Detecting Non-Alcoholic Fatty Liver Disease (NAFLD) using Clinical Reports

Lakshmi Warrier¹, M Devika¹, Perumalla Raghavendra¹, and Dr. Remya S²

¹amenu4aie20143@am.students.amrita.edu, ²amenu4aie20144@am.students.amrita.edu

¹,²Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, India.

Abstract

Non-Alcoholic Fatty Liver Disease (NAFLD) is a prevalent liver condition that necessitates accurate and non-invasive diagnostic approaches for effective treatment. This research addresses the challenges associated with present invasive procedures, such as liver biopsies, and proposes a novel diagnostic tool. Inspired by the limitations of existing methods, our project focuses on revolutionizing routine check-ups for middle-aged individuals at risk of NAFLD. Instead of traditional invasive biopsies, our diagnostic tool recommends a blood test, ensuring accurate identification and timely intervention. The conventional diagnostic methods for NAFLD involve imaging and invasive procedures, leading to accessibility and accuracy issues. In response, our user-friendly web application utilizes standard blood test findings to provide a quick and painless NAFLD diagnosis. This approach aims to create an affordable, easily accessible tool that minimizes patient discomfort. Leveraging a dataset of 3,237 individuals from NHANES III, our model achieves an outstanding accuracy rate of 89%. The dataset includes both NAFLD-positive and NAFLD-negative cases, ensuring a robust and representative model. In summary, this work makes significant strides in developing a blood-based, non-invasive method that enhances accessibility to NAFLD diagnostics through a user-friendly web application. The proposed tool offers a convenient option for patients and equips healthcare providers with an effective NAFLD diagnostic tool, fostering better patient care outcomes through early detection and intervention.

Keywords: NAFLD, Non-Invasive technique, Hepatic steatosis, Ultrasound imaging, Deep Learning, Biomarkers

1 Introduction

Non-alcoholic fatty Liver Disease (NAFLD) encompasses a spectrum of liver conditions characterized by fat accumulation, ranging from benign fatty liver to severe non-alcoholic steato-hepatitis (NASH). The prevalence and characteristics of NAFLD vary globally, influenced by factors such as genetics, insulin resistance, and lifestyle. In India, NAFLD affects a significant portion of the population, with around 38% of adults and 52% of high-risk groups impacted, underscoring the need for increased healthcare awareness and integration [1, 2].

Machine learning has emerged as a valuable tool in predicting and understanding diseases like Type 2 Diabetes, which often coexists with NAFLD. Studies utilizing Goldberg’s Genetic
Algorithm and Multi-Objective Evolutionary Fuzzy Classifier have shown promise in this regard. NAFLD, prevalent in Western societies, presents substantial cardiovascular risks and is linked with various metabolic comorbidities. Notably, the incidence and prevalence of NAFLD are higher among men than women [3, 4].

Detection and diagnosis of liver steatosis can be accomplished from invasive liver biopsy to non-invasive imaging modalities. Radiologic approaches, including sophisticated computer-based techniques for analyzing liver diseases from ultrasound images, are continuously advancing, emphasizing the importance of ultrasound-based detection methods [5, 6, 7].

Management of NAFLD often involves lifestyle modifications and pharmacological interventions. Artificial Intelligence (AI) is revolutionizing healthcare delivery, extending its impact to liver transplantation phases. AI models like Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forests are being explored. Despite promising advancements, challenges such as ethical concerns, data standardization, and financial costs persist [8, 9, 10].

Moreover, AI and machine learning play a crucial role in the early and accurate diagnosis of liver cancer, another global health concern. Advanced techniques like Deep Belief Networks (DBN) and transfer learning have significantly improved diagnostic accuracy. These algorithms can predict disease progression, complications, and mortality related to hepatic disorders. Integration of AI with conventional diagnostic methods enhances overall diagnostic performance, offering insights into conditions like NAFLD, hepatocellular carcinoma, and liver cirrhosis [11, 12, 13].

2 Related Works

Liver fibrosis assessment traditionally relied on invasive biopsy methods, presenting challenges such as cost, sampling errors, and complications. Various studies have explored non-invasive approaches aiming to overcome these limitations. For instance, Tsipakidou et al. [14] proposed a methodology for liver steatosis detection from biopsy images in 2016, but faced issues with false-positive detections. Similarly, Byra et al. [15] and U Rajendra Acharya et al. [16] investigated novel approaches for liver disease detection using advanced imaging techniques, such as transfer learning and 2D contourlet transform with texture features.

In parallel, ML models have gained traction for predicting liver disease outcomes. Studies by Wu et al. [17], Atabaki et al. [18], and Chen et al. [19] have explored the development and comparison of ML models for early prediction of liver disease, demonstrating promising results with high AU-ROC values.

Systematic reviews, such as the one conducted by Miller et al. [20], have highlighted the potential of non-invasive markers for assessing disease severity in NAFLD. In 2021, Lanthier et al. [21] provided insights into the progression of metabolic-associated NAFLD. The abnormal expression of peroxisome proliferator-activated receptor (PPAR) isotypes in NAFLD underscores the molecular mechanisms that contribute to insulin resistance and hepatic steatosis [22]. Leveraging this understanding, biomarker-based approaches enable earlier diagnosis, personalized treatment strategies, and improved prognostic outcomes for patients with the disease.


Furthermore, the article review by YN Zhang, KJ Fowler et al. [25] explores various imaging methods for assessing liver steatosis, ranging from traditional ultrasound to advanced techniques like MRI PDFF. Additionally, advancements in imaging techniques, such as radiologic
approaches [6], have played a significant role in liver disease diagnosis. Computer-based techni-ques for analyzing liver diseases from ultrasound images have shown promise[7].

Lifestyle modifications, pharmacological interventions, and AI-driven approaches are being explored for NAFLD prevention and management. AI and ML have demonstrated remarkable efficacy in reducing the time required for microbial detection using biosensors [11]. Additionally, AI integration with conventional diagnostic methods enhances overall diagnostic performance, offering insights into conditions like NAFLD, hepatocellular carcinoma, and liver cirrhosis [13].

Collectively, these studies highlight the evolving landscape of computer-aided non-invasive methods for diagnosing and assessing liver conditions. However, there is a need for broader validation and refinement of diagnostic accuracy. Collaborative efforts across healthcare institutions are crucial for advancing our understanding and management of liver diseases in diverse patient populations.

3 Proposed Methodology

This study employs a Convolutional Neural Network (CNN) architecture, primarily due to its ability to extract and learn complex patterns from input data. CNNs are known for their superior performance in tasks involving structured data, making them well-suited for the analysis of clinical records. The architecture of the study is as shown in Figure 1.

![Figure 1: System Architecture](image)

3.1 Dataset

The dataset utilized in this study comprises 31 clinical characteristics sourced from NHANES III (CDC/NCHS), offering insights into parameters like age, gender, BMI, waist circumference, race, and ultrasound results indicating NAFLD presence or absence as illustrated in 1. With a total of 3,235 individuals, the dataset provides a comprehensive overview of health-related traits, including 2,418 negative and 817 positive NAFLD cases. Ultrasound findings play a
pivotal role as definitive indicators for NAFLD detection, contributing crucial information to the study’s analysis and conclusions.

Table 1: Sample rows of the important parameters of the dataset

<table>
<thead>
<tr>
<th>us</th>
<th>gender</th>
<th>age</th>
<th>bmi</th>
<th>waist</th>
<th>ghp</th>
<th>c1p</th>
<th>fglu</th>
<th>ins</th>
<th>trig</th>
<th>alt</th>
<th>ggt</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>28</td>
<td>27.9</td>
<td>97.4</td>
<td>4.9</td>
<td>0.345</td>
<td>94.6</td>
<td>8.52</td>
<td>96</td>
<td>38</td>
<td>20</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>21</td>
<td>17.9</td>
<td>64.2</td>
<td>9.9</td>
<td>0.021</td>
<td>142.6</td>
<td>17.41</td>
<td>76</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>25</td>
<td>18.9</td>
<td>71.5</td>
<td>4.6</td>
<td>0.063</td>
<td>82.8</td>
<td>4.59</td>
<td>56</td>
<td>8</td>
<td>17</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>41</td>
<td>23.9</td>
<td>77.2</td>
<td>4.4</td>
<td>0.409</td>
<td>85.3</td>
<td>6.03</td>
<td>71</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>55</td>
<td>25.9</td>
<td>97</td>
<td>5.3</td>
<td>0.734</td>
<td>90.1</td>
<td>13.75</td>
<td>162</td>
<td>19</td>
<td>28</td>
</tr>
</tbody>
</table>

Units: BMI (kg/m²), waist (cm), ghp (mg/dL), fglu (mg/dL), ins (mU/L), trig (mg/dL), alt (U/L), ggt (U/L)

3.2 Feature Selection

Our dataset comprises a range of biomarkers indicative of liver steatosis, including demographic factors, anthropometric measures, and blood profile markers. Acknowledging potential variations in disease prevalence among different ethnic groups, we included racial categories in our analysis. Our research focuses on eight key features considered essential for predicting NAFLD. To identify the most significant features for NAFLD prediction, we utilized a comprehensive approach involving correlation matrix analysis, Random Forest Classifier (RFC), and Principal Component Analysis (PCA). As shown in Table 2, both PCA and Random Forest Classifier assign varying importance to different features in predicting NAFLD. While PCA highlights waist circumference as the most significant feature, closely followed by variables such as glycohaemoglobin (GHP), BMI, and Ceramide-1-phosphate (C1P), the Random Forest Classifier places greater emphasis on waist circumference, GHP, and C1P. Notably, features like insulin (INS), triglycerides (Trig), and age are significant in PCA but less influential in the Random Forest Classifier. These findings underscore the nuanced considerations in developing predictive models for NAFLD. The prioritization of features such as waist circumference, GHP, BMI, C1P, Fasting Glucose (FGLU), INS, Trig, and Alanine Aminotransferase (ALT) is supported by their consistently high importance across PCA and Random Forest Classifier analyses, indicating their robust predictive capability for NAFLD.

Table 2: Feature importance according to PCA and Random Forest Classifier

<table>
<thead>
<tr>
<th></th>
<th>waist</th>
<th>ghp</th>
<th>bmi</th>
<th>c1p</th>
<th>fglu</th>
<th>ins</th>
<th>trig</th>
<th>appsi</th>
<th>uap</th>
<th>gbp</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>0.332</td>
<td>0.295</td>
<td>0.292</td>
<td>0.264</td>
<td>0.259</td>
<td>0.245</td>
<td>0.211</td>
<td>0.200</td>
<td></td>
<td>0.192</td>
<td>0.172</td>
</tr>
<tr>
<td>RFC</td>
<td>0.08</td>
<td>0.03</td>
<td>0.054</td>
<td>0.074</td>
<td>0.066</td>
<td>0.055</td>
<td>0.044</td>
<td>0.043</td>
<td>0.042</td>
<td>0.037</td>
<td>0.037</td>
</tr>
</tbody>
</table>

3.3 Data Preprocessing

To ensure data completeness and handle missing values, a combination of imputation and removal techniques was employed, with standard scaling and normalisation to enhance the model’s ability to discern meaningful patterns within the data.
3.4 Architecture

The CNN architecture comprises three convolutional layers, each with 32 filters activated by a Rectified Linear Unit (ReLU), followed by a max-pooling layer. The flattened vector then enters dense layers with 64, 32, and 8 neurons, utilizing ReLU activation. The final sigmoid activation function classifies to align with the binary results of ultrasound scans in the dataset.

Algorithm 1 Convolutional Neural Network (CNN)

1: **Input:** Training data: \texttt{train.data}, Training labels: \texttt{train_l}, Validation data: \texttt{val.data}, Validation labels: \texttt{val_l}, Number of convolutional layers: \( L \), Number of filters in each layer: \( F \), Filter sizes: \( K \), Pooling size: \( P \), Number of fully connected layers: \( FC \), Learning rate: \( \alpha \), Number of epochs: \( N \), Batch size: \( B \)
2: **Output:** Trained CNN model
3: Initialize CNN model as \texttt{model}
4: Add \( L \) convolutional layer, ReLU activation, and max-pooling layer
5: Flatten the output of the last convolutional layer
6: for \( j \) in range \( FC \) do
7: Add a fully connected layer and ReLU activation
8: end for
9: Add output layer with sigmoid activation for binary classification
10: Compile \texttt{model} with binary cross-entropy loss and Adam optimizer
11: Train \texttt{model} on \texttt{train.data} and \texttt{train_l} for \( N \) epochs with batch size \( B \)
12: Evaluate \texttt{model} on \texttt{val.data} and \texttt{val_l}

For training and optimization, binary cross-entropy was used as the loss function and the Adam optimizer. To prevent overfitting, five-fold stratified cross-validation was done with the model trained across 50 epochs and a batch size of 32. Stratified cross-validation was important because of the evident imbalance in the dataset.

The thorough evaluation included various metrics such as accuracy, precision, recall, F1-score, and AUC-ROC on both training and validation sets. The architecture, coupled with cross-validation, ensures a balance between complexity and generalizability, mitigating overfitting risks while allowing effective pattern capture.

4 Experimental Results

In predicting NAFLD, a CNN model was utilized. While the model achieved high accuracy, there was a recognized need for improvement in specificity. Enhancements to the model architecture, including the exploration of pre-trained deep learning models, training on new datasets, and employing K-fold Cross-Validation, were implemented.

Additionally, the analysis included the estimation of scores such as the Fatty Liver Index score and Child-Pugh score, offering detailed insights into disease severity. These experiments contribute to a comprehensive and user-focused strategy for managing and predicting NAFLD.

4.1 Analysis of CNN and ANN Model

After thorough analysis, the CNN model outperformed the ANN model across key metrics such as precision, recall, F1-Score, and AUC-ROC (Table 3). Notably, the CNN model exhibited superior validation and training accuracy, emphasizing its enhanced predictive capabilities.
### Table 3: Analysis of ANN and CNN Models

<table>
<thead>
<tr>
<th>Models</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-Score</th>
<th>AUC-ROC</th>
<th>Validation Accuracy</th>
<th>Training Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN Model</td>
<td>0.706</td>
<td>0.529</td>
<td>0.601</td>
<td>0.724</td>
<td>0.822</td>
<td>0.853</td>
</tr>
<tr>
<td>CNN Model</td>
<td>0.796</td>
<td>0.737</td>
<td>0.764</td>
<td>0.839</td>
<td>0.890</td>
<td>0.959</td>
</tr>
</tbody>
</table>

#### 4.2 Analysis of CNN Model

The proposed CNN model depicted through the ROC curve (Figure 2a), demonstrates commendable sensitivity with a minimal bias. While the confusion matrix (Figure 2b) reveals a false positive rate of 110, it’s essential to note that the model maintains a strong precision in identifying NAFLD cases, supported by a considerable true positive count. These results suggest promising implications for the CNN model’s generalizability across diverse populations.

![ROC curve](image1.png)  
![Confusion matrix](image2.png)  

Figure 2: Result analysis of CNN model

Compared to existing scoring systems like Child-Pugh, F1, and NAFLD-LFS, our CNN model exhibits superior disease detection. With an average AUC-ROC of 0.839, our model surpasses current web-based NAFLD prediction tools in accuracy. A model comparison (Table 4) with various ML models highlights competitive performance. Recognizing scenarios where other models excel provides valuable insights for refining NAFLD prediction models, ensuring continuous improvement for enhanced clinical utility and patient outcomes.

#### 5 Conclusion and Future Scope

In conclusion, our focus on enhancing the CNN model for NAFLD detection has yielded significant progress. Through meticulous refinement of the architecture and exploration of advanced
Table 4: Summary of Evaluation Metrics for Machine Learning Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Evaluation Metrics</th>
<th>Performance Highlights</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression with 10-fold cross validation [26]</td>
<td>Accuracy = 76.3%</td>
<td>Susceptible to variations in data quality.</td>
<td>Specificity around 64% and reliance on limited EMR data.</td>
</tr>
<tr>
<td>LASSO and stochastic gradient boosting [18]</td>
<td>Cross-validated ROCAUC = 0.84</td>
<td>Outperformed existing tools using 18 models.</td>
<td>Focused on European adults and significantly reduced sample size</td>
</tr>
<tr>
<td>ANN model using 8 features from medical checkups [19]</td>
<td>AUROC = 0.908</td>
<td>Outperformed established FLD indices in terms of accuracy.</td>
<td>Relies on ultrasound and accuracy decreases across diverse populations.</td>
</tr>
<tr>
<td>Our ANN model: 5-fold CV</td>
<td>Validation accuracy = 85.3%</td>
<td>Uses easily available blood results</td>
<td>High false negatives</td>
</tr>
<tr>
<td>Our CNN model: 5-fold CV</td>
<td>Validation accuracy = 88%</td>
<td>Uses easily available blood results</td>
<td>High false negatives</td>
</tr>
</tbody>
</table>

ML techniques, we have enhanced overall prediction accuracy. Our innovative approach, utilizing blood test results, offers a non-invasive, easily accessible, and cost-effective diagnostic method with potential for widespread adoption.

Moving forward, our scope involves further refining the CNN model’s architecture, exploring pre-trained models, and incorporating diverse datasets to address specificity issues. To improve performance, we plan to implement techniques such as data augmentation, class weighting, and ensemble learning to mitigate the impact of imbalanced data. Additionally, interpretability measures will enhance trust in the model’s predictions.

Extending our model to include ultrasound scans aims to improve accessibility and accuracy in liver health assessments. Collaboration with healthcare professionals and institutions remains crucial for validating and implementing these enhancements. Our trajectory seeks to revolutionize early NAFLD detection and management, ultimately improving patient outcomes and healthcare efficacy.

References


Detecting NAFLD using Clinical Reports

L. Warrier et al.


