Gupta et al. [29]	Light GB	63.00
Altaf et al. [28]	Voting	73.56
D. Elias et al. [30]	Voting	80.10
R. Amin et al. [10]	Random Forest	88.10
Proposed DNN	DNN	90.12

5. Conclusion

Liver disease (LD) is a severe condition that poses significant risks to human health and requires timely medical intervention. Healthcare professionals rely on neurological techniques for evaluating and diagnosing affected individuals. In this paper, we explored the prediction of chronic LD using a deep neural network algorithm enhanced with integrated and optimized features. Our model used several techniques to assess feature importance, including PCA, LDA, gain ratio, Pearson correlation coefficient, and ReliefF algorithm. We also employed SHAP analysis to interpret complex features and select the most relevant ones for accurate LD prediction. Our novel approach demonstrated superior performance compared to existing models, which can be attributed to the implementation of efficient feature extraction methods, advanced compression techniques, and an optimized feature selection process. The results show that our model enhances discriminative ability, offering a more reliable tool for the early detection of chronic liver disease. Integrating AI and machine learning into clinical settings holds great potential for advancing disease detection and improving patient outcomes.

In the future, we plan to create an accessible web platform for biologists to use this model. Additionally, we aim to expand the dataset, explore new features, and implement more advanced algorithms to refine further and validate our model's predictive capabilities.

Acknowledgments

This work was supported by Research Supporting Project Number (RSPD2025R585), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

All authors contribute equally.

References

1. Singh HR, Rabi S (2019) Study of morphological variations of liver in human. *Transl Res Anat* 14: 1–5. https://doi.org/10.1016/j.tria.2018.11.004
2. Razavi H (2020) Global epidemiology of viral hepatitis. *Gastroenterol Clin North Am* 49: 179–189. https://doi.org/10.1016/j.gtc.2020.01.001
3. Miao JH, Miao KH (2018) Cardiotocographic diagnosis of fetal health based on multiclass morphologic pattern predictions using deep learning classification. *Int J Adv Comput Sc* 9: 1–11. https://doi.org/10.14569/IJACSA.2018.090501

4. Dritsas E, Alexiou S, Moustakas K (2022) COPD severity prediction in elderly with ML techniques, *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, 2022: 185–189. https://doi.org/10.1145/3529190.3534748
5. Babu MSP, Ramjee M, Katta S, et al. (2016) Implementation of partitional clustering on ILPD dataset to predict liver disorders, *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS),* IEEE, 2016: 1094–1097. https://doi.org/10.1109/ICSESS.2016.7883256
6. Gan D, Shen J, An B, et al. (2020) Integrating TANBN with cost sensitive classification algorithm for imbalanced data in medical diagnosis. *Comput Ind Eng* 140: 106266. https://doi.org/10.1016/j.cie.2019.106266
7. Anagaw A, Chang YL (2019) A new complement naïve Bayesian approach for biomedical data classification. *J Amb Intel Hum Comp* 10: 3889–3897. https://doi.org/10.1007/s12652-018-1160-1
8. Sreejith S, Nehemiah HK, Kannan A (2020) Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection. *Comput Biol Med* 126: 103991. https://doi.org/10.1016/j.compbiomed.2020.103991
9. Kuzhippallil MA, Joseph C, Kannan A (2020) Comparative analysis of machine learning techniques for indian liver disease patients, *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS),* IEEE, 2020: 778–782. https://doi.org/10.1109/ICACCS48705.2020.9074368
10. Amin R, Yasmin R, Ruhi S, et al. (2023) Prediction of chronic liver disease patients using integrated projection based statistical feature extraction with machine learning algorithms. *Informatics in Medicine Unlocked* 36: 101155. https://doi.org/10.1016/j.imu.2022.101155
11. Yousefian-Jazi A, Ryu JH, Yoon S, et al. (2014) Decision support in machine vision system for monitoring of TFT-LCD glass substrates manufacturing. *J Process Contr* 24: 1015–1023. https://doi.org/10.1016/j.jprocont.2013.12.009
12. Jain D, Singh V (2018) Feature selection and classification systems for chronic disease prediction: a review. *Egypt Inform J* 19: 179–189. https://doi.org/10.1016/j.eij.2018.03.002
13. Khan S, AlQahtani SA, Noor S, et al. (2024) PSSM-Sumo: deep learning based intelligent model for prediction of sumoylation sites using discriminative features. *BMC Bioinformatics* 25: 284. https://doi.org/10.1186/s12859-024-05917-0
14. Khan S, Khan M, Iqbal N, et al. (2023) Enhancing sumoylation site prediction: a deep neural network with discriminative features. *Life* 13: 2153. https://doi.org/10.3390/life13112153
15. Naeem M, Qiyas M (2023) Deep intelligent predictive model for the identification of diabetes. *AIMS Math* 8: 16446–16462. https://doi.org/10.3934/math.2023840
16. Lu J, Kerns RT, Peddada SD, et al. (2011) Principal component analysis-based filtering improves detection for Affymetrix gene expression arrays. *Nucleic Acids Res* 39: e86–e86. https://doi.org/10.1093/nar/gkr241
17. Heuillet A, Couthouis F, Díaz-Rodríguez N (2022) Collective explainable AI: Explaining cooperative strategies and agent contribution in multiagent reinforcement learning with shapley values. *IEEE Comput Intell M* 17: 59–71. https://doi.org/10.1109/MCI.2021.3129959

18. Khan S, Khan M, Iqbal N, et al. (2022) Deep-PiRNA: Bi-layered prediction model for PIWI-interacting RNA using discriminative features. *Comput Mater Contin* 72: 2243–2258. https://doi.org/10.32604/cmc.2022.022901
19. Uddin I, Awan HH, Khalid M, et al. (2024) A hybrid residue based sequential encoding mechanism with XGBoost improved ensemble model for identifying 5-hydroxymethylcytosine modifications. *Sci Rep* 14: 20819. https://doi.org/10.1038/s41598-024-71568-z
20. Bibi N, Khan M, Khan S, et al. (2024) Sequence-Based intelligent model for identification of tumor t cell antigens using fusion features. *IEEE Access* 12: 155040–155051. https://doi.org/10.1109/ACCESS.2024.3481244
21. Noor S, Naseem A, Awan HH, et al. (2024) Deep-m5U: a deep learning-based approach for RNA 5-methyluridine modification prediction using optimized feature integration. *BMC Bioinformatics* 25: 360. https://doi.org/10.1186/s12859-024-05978-1
22. Khan S, Khan MA, Khan M, et al. (2023) Optimized feature learning for anti-inflammatory peptide prediction using parallel distributed computing. *Appl Sci* 13: 7059. https://doi.org/10.3390/app13127059
23. Fawagreh K, Gaber MM, Elyan E (2014) Random forests: from early developments to recent advancements. *Syst Sci Control Eng* 2: 602–609. https://doi.org/10.1080/21642583.2014.956265
24. Yue S, Li P, Hao P (2003) SVM classification: Its contents and challenges. *Appl Math* 18: 332–342. https://doi.org/10.1007/s11766-003-0059-5
25. Cheng D, Zhang S, Deng Z, et al. (2014) k NN algorithm with data-driven k value, *Advanced Data Mining and Applications: 10th International Conference*, Guilin, China, 2014: 499–512.
26. Khan S, Uddin I, Khan M, et al. (2024) Sequence based model using deep neural network and hybrid features for identification of 5-hydroxymethylcytosine modification. *Sci Rep* 14: 9116. https://doi.org/10.1038/s41598-024-59777-y
27. Basit A, Fawwad A, Qureshi H, et al. (2018) Prevalence of diabetes, pre-diabetes and associated risk factors: second National Diabetes Survey of Pakistan (NDSP), 2016–2017. *BMJ Open* 8: e020961. https://doi.org/10.1136/bmjopen-2017-020961
28. Altaf I, Butt MA, Zaman M (2022) Hard voting meta classifier for disease diagnosis using mean decrease in impurity for tree models. *Rev Comput Eng Res* 9: 71–82. https://doi.org/10.18488/76.v9i2.3037
29. Gupta K, Jiwani N, Afreen N, et al. (2022) Liver disease prediction using machine learning classification techniques, *2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT),* IEEE, 2022: 221–226. https://doi.org/10.1109/CSNT54456.2022.9787574
30. Dritsas E, Trigka M (2023) Supervised machine learning models for liver disease risk prediction. *Computers* 12: 19. https://doi.org/10.3390/computers12010019

AIMS Press