LIVER DISEASE DETECTION USING DEEP LEARNING

*Abdul Rahman,*

[abdulrahman.college02@gmail.com,](mailto:abdulrahman.college02@gmail.com,)

Student of Computer Engineering Department at Maulana Mukhtar Ahmad Nadvi Technical Campus Malegaon Maharashtra India 423203

*Dr. Salman Baig,*

[salmanbaig@mmantc.edu.in](mailto:salmanbaig@mmantc.edu.in)

Head of Computer Engineering Department Maulana at Mukhtar Ahmad Nadvi Technical Campus Malegaon Maharashtra India 423203.

*Ansari Farzan,*

[ansarifarzan771@gmail.com,](mailto:ansarifarzan771@gmail.com,)

Student of Computer Engineering Department at Maulana Mukhtar Ahmad Nadvi Technical Campus Malegaon Maharashtra India 423203

*Isar Ahmed,*

[isarfaridi@gmail.com,](mailto:isarfaridi@gmail.com,)

Student of Computer Engineering Department at Maulana Mukhtar Ahmad Nadvi Technical Campus Malegaon Maharashtra India 423203

*Mohammed Hassan,*

[mdhassankm0011@gmail.com*,*](mailto:mdhassankm0011@gmail.com,)

Student of Computer Engineering Department at Maulana Mukhtar Ahmad Nadvi Technical Campus Malegaon Maharashtra India 423203

*Asst. Professor Master Simmi*

[simmimaster@gmail.com](mailto:simmimaster@gmail.com)

Assistant Professor Department of Computer Engineering at Maulana Mukhtar Ahmad Nadvi Technical Campus Malegaon Maharashtra India 423203

## Abstract

Early diagnosis of liver disease plays a crucial role in successful treatment and better patient outcomes. This paper proposes a deep learning framework with Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) for the binary categorization of liver disease. The data set consists of 2,391 medical records obtained from Kaggle, local diagnostic centers of Maharashtra, and AI-synthesized samples to overcome class imbalance. The important features are demographic and biochemical liver function tests. Preprocessing steps of removal of outliers, feature scaling (z-score normalization), and SMOTE for class balancing were performed. The ANN with dropout regularization obtained a testing accuracy of 92% with an AUC of 0.98, and the CNN model performed similarly.‎ The findings indicate that deep learning provides a suitable, non-invasive decision-support mechanism for detection of liver disease in clinical practice.‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎ ‎  
  
Keywords: Liver Disease, Deep Learning, ANN, CNN, SMOTE, Healthcare Analytics

## Introduction

This chapter gives a general overview of liver disease, its prevalence, and the difficulties of early detection. It provides an introduction to the motivation of applying Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) to the diagnosis of liver disease and describes the objective of the project, which is to create a non-invasive data-driven model of detection. The chapter also provides the problem statement, project scope, and particular objectives for classification accuracy. By analyzing various patient indicators, this project demonstrates how ANN and CNN models effectively classify individuals based on liver health [1][2].

The liver, located in the upper gastrointestinal tract, weighs between 1,200–1,800 grams and is vital for digestion, metabolism, detoxification, immune function, and nutrient storage [1]. Liver diseases, which can lead to serious health issues and even death, are categorized based on their causes—such as infections, injuries, genetic abnormalities, or exposure to toxins [3]—and their effects on liver function [4]. Key liver diseases include non-alcoholic fatty liver disease (NAFLD) [6], cirrhosis [5], and various forms of hepatitis [12]. NAFLD is characterized by lipid accumulation, often linked to obesity and diabetes [20]. Cirrhosis involves the replacement of healthy liver tissue with scar tissue, frequently caused by alcoholism or chronic hepatitis [17]. Hepatitis can be acute or chronic and is primarily caused by viral infections from a group of hepatitis viruses (A, B, C, D, and E) [13], each with distinct transmission methods [15]. Common liver diseases such as NAFLD, cirrhosis, and hepatitis have different health implications [3][4]. Advanced stages of fatty liver disease can significantly elevate mortality risk [19]. Preventative measures include reducing alcohol consumption [18], practicing safe sex [14], avoiding shared needles [16], getting vaccinated against hepatitis A and B [11], and maintaining a healthy lifestyle [10]. Traditional methods for diagnosing liver disease are often invasive and costly, emphasizing the need for non-invasive solutions like deep learning models [6][22]. Increasingly, liver disease is linked to obesity and diabetes [21], posing a silent but severe health threat. With advancements in artificial intelligence (AI) and machine learning (ML) [23], healthcare professionals are improving disease detection through better data collection and analysis [24]. Emerging research suggests that gut microbiota may play a significant role in liver disease progression and management [26]. Dysbiosis, or an imbalance in gut bacteria, has been linked to NAFLD and cirrhosis, highlighting potential therapeutic interventions through probiotics and dietary modifications [27]. Additionally, genetic predisposition is increasingly recognized as a contributing factor to liver diseases, with genome-wide association studies (GWAS) identifying specific genes linked to liver dysfunction [28].

This paper also highlights the potential of ANN and CNN for liver disease detection, focusing on classifying patients based on crucial features such as age [39], bilirubin levels [40], liver enzyme levels [42], and other liver indicators [45]. By integrating these technologies, clinicians can enhance diagnostic accuracy and improve patient outcomes in liver health management [25].

### Materials and Methods

* **Dataset Description:**

The dataset utilized in this study comprises 2,391 patient records, derived from three sources:

* 1. A publicly available dataset from Kaggle [7].
  2. Collected from local Laboratories from Maharashtra.
  3. Synthetic data generated using AI techniques to ensure diversity and balance in class distribution.
* Each record contains the following features:
  + **Demographics:** Age, Gender

### Liver Function Indicators:

* + - Bilirubin levels (Total and Direct)
    - Enzyme concentrations (e.g., Alanine Transaminase (ALT), Aspartate Transaminase (AST))
    - Protein levels (e.g., Albumin, Globulin ratio)

**Table:1 Dataset Features**

T

|  |  |  |
| --- | --- | --- |
| **Feature** | **Type** | **Description** |
| Gender [38] | Nominal | This feature illustrates the participant’s gender. |
| Age (years) [39] | Numeric | The age range of the participants is 4– 90 years. |
| Total Bilirubin—TB (mg/dL) [40] | Numeric | This feature captures the participant’s |
| Direct Bilirubin— DB (mg/dL) [40] | Numeric | This feature captures the participant’s direct bilirubin. |
| Alkaline Phosphatase—ALP (IU/L) [41] | Numeric | This feature captures the participant’s alkaline phosphatase. |
| Alanine Aminotransferase— SGPT (U/L) [42] | Numeric | This feature captures the participant’s alanine aminotransferase. |
| Aspartate Aminotransferase— SGOT (U/L) [42] | Numeric | This feature captures the participant’s aspartate aminotransferase. |
| Total Protein—TP (g/L) [43] | Numeric | This feature captures the participant’s total protein. |
| Albumin—ALB (g/dL) [44] | Numeric | This feature captures the participant’s albumin. |
| Albumin and Globulin Ratio—AGR [45] | Numeric | This feature captures the participant’s albumin and globulin Ratio. |
| Liver Disease | Nominal | This feature stands for whether the participant has been diagnosed with liver disease or not. |

* **Data Collection and Preprocessing:** The dataset for this research comprises 2391 patient records. Each record contains demographic and clinical attributes related to liver function, including age, gender, bilirubin levels, enzyme levels, and protein ratios. To maintain data quality, the preprocessing process included:

## Handling Missing Values:

One of the initial tasks in data preprocessing is to address missing or incomplete data. Missing data may be due to numerous reasons, e.g., errors during data collection or patients not receiving certain tests. Missing values need to be addressed since most machine learning algorithms, including ANNs, do not directly work with missing data. In our dataset, we performed a rigorous check for missing values with pandas functions. Thankfully, the data set contained no missing values across any of the features. If missing values did exist, multiple imputation approaches could have been used:

* **Mean/Median Imputation**: For continuous variables, replacing missing values with the mean or median of the available data.
* **Mode Imputation**: For categorical variables like gender, missing values could be replaced with the most frequent category.

## Removing Outliers:

For detection of outliers, we employed the use of a Box Plot that plotted the distribution of data as well as searched for extreme values. We applied the Interquartile Range (IQR) method next in order to delete outliers consistently. The use of the IQR method would see us measuring the range of Q1 to Q3, then using it to detect as well as discard values that exist beyond the threshold acceptable limits of say 1.5 multiples of the IQR. Once outliers had been eliminated, missing values were imputed to make the dataset complete and suitable for further processing. Based on the type of data, varying imputation strategies like mean, median, or regression-based imputation were utilized to maintain consistency and integrity of the data.

These preprocessing operations are critical to make sure that the data is clean, trustworthy, and prepared for the following analysis and modelling processes. By dealing with outliers and missing values, we enhance the quality of the data, eliminate noise, and increase the accuracy of any machine learning model or statistical analysis carried out.

## Feature Scaling:

Artificial Neural Networks (ANNs) are sensitive to input feature scale. If a feature has a significantly larger range than other features, it may overwhelm the learning process and result in poor performance. For instance, the Age feature varies from 0 to 90, while Albumin levels are in a very small range, approximately 1-5. To make all features contribute equally, feature scaling was used.

* **Standardization**: In this project, standardization (z-score normalization) was applied, which transforms each feature to have a mean of 0 and a standard deviation of 1. This is done using the formula:
  + z=σ(X−μ)

Where,

* x is the original value,
* μ is the mean of the feature,
* σ is the standard deviation of the feature.

Standardization guarantees that all the features are of the same scale, irrespective of their original ranges. This is especially necessary for distance-based algorithms and neural networks, where feature magnitudes can influence the learning process. The Standard Scaler of the scikit-learn library was utilized to scale the dataset. The scaler was applied to both the training data and then consistently to the training and test data to make sure that the test data were treated identically.

## Data Splitting:

To test the performance of the **Artificial Neural Network (ANN)** and **Convolutional Neural Network (CNN)** models, the dataset was divided into training and test sets. This is the common machine learning practice to ascertain that the model will perform well on unseen data.

* + - **Train-Test Split:** We split the dataset into 80% training data and 20% test data using the train\_test\_split function from the scikit-learn library. This 80/20 proportion is widely embraced in machine learning projects since it offers sufficient data for training and still maintains an adequate amount of data for testing the generalization capability of the model.

The training set is utilized to train the model, either the ANN or CNN, by learning the relationships and patterns of the data. The test set is held back to test how well the trained model can perform on new, unseen data so that it can generalize to new, real-world instances.

The CNN is especially useful for tasks like image classification, object detection or any type of task where spatial hierarchies and patterns play a significant role. CNNs utilize convolutional filters to identify local features like edges, textures, and shapes and hence are extremely powerful in visual data analysis.

By dividing the data into training and test sets, we can avoid overfitting of the models, in which they could learn the data by heart rather than the underlying patterns. The test set gives us a real-world estimate of how the model would behave when put to use.

Moreover, employing both ANN and CNN enables us to compare how well general-purpose neural networks (ANNs) and more specific deep learning architectures (CNNs) perform in dealing with the dataset, and thus enable us to select the best model for the task.

## Handling Imbalanced Data (SMOTE)

In most machine learning problems, particularly classification, the data set might possess an imbalanced class distribution where one of the classes contains a much larger number of instances than the other. The imbalance, in this case, causes biased model performance since the model will end up being biased towards predicting the majority class and performing poorly in predicting the minority class.

To deal with this problem, we employed the use of the Synthetic Minority Over-sampling Technique (SMOTE). SMOTE is a powerful method for dealing with imbalanced data through the creation of synthetic examples of the minority class. In contrast to naive oversampling, where minority class instances are simply copied, SMOTE generates new, reasonable instances by interpolating between current points in the feature space.

* How SMOTE works:
  + - * SMOTE selects instances from the minority class.
      * For each selected instance, it finds its nearest neighbors.
      * A synthetic instance is created by taking a random point along the line segment between the selected instance and one of its neighbors.
      * This process is repeated until the minority class is sufficiently balanced with the majority class.

This approach helps in several ways:

1. **Better Model Performance**: Through dataset balancing, SMOTE eliminates the tendency of the model to be skewed toward the majority class and enhances its performance in identifying patterns pertaining to the minority class.
2. **Improved Generalization:** The artificially created samples by SMOTE assist the model in generalizing more accurately to new, unseen data, enhancing its performance in real-world applications where class distributions are often imbalanced.
3. **Less Overfitting:** In contrast to basic oversampling, SMOTE minimizes the chances of overfitting because it generates new instances rather than simply duplicating the available data. We used SMOTE on the training set following the initial train-test split to make sure that the model learns a well-balanced representation of both classes. It was a crucial step in those instances where the model initially performed badly in predicting the minority class, making for more balanced and accurate classification results.

### Model Development

In this study, we introduced a deep learning-based methodology for data classification. The process is a structured pipeline consisting of data preprocessing, feature selection, training of a deep learning model, classification, and evaluation of results.

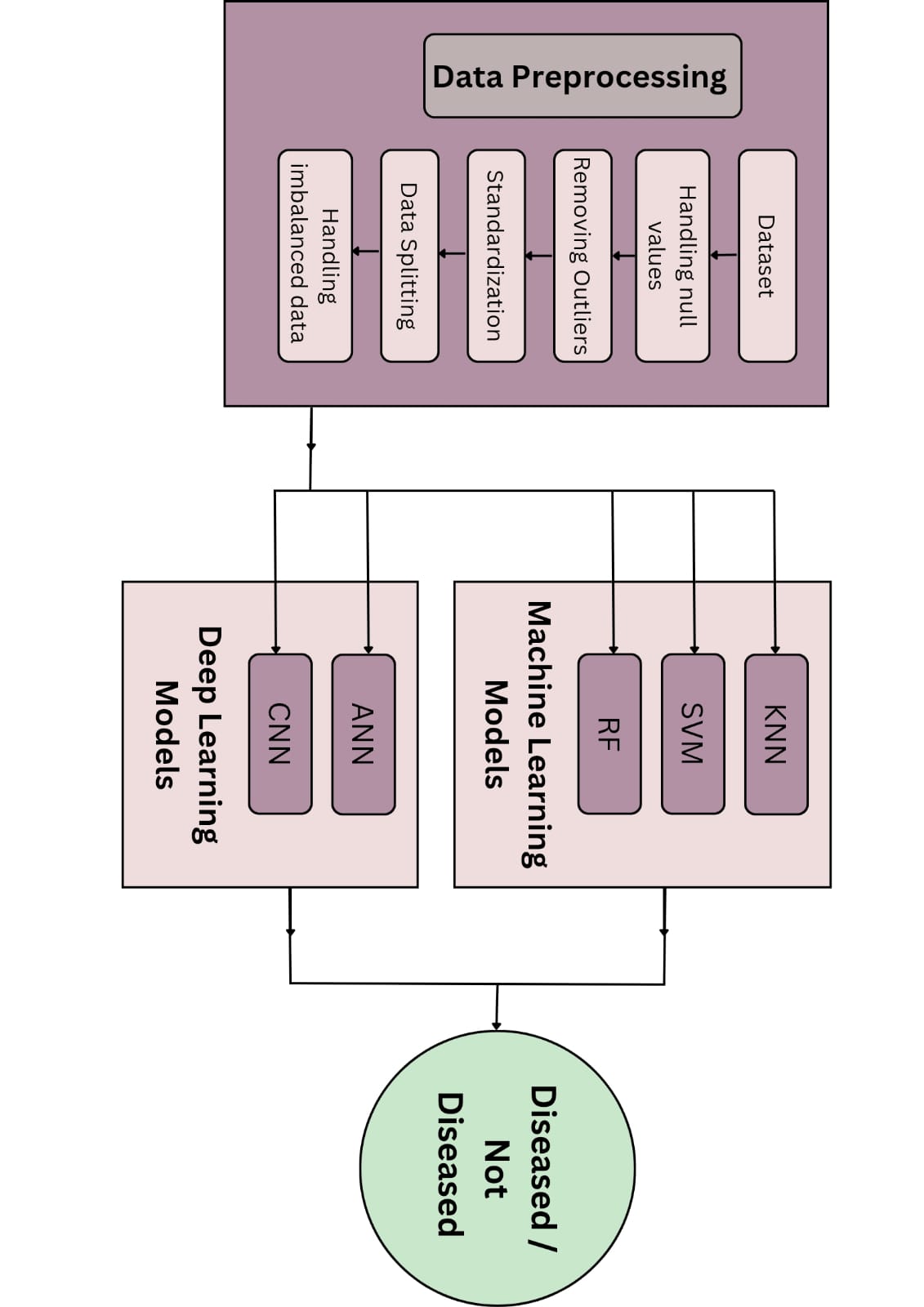


Fig.1 Methodology

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Training Accuracy** | **Testing Accuracy** | **Training Precision** | **Testing Precision** | **Training Recall** | **Testing Recall** | **Training F1 Score** | **Testing F1 Score** |
| KNN | 94% | 91% | 95% | 92% | 96% | 94% | 96% | 90% |
| SVM | 91% | 92% | 94% | 93% | 94% | 94% | 94% | 94% |
| RF | 100% | 92% | 100% | 93% | 100% | 95% | 100% | 94% |

**Table: 2 Performance Comparison of Machine Learning Models**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Training Accuracy** | **Testing Accuracy** | **Training Precision** | **Testing Precision** | **ROC Area** |
| ANN | 93% | 91% | 95% | 93% | 0.97 |
| ANN (L2 Regularization) | 91% | 91% | 95% | 93% | 0.97 |
| ANN (L1L2 Regularization) | 91% | 91% | 93% | 93% | 0.97 |
| ANN (Dropout 20%) | 92% | 92% | 93% | 92% | 0.97 |
| CNN (2 Conv Layers) | 92% | 91% | 95% | 92% | 0.97 |
| CNN (3 Conv Layers) | 92% | 91% | 96% | 93% | 0.98 |

**Table:3 Performance Comparison of Deep Learning Models**

## Model Architecture:

The Artificial Neural Network (ANN) model has been created with an aim to detect intricate patterns within the data. The architecture of the network is as follows:

* + - Input Layer: Handles 10 features that comprise patient demographics and clinical biomarkers.
    - Hidden Layers:
      * First hidden layer: 64 neurons and Rectified Linear Unit (ReLU) activation.
      * Second hidden layer: 32 neurons with ReLU activation.
    - Output Layer: Single neuron having sigmoid activation for the classification of patients into two categories: with and without liver disease.

**Table. 04 Artificial Neural Network Model Architecture**

|  |  |
| --- | --- |
| **Layer Type** | **Layer Details** |
| Dense (Layer 1) | Units: 10, Input Shape: (num\_features,), Activation: ReLU |
| Dense (Layer 2) | Units: 20, Activation: ReLU |
| Dense (Layer 3) | Units: 40, Activation: ReLU |
| Dense (Layer 4) | Units: 80, Activation: ReLU |
| Dense (Layer 5) | Units: 1, Activation: Sigmoid |

**Table. 05 Artificial Neural Network Model Architecture with L1-L2 Regularization**

|  |  |
| --- | --- |
| **Layer Type** | **Layer Details** |
| Dense (Layer 1) | Units: 10, Activation: ReLU, Input Shape: (num\_features,), Regularization: L1=0.01, L2=0.01 |
| Dense (Layer 2) | Units: 20, Activation: ReLU, Regularization: L1=0.01, L2=0.01 |
| Dense (Layer 3) | Units: 40, Activation: ReLU, Regularization: L1=0.01, L2=0.01 |
| Dense (Layer 4) | Units: 80, Activation: ReLU, Regularization: L1=0.01, L2=0.01 |
| Dense (Output Layer) | Units: 1, Activation: Sigmoid |

**Table. 06 Artificial Neural Network Model Architecture with Dropout**

|  |  |
| --- | --- |
| **Layer Type** | **Layer Details** |
| Dense (Layer 1) | Units: 10, Activation: ReLU, Input Shape: (num\_features,) |
| Dropout (Layer 1) | Rate: 0.2 |
| Dense (Layer 2) | Units: 20, Activation: ReLU |
| Dropout (Layer 2) | Rate: 0.2 |
| Dense (Layer 3) | Units: 40, Activation: ReLU |
| Dropout (Layer 3) | Rate: 0.2 |
| Dense (Layer 4) | Units: 80, Activation: ReLU |
| Dropout (Layer 4) | Rate: 0.2 |
| Dense (Output Layer) | Units: 1, Activation: Sigmoid |

**Table. 07 Convolutional Neural Network Model Architecture**

|  |  |
| --- | --- |
| **Layer Type** | **Layer Details** |
| Conv1D (Layer 1) | Filters: 32, Kernel Size: 3, Activation: ReLU, Input Shape: (X\_train\_cnn.shape[1], 1) |
| MaxPooling1D (Layer 1) | Pool Size: 2 |
| Conv1D (Layer 2) | Filters: 64, Kernel Size: 2, Activation: ReLU, Padding: 'same' |
| MaxPooling1D (Layer 2) | Pool Size: 2 |
| Flatten | - |
| Dense | Units: 1, Activation: Sigmoid |

## Model Training and Optimization:

The ANN model was trained using the following configurations:

* + **Loss Function:** Binary Cross-Entropy, appropriate for binary classification problems.
  + **Optimizer:** Adam optimizer, selected due to its adaptive learning rate feature.
  + **Batch Size:** 32, for balancing training efficiency and stability.
  + **Epochs:** 100, based on the convergence behavior of the model.
  + **Regularization Techniques:** Dropout layers were added to prevent overfitting by randomly disabling neurons during training.

## Model Evaluation:

The model that was trained was evaluated with several performance measures to verify its efficacy:

**• Accuracy**: The ratio of well-classified instances.

**• Precision**: The proportion of correct positive predictions to all predicted positives.

**• Recall**: The model's capacity to detect actual instances of liver disease.

**• F1-score**: The harmonic mean between precision and recall, balancing both scores.

**• Receiver Operating Characteristic (ROC) Curve and Area Under the Curve (AUC)**: Tested the model's discriminative power to distinguish between positive and negative examples. The model obtained an AUC value of 0.92, representing high classification performance.

## Testing and Validation:

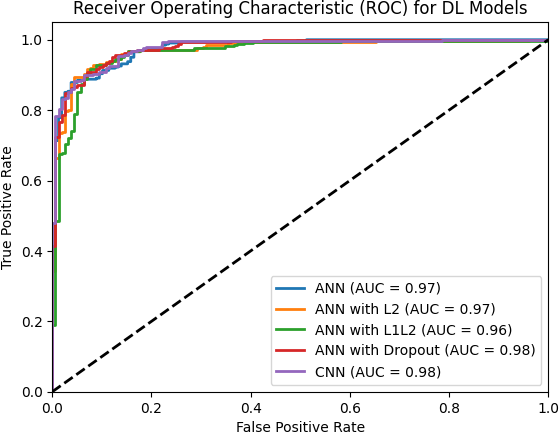
The trained model was also evaluated against the unseen 20% test set to see if it is generalizable. Trends in the validation loss and accuracy were kept track of while observing epochs for overfitting if any.

### Results and discussion

**Table:8 Evaluation of Deep Learning Models**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Training Accuracy** | **Testing Accuracy** | **Training Precision** | **Testing Precision** | **ROC**  **Area** |
| ANN | 93% | 91% | 95% | 93% | 0.97 |
| ANN (L2 Regularization) | 91% | 91% | 95% | 93% | 0.97 |
| ANN (L1L2 Regularization) | 91% | 91% | 93% | 93% | 0.97 |
| ANN (Dropout 20%) | 92% | 92% | 93% | 92% | 0.97 |
| CNN (2 Conv Layers) | 92% | 91% | 95% | 92% | 0.97 |
| CNN (3 Conv Layers) | 92% | 91% | 96% | 93% | 0.98 |

* + Training Accuracy refers to how well the model represents the training data. The majority of models were able to achieve high training accuracy, with the ANN models faring slightly better in this aspect.
  + Testing Accuracy indicates the generalizability of the model to new, unseen data. Both the ANN and CNN models performed about the same testing accuracy, with the majority of models ranging from 91% to 92% on the test set. This implies that the models are generalizable.
  + Training Precision and Testing Precision give an idea of the performance of the models when it comes to predicting the positive class (in classification). Precision values in all models are very high, between 93% and 96%, indicating that the models are excellent at getting the right predictions for the positive class.
  + ROC Area is a significant measure for assessing the capacity of the model to distinguish between classes. The closer the value to 1, the better the performance. The majority of models had an ROC area of approximately 0.97, with the CNN model with 3 Conv Layers performing slightly better than the others with an ROC area of 0.98.

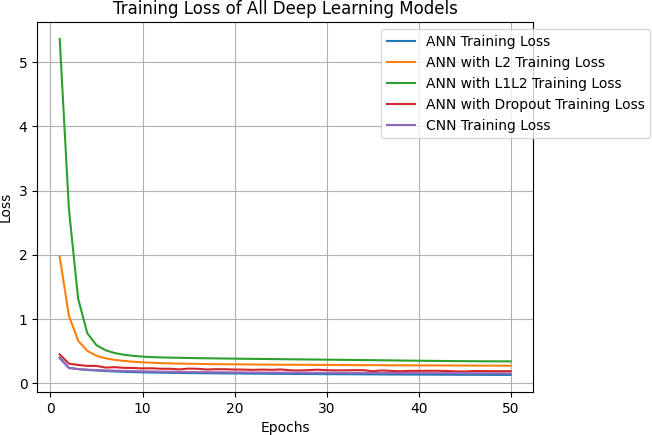


### Fig.2 Receiver Operating Characteristic (ROC) for DL Models

This is a Receiver Operating Characteristic (ROC) curve for different Deep Learning (DL) models employed in classification. The ROC curve plots the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) at different thresholds of classification.

* Key Elements of the Graph

1. X-Axis (False Positive Rate - FPR):
   * Measures the proportion of incorrectly classified negative samples as positive.
   * Lower values indicate better model performance.
2. Y-Axis (True Positive Rate - TPR):
   * Measures the proportion of correctly classified positive samples.
   * Higher values indicate better model performance.
3. Diagonal Line (Black Dashed Line):
   * Represents a random classifier with no discrimination capability (AUC = 0.5).
   * Any model above this line is performing better than random chance.
4. Colored ROC Curves:
   * Each curve represents a different DL model's performance.
   * Models closer to the top-left corner indicate better performance.

* Models and Their AUC Scores
* ANN (Artificial Neural Network) (AUC = 0.97): A standard ANN model.
* ANN with L2 Regularization (AUC = 0.97): Regularization helps prevent overfitting.
* ANN with L1L2 Regularization (AUC = 0.96): Uses both L1 and L2 norms to penalize weights.
* ANN with Dropout (AUC = 0.98): Dropout is used to prevent overfitting by randomly deactivating neurons.
* CNN (Convolutional Neural Network) (AUC = 0.98): A CNN model that slightly outperforms the standard ANN.
* Observations
* The CNN and ANN with Dropout achieved the highest AUC (0.98), indicating better classification performance.
* The standard ANN and ANN with L2 regularization performed slightly lower (AUC = 0.97).
* The ANN with L1L2 regularization has the lowest AUC (0.96), but still performs well.

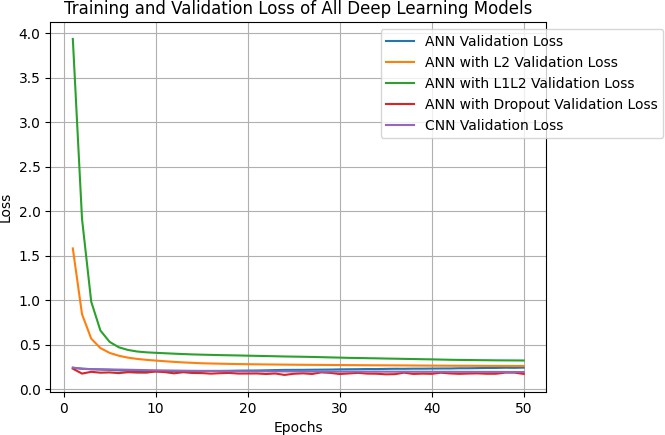
### Fig.3 Training Loss of All Deep Learning Models

This chart plots the training loss curves of various Deep Learning (DL) models across 50 epochs. Training loss tracks a model's ability to learn by comparing predicted and actual values.

* Key Elements of the Graph

1. X-Axis (Epochs):
   * Represents the number of training iterations.
   * As epochs increase, the model adjusts weights to reduce the error.
2. Y-Axis (Loss):
   * Represents the loss value (e.g., categorical cross-entropy or MSE).
   * A lower loss indicates better model performance.
3. Colored Loss Curves:
   * Each curve represents the training loss of a different DL model.
   * A steep drop at the beginning is common, as models learn patterns quickly in early epochs.

* Models and Their Loss Curves
* ANN Training Loss (Blue Curve): A standard ANN that converges smoothly.
* ANN with L2 Regularization (Orange Curve): Has a slightly higher initial loss but stabilizes.
* ANN with L1L2 Regularization (Green Curve): Starts with the highest loss (~5.5) but eventually converges.
* ANN with Dropout (Red Curve): Shows a stable loss curve.
* CNN Training Loss (Purple Curve): Has the lowest loss among all models, indicating better learning.
* Observations
* All models show a decreasing trend, which means they are learning properly.
* The ANN with L1L2 regularization started with the highest loss, but it eventually converges, showing that the model required more updates to optimize.
* CNN has the lowest loss, indicating that it learns patterns more effectively.
* Regularized models (L2, L1L2, Dropout) tend to have slightly higher loss values due to the penalty applied to weights, which helps prevent overfitting.



### Fig.4 Testing Loss of All Deep Learning Models

This graph represents the training and validation loss of multiple deep learning models over 50 epochs. Here’s what the graph indicates:

### Axes:

* X-axis (Epochs): Represents the number of training iterations (1 to 50).
* Y-axis (Loss): Measures the error in predictions, with lower values indicating better model performance.

### Curves in the Graph:

* Blue Line: Represents the validation loss of a standard Artificial Neural Network (ANN).
* Orange Line: Represents the validation loss of an ANN with L2 regularization (helps reduce overfitting by adding a penalty to large weights).
* Green Line: Represents the validation loss of an ANN with L1L2 regularization (combination of L1 and L2 penalties).
* Red Line: Represents the validation loss of an ANN with Dropout regularization (randomly drops neurons to prevent overfitting).
* Purple Line: Represents the validation loss of a Convolutional Neural Network (CNN).

### Observations:

1. Initial High Loss:
   * All models start with a high loss, but it quickly decreases within the first 10 epochs.
2. Stabilization:
   * After around 10–20 epochs, the loss values flatten, indicating that models are converging.
3. Performance Comparison:
   * The green curve (ANN with L1L2) starts with the highest loss but eventually stabilizes.
   * The red and purple curves (Dropout ANN and CNN) show the lowest loss, indicating that these models generalize better.
   * The CNN (purple) seems to perform better than all ANN models in terms of maintaining a low loss.

* The proposed model highlights the feasibility of using deep learning for liver disease detection. Key strengths include:
* Effective handling of class imbalance through synthetic data generation.
* Robust performance metrics, with high precision and recall.

### Conclusion

This study proves the great potential of Artificial Neural Networks (ANNs) in the early detection of liver diseases as a non-invasive and effective alternative to conventional diagnostic techniques. The 92% accurate model reveals great promise for the early-stage diagnosis that is essential to enhance patient outcome. The dropout-regularized ANN used in this study proved high in accuracy and generalizability, promising its potential to be integrated into health care systems to assist doctors in decision-making.

Future developments may involve extending the model to multi-class classification for distinguishing between different liver conditions, incorporating real-time data via IoT devices for ongoing monitoring, and improving the model with larger, more diverse datasets to enhance its usability across heterogeneous populations. Furthermore, implementing the model in clinical environments for field testing will be necessary to confirm its efficacy in actual use.

With additional research and refinement, deep learning models such as the one described in this study could greatly improve early diagnosis and transform healthcare practices, ultimately leading to better patient care and outcomes.

### REFERENCES

1. *Arias, I.M.; Alter, H.J.; Boyer, J.L.; Cohen, D.E.; Shafritz, D.A.; Thorgeirsson, S.S.; Wolkoff, A.W. The Liver: Biology and Pathobiology; John Wiley & Sons: Hoboken, NJ, USA, 2020.*
2. *Singh, H.R.; Rabi, S. Study of morphological variations of liver in human. Transl. Res. Anat. 2019, 14, 1–5.*
3. *Razavi, H. Global epidemiology of viral hepatitis. Gastroenterol. Clin. 2020, 49, 179–189.*
4. *Ginès, P.; Krag, A.; Abraldes, J.G.; Solà, E.; Fabrellas, N.; Kamath, P.S. Liver cirrhosis. Lancet 2021, 398, 1359–1376.*
5. *Ringehan, M.; McKeating, J.A.; Protzer, U. Viral hepatitis and liver cancer. Philos. Trans. R. Soc. B Biol. Sci. 2017, 372, 20160274.*
6. *Powell, E.E.; Wong, V.W.S.; Rinella, M. Non-alcoholic fatty liver disease. Lancet 2021, 397, 2212–2224.*
7. *Smith, A.; Baumgartner, K.; Bositis, C. Cirrhosis: Diagnosis and management. Am. Fam. Physician 2019, 100, 759–770.*
8. *Rycroft, J.A.; Mullender, C.M.; Hopkins, M.; Cutino-Moguel, T. Improving the accuracy of clinical interpretation of serological testing for the diagnosis of acute hepatitis A infection. J. Clin. Virol. 2022, 155, 105239.*
9. *Thomas, D.L. Global elimination of chronic hepatitis. N. Engl. J. Med. 2019, 380, 2041–2050.*
10. *Rasche, A.; Sander, A.L.; Corman, V.M.; Drexler, J.F. Evolutionary biology of human hepatitis viruses. J. Hepatol. 2019, 70, 501–520.*
11. *Gust, I.D. Hepatitis A; CRC Press: Boca Raton, FL, USA, 2018.*
12. *Yuen, M.F.; Chen, D.S.; Dusheiko, G.M.; Janssen, H.L.; Lau, D.T.; Locarnini, S.A.; Peters, M.G.; Lai, C.L. Hepatitis B virus infection. Nat. Rev. Dis. Prim. 2018, 4, 1–20.*
13. *Manns, M.P.; Buti, M.; Gane, E.; Pawlotsky, J.M.; Razavi, H.; Terrault, N.; Younossi, Z. Hepatitis C virus infection. Nat. Rev. Dis. Prim. 2017, 3, 1–19.*
14. *Mentha, N.; Clément, S.; Negro, F.; Alfaiate, D. A review on hepatitis D: From virology to new therapies. J. Adv. Res. 2019, 17, 3–15.*
15. *Kamar, N.; Izopet, J.; Pavio, N.; Aggarwal, R.; Labrique, A.; Wedemeyer, H.; Dalton, H.R. Hepatitis E virus infection. Nat. Rev. Dis. Prim. 2017, 3, 1–16.*
16. *Marchesini, G.; Moscatiello, S.; Di Domizio, S.; Forlani, G. Obesity-associated liver disease. J. Clin. Endocrinol. Metab. 2008, 93, s74–s80.*
17. *Seitz, H.K.; Bataller, R.; Cortez-Pinto, H.; Gao, B.; Gual, A.; Lackner, C.; Mathurin, P.; Mueller, S.; Szabo, G.; Tsukamoto, H. Alcoholic liver disease. Nat. Rev. Dis. Prim. 2018, 4, 1–22.*
18. *Åberg, F.; Färkkilä, M. Drinking and obesity: Alcoholic liver disease/nonalcoholic fatty liver disease interactions. In Seminars in Liver Disease; Thieme Medical Publishers: New York, NY, USA, 2020; Volume 40, pp. 154–162.*
19. *Bae, M.; Park, Y.K.; Lee, J.Y. Food components with antifibrotic activity and implications in prevention of liver disease. J. Nutr. Biochem. 2018, 55, 1–11.*
20. *Cai, J.; Zhang, X.J.; Li, H. Progress and challenges in the prevention and control of nonalcoholic fatty liver disease. Med. Res. Rev. 2019, 39, 328–348.*
21. *Fazakis, N.; Kocsis, O.; Dritsas, E.; Alexiou, S.; Fakotakis, N.; Moustakas, K. Machine learning tools for long-term type 2 diabetes risk prediction. IEEE Access 2021, 9, 103737–103757.*
22. *Dritsas, E.; Trigka, M. Data-Driven Machine-Learning Methods for Diabetes Risk Prediction. Sensors 2022, 22, 5304.*
23. *Alexiou, S.; Dritsas, E.; Kocsis, O.; Moustakas, K.; Fakotakis, N. An approach for Personalized Continuous Glucose Prediction with Regression Trees. In Proceedings of the 2021 6th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Preveza, Greece, 24–26 September 2021; pp. 1–6.*
24. *Dritsas, E.; Alexiou, S.; Konstantoulas, I.; Moustakas, K. Short-term Glucose Prediction based on Oral Glucose Tolerance Test Values. In Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies—HEALTHINF, Lisbon, Portugal, 9–11 February 2022; Volume 5, pp. 249–255.*
25. *Fazakis, N.; Dritsas, E.; Kocsis, O.; Fakotakis, N.; Moustakas, K. Long-Term Cholesterol Risk Prediction with Machine Learning Techniques in ELSA Database. In Proceedings of the 13th International Joint Conference on Computational Intelligence (IJCCI), Online, 24–26 October 2021; pp. 445–450.*
26. *Dritsas, E.; Fazakis, N.; Kocsis, O.; Fakotakis, N.; Moustakas, K. Long-Term Hypertension Risk Prediction with ML Techniques in ELSA Database. In Proceedings of the International Conference on Learning and Intelligent Optimization, Athens, Greece, 20–25 June 2021; pp. 113–120.*
27. *Dritsas, E.; Alexiou, S.; Moustakas, K. Efficient Data-driven Machine Learning Models for Hypertension Risk Prediction. In Proceedings of the 2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Biarritz, France, 8–10 August 2022; pp. 1–6.*
28. *Dritsas, E.; Trigka, M. Machine Learning Methods for Hypercholesterolemia Long-Term Risk Prediction. Sensors 2022, 22, 5365.*
29. *Dritsas, E.; Alexiou, S.; Moustakas, K. COPD Severity Prediction in Elderly with ML Techniques. In Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, Corfu Island, Greece, 29 June–1 July 2022; pp. 185–189.*
30. *Dritsas, E.; Trigka, M. Supervised Machine Learning Models to Identify Early-Stage Symptoms of SARS-CoV-2. Sensors 2023, 23, 40.*
31. *Dritsas, E.; Trigka, M. Stroke Risk Prediction with Machine Learning Techniques. Sensors 2022, 22, 4670.*
32. *Dritsas, E.; Trigka, M. Machine Learning Techniques for Chronic Kidney Disease Risk Prediction. Big Data Cogn. Comput. 2022, 6, 98.*
33. *Dritsas, E.; Trigka, M. Lung Cancer Risk Prediction with Machine Learning Models. Big Data Cogn. Comput. 2022, 6, 139.*
34. *Konstantoulas, I.; Kocsis, O.; Dritsas, E.; Fakotakis, N.; Moustakas, K. Sleep Quality Monitoring with Human Assisted Corrections. In Proceedings of the International Joint Conference on Computational Intelligence (IJCCI), Online, 24–26 October 2021; pp. 435–444.*
35. *Konstantoulas, I.; Dritsas, E.; Moustakas, K. Sleep Quality Evaluation in Rich Information Data. In Proceedings of the 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA), Corfu, Greece, 18–20 July 2022; pp. 1–4.*
36. *Dritsas, E.; Alexiou, S.; Moustakas, K. Cardiovascular Disease Risk Prediction with Supervised Machine Learning Techniques. In Proceedings of the ICT4AWE, Online, 23–25 April 2022; pp. 315–321.*
37. *Indian Liver Patient Records. Available online: https://www.kaggle.com/datasets/uciml/indian-liver-patient-records.*
38. *Mauvais-Jarvis, F.; Merz, N.B.; Barnes, P.J.; Brinton, R.D.; Carrero, J.J.; DeMeo, D.L.; De Vries, G.J.; Epperson, C.N.; Govindan, R.; Klein, S.L.; et al. Sex and gender: Modifiers of health, disease, and medicine. Lancet 2020, 396, 565–582.*
39. *Lin, H.; Yip, T.C.F.; Zhang, X.; Li, G.; Tse, Y.K.; Hui, V.W.K.; Liang, L.Y.; Lai, J.C.T.; Chan, S.L.; Chan, H.L.Y.; et al. Age and the relative importance of liver-related deaths in nonalcoholic fatty liver disease. Hepatology 2022.*
40. *Ruiz, A.R.G.; Crespo, J.; Martínez, R.M.L.; Iruzubieta, P.; Mercadal, G.C.; Garcés, M.L.; Lavin, B.; Ruiz, M.M. Measurement and clinical usefulness of bilirubin in liver disease. Adv. Lab. Med. Med. Lab. 2021, 2, 352–361.*
41. *Liu, Y.; Cavallaro, P.M.; Kim, B.M.; Liu, T.; Wang, H.; Kühn, F.; Adiliaghdam, F.; Liu, E.; Vasan, R.; Samarbafzadeh, E.; et al. A role for intestinal alkaline phosphatase in preventing liver fibrosis. Theranostics 2021, 11, 14.*
42. *Goodarzi, R.; Sabzian, K.; Shishehbor, F.; Mansoori, A. Does turmeric/curcumin supplementation improve serum alanine aminotransferase and aspartate aminotransferase levels in patients with nonalcoholic fatty liver disease? A systematic review and meta-analysis of randomized controlled trials. Phytother. Res. 2019, 33, 561–570.*
43. *He, B.; Shi, J.; Wang, X.; Jiang, H.; Zhu, H.J. Genome-wide pQTL analysis of protein expression regulatory networks in the human liver. BMC Biol. 2020, 18, 1–16.*
44. *Carvalho, J.R.; Machado, M.V. New insights about albumin and liver disease. Ann. Hepatol. 2018, 17, 547–560.*
45. *Ye, Y.; Chen, W.; Gu, M.; Xian, G.; Pan, B.; Zheng, L.; Zhang, Z.; Sheng, P. Serum globulin and albumin to globulin ratio as potential diagnostic biomarkers for periprosthetic joint infection: A retrospective review. J. Orthop. Surg. Res. 2020, 15, 1–7.*