JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

Enhanced Liver Disease Prediction using Multi-Dataset Integration and CNN Optimized with Pelican Algorithm

Najjia Karim[1*], Naeem Aslam[1], Muhammad Fuzail[1], and Ahmad Naeem[1]

[1]Department of Computer Science, NFC-IET, Multan, Pakistan.
[*]Corresponding Author: Najjia Karim. Email: najjiakarim815@gmail.com

Received: March 11, 2025 **Accepted:** May 10, 2025

Abstract: Several prediction techniques for liver disorders have been developed. However, they are more costly and sophisticated. This endeavor aims to develop an effective approach for detecting liver disorders in their early stages. This research describes convolutional neural network (CNN) infrastructure for harmless he-patic failure forecast. The Pelican Optimisation Algorithm (POA) balances bounding box regression and branching training losses for the CNN model. The liver disease characteristics were taken from three da-tasets: Indian liver patient records(ILPR), Hepatitis C, and Cirrhosis Prediction dataset. The POA-modified CNN model mostly identifies relationships between various laboratory values and diag-noses. The proposed model outperforms SOTA methods, including Opposition-based Laplacian Equi-librium Optimiser, Adaptive Hybridised Deep CNN, SVM, and Tree-based classifiers, in terms of accuracy, precision, recall, F-measure, and Mathews Correlation Coefficient. The proposed model has an MCC value of 94.8945, accuracy of 98.6743%, precision of 96.2436%, F1-measure of 97.5524%, and recall of 95.7887%, respectively. The findings show that the suggested strategy effectively predicts liver illness early on through automated screening, reducing strain on caregivers.

Keywords: Cirrhosis; Deep Learning; CNN; MCC; Pelician Optimization.

1. Introduction

The liver is an important organ responsible for removing poisons from body [1]. When the circulatory system ceases serving its purpose adequately, it can cause harm to the body and prevent it from performing certain activities [2]. The virus infects the liver and assaults the immune system. Hepato-tropic viruses HBV cause liver disorders. Its illness impacts 237 mil. Individuals globally, with 1 million chronically afflicted and dying from liver disorders[3]. Hepatitis refers to many liver illnesses. Five types of hepatitis cause liver disease[4]. Viral strains causing chronic illness include hepatitis C, and B. Hepatitis C or B affects an estimated 325 million individuals globally. Cirrhosis is the deformation or fibrosis of the liver[5]. Hepatitis causes inflammation in the liver due to the development of viruses such B, C, and A [6]. Bodily substances transfer hepatitis B and can cause illness. Hepatitis C is transmitted by contaminated blood. The liver illness causes no symptoms at first and might persist in the liver for several years. Increased fat deposition can lead to fatty liver disease[7].

Research suggests that smoking and alcohol intake might worsen disease severity [8]. Machine learning (ML) is useful in predicting liver illness at an early stage[9]. Liver function and imaging tests are used to assess liver damage and aid in disease diagnosis [10]. While liver transplantation is the most effective option, it is also the most expensive [11].

The likelihood of liver failure will be reduced if liver disease is detected early [12]. Due to their ongoing function, even when partially impacted, liver problems are typically dis-covered later on [13]. In general, liver illness only affects the liver's functions when it affects 75%. of its tissues. The signs of liver

illness can vary from case to case [14]. They may not always exhibit any overt signs. Early diagnosis has the potential to save more lives. Due to the liver's vital role in manufacturing triglycerides, proteins, and blood clotting factors, among other things, an early diagnosis is essential [15]. Another difficulty is that, despite seeing the early signs, even medical professionals could fail to diagnose the illness. Therefore, creating an effective model to diagnose the illness automatically will considerably aid medical professionals in their decision-making [16].

CNN focuses on picture segmentation and ignores the form of an object while de-termining the actual boundaries of the item, in contrast to other approaches that provide single bounding boxes surrounding the object of interest [17]. It avoids further image processing work by directly measuring the object's area, length, and axis. This data facilitates the creation of algorithms to determine the object's precise properties [18]. The high-level features that doctor needs from the raw images. Researchers from various disciplines have noticed CNN's precision and compu-ting efficiency [19]. This encourages us to choose pelican optimization and CNN to predict complications of liver disease. Additional benefits of CNN are decreased computational complexity and storage space [20].

The necessity to guarantee a varied representation of liver illnesses, including cirrhosis, hepatitis C, and general liver abnormalities, motivated the dataset selection in this work. The Cirrhosis Prediction Dataset, Hepatitis C Prediction Dataset, and ILPR Dataset were se-lected because of their comprehensive features, which include biochemical indicators, liver function tests, and demographic information. Additionally, all datasets were openly acces-sible, guaranteeing reproducibility and comparison with current techniques. Datasets were merged to increase generalizability and model robustness across various liver disease situ-ations. Traditional models trained on a single dataset are frequently biased and less effective when applied to new, unseen data. By integrating multiple datasets, we aimed to create a comprehensive model that performs well across different patient populations and diag-nostic markers.

We have optimized the CNN architecture's training process using the Pelican Opti-mization Algorithm (POA) to provide the best possible solution. The POA method considers pelicans' social behavior to optimize the CNN's hyperparameters, such as an epoch, learning'srate, and speed or momentum. The ability of pelicans to hunt intelligently inspires CNN to identify the afflicted cases. As a result, it reduces function loss and improves liver disease prediction accuracy. In this work, a unique liver disease prediction model is developed. The main features of our suggested framework are outlined below.

1. It is suggested that a Convolutional Neural Network (CNN) model be used to identify precisely if a patient has a liver illness.

2. Use a POA technique to optimize the CNN model's hyperparameters to improve disease prediction accuracy. The hyperparameter optimization procedure reduces the loss of training that happens during branching and bounding box regression.

3. Three distinct liver illness datasets—the Hepatitis C, Indian liver patient records, and the Cirrhosis Prediction dataset—are used to assess the suggested model's effec-tiveness using various performance criteria (accuracy, precision, recall, efficiency, ar-bitration time, etc.).

2. Literature Review

Chen et al. [20] introduced (AHDCNN) to investigate the potential of different deep-learning techniques for the effective and efficient early diagnosis of kidney illness. A Machine Learning (ML) method was presented by Khan et al. [21] to classify dataset as either having or not having CKD. The experimental results showed that the MAE for LR was 0.035, for J48 it was 0.0229, for NBTree it was 0.0158, for Naïve Bayes (NB) it was 0.0419, for Multiple layer it was 0.265, and SVM at 0.015 [21].

Vasquez-Morales et al.[22] suggested a classifier based on a neural network to deter-mine if an individual is at risk of developing CKD[23]. It is used and verified to explain the predictions of CKD. Malathi et al. [24] introduced a hybrid reasoning-based model for disease prediction. Fuzzy set theory, case-based reasoning, and K-nearest neighbor combine to improve prediction results.

An effective HCS approach for CKD detection was introduced by Aswini et al. [25]. The plan increased productivity by 6%. The Opposition-based-Laplacian-Equilibrium-Optimizer(O-LEO) based Cloud-Sim method decreased the task's execution time. Senan et al. [26] developed a Recursive Feature Elimination (RFE) method for diagnosing CKD. This study uses the Random Forest (RF), Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) algorithms.

Recent liver-specific CNN-based models have shown promising results. For instance, Aswini et al. [25] used a CNN with POA for liver disease prediction with high accuracy. Outcome showed design achieved a 98.08% accuracy rate. Hashem et al. [29] presented a machine learning-based prediction model for the detection of hepatocellular carcinoma (HCC) associated with chronic hepatitis C (CHC). According to the experimental results, the accuracy of the scheme ranged from 93.2% to 95.6%.

To forecast liver disorders, Assegie et al. [30] created a hybrid model that combines SVM and RF. The SVM trained the feature sets, while the RF approach performed the recursive feature elimination. It attained the highest accuracy of 78.3% and increased suggested accuracy by 12.2% compared to the contrary that governs forecasting algorithms [31]. A sophisticated Gaussian SVM learning system was introduced by Ghazal [32] to forecast the chronic infectious disease Hepatitis C. The disease stages were divided into four stages by the Hep-Pred model. In contrast to the current biopsy procedure, which was painful and time-consuming, the method proved to be effective. The accuracy rate of the approach was 97.9%.

We studied CNN architecture from the ground up and evaluated its performance against pretrained data augmentation algorithms[33]. They suggest a safe and noninvasive automatic classification method based on blood tests and CT images for detecting liver disease. The classifier's performance metrics are then assessed using a confusion matrix for optimal sensitivity, specificity, and accuracy scores. Even partial detection can be quite helpful in the case of conditions like cirrhosis[34]. Thus, a reliable diagnostic approach is provided by the ability to detect and categorize liver illnesses early on before they progress and necessitate more stringent surveillance[35]. The study aims to learn a convolution neural network (CNN) and compare its detection performance to get strong generalization capabilities of fatty liver features compared to alternative pre-trained deep CNN designs. CNNs are very useful tools in medical image analysis, particularly for detecting conditions such as fatty liver disease [36].

CNNs are helpful for the automated diagnosis of fatty liver because of their excep-tional ability to recognize intricate patterns and characteristics in medical images[37]. Be-cause these networks have been trained on massive liver image datasets, they can detect subtle variations in tissue density and texture associated with fatty infiltration [38]. CNNs can effectively use input images to extract relevant information and spatial correlations. This paper aims to identify impaired liver function and cirrhosis. The liver's capacity to carry out vital processes, including detoxification, protein synthesis, and nutrition and medication metabolism, is severely compromised by cirrhosis[39].

Many existing studies use machine learning models to predict liver disease but often rely on just one dataset. This can introduce biases specific to that dataset, limiting the model's effectiveness in real-world scenarios. Additionally, traditional models like SVM and decision trees (DT) struggle to grasp the complex relationships between medical parameters fully. Our study combines three datasets to address these challenges and employs a mod-ified CNN, enhanced with the POA. This approach improves feature extraction and en-hances prediction accuracy, significantly outperforming conventional machine learning techniques.

3. Proposed Method

This prediction model efficiently manages patients' liver illness data and will track their health to determine their condition's severity. The CNN framework is used to examine and forecast whether or not a patient has liver illnesses, including cirrhosis of the liver and hepatitis C, using sample data from liver patients gathered suffering from three liver con-ditions databases. The POA and the CNN model are combined to create the CNN model. Even while CNN model can localize data accurately, it has several drawbacks, namely loss functions. The model's prediction performance will suffer greatly if these loss functions are not weighed while training. To balance the loss functions, POA is used to optimize the CNN model's configured hyperparameters, including learning rate, momentum, batch size, and epochs. Consequently, the suggested CNN model can accurately and efficiently identify liver disorders from the datasets, and anticipated findings are saved on servers. The proposed model used a dataset to train modified CNN architecture, and probes are categorized once the body area sensor network has acquired feature vectors following the data preparation stage. The resource manager is given the task, and using data collected from a gateway's devices, the model is created and forecasts outcomes.

Several preprocessing steps were applied to ensure consistency in the merged dataset, including feature alignment, normalization, handling missing values, and feature selection [40]. First, feature alignment helped unify the datasets by resolving differences in attributes and data types. Next, normalization was used to standardize numerical values, ensuring a balanced dataset that wouldn't introduce biases during model training. Missing values were carefully handled by using mean imputation for continuous variables and mode imputation for categorical ones, maintaining the integrity of the data. Lastly, feature selection was made to keep only the most relevant attributes for liver disease prediction, reducing complexity and improving computational efficiency. These steps are crucial in medical data analysis, as they enhance the accuracy and reliability of predictive models [41].

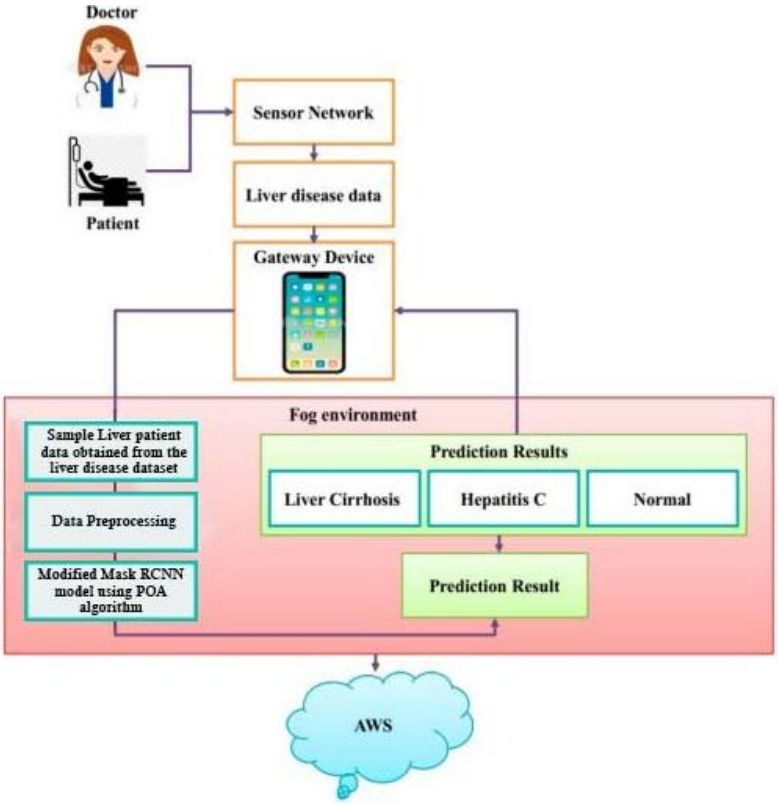

Figure 1. Proposed liver disease detection model's architecture.

Our model meticulously aligned the features across several datasets to guarantee consistency in the training process. We eliminated features in only one dataset, retained only the features shared by all datasets, and used statistical imputation to fill in any missing values. A uniform and well-organized dataset for model training was produced using these procedures. Balance is crucial in this approach to avoid bias toward a larger sample size dataset. Resampling methods like undersampling the majority class or oversampling the minority class (e.g., SMOTE) can be used to accomplish this. Furthermore, domain adaptation techniques like normalization and adversarial training align feature distributions, and weighted loss functions ensure equitable learning across datasets. By dynamically altering learning rates and weights to reduce bias, the Pelican Algorithm can further optimize hyperparameters.

By strengthening the model's generalizability and robustness, these techniques improve the prediction of liver disease across various datasets. The proposed model uses the CNN model to determine whether or not patients have liver problems after receiving identification data from many designs. The IOT model, utilized in the medical field, is another term for the liver disease prediction model. The liver disease prediction model integrates the hardware tools through the software tool, allowing for a smooth and well-thought-out end-to-end integration that yields precise and quick results. Figure 1 shows the general design of the suggested paradigm.

3.1. Convolutional Neural Network(CNN)

In illness prediction tasks, CNN, an enhanced variant of the faster CNN framework, performs better. [42]. The framework comprise CNN model to achieve the intended out-come [43]. The following is a discussion of how these components operate.

To properly utilize the multi-dimensional information, the framework exemplifies the traditional CNN model (usually ResNet) for extracting characteristics. Model extracts features from subsequent layers, whereas low features are extracted before. When the input data is introduced, CNN model, which acts as the centerpiece. The feature maps produced are used as input by the network layers that follow. The FPN supports ResNet and facilitates an efficient feature extraction procedure. FPN can provide a more comprehensive description of the identified sick cases on multiple dimensions. At every level, FPN effectively supports both high-degree and low-degree features.

Region proposal network: RPNs suggest potential areas for creating boxes bound based on data attributes from the backbone. To eliminate the sluggish computing process associated with region proposals utilizing a selective search method, the RPN is proposed in place of selective search. RPN screens all of the regions known as anchors and uses feature maps to determine the position of targeted instances in the dataset. The anchors chosen for the supplemental phase are more likely to receive RPN proposals. The non-maximum-suppression technique is applied based on grades if the anchors overlap to reduce data redundancy between RPN proposals. The RPN loss function can be found in Equation (1).

$$\Gamma_{RP} = \frac{1}{N_{\mathrm{cl}}} \sum_{x} \ell_{\mathrm{cl}} \left(P_x, P_x^*\right) + \mu \frac{1}{N_{\mathrm{rel}}} \sum_{x} P_x^* \, \ell_{\mathrm{re}} (\beta_x, \beta_x^*) \tag{1}$$

In this case, regression loss is represented by the last word, whereas classification loss is represented by the first. β_x indicates the reality, μ stands for the weight balancing factor, and P_x Indicates the. P_x^* represents the labeling, either 0 or 1, corresponding to negative & positive anchors.

ROI alignment: The CNN model requires intriguing areas based on data features to accurately retain the spatial correlation for each pixel to identify the pixels. The alignment layer uses Bi-linear approximation. Approach to compute an accurate localization of the input features[27]. The previously acquired data features are also subjected to average or max pooling techniques for improved refinement.

A completely linked layer that accurately anticipates instances and handles categorization process. The features retrieved from the ROI alignment levels are supplied into the network head. The network head simultaneously performs the tasks of feature extraction. The fully linked layer forecasts the n × n-dimensional against each region of interest. The intended data dimension must be maintained for effective classification results. Additionally, multi-task loss function and sigmoidal function's mathematical representations are explained in Equations (2) and (3).

$$f(\sigma) = \frac{1}{1 + e^{-x}} \tag{2}$$

$$\ell = \ell_{\mathrm{cl}} + \ell_{\mathrm{re}} + \ell_{\mathrm{mask}} \tag{3}$$

3.2. CNN: Using the Pelican Optimization Algorithm

This section contains an empirical framework for creating swarm-based POA[44]. Figure 2 displays the POA algorithm's flow chart.

Population members are randomly selected using Equation (4).

$$y_{p,q} = k_q + \Re \cdot \left(v_q - k_q\right), p = 1,2,3,4, \ldots, M, q = 1,2,3,4, \ldots, n \tag{4}$$

Where y_(p,q) is the value of the qth variable as determined by the pth candidate solution; M: stands for the population size; n for the number of variables in the problem; R is a random number between 0 and 1; k_q is the qth lower bound of the problem variables, and v_q is the qth upper bound, respectively. The presented POA determined in Equation (5) using the matrix, sometimes referred to as the population matrix.

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_p \\ \vdots \\ Y_M \end{bmatrix}_{M \times 1} = \begin{bmatrix} y_{1,1} & \cdots & y_{q,1} & \cdots & y_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_{p,1} & \cdots & y_{p,q} & \cdots & y_{p,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ y_{M,1} & \cdots & y_{M,q} & \cdots & y_{M,n} \end{bmatrix}_{M \times n} \tag{5}$$

Y stands for the pelican population matrix, and Y_p for the pth pelican. The values attained for the function's objective constitute the objective function vector in Equation (6).

$$Fitness = similarity(x, y) = \frac{1}{|h_x - h_y|} \quad (6)$$

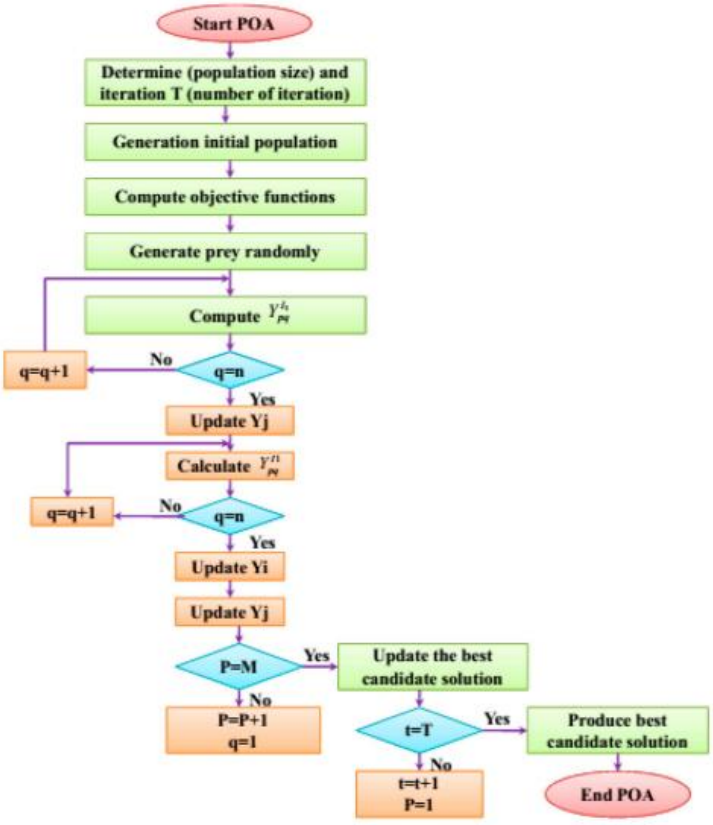

Figure 2. Flowchart of POA.

According to stage 1, y_(p,q) indicates the pth pelican's new state in the qth dimension. P is a random number between one and two, i_q is the prey's location in the qth dimension, and E_i is the value of the objective function. P is an integer parameter that can be either one or two. This parameter is selected at random for every individual iteration. A member experiences greater displacement when the value of this parameter equals 2, which encourages them to explore more recent areas of the search region. Consequently, the parameter P affects how accurately the POA can explore the search space. When the value of the objective function improves, the new location of a pelican is accepted in the evolved POA. This type of updating, called effective updating, prevents the algorithm from moving to non-optimal regions. This process is modeled by the Equation (7).

$$\boldsymbol{Y_p} = \begin{cases} \boldsymbol{Y_p^{I_1}}, & \boldsymbol{E_p^{I_1}} < \boldsymbol{E_p}; \\ \boldsymbol{Y_p}, & \textbf{else} \end{cases} \quad (07)$$

Therefore, the POA is used to fine-tune the CNN's hyperparameters, including the learning rate, number of epochs, momentum, and batch size. The best values for these hyperparameters are set: learning rate = 0.005, epochs = 2000, momentum = 0.0005, and batch size = 12.

4. Results and Discussion

The CNN performance is assessed in this part, and the findings are explained below. The model was trained using the Adam optimizer with ReLU activation in hidden layers and Softmax in the output. All training was done on a system with an NVIDIA RTX 3060 GPU (CUDA 11.7), TensorFlow and Keras frameworks. The public cloud is utilized in this. Separate implementations are used to calculate the POA optimization algorithm's efficiency. A thousand iterations are involved in solving the goal function for each independent implementation. The learning rate, epoch count, and batch size are the ideal CNN

parameters. The POA algorithm yielded the following results: 8, 40, and 0.001 for batch size, epoch count, and learning rate. Table 1 displays the parameter settings.

Table 1. Configuring parameters.

Parameter	Value
Learning-Rate	0.0001
Learning- momentum	0.10
Weight-decay	0.00018
pool-size	15
Totalnoofepoch	100
Validation-Steps	55
Class-No's	3

4.1. Dataset Description

Table 2. summarizes the key attributes of each dataset used in this study:

Dataset	Sample	Attribute	Source
ILPD	583	10	Kaggle[45]
Cirrhosis	424	13	Kaggle[46]
Heapatitis C	615	14	Kaggle[47]

ILPD Dataset [45], the Cirrhosis Prediction Dataset [46], and the Hepatitis C predic-tion dataset[47] are three datasets used in this study. This subsection contains a description of these datasets. Learning paradigm solely focuses on train data once testing & training datasets have been separated, and the testing data is used to calculate performance. Training data makes up 70% of the dataset, while testing and validation data make up 15% each. All of the data are combined in the beginning.

4.1.1. Dataset for predicting Cirrhosis

Information regarding liver cirrhosis was gathered from PBC's Mayo Clinic Trial (Primary-Binary-Cirrhosis) and is included in the data [48]. Data is gathered from 424 PBC patients at 10-year intervals via the Mayo Clinic.

4.1.2. Record of Indian Liver Patients

The dataset consists of 416 liver patient records, and 167 non-liver patient records are included in the dataset [49], which was gathered from the northeast region of Andra Pradesh, India. Gender, age, Alkaline phosphatase, aspartate aminotransferase, and total protein levels and bilirubin are among the statistics included in each of the dataset's columns.

4.1.3. Hepatitis C Prediction Dataset

There are 14 attributes in this dataset, which will be divided into two categories. The patient's data is used to get the first four qualities, while laboratory data is used to gather the remaining attributes. From patient data, characteristics such as x (patient-id), age (in decades), sex (M, F), and diagnosis categories are gathered. The characteristics are GGT (glutamyl-transferase), PROT, CHOLINES (cho-linesterase), CHOLES (cholesterol), CREATI (creatinine), ALBU (Albumin), ALKAL (al-kaline phosphatase), ALAMTF (alanine aminotransferase), ASMTF (aspartate ami-notransferase), and BILIRU (bilirubin)[50].

4.2. Model Training

Assuming that the basic characteristics of additional data are the same, different liver disease datasets used for pretraining in this study. Trained weight files were then moved to datasets to fine-tune network parameters and training. This helps CNN over-come the problem of limited data by achieving decent results with small datasets. The following lists the benefits of transfer learning. (1) The model's performance improves significantly during training. (2) At first, it performs well. (3) High performance is achieved via trained models.

4.3. Performance Evaluation:

Matthew's correlation coefficient (MCC's) evaluation using different approaches is shown in Figure 3. The MCC values of different models are compared with MCC value of the suggested CNN. The outcome

demonstrates that the CNN outperforms current techniques regarding MCC value, with an MCC value of roughly 94.89%.

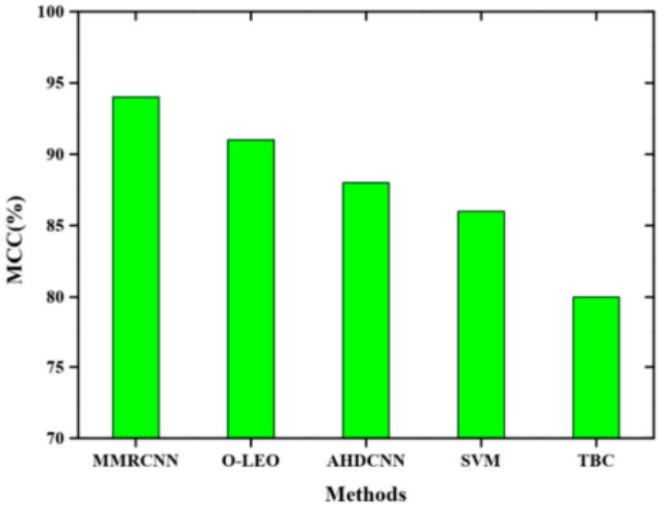

Figure 3. MCC Analysis on various models.

Figure 4 illustrates CNN's performance rate evaluation using several performance indicators, including accuracy, precision, recall, and F-measure. According to the figure, the CNN attains 98% accuracy, 96.2% precision, 97.3% F-measure, and 95% recall. Table 3 discusses the performance evaluation of suggested CNN using various crossfolds. The suggested CNN outperforms the other two lower folds in metrics by a factor of 10.

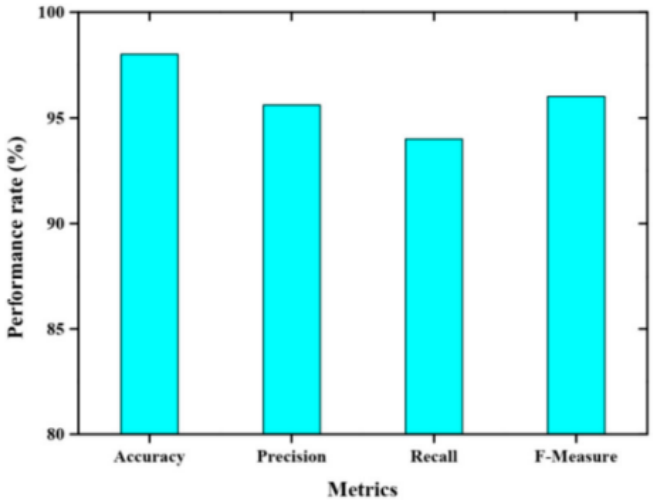

Figure 4. Evaluation of the suggested CNN method's performance.

Table 3. Assessment of CNN's performance with several cross folds.

Cross-Folds No's	Accur(%)	Prec(%)	Recall(%)	F1-Score(%)	MCC(%)
3Folds	93.24	94.13	94.53	95.72	94.82
5Folds	94.85	95.34	95.23	96.82	96.61
10Folds	96.16	97.47	97.38	97.89	97.99

The examination of the execution times of different approaches is shown in Figure 5. The CNN execution time is contrasted with the execution times of the O-LEOO, AHCNN, SVM, and TBC strategies. Analysis demonstrates CNN's duration of execution is shorter rather than current techniques, with a final execution time of 530 ms.

Table 4 shows how the CNN model, optimized with the POA, predicts liver disease using different datasets. It compares the model's effectiveness across three separate da-tasets—ILPR Cirrhosis Prediction Dataset, and Hepatitis C Prediction Dataset—and a merged dataset that combines all three. The ILPR dataset achieves 92.45% accuracy, while the Cirrhosis Prediction Dataset and Hepatitis C Prediction Dataset perform slightly better at

94.78% and 96.12% accuracy, respectively. However, when these datasets are merged, the accuracy jumps to 98.67%. These findings highlight how multi-dataset integration and advanced optimization techniques like POA enhance the accuracy and reliability of liver disease prediction, making it a powerful approach for real-world medical diagnosis.

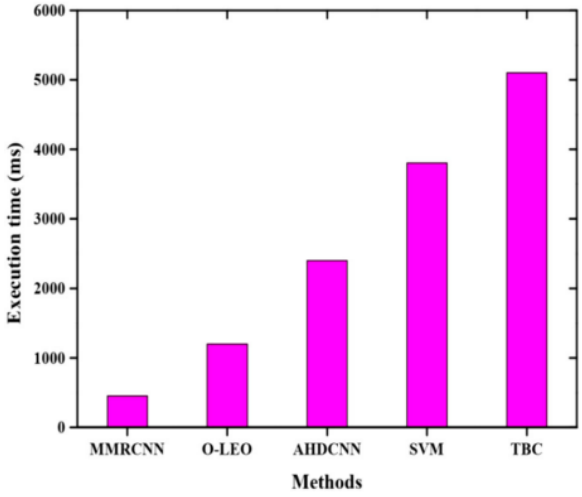

Figure 5. Evaluation of execution duration.

Table 4. Comparative Dataset Analysis.

Dataset	Accur(%)	Prec(%)	Recall(%)	F1-score(%)	MCC(%)
Indian Liver Patient	92.45	91.30	93.20	92.24	90.89
Cirrhosis Prediction Dataset	94.78	93.90	95.50	94.69	93.32
Hepatitis C Prediction Dataset	96.12	95.80	96.70	96.25	95.45
Merged Dataset	98.67	96.25	95.90	97.33	94.89

4.4. Ablation Study

Table 5 shows that the CNN model predicts liver disease when integrated with mul-tiple datasets, comparing results with and without the POA. The findings show a clear advantage of using POA, as it boosts accuracy from 94.52% to 98.67%, precision from 93.70% to 96.25%, recall from 94.00% to 95.90%, and F1-score from 93.85% to 97.33%..

Table 5. Performance evaluation of CNN with and without POA.

Model Variant	Accur(%)	Prec(%)	Recall(%)	F1- Score(%)
CNN without POA	94.52	93.70	94.00	93.85
CNN with POA	98.67	96.25	95.90	97.33

With the increasing prevalence of liver disorders, it has become crucial to develop novel methods for predicting liver illness. We introduced an CNN with combinataioan of CNN and POA architecture to enable earlier detection. The preprocessed images are ex-tracted for feature, followed by the CNN technique to generate bounding boxes. Hy-perparameters such as learning-rate, epoches, and momentum were fine-tuned using the POA.

Table 6 highlights the higher performance of the proposed methodology in predicting illness of liver by comparing several machine learning methods for predictive accuracy, preci-sion, and sensitivity. While conventional models such as SVM, LR, DT, RF, and NB demonstrate moderate accuracy (between 69.26% and 83%), the suggested CNN with POA outperforms all current techniques with 98.67% accuracy, 96.25% precision, and 95.90% sensitivity. The POA, which improves feature extraction and classification, and mul-ti-dataset integration, which permits a more generic model, are credited with this devel-opment. According to the findings, CNN with POA is an advantageous model for de-tecting liver illness early and accurately, reducing false negatives, and enhancing diagnostic dependability.

Table 6. Comparisons with other existing methods.

Ref	Method	Accuracy (%)	Precision (%)	Sensitivity (%)
[51]	SVM	71	64.1	71.5
	Back Propagation	73.2	65.7	73.3
[52]	Logistic	73.39	57.69	22.73
	Linear LR	72.10	52.17	18.18
	SVM	71.24	47.83	16.67
[53]	LR	74.36	72.33	77.11
	RF	71.87	73.44	77.25
[54]	DT	81	84.25	85.32
	RF	77	78	75
	SVM	77	74.25	77.98
[55]	LR	73.97	75.25	80.22
	SVM	71.97	75.25	74.68
	K-NN	73.97	77.39	77.33
[56]	AdaBoost	70.25	76.3	77.69
	RF	69.26	79.66	78.22
[57]	Naive Bayes	61.28	55.8	74.5
	SVM	79.66	76.6	75.7
Proposed	CNN with POA	98.67	96.25	95.90

Several factors influence the evolutionary significance of liver disease in this analysis. The CNN Classifier computations are mastered in multiple instances. Once the evaluation workflow is completed, clinician-given selectivity is employed appropriately, thanks to the increasing number of predefined classes. With the help of our technique, the initial phase of antifungal medicine therapeutic assurances is carefully planned with the patients at risk in mind. This allows physicians and patients in the final stage to manage themselves appro-priately. Compared to other machine learning techniques, the classification method solves the issue more successfully and produces better outcomes when considering the regression strategy. As mentioned earlier, the CNN is crucial to achieving a favorable outcome while striving for the classification algorithm. By contemplating a suitable and straightforward model that allows clinical position patients to have certain qualities, the CNN model complexity is avoided. A successful pattern investigation requires additional data from the time series, but this data is not available in the clinical setting. It is challenging to locate the patient's data in the hospital database since patient's text from medical history is not organized in table form. For this investigation, text mining techniques based on key-word searches are employed.

5. Conclusion

The modified CNN model architecture for liver disease prediction is presented in this research. Liver disease is successfully localized within datasets by use of the CNN model. Nevertheless, CNN model's loss functions introduce significant flaws that eventually affect prediction accuracy. Consequently, a POA approach is used to balance and reduce loss functions that arise throughout the training process for data to improve prediction accuracy. An CNN is an enhanced version of the CNN model created by a POA. Three different dataset types—the liver disease patient Prediction Dataset, ILPD Dataset, and (HCIFPD) Dataset with multiple features are used in this study to forecast liver illness. The effec-tiveness of the suggested CNN approach in identifying liver illness in patients, such as cirrhosis or hepatitis, is evaluated using several performance criteria. Better accuracy of 98.78%, precision of 96.26%, Fmeasure of 97.32%, and recall of 95.33% are all attained by the CNN. The suggested CNN approach also achieves reduced execution, arbitration, and latency times. Limitations of this study include lack of clinical validation and potential generalization issues to other organs or imaging modalities. To detect different diseases in different settings, we hope to further study different CNN parameters in the future, such as ROI, anchor scale, and backbone stride. It may be possible to improve diagnosis by gathering additional inputs, such as nephritis end-stage

illness, gastrointestinal bleeding, osteoporosis, recent bone fractures, and so forth. This work must be expanded to include further clinical trials with photographs and applied to various potential uses.

References

1. S. Ambesange, V. A, R. Uppin, S. Patil, and V. Patil, "Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques," in 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Nov. 2020, pp. 98–102. doi: 10.1109/CCEM50674.2020.00030.
2. N. Li, Y. Jiang, G. Gong, G. Han, and J. Ma, "Non-Invasive Assessment Model of Liver Disease Severity by Serum Markers Using Cloud Computing and Internet of Things," IEEE Access, vol. 6, pp. 33969–33976, 2018, doi: 10.1109/ACCESS.2018.2849016.
3. A. Sivasangari, B. J. Krishna Reddy, A. Kiran, and P. Ajitha, "Diagnosis of Liver Disease using Machine Learning Models," in 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Oct. 2020, pp. 627–630. doi: 10.1109/I-SMAC49090.2020.9243375.
4. 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," J. Futur. Artif. Intell. Technol., vol. 1, no. 1, pp. 51–63, Jun. 2024, doi: 10.62411/faith.2024-12.
5. J. Aswini, B. Yamini, K. Venkata Ramana, and J. Jegan Amarnath, "An Efficient Liver Disease Prediction Using Mask-Regional Convolutional Neural Network and Pelican Optimization Algorithm," IETE J. Res., vol. 70, no. 2, pp. 1985–1996, Feb. 2024, doi: 10.1080/03772063.2022.2160839.
6. N. Nahar, F. Ara, M. A. I. Neloy, V. Barua, M. S. Hossain, and K. Andersson, "A Comparative Analysis of the Ensemble Method for Liver Disease Prediction," in 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), Dec. 2019, pp. 1–6. doi: 10.1109/ICIET48527.2019.9290507.
7. G. Shobana and K. Umamaheswari, "Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling," in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Apr. 2021, pp. 1223–1229. doi: 10.1109/ICCMC51019.2021.9418333.
8. D. Devikanniga, A. Ramu, and A. Haldorai, "Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm," EAI Endorsed Trans. Energy Web, vol. 7, no. 29, p. 164177, Jul. 2018, doi: 10.4108/eai.13-7-2018.164177.
9. k. Thirunavukkarasu, A. S. Singh, M. Irfan, and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), Dec. 2018, pp. 1–3. doi: 10.1109/CCAA.2018.8777655.
10. Z. Yao, J. Li, Z. Guan, Y. Ye, and Y. Chen, "Liver disease screening based on densely connected deep neural networks," Neural Networks, vol. 123, pp. 299–304, Mar. 2020, doi: 10.1016/j.neunet.2019.11.005.
11. N. Nahar and F. Ara, "Liver Disease Prediction by Using Different Decision Tree Techniques," Int. J. Data Min. Knowl. Manag. Process, vol. 8, no. 2, pp. 01–09, Mar. 2018, doi: 10.5121/ijdkp.2018.8201.
12. M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using VGG19: Caltech-101 image data set," J. Ambient Intell. Humaniz. Comput., vol. 14, no. 4, pp. 3609–3620, Apr. 2023, doi: 10.1007/s12652-021-03488-z.
13. K. Shaheed, A. Mao, I. Qureshi, M. Kumar, S. Hussain, and X. Zhang, "Recent advancements in finger vein recognition technology: Methodology, challenges and opportunities," Inf. Fusion, vol. 79, pp. 84–109, Mar. 2022, doi: 10.1016/j.inffus.2021.10.004.
14. S. Gupta, N. Mohan, and M. Kumar, "A Study on Source Device Attribution Using Still Images," Arch. Comput. Methods Eng., vol. 28, no. 4, pp. 2209–2223, Jun. 2021, doi: 10.1007/s11831-020-09452-y.
15. S. Walia, K. Kumar, M. Kumar, and X.-Z. Gao, "Fusion of Handcrafted and Deep Features for Forgery Detection in Digital Images," IEEE Access, vol. 9, pp. 99742–99755, 2021, doi: 10.1109/ACCESS.2021.3096240.
16. H. Kaur and M. Kumar, "Offline handwritten Gurumukhi word recognition using eXtreme Gradient Boosting methodology," Soft Comput., vol. 25, no. 6, pp. 4451–4464, Mar. 2021, doi: 10.1007/s00500-020-05455-w.
17. F. Mazhar, W. Akbar, M. Sajid, N. Aslam, M. Imran, and H. Ahmad, "Boosting Early Diabetes Detection: An Ensemble Learning Approach with XGBoost and LightGBM," J. Comput. Biomed. Informatics, vol. 6, no. 02, pp. 127–138, 2024.
18. F. Mazhar, N. Aslam, A. Naeem, H. Ahmad, M. Fuzail, and M. Imran, "Enhanced Diagnosis of Skin Cancer from Dermoscopic Images Using Alignment Optimized Convolutional Neural Networks and Grey Wolf Optimization," J. Comput. Theor. Appl., vol. 2, no. 3, pp. 368–382, Jan. 2025, doi: 10.62411/jcta.11954.

19. A. Ben Slama, H. Sahli, Y. Amri, and S. Labidi, "V-NET-VGG16: Hybrid deep learning architecture for optimal segmentation and classification of multi-differentiated liver tumors," Intell. Med., vol. 11, p. 100210, 2025, doi: 10.1016/j.ibmed.2025.100210.
20. A. Naeem, T. Anees, M. Khalil, K. Zahra, R. A. Naqvi, and S.-W. Lee, "SNC_Net: Skin Cancer Detection by Integrating Handcrafted and Deep Learning-Based Features Using Dermoscopy Images," Mathematics, vol. 12, no. 7, p. 1030, Mar. 2024, doi: 10.3390/math12071030.
21. A. Naeem and T. Anees, "DVFNet: A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images," PLoS One, vol. 19, no. 3, p. e0297667, Mar. 2024, doi: 10.1371/journal.pone.0297667.
22. G. R. Vasquez-Morales, S. M. Martinez-Monterrubio, P. Moreno-Ger, and J. A. Recio-Garcia, "Explainable Prediction of Chronic Renal Disease in the Colombian Population Using Neural Networks and Case-Based Reasoning," IEEE Access, vol. 7, pp. 152900–152910, 2019, doi: 10.1109/ACCESS.2019.2948430.
23. A. Naeem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," IEEE Access, vol. 8, pp. 110575–110597, 2020, doi: 10.1109/ACCESS.2020.3001507.
24. M. D., L. R., S. V., V. V., and A. K. Sangaiah, "Hybrid Reasoning-based Privacy-Aware Disease Prediction Support System," Comput. Electr. Eng., vol. 73, pp. 114–127, Jan. 2019, doi: 10.1016/j.compeleceng.2018.11.009.
25. J. Aswini, B. Yamini, R. Jatothu, K. S. Nayaki, and M. Nalini, "An efficient cloud-based healthcare services paradigm for chronic kidney disease prediction application using boosted support vector machine," Concurr. Comput. Pract. Exp., vol. 34, no. 10, p. e6722, May 2022, doi: 10.1002/cpe.6722.
26. E. M. Senan et al., "Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques," J. Healthc. Eng., vol. 2021, no. 1, pp. 1–10, Jun. 2021, doi: 10.1155/2021/1004767.
27. N. Bhaskar, M. Suchetha, and N. Y. Philip, "Time Series Classification-Based Correlational Neural Network With Bidirectional LSTM for Automated Detection of Kidney Disease," IEEE Sens. J., vol. 21, no. 4, pp. 4811–4818, Feb. 2021, doi: 10.1109/JSEN.2020.3028738.
28. A. Naeem, T. Anees, M. Fiza, R. A. Naqvi, and S.-W. Lee, "SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images," Sensors, vol. 22, no. 15, p. 5652, Jul. 2022, doi: 10.3390/s22155652.
29. S. Hashem et al., "Machine Learning Prediction Models for Diagnosing Hepatocellular Carcinoma with HCV-related Chronic Liver Disease," Comput. Methods Programs Biomed., vol. 196, p. 105551, Nov. 2020, doi: 10.1016/j.cmpb.2020.105551.
30. T. Admassu Assegie, R. Subhashni, N. Komal Kumar, J. Prasath Manivannan, P. Duraisamy, and M. Fentahun Engidaye, "Random forest and support vector machine based hybrid liver disease detection," Bull. Electr. Eng. Informatics, vol. 11, no. 3, pp. 1650–1656, Jun. 2022, doi: 10.11591/eei.v11i3.3787.
31. A. Naeem, T. Anees, R. A. Naqvi, and W.-K. Loh, "A Comprehensive Analysis of Recent Deep and Federated-Learning-Based Methodologies for Brain Tumor Diagnosis," J. Pers. Med., vol. 12, no. 2, p. 275, Feb. 2022, doi: 10.3390/jpm12020275.
32. T. M. Ghazal et al., "Hep-Pred: Hepatitis C Staging Prediction Using Fine Gaussian SVM," Comput. Mater. Contin., vol. 69, no. 1, pp. 191–203, 2021, doi: 10.32604/cmc.2021.015436.
33. A. Kiran, V. Neerugatti, D. S. Kumar, J. Somasekar, S. Boddupalli, and B. Sai Praneetha, "Revolutionizing cirrhosis liver disorder diagnosis: A deep learning approach with convolutional neural networks," in Challenges in Information, Communication and Computing Technology, London: CRC Press, 2024, pp. 449–454. doi: 10.1201/9781003559085-79.
34. H. S. Yadav and R. K. Singhal, "Classification and Prediction of Liver Disease Diagnosis Using Machine Learning Algorithms," in 2023 2nd International Conference for Innovation in Technology (INOCON), Mar. 2023, pp. 1–6. doi: 10.1109/INOCON57975.2023.10101221.
35. S. Mohamed, R. Ezzat, S. Ghorab, R. Bhatnagar, and M. Y. Shams, "Liver Disease Identification Based on Machine Learning Algorithms," in 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Nov. 2023, pp. 1062–1067. doi: 10.1109/ICTACS59847.2023.10390061.
36. S. Chauhan, A. Bhardwaj, T. Jain, and A. Pandey, "Advancements in non-alcoholic fatty liver disease (NAFLD) detection: A machine learning approach," in AIP Conference Proceedings, 2025, vol. 3253, no. 1, p. 030008. doi: 10.1063/5.0250281.
37. Sajid, M., Malik, K. R., Almogren, A., Malik, T. S., Khan, A. H., Tanveer, J., & Rehman, A. U. (2024). Enhancing intrusion detection: a hybrid machine and deep learning approach. Journal of Cloud Computing, 13(1), 123..

38. J. G. Suárez-García, B. de C. Alonso, J. M. Hernández-López, S. S. Hidalgo-Tobón, P. Dies-Suárez, and P.-W. So, "Automated MRI quantification of pediatric abdominal adipose tissue using convolutional neural networks and novel total intensity maps," Biomed. Signal Process. Control, vol. 102, p. 107250, Apr. 2025, doi: 10.1016/j.bspc.2024.107250.
39. K. Mohit, R. Gupta, and B. Kumar, "Contrastive Learned Self-Supervised Technique for Fatty Liver and Chronic Liver Identification," Biomed. Signal Process. Control, vol. 100, p. 106950, Feb. 2025, doi: 10.1016/j.bspc.2024.106950.
40. A. Q. 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, 2023, doi: 10.3390/biomedicines11020581.
41. D. Herrera-Sánchez et al., "Imaging Estimation for Liver Damage Using Automated Approach Based on Genetic Programming," Math. Comput. Appl., vol. 30, no. 2, p. 25, Feb. 2025, doi: 10.3390/mca30020025.
42. Z. U. Rehman et al., "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on CNN and transfer learning: An application for smart agriculture," IET Image Process., vol. 15, no. 10, pp. 2157–2168, Aug. 2021, doi: 10.1049/ipr2.12183.
43. Y. Zhan, W. Liu, and Y. Maruyama, "Damaged Building Extraction Using Modified CNN Model Using Post-Event Aerial Images of the 2016 Kumamoto Earthquake," Remote Sens., vol. 14, no. 4, p. 1002, Feb. 2022, doi: 10.3390/rs14041002.
44. P. Trojovský and M. Dehghani, "Pelican Optimization Algorithm: A Novel Nature-Inspired Algorithm for Engineering Applications," Sensors, vol. 22, no. 3, p. 855, Jan. 2022, doi: 10.3390/s22030855.
45. UCI Machine Learning, "Indian Liver Patient Records," Kaggle.com, 2018. https://www.kaggle.com/datasets/uciml/indian-liver-patient-records
46. Fedesoriano, "Cirrhosis Prediction Dataset," Kaggle.com, 2021. https://www.kaggle.com/datasets/fedesoriano/cirrhosis-prediction-dataset
47. Fedosriano, "Hepatitis C Prediction Dataset," Kaggle.com, 2021. https://www.kaggle.com/datasets/fedesoriano/hepatitis-c-dataset
48. T. R. Fleming and D. P. Harrington, Counting Processes and Survival Analysis, vol. 625. Wiley, 2005. doi: 10.1002/9781118150672.
49. A. Iyer and S. Narayanaswami, "A Novel Model Using ML Techniques for Clinical Trial Design and Expedited Patient Onboarding Process," Clin. Outcomes Res., vol. Volume 17, pp. 1–18, Jan. 2025, doi: 10.2147/CEOR.S479603.
50. G. Hoffmann, A. Bietenbeck, R. Lichtinghagen, and F. Klawonn, "Using machine learning techniques to generate laboratory diagnostic pathways—a case study," J. Lab. Precis. Med., vol. 3, no. 6, pp. 58–58, Jun. 2018, doi: 10.21037/jlpm.2018.06.01.
51. M. Ghosh et al., "A Comparative Analysis of Machine Learning Algorithms to Predict Liver Disease," Intell. Autom. Soft Comput., vol. 30, no. 3, pp. 917–928, 2021, doi: 10.32604/iasc.2021.017989.
52. B. Yang et al., "Correlation between liver fibrosis in non-alcoholic fatty liver disease and insulin resistance indicators: a cross-sectional study from NHANES 2017–2020," Front. Endocrinol. (Lausanne)., vol. 16, p. 1514093, Jan. 2025, doi: 10.3389/fendo.2025.1514093.
53. P. Pavani, P. Sravanthi, C. Suneetha, M. Azhar, L. H. supriya Reddy, and T. Maddineni, "Advanced Machine Learning Approaches for Liver Disease Prediction: A Focus on Feature Selection and Hybrid Ensemble Model," in 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Jan. 2025, pp. 1003–1010. doi: 10.1109/ICMCSI64620.2025.10883633.
54. P. Theerthagiri, "Liver disease classification using histogram-based gradient boosting classification tree with feature selection algorithm," Biomed. Signal Process. Control, vol. 100, p. 107102, Feb. 2025, doi: 10.1016/j.bspc.2024.107102.
55. P. Le et al., "Estimated Burden of Metabolic Dysfunction–Associated Steatotic Liver Disease in US Adults, 2020 to 2050," JAMA Netw. Open, vol. 8, no. 1, p. e2454707, Jan. 2025, doi: 10.1001/jamanetworkopen.2024.54707.
56. N. K. Telugu and T. F. Fernandez, "The prediction of liver cirrhosis using logistic regression algorithm by comparing accuracy with the Ada boost algorithm," in AIP Conference Proceedings, 2025, vol. 3252, no. 1.
57. N. Arora, S. Srivastava, A. Tripathi, and V. Gupta, "An enhanced predictive modelling framework for highly accurate non-alcoholic fatty liver disease forecasting," Indones. J. Electr. Eng. Comput. Sci., vol. 37, no. 1, p. 214, Jan. 2025, doi: 10.11591/ijeecs.v37.i1.pp214-222.