SSR Journal of Artificial Intelligence (SSRJAI)

Homepage: https://ssrpublisher.com/ssrjai/

Volume 3, Issue 8, 2026 ISSN: 3049-0413

Enhancing Nonalcoholic Fatty Liver Disease (NAFLD) Detection: AI-Enhanced Ultrasound Validation and Performance Improvement

Linus Mathias[1], Giliki Jerison[2] and Ishaku Kpapisom[3]

[1,2]Department of Computer Science, College of Education, Hong, Adamawa State, Nigeria

[3]Adamawa State College of Nursing Sciences, Yola. Department of General Studies

Received: *05.07.2026* | ***Accepted:*** *07.08.2026* | ***Published:*** *25.08.2026*

***Corresponding author:** Linus Mathias

DOI: *10.5281/zenodo.22091730*

Abstract **Original Research**

Nonalcoholic Fatty Liver Disease (NAFLD) is one of the most common chronic liver diseases worldwide and is strongly associated with obesity, type 2 diabetes mellitus, metabolic syndrome, and other cardiovascular risk factors. If left undiagnosed, NAFLD may progress to nonalcoholic steatohepatitis (NASH), liver fibrosis, cirrhosis, and hepatocellular carcinoma, making early detection essential for effective clinical management. Although magnetic resonance-based fat quantification is regarded as the noninvasive reference standard for assessing hepatic steatosis, its high cost, limited accessibility, and lengthy examination time restrict its widespread use, particularly in low-resource healthcare settings. Ultrasound imaging provides an affordable, safe, and widely available alternative; however, its diagnostic accuracy is often affected by operator dependency and subjective interpretation. This study investigates the feasibility of integrating artificial intelligence (AI) with ultrasound imaging to improve the accuracy and reliability of NAFLD detection. The proposed framework employs advanced deep learning algorithms, including convolutional neural networks (CNNs) and transfer learning, to automatically extract and classify discriminative imaging features associated with hepatic fat accumulation. To enhance model robustness and generalization, the training dataset was expanded to include a larger and more diverse patient population. Furthermore, multimodal data comprising ultrasound images, demographic information, clinical history, and laboratory findings were integrated to improve diagnostic performance and support more informed clinical decision-making. Model performance was assessed using accuracy, precision, recall, F1-score, receiver operating characteristic–area under the curve (ROC-AUC), k-fold cross-validation, and external validation to ensure reliability and clinical applicability. The findings indicate that AI-enhanced ultrasound can significantly improve diagnostic consistency, reduce observer variability, and provide a cost-effective, noninvasive solution for routine NAFLD screening and early detection. The proposed approach demonstrates strong potential for supporting clinical decision-making and improving patient outcomes, particularly in resource-constrained healthcare environments where access to advanced imaging modalities remains limited.

Keywords: Nonalcoholic Fatty Liver Disease (NAFLD), Artificial Intelligence, Ultrasound Imaging, Deep Learning, Convolutional Neural Networks, Transfer Learning, Multimodal Data, Early Detection.

INTRODUCTION

Background of the study

Fatty liver is a condition characterized by the accumulation of excess fat in the liver cells. This condition is divided into two main types: nonalcoholic fatty liver disease (NAFLD) and alcoholic fatty liver disease (Media, 2018).

Nonalcoholic fatty liver disease (NAFLD) is not associated with heavy alcohol consumption and is the most prevalent type, affecting approximately 25% of the global population. NAFLD can begin as a simple fatty liver but can progress to a more severe form known as nonalcoholic steatohepatitis (NASH). NASH involves liver inflammation and cellular damage, which can further lead to fibrosis (the formation of scar tissue), cirrhosis (severe scarring of the liver), and even liver cancer. This progression highlights the potential seriousness of NAFLD if left untreated (Pafili, & Roden, 2021).

In contrast, alcoholic fatty liver disease results from prolonged heavy alcohol use. It represents the initial stage of alcohol-related liver disease, which can advance to alcoholic hepatitis—a condition marked by liver inflammation—and ultimately to cirrhosis. The development of alcoholic fatty liver disease is directly linked to the toxic effects of alcohol on the liver over time (Sharma, & Arora, 2020).

Often, fatty liver disease does not present noticeable symptoms. However, when symptoms do occur, they may include fatigue, abdominal discomfort, and abnormalities in liver function tests. Various factors increase the risk of developing fatty liver, including obesity, type 2 diabetes, high cholesterol, elevated triglycerides, and the use of certain medications (Media, 2018).

Diagnosis of fatty liver disease typically involves a combination of medical history review, physical examination, blood tests, imaging studies like ultrasound, and sometimes a liver biopsy to assess the extent of liver damage. Treatment primarily focuses on lifestyle modifications such as weight loss, regular physical activity, and, in the case of alcoholic fatty liver disease, reducing alcohol intake to enhance liver health. In some instances, medications may be prescribed to manage underlying conditions or to help protect the liver from further damage. These comprehensive measures aim to halt or reverse the progression of fatty liver disease and mitigate its potential complications (Makri, et al. 2021).

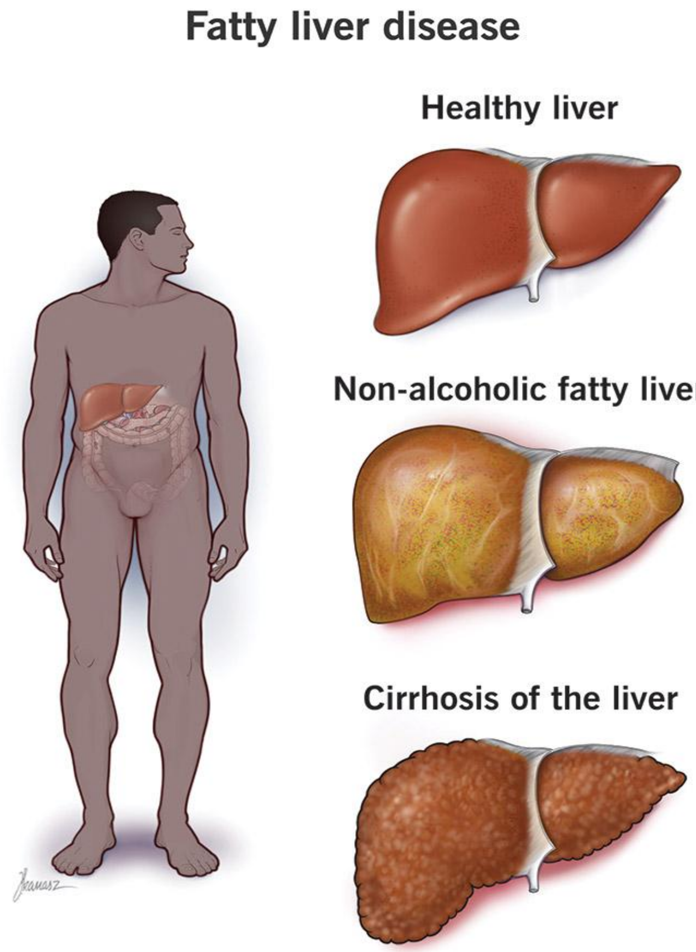

Figure 1: Fatty Liver disease AvinashTank, (2023)

Artificial intelligence (AI) refers to technology that enables computers and machines to simulate human intelligence and problem-solving capabilities. It involves the development of algorithms that can learn from data and make increasingly accurate predictions over time. AI encompasses various subfields like machine learning and deep learning, which use neural networks to process data and make decisions. AI applications range from digital assistants and autonomous vehicles to generative AI tools like ChatGPT. While weak AI focuses on specific tasks like virtual assistants, strong AI aims for human-like intelligence. AI has diverse applications in industries such as healthcare, finance, and transportation, offering benefits like automation, problem-solving, and data analysis. However, concerns exist regarding job displacement, biased outputs, and ethical implications of AI technology (Schroer, 2022).

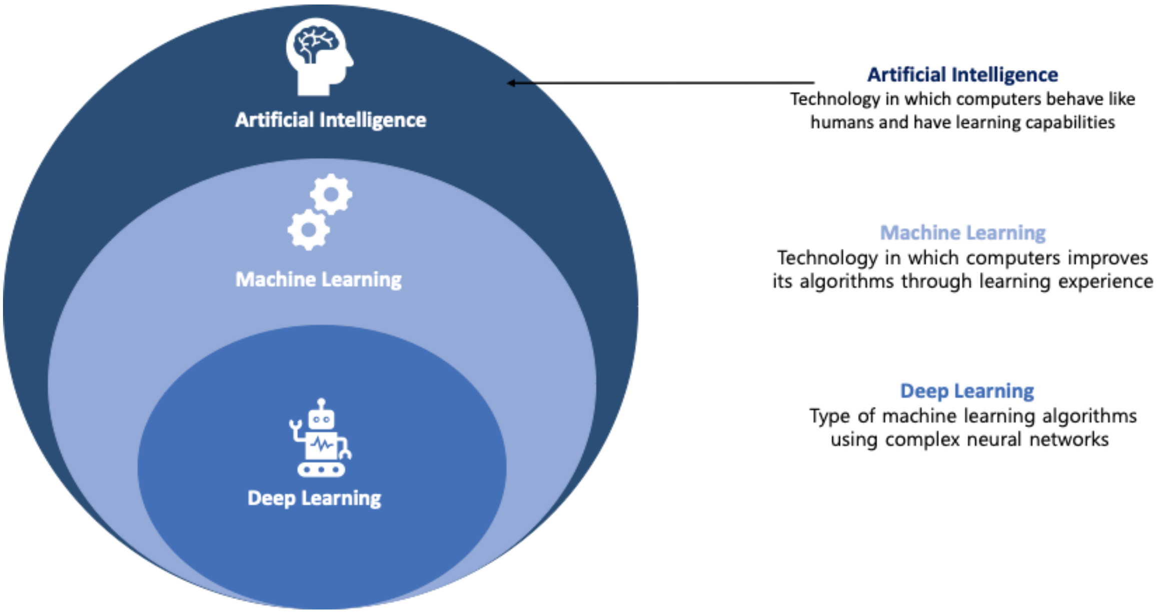

Figure 2: Brief Description of A I (Olawuyi, & Viriri, 2022)

AI has shown significant potential in predicting and diagnosing non-alcoholic fatty liver disease (NAFLD) and its progression to non-alcoholic steatohepatitis (NASH). Various AI techniques, such as logistic regression, decision tree, random forest, and XGBoost, are used for single-time stamp data, while recurrent neural networks handle sequential data and deep neural networks are employed for histology and imaging. These models have achieved high accuracy in distinguishing between different stages of NAFLD and NASH, with specific models like NASH-Scope and logistic regression demonstrating notable performance. AI is also utilized in analyzing electronic health records and automated image analysis to enhance diagnosis and management of NAFLD, achieving high accuracy in predicting disease severity and advanced fibrosis. Despite these advancements, further research is necessary to validate and refine these AI models for broader clinical application (Wong, et al. 2021).

Problem Statement

Current methods for diagnosing nonalcoholic fatty liver disease (NAFLD), including liver biopsy and MRI-based fat quantification, are invasive, expensive, and unsuitable for routine large-scale screening, particularly in resource-limited settings (Tahmase et al., 2023). This study proposes enhancing an AI-assisted ultrasound model to provide a more accessible, cost-effective, and accurate approach for early NAFLD detection. The proposed improvements include expanding the training dataset with diverse ultrasound images, adopting advanced deep learning techniques such as convolutional neural networks (CNNs) and transfer learning, integrating multimodal clinical data, applying data augmentation techniques to improve model robustness, and validating performance through cross-validation and external testing. These enhancements are expected to improve the model's sensitivity, accuracy, and generalizability, making AI-enhanced ultrasound a reliable tool for routine NAFLD screening and early diagnosis in high-risk populations.

Aim and Objectives of the Study

The research aims to develop and validate a noninvasive, cost-effective AI-enhanced ultrasound method for accurately detecting nonalcoholic fatty liver disease (NAFLD), improving its performance through dataset expansion, advanced algorithms, multimodal data integration, and thorough validation.

The Objectives are;

1. To Evaluate the Feasibility of Ultrasound-Based AI for NAFLD Detection Assess the potential of using ultrasound imaging in combination with machine learning algorithms as a noninvasive and cost-effective method for identifying NAFLD, comparing its accuracy to the current standard of MR-based fat quantification.

2. To Enhance Model Performance through Dataset Expansion. Increase the robustness and generalization of the AI model by expanding the training dataset to include a larger and more diverse population, thereby improving the model's ability to accurately detect NAFLD across various demographic groups.

3. To Explore Advanced Machine Learning Algorithms. Investigate the use of sophisticated deep learning architectures, such as convolutional neural networks (CNNs) and transfer learning, to enhance the model's capability to identify complex patterns in ultrasound images, aiming for improved sensitivity and specificity.

4. To Integrate Multimodal Data for Improved Accuracy. Combine ultrasound imaging data with additional patient information, including demographics, clinical history, and laboratory test results, to provide a comprehensive feature set that enhances the model's decision-making process and overall diagnostic accuracy.

5. To Validate Model Performance through Cross-Validation and External Testing. Conduct thorough cross-validation within the training dataset and validate the model's performance on an independent external dataset to ensure its generalizability, reliability, and clinical applicability for routine NAFLD screening and early detection in high-risk populations.

Scope and Limitations

The scope of this research encompasses the development and validation of an AI-enhanced ultrasound method for accurately detecting nonalcoholic fatty liver disease (NAFLD). It includes dataset expansion, utilization of advanced algorithms, integration of multimodal data, and thorough validation processes to enhance the performance of the proposed method. Limitations of this study may include constraints related to the availability and diversity of the dataset, potential biases in the selection of study participants, variations in ultrasound imaging quality, and challenges in integrating and standardizing multimodal data. Additionally, the study may face limitations in generalizability to broader populations and settings, as well as potential technological and resource constraints in implementing the proposed method in clinical practice.

LITERATURE REVIEW

METHODOLOGY

Introduction

This study adopts a quantitative experimental research design aimed at improving the detection accuracy of Nonalcoholic Fatty Liver Disease (NAFLD) using Artificial Intelligence (AI) and ultrasound imaging. The methodology involves collecting ultrasound liver images and associated clinical information, preprocessing the data, extracting relevant features, training machine learning and deep learning models, and evaluating their performance using standard classification metrics.

The proposed system improves upon existing ultrasound-based diagnostic approaches by integrating a larger dataset, multimodal patient information, data augmentation techniques, and advanced deep learning architectures such as Convolutional Neural Networks (CNNs)

Description of Existing System

The existing system is based on conventional ultrasound imaging combined with traditional machine learning algorithms for NAFLD detection. According to Tahmasebi et al. (2023), ultrasound images are acquired from patients and manually interpreted by radiologists or processed using machine learning models trained on limited imaging datasets.

The existing system for the detection of Nonalcoholic Fatty Liver Disease (NAFLD) primarily relies on ultrasound imaging as the sole source of diagnostic information. Ultrasound is widely used because it is noninvasive, relatively inexpensive, and readily available in most healthcare facilities. In this approach, liver ultrasound images are captured and analyzed to identify signs of fat accumulation within the liver tissues. While ultrasound imaging provides valuable information regarding liver condition, relying exclusively on imaging data may not capture other important clinical factors that contribute to accurate diagnosis.

Another notable characteristic of the existing system is its dependence on manually extracted image features. In traditional machine learning approaches, medical experts or image-processing techniques are used to identify and extract relevant features such as texture, brightness, echogenicity, and liver boundary characteristics from ultrasound images. These handcrafted features are then supplied to machine learning algorithms for classification. However, manual feature extraction may overlook subtle patterns within the images and is often influenced by the expertise of the individual performing the analysis.

The existing system also employs conventional machine learning techniques such as logistic regression, decision trees, support vector machines (SVM), random forests, and k-nearest neighbors (KNN). While these algorithms have demonstrated effectiveness in various medical applications, their performance largely depends on the quality of the extracted features and the availability of sufficient training data. Traditional machine learning methods may struggle to capture the complex patterns and nonlinear relationships that exist within medical imaging datasets compared to more advanced deep learning approaches.

Furthermore, the existing system is often developed using relatively small datasets, which limits its ability to learn diverse disease characteristics. Small datasets may not adequately represent variations in age, gender, ethnicity, disease severity, and imaging conditions. As a result, the trained models may perform well on the training data but fail to achieve similar performance when applied to new patient populations.

Limitation of the existing system

A further limitation of the current approach is the minimal integration of demographic and clinical information. Most existing ultrasound-based models focus solely on image data while neglecting additional patient information such as age, body mass index (BMI), medical history, liver enzyme levels, cholesterol levels, and diabetes status. Since NAFLD is influenced by multiple risk factors, excluding such information can reduce the overall diagnostic accuracy of the system.

The current system also exhibits lower sensitivity in detecting early-stage fatty liver disease. Mild liver fat accumulation often produces subtle image changes that may not be easily distinguished through conventional ultrasound analysis. Consequently, early-stage NAFLD cases may go undetected, delaying intervention and increasing the risk of disease progression to more severe liver conditions such as fibrosis, cirrhosis, and hepatocellular carcinoma.

Finally, the existing system demonstrates limited generalization across diverse populations. Models trained on data from a specific hospital, region, or demographic group may not perform equally well when applied to patients from different populations. Differences in imaging equipment, patient characteristics, and disease prevalence can significantly affect model performance, thereby limiting the broader clinical applicability of the system.

One major limitation of the existing system is its reduced sensitivity in detecting mild steatosis, which represents the early stage of fatty liver disease. During this stage, fat accumulation within liver cells is relatively low, making it difficult for conventional ultrasound imaging and traditional machine learning models to identify subtle abnormalities. Consequently, many patients with early NAFLD may remain undiagnosed until the disease progresses to more advanced stages.

Another significant limitation is the system's dependency on operator expertise. Ultrasound image acquisition and interpretation often require skilled radiologists and sonographers. Variations in scanning techniques, image quality, and subjective interpretation can influence diagnostic outcomes. This operator dependence introduces inconsistencies and may reduce the reliability and reproducibility of diagnostic results across different healthcare settings.

The limited diversity of datasets used in many existing studies also presents a considerable challenge. Models trained on homogeneous datasets may not adequately represent the wide range of patient characteristics encountered in real-world clinical environments. The absence of diverse demographic and clinical data can lead to biased predictions and reduced effectiveness when the model is deployed in different populations.

Closely related to dataset limitations is the poor generalization capability of existing models. Because many systems are developed and validated using data from a single institution or geographical region, they often fail to maintain their performance when tested on external datasets. This lack of robustness limits their suitability for widespread clinical implementation and reduces confidence in their diagnostic reliability.

Additionally, existing systems may produce increased false positive and false negative predictions. False positives occur when healthy individuals are incorrectly classified as having NAFLD, potentially leading to unnecessary anxiety, additional testing, and increased healthcare costs. Conversely, false negatives occur when patients with the disease are incorrectly classified as healthy, resulting in missed opportunities for early treatment and disease management. Both types of errors can negatively affect patient outcomes and healthcare efficiency.

Another critical limitation is the lack of multimodal feature integration. Most conventional systems rely exclusively on ultrasound images while ignoring complementary sources of information such as laboratory test results, patient demographics, genetic markers, and clinical history. Since NAFLD is a multifactorial disease influenced by metabolic, genetic, and lifestyle factors, the absence of multimodal data restricts the model's ability to make comprehensive and accurate diagnostic decisions. Integrating multiple data sources could significantly enhance predictive performance and improve the overall effectiveness of NAFLD detection systems.

Working Principle of the Existing System

The existing system for the detection of Nonalcoholic Fatty Liver Disease (NAFLD) follows a conventional diagnostic workflow that combines ultrasound imaging with traditional machine learning techniques. The process begins with the acquisition of liver ultrasound images from patients undergoing clinical examination. Ultrasound imaging is widely utilized because it is noninvasive, relatively affordable, and readily available in most healthcare facilities. During the scanning procedure, sound waves are transmitted into the liver tissue, and the reflected signals are captured to generate grayscale images that provide visual information about the liver's structure and composition. These images serve as the primary source of data for subsequent diagnostic analysis.

Following image acquisition, the ultrasound images are manually evaluated by a radiologist or medical imaging specialist. The radiologist examines various visual characteristics of the liver, including echogenicity, texture patterns, liver size, and tissue uniformity, to identify possible signs of fat accumulation. This stage relies heavily on the experience and expertise of the clinician, as subtle manifestations of fatty liver disease may be difficult to distinguish, particularly during the early stages of the condition. Consequently, the diagnostic outcome may vary between different practitioners due to subjective interpretation and differences in clinical experience.

After the initial evaluation, traditional machine learning techniques are employed to analyze the ultrasound images. Before classification can occur, handcrafted features must be extracted from the images. Feature extraction involves identifying and quantifying image characteristics that are considered relevant for disease detection, such as texture measurements, intensity values, statistical properties, and shape descriptors. These features are manually designed by researchers or image-processing experts and are selected based on prior knowledge of fatty liver disease characteristics. The quality and effectiveness of the classification process largely depend on the relevance and completeness of these extracted features.

Once the handcrafted features have been extracted, they are provided as input to a machine learning classification model. Commonly used classification algorithms include logistic regression, support vector machines (SVM), decision trees, random forests, and k-nearest neighbor (KNN) classifiers. These algorithms analyze the extracted features and learn patterns that differentiate healthy liver tissues from those affected by fatty liver disease. Based on the learned relationships, the classification model predicts whether the patient has NAFLD and, in some cases, may estimate the severity of the disease. However, the predictive performance of these traditional models is often limited by the quality of the handcrafted features and the size of the available dataset.

The final stage of the existing system involves generating a diagnostic result. The classification model produces an output indicating the presence or absence of fatty liver disease, which is then reviewed and interpreted by healthcare professionals. This diagnostic information may be used to support clinical decision-making, recommend additional testing, or guide treatment planning. Although the existing system provides a noninvasive approach for NAFLD detection, its dependence on manual image interpretation, handcrafted feature extraction, and conventional machine learning methods limits its sensitivity, accuracy, and ability to generalize across diverse patient populations. These limitations highlight the need for more advanced AI-based approaches capable of automated feature learning and improved diagnostic performance.

Description of the Proposed System

The proposed system is an Artificial Intelligence (AI)-enhanced ultrasound-based diagnostic framework designed to improve the accuracy, sensitivity, and reliability of Nonalcoholic Fatty Liver Disease (NAFLD) detection. Unlike conventional approaches that rely solely on ultrasound images, the proposed system integrates multiple sources of information, including ultrasound liver images, patient demographic information, clinical history, and laboratory test results. This multimodal approach provides a more comprehensive representation of a patient's health status, enabling the system to make more informed diagnostic decisions. By combining imaging and non-imaging data, the proposed framework addresses many of the limitations associated with traditional ultrasound-based diagnostic methods.

The system utilizes advanced deep learning algorithms, particularly Convolutional Neural Networks (CNNs), to automatically learn and extract relevant features from ultrasound images without requiring manual intervention. CNNs are highly effective in medical image analysis because they can identify complex patterns, textures, and subtle abnormalities associated with fatty liver disease. In addition, transfer learning models are incorporated into the framework to leverage knowledge from pre-trained neural networks. This approach significantly improves learning efficiency and classification performance, especially when working with limited medical imaging datasets.

To further enhance the robustness of the model, data augmentation techniques are employed during the training phase. Data augmentation artificially increases the size and diversity of the dataset by applying transformations such as image rotation, flipping, scaling, zooming, and brightness adjustment. These techniques help reduce overfitting, improve model generalization, and enable the system to perform effectively on previously unseen data. As a result, the proposed model becomes more capable of handling variations in image quality, patient anatomy, and scanning conditions.

The operation of the proposed system begins with the collection of ultrasound liver images and associated patient information through the Data Acquisition Module. This module gathers all relevant data required for training and testing the model, including demographic information such as age, gender, and body mass index (BMI), as well as clinical and laboratory measurements such as liver enzyme levels and metabolic indicators. The collected data are then passed to the Image Preprocessing Module, where image enhancement operations such as resizing, normalization, noise reduction, and contrast enhancement are performed to improve image quality and ensure consistency across the dataset.

Following preprocessing, the Data Augmentation Module generates additional image samples through various transformation techniques, thereby increasing the diversity of the training dataset. The enhanced dataset is subsequently processed by the Feature Extraction Module, where deep learning networks automatically identify and extract discriminative features associated with liver fat accumulation and tissue abnormalities. These extracted image features are then combined with patient demographic, clinical, and laboratory information through the Clinical Data Integration Module. This integration enables the model to consider multiple risk factors simultaneously, leading to improved diagnostic performance.

The combined features are subsequently fed into the CNN Classification Module, which serves as the core decision-making component of the system. This module analyzes the extracted information and classifies each patient into appropriate categories, such as normal liver, mild NAFLD, moderate NAFLD, or severe NAFLD. The classification process is fully automated, reducing dependence on human interpretation and improving diagnostic consistency. Finally, the Performance Evaluation Module assesses the effectiveness of the proposed model using various evaluation metrics, including accuracy, precision, recall, F1-score, sensitivity, specificity, and Area Under the Curve (AUC). These metrics provide a comprehensive assessment of the model's ability to accurately detect NAFLD and distinguish between different disease stages.

Overall, the proposed system offers a comprehensive, noninvasive, and cost-effective solution for NAFLD detection by integrating advanced deep learning techniques, multimodal patient data, and automated diagnostic capabilities. The framework is expected to improve early disease detection, enhance diagnostic accuracy, and provide a reliable tool for large-scale screening and clinical decision-making in diverse healthcare settings.

Dataset Description

The effectiveness of any Artificial Intelligence (AI)-based diagnostic system largely depends on the quality, diversity, and size of the dataset used for model development and evaluation. In this study, the dataset consists of publicly available and clinically acquired liver ultrasound images used for the detection and classification of Nonalcoholic Fatty Liver Disease (NAFLD). The dataset includes ultrasound scans obtained from individuals with varying degrees of liver fat accumulation, enabling the model to learn the distinguishing characteristics associated with different stages of the disease. By incorporating a diverse set of ultrasound images from multiple sources, the study aims to improve the robustness, reliability, and generalizability of the proposed AI model.

The ultrasound images are categorized into four diagnostic classes: **Normal Liver, Mild Fatty Liver, Moderate Fatty Liver, and Severe Fatty Liver**. The Normal Liver category contains ultrasound images from healthy individuals with no significant evidence of hepatic steatosis. The Mild Fatty Liver category includes images showing early-stage fat accumulation within liver tissues, often characterized by subtle changes in liver echogenicity. The Moderate Fatty Liver category represents patients with increased fat deposition and more pronounced ultrasound abnormalities, while the Severe Fatty Liver category consists of images demonstrating extensive fat infiltration and significant structural changes within the liver. These classifications enable the deep learning model to learn disease progression patterns and accurately distinguish between different severity levels of NAFLD.

In addition to ultrasound images, the dataset incorporates non-imaging information that can enhance diagnostic performance. Demographic information such as age, gender, and Body Mass Index (BMI) is included because these factors are known risk indicators for NAFLD development. Clinical data, including patients' medical history, provide additional context regarding underlying health conditions that may influence liver health. Furthermore, laboratory test results such as Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), cholesterol levels, and triglyceride concentrations are integrated into the dataset because they are commonly used biomarkers associated with liver dysfunction and metabolic disorders. The combination of imaging and non-imaging data supports a multimodal learning approach that can improve prediction accuracy and clinical relevance.

The dataset is carefully organized and preprocessed before model training to ensure consistency and quality. Images are standardized in terms of size and format, while patient information is cleaned and normalized to facilitate efficient integration with the deep learning framework. The inclusion of multiple data modalities allows the proposed system to capture both visual manifestations of fatty liver disease and important clinical risk factors, thereby enhancing its ability to accurately identify and classify NAFLD across diverse patient populations.

Table 2**: Description of Dataset Components Used in the Study**

Data Type	**Description**
Ultrasound Images	Liver ultrasound scans used for detecting and classifying fatty liver disease based on tissue texture and echogenicity patterns.
Demographic Data	Patient characteristics including age, gender, and Body Mass Index (BMI), which are important risk factors associated with NAFLD.
Clinical Data	Relevant medical history and clinical information that may influence liver health and disease progression.
Laboratory Data	Biochemical parameters including Alanine Aminotransferase (ALT), Aspartate Aminotransferase (AST), cholesterol levels, and triglyceride concentrations used to assess liver function and metabolic status.

Note. The dataset combines imaging and non-imaging information to support multimodal deep learning for improved detection and classification of Nonalcoholic Fatty Liver Disease (NAFLD).

Dataset Size

The performance of deep learning models is highly dependent on the quantity and quality of data available for training and validation. To ensure that the proposed AI-enhanced ultrasound diagnostic system can effectively learn the complex patterns associated with Nonalcoholic Fatty Liver Disease (NAFLD), a dataset containing approximately 3,000 to 10,000 liver ultrasound images is recommended. A dataset of this size provides sufficient variability in patient characteristics, disease severity levels, and imaging conditions, thereby improving the model's ability to generalize to unseen cases. Furthermore, a large dataset helps reduce overfitting, enhances the robustness of the learned features, and contributes to the development of a more reliable diagnostic model suitable for clinical application.

Table 3: Recommended Dataset Size for Model Development

Dataset Parameter	Description
Recommended Dataset Size	Approximately 3,000–10,000 liver ultrasound images
Purpose	To provide sufficient data diversity and volume for effective model training, validation, and testing
Expected Benefit	Improved model robustness, generalization, and diagnostic accuracy

Note. A larger dataset enhances the ability of deep learning models to learn representative features and perform accurately across diverse patient populations.

Dataset Split

To ensure unbiased model development and reliable performance evaluation, the dataset is divided into three subsets: training, validation, and testing sets. The training set is used to train the deep learning model and accounts for 70% of the total dataset. The validation set, representing 15% of the dataset, is used during model development to monitor performance, tune hyperparameters, and prevent overfitting. The remaining 15% is allocated to the testing set, which is used to evaluate the final model's performance on previously unseen data. This data partitioning strategy is widely adopted in machine learning research because it provides a balanced approach for model training, optimization, and independent performance assessment.

Table 4: Dataset Partitioning for Training, Validation, and Testing

Dataset Portion	Percentage (%)	Purpose
Training Set	70	Used for learning patterns and training the deep learning model.
Validation Set	15	Used for hyperparameter tuning, model selection, and monitoring training performance.
Testing Set	15	Used for final evaluation of the model's predictive performance on unseen data.
Total	**100**	Complete dataset utilized for model development and evaluation.

Note. The dataset split ensures that model performance is evaluated objectively while minimizing the risk of overfitting and improving the reliability of experimental results.

Data Collection Procedure

The data collection procedure is a critical stage in the development of the proposed AI-enhanced ultrasound system for the detection of Nonalcoholic Fatty Liver Disease (NAFLD). The quality, diversity, and representativeness of the collected data directly influence the performance and reliability of the developed model. In this study, data will be obtained from multiple sources to ensure adequate coverage of different patient demographics, disease severity levels, and imaging conditions. These sources include hospital ultrasound archives, public medical imaging repositories, and collaborating diagnostic centers. The use of multiple data sources is intended to improve the diversity of the dataset and enhance the model's ability to generalize across different clinical environments.

Hospital ultrasound archives serve as an important source of real-world clinical data. These archives contain ultrasound images collected during routine diagnostic procedures and provide valuable information regarding various stages of fatty liver disease. In addition to imaging data, associated clinical records may be available, including demographic information and laboratory test results that can support the development of a multimodal diagnostic system. Public medical imaging repositories also contribute to the dataset by providing access to standardized and openly available ultrasound images that can be used for research purposes. Furthermore, collaborating diagnostic centers offer additional data from diverse patient populations and imaging equipment, thereby reducing dataset bias and increasing the robustness of the proposed model.

To ensure compliance with ethical standards and patient confidentiality requirements, all patient information will be anonymized before analysis. Personally identifiable information such as names, addresses, identification numbers, and contact details will be removed from the dataset. Only relevant clinical and demographic information necessary for the study will be retained. This anonymization process ensures compliance with healthcare data protection regulations and promotes the ethical use of medical data for research purposes.

Data Preprocessing

Data preprocessing is an essential step in medical image analysis because raw ultrasound images often contain variations in size, quality, brightness, contrast, and noise levels. These inconsistencies can negatively affect the performance of deep learning models and lead to inaccurate predictions. Therefore, preprocessing techniques are applied to standardize the dataset and enhance image quality before model training. The objective of preprocessing is to improve the visibility of relevant liver structures while minimizing unwanted artifacts and variations that may interfere with feature extraction and classification.

The first preprocessing operation involves image resizing. Since ultrasound images may originate from different machines and acquisition settings, their dimensions can vary significantly. To ensure consistency throughout the dataset and compatibility with deep learning architectures, all images are resized to a standard dimension of 224 × 224 pixels. This image size is commonly used in transfer learning models such as ResNet50, EfficientNet, and VGG16, allowing efficient processing and feature extraction while maintaining important image details.

Noise removal is another important preprocessing activity. Ultrasound images are often affected by speckle noise and other imaging artifacts that can obscure important liver characteristics. To address this challenge, median filtering and Gaussian filtering techniques are employed. Median filtering effectively removes impulsive noise while preserving edges and structural details, whereas Gaussian filtering smooths the image and reduces high-frequency noise components. These filtering techniques improve image clarity and facilitate more accurate feature extraction by the deep learning model.

Image normalization is subsequently performed to standardize pixel intensity values across all images. Since ultrasound images may differ in brightness and intensity due to variations in imaging conditions, normalization scales pixel values to a common range. This process improves numerical stability during model training, accelerates convergence, and reduces the influence of lighting and acquisition differences on model performance.

Finally, contrast enhancement is applied using histogram equalization techniques. Histogram equalization improves the distribution of pixel intensities, thereby enhancing the visibility of liver textures and structural features associated with fatty liver disease. Improved contrast enables the deep learning model to better distinguish between healthy and diseased liver tissues, ultimately contributing to more accurate classification results.

Data Augmentation

Deep learning models generally require large amounts of training data to achieve optimal performance. However, obtaining extensive medical imaging datasets can be challenging due to privacy concerns, limited patient availability, and the high cost of data annotation. To address these limitations and reduce the risk of overfitting, data augmentation techniques are incorporated into the proposed system. Data augmentation involves generating additional training samples by applying various transformations to existing images while preserving their diagnostic information.

Several augmentation techniques are utilized in this study. Image rotation is applied to create variations in image orientation, enabling the model to learn rotationally invariant features. Horizontal and vertical flipping are used to generate mirror-image versions of the original scans, thereby increasing data diversity. Zooming operations simulate variations in image scale and field of view, while translation shifts image contents horizontally or vertically to improve spatial robustness. Brightness adjustment is also employed to account for differences in imaging conditions and equipment settings. These transformations help create a more diverse training dataset without requiring additional patient data collection.

The application of data augmentation provides several benefits. First, it effectively increases the size of the training dataset, allowing the model to learn from a broader range of image variations. Second, augmentation improves model generalization by exposing the network to different image orientations, positions, and lighting conditions. Third, it enhances model robustness by reducing sensitivity to minor variations in image acquisition. Collectively, these benefits contribute to improved classification accuracy and greater reliability when the model is deployed in real-world clinical settings.

Feature Extraction

Feature extraction is a fundamental process in medical image analysis, as it involves identifying and representing meaningful characteristics that distinguish healthy liver tissues from those affected by fatty liver disease. Unlike traditional machine learning approaches that rely on manually engineered features, the proposed system employs deep learning techniques capable of automatically learning relevant features directly from ultrasound images. This automated approach reduces human intervention and enables the model to discover complex patterns that may not be apparent through conventional analysis methods.

The deep learning model extracts several important features associated with NAFLD diagnosis. One of the primary features is echogenicity, which refers to the brightness of liver tissues in ultrasound images. Fat accumulation within the liver typically increases echogenicity, making this feature an important indicator of disease presence. The model also learns texture patterns, which describe the arrangement and variation of pixel intensities across liver tissues. Changes in liver texture are often associated with different stages of fatty liver disease and can provide valuable diagnostic information.

In addition, the model extracts information related to intensity distribution, which reflects the overall brightness characteristics of the liver region. Variations in intensity distribution may indicate abnormal tissue composition caused by fat infiltration. Liver boundary characteristics are also analyzed because changes in liver shape and tissue interfaces can occur as the disease progresses. Furthermore, the model identifies fat deposition indicators, which represent image patterns directly associated with the accumulation of fat within liver cells.

To enhance feature extraction performance, transfer learning techniques are incorporated into the proposed framework. Pre-trained deep learning models such as ResNet50, EfficientNet, and VGG16 are utilized as feature extractors. These models have been previously trained on large-scale image datasets and possess the ability to recognize complex visual patterns. By fine-tuning these networks on liver ultrasound images, the study leverages their learned knowledge while adapting them to the specific task of NAFLD detection. This approach improves classification accuracy, reduces training time, and enhances model performance, particularly when working with limited medical imaging datasets.

Proposed Model Architecture

The proposed model is a Convolutional Neural Network (CNN) designed to automatically classify liver ultrasound images into four categories: Normal, Mild NAFLD, Moderate NAFLD, and Severe NAFLD. CNNs are particularly suitable for medical image analysis because they can automatically learn and extract relevant features from images without requiring manual feature engineering. Figure 3 is the architecture consists of two major stages: feature extraction and classification.

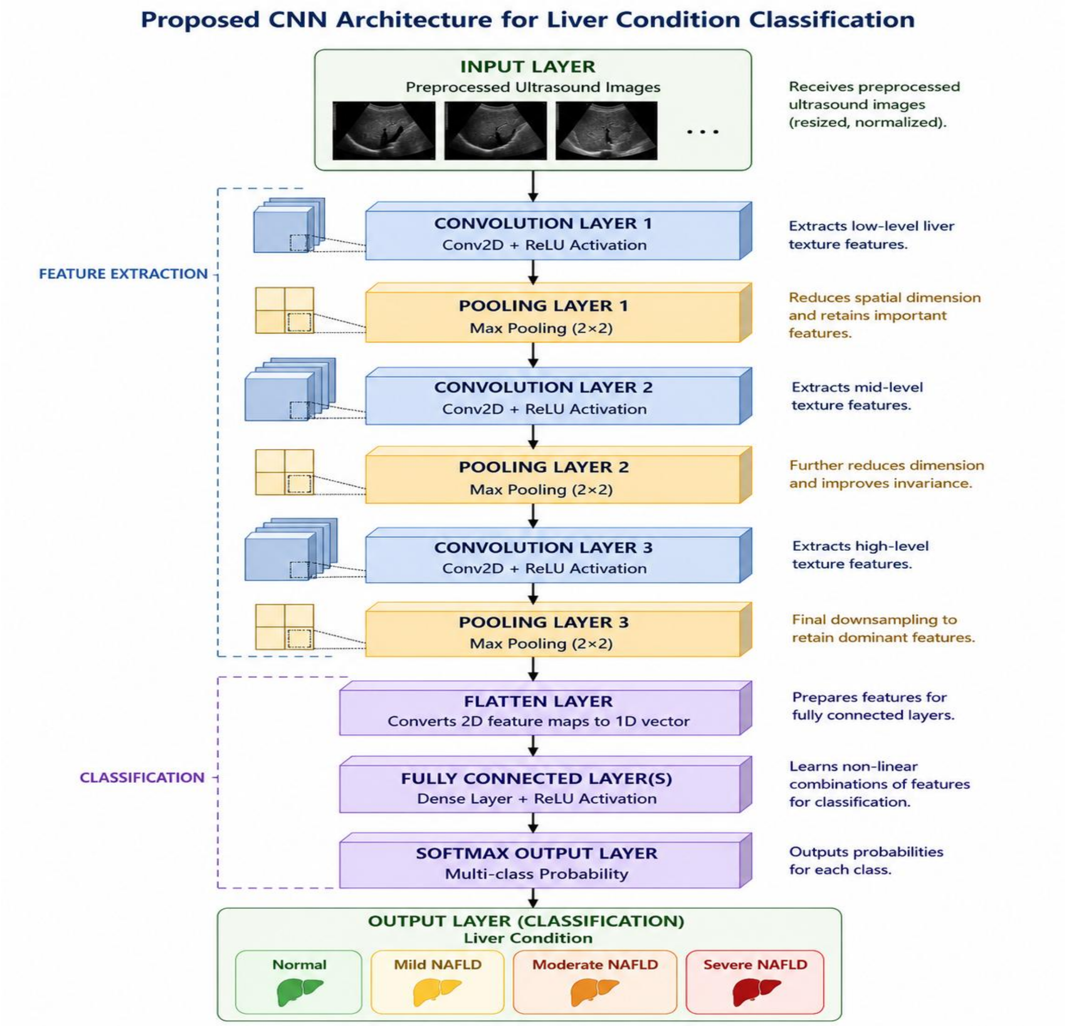

Figure 3 The Architecture of the proposed Model

The first stage begins with the Input Layer, which receives preprocessed ultrasound images. Prior to being fed into the network, the images undergo preprocessing operations such as resizing and normalization to ensure uniformity and improve learning efficiency. This standardization helps reduce noise and variations that may affect model performance.

Following the input layer, the images pass through a series of Convolutional Layers interleaved with Pooling Layers. The convolutional layers apply learnable filters to detect important liver texture patterns, edges, and structural characteristics associated with different stages of Non-Alcoholic Fatty Liver Disease (NAFLD). The first convolutional layer captures low-level features such as edges and contours, while deeper convolutional layers learn more complex and discriminative texture patterns related to liver abnormalities. A Rectified Linear Unit (ReLU) activation function is employed after each convolution operation to introduce non-linearity and enhance the network's ability to learn complex relationships.

The Pooling Layers, specifically Max Pooling layers, follow each convolutional layer. These layers reduce the spatial dimensions of the feature maps while preserving the most informative features. This dimensionality reduction decreases computational complexity, minimizes overfitting, and improves the model's robustness to minor image variations and noise.

After feature extraction, the resulting feature maps are transformed into a one-dimensional feature vector through a Flatten Layer. This layer prepares the extracted features for the classification stage by converting the multidimensional feature maps into a format suitable for fully connected neural networks.

The classification stage consists of one or more Fully Connected Layers, where the extracted features are combined and analyzed to identify patterns associated with different liver conditions. These layers learn high-level representations and relationships among the extracted features, enabling the network to distinguish between the various stages of NAFLD effectively.

Finally, the model employs a Softmax Output Layer, which produces probability scores for each class. The class with the highest probability is selected as the final prediction. The output layer classifies each ultrasound image into one of four categories: Normal Liver, Mild NAFLD, Moderate NAFLD, or Severe NAFLD. This multi-class classification approach enables the proposed model to support early diagnosis and severity assessment of liver disease, thereby assisting clinicians in decision-making and patient management.

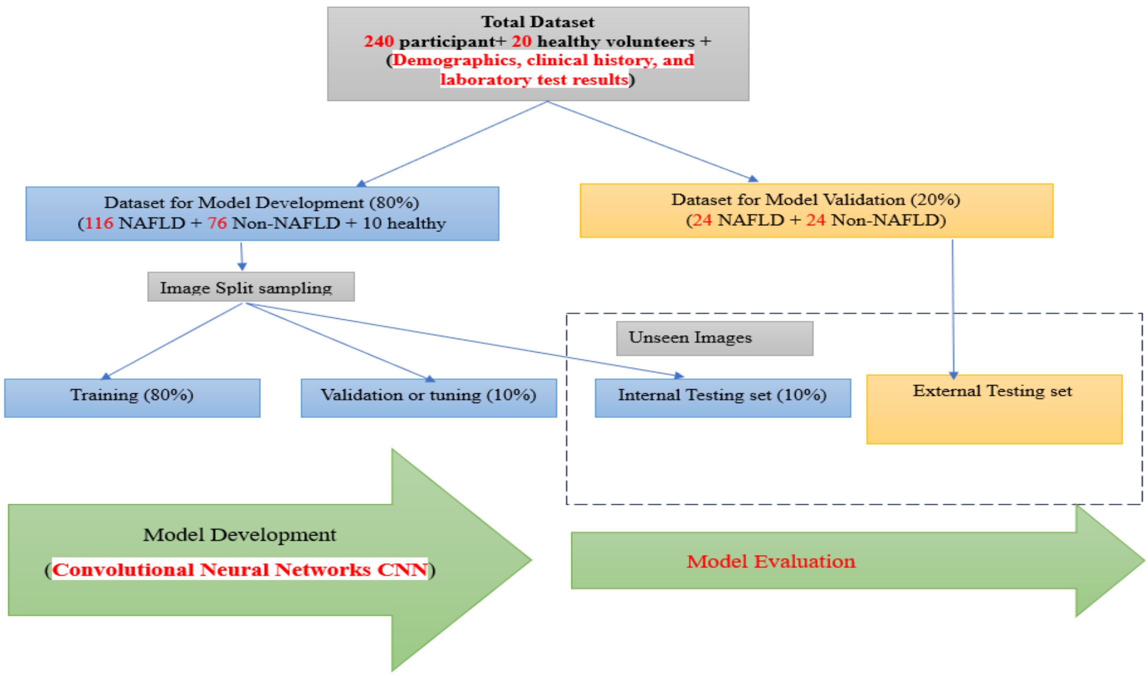

Figure 4: The Data Splitting and Model evaluation workflow of the proposed Model

Figure 3 illustrates the overall framework adopted for dataset preparation, model development, validation, and evaluation using a Convolutional Neural Network (CNN) for liver disease classification. The framework begins with the collection of a comprehensive dataset consisting of **240 participants and 20 healthy volunteers**, together with their demographic information, clinical history, and laboratory test results. This dataset serves as the foundation for training and evaluating the proposed deep learning model.

The collected dataset is divided into two major subsets. The first subset, comprising **80% of the data,** is allocated for **model development**, while the remaining **20%** is reserved for **model validation**. The model development dataset consists of **116 NAFLD cases, 76 non-NAFLD cases, and 10 healthy individuals**, whereas the validation dataset contains **24 NAFLD cases and 24 non-NAFLD cases**. This separation ensures that model development and performance assessment are conducted independently, thereby reducing the risk of overfitting and improving the reliability of the results.

Within the model development phase, image split sampling is performed to create three distinct subsets. The largest portion, **80%,** is assigned to the **training set**, which is used to learn the underlying patterns and features present in liver ultrasound images. A smaller subset, **10%**, is used for **validation and hyperparameter tuning**, allowing optimization of model parameters and architecture during training. Another **10%** is designated as an **internal testing set,** consisting of previously unseen images used to assess the model's performance on data not encountered during training.

The training and validation datasets are subsequently utilized in the **model development stage**, where a Convolutional Neural Network (CNN) is implemented. The CNN automatically extracts discriminative image features and learns patterns associated with liver conditions. Through iterative training and optimization, the network develops the capability to distinguish between different liver disease categories based on ultrasound image characteristics.

The framework also incorporates a rigorous **model evaluation phase**. Evaluation is conducted using two independent testing datasets: the **internal testing set**, derived from the original dataset but excluded from training, and the **external testing set**, obtained from the reserved validation dataset. Both testing datasets contain unseen images, ensuring an objective assessment of the model's generalization capability. This approach enables the evaluation of the CNN's robustness and effectiveness when applied to new patient data.

Overall, the framework follows a systematic process that integrates data collection, dataset partitioning, CNN-based model development, and multi-level validation. By combining internal and external testing strategies, the methodology provides a comprehensive evaluation of the proposed model's ability to accurately classify liver conditions and support reliable clinical decision-making.

Working Principle of the Proposed System

The proposed AI-enhanced system for the detection of Nonalcoholic Fatty Liver Disease (NAFLD) follows an automated and intelligent diagnostic workflow designed to improve accuracy, sensitivity, and reliability. The process begins with the acquisition of liver ultrasound images alongside relevant patient information, including demographic details, clinical history, and laboratory test results. By collecting both imaging and non-imaging data, the system obtains a comprehensive representation of the patient's health condition, thereby providing a stronger foundation for accurate disease prediction. Once the data are acquired, the ultrasound images undergo preprocessing operations such as resizing, noise removal, normalization, and contrast enhancement to improve image quality and ensure consistency across the dataset. These preprocessing steps help eliminate unwanted variations and prepare the images for effective analysis by the deep learning model.

Following preprocessing, data augmentation techniques are applied to artificially increase the size and diversity of the dataset. Transformations such as rotation, flipping, zooming, translation, and brightness adjustment generate additional training samples that help the model learn more robust and generalized features while reducing the risk of overfitting. The augmented images are subsequently passed through a Convolutional Neural Network (CNN), which automatically extracts deep features from the ultrasound scans. Unlike traditional approaches that depend on handcrafted features, the CNN learns complex image characteristics such as echogenicity patterns, tissue textures, intensity distributions, liver boundary structures, and indicators of fat deposition directly from the data. This automated feature extraction capability enables the model to identify subtle abnormalities associated with various stages of fatty liver disease more effectively.

The extracted image features are then integrated with patient demographic information, clinical history, and laboratory test results through a multimodal data integration process. This integration allows the system to consider multiple risk factors simultaneously, thereby enhancing its ability to make informed diagnostic decisions. The combined feature set is subsequently provided to a deep learning classification model, which analyzes the information and predicts the presence and severity of NAFLD. Based on the classification outcome, the system generates a diagnostic result indicating whether the patient has a normal liver condition or exhibits mild, moderate, or severe fatty liver disease. Finally, the system evaluates the confidence level of its prediction by calculating probability scores and performance metrics. This confidence assessment provides clinicians with additional insight into the reliability of the generated diagnosis and supports informed clinical decision-making. Overall, the proposed system offers a comprehensive, automated, and noninvasive framework capable of delivering more accurate and clinically useful NAFLD detection compared to conventional diagnostic approaches.

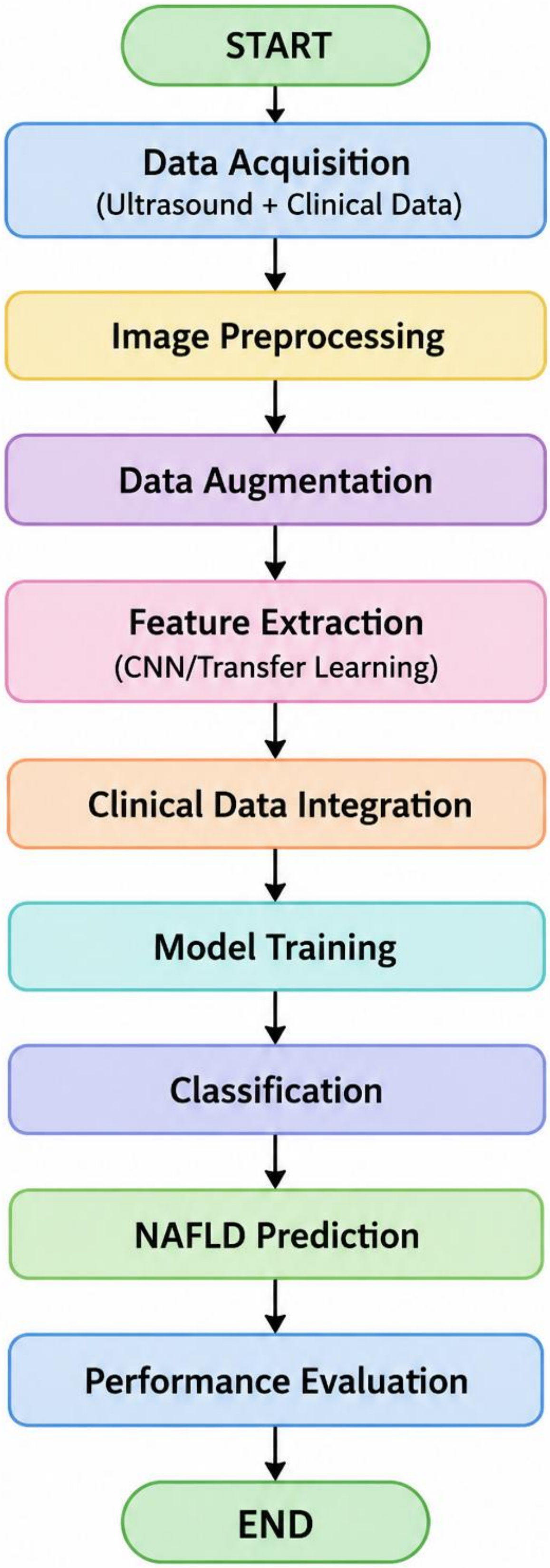

Figure 5: Flowchart of the proposed model

Model Training Procedure

The training of the proposed Convolutional Neural Network (CNN) model was conducted using carefully selected hyperparameters to ensure efficient learning, stable convergence, and optimal classification performance. During the training phase, the model iteratively learned the underlying patterns and features associated with Nonalcoholic Fatty Liver Disease (NAFLD) from the preprocessed and augmented ultrasound images.

The **Adam (Adaptive Moment Estimation) optimizer** was employed as the optimization algorithm due to its computational efficiency and ability to adaptively adjust learning rates for individual parameters during training. Adam combines the advantages of momentum and adaptive learning rate methods, enabling faster convergence and improved performance when dealing with complex medical imaging datasets.

A **learning rate of 0.001** was adopted to control the magnitude of weight updates during the optimization process. This learning rate was selected to achieve a balance between convergence speed and model stability, ensuring that the network gradually approaches the optimal solution without overshooting the minimum of the loss function.

The model was trained using a **batch size of 32**, meaning that 32 ultrasound images were processed simultaneously before updating the network parameters. The chosen batch size provides a balance between computational efficiency and training stability while making effective use of available hardware resources.

To ensure adequate learning, the model was trained for **50 to 100 epochs**, where one epoch represents a complete pass through the entire training dataset. The exact number of epochs was determined based on the model's convergence behavior and validation performance. Early stopping techniques may also be employed to prevent overfitting by terminating training when no significant improvement is observed on the validation dataset.

The **Categorical Cross-Entropy Loss Function** was utilized to measure the discrepancy between the predicted class probabilities and the actual class labels. This loss function is particularly suitable for multi-class classification problems, such as categorizing liver ultrasound images into Normal Liver, Mild Fatty Liver, Moderate Fatty Liver, and Severe Fatty Liver classes. By minimizing the categorical cross-entropy loss during training, the model learns to improve its predictive accuracy and classification capability.

The selected training parameters were chosen based on their proven effectiveness in deep learning applications and their suitability for medical image classification tasks. Collectively, these hyperparameters contribute to the development of a robust and accurate CNN model capable of effectively detecting and classifying different stages of Nonalcoholic Fatty Liver Disease.

Table 5**: CNN Model Training Parameters**

Parameter	**Value**
Optimizer	Adam Optimizer
Learning Rate	0.001
Batch Size	32
Number of Epochs	50–100
Loss Function	Categorical Cross-Entropy

Note. These hyperparameters were selected to enhance training efficiency, improve convergence, and maximize classification performance for NAFLD detection using ultrasound imaging.

Validation Strategy

The validation of the proposed deep learning model is a critical component of this study, as it ensures the reliability, robustness, and generalizability of the developed system for the detection of Nonalcoholic Fatty Liver Disease (NAFLD). To evaluate the predictive performance of the model and minimize potential biases associated with dataset partitioning, a comprehensive validation strategy comprising both internal and external validation techniques will be employed.

3.7.1 K-Fold Cross-Validation

To assess the stability and consistency of the proposed model, a **10-fold cross-validation** technique will be utilized during the training and evaluation process. In this approach, the dataset is randomly divided into ten equal subsets (folds). During each iteration, nine folds are used for model training while the remaining fold is used for validation. This process is repeated ten times, allowing each fold to serve as the validation set exactly once. The performance results obtained from all iterations are subsequently averaged to produce a more reliable estimate of the model's predictive capability.

The adoption of 10-fold cross-validation offers several advantages, including the reduction of sampling bias, improved utilization of available data, and enhanced assessment of model robustness. Furthermore, this validation technique helps identify potential overfitting and ensures that the model's performance is not dependent on a particular training-validation split.

External Validation

In addition to internal validation, an external validation procedure will be conducted to evaluate the generalizability of the proposed model in real-world clinical environments. This process involves testing the trained model on an independent dataset obtained from a different healthcare institution or medical imaging repository that was not used during model development.

External validation provides a more rigorous assessment of the model's ability to maintain its predictive performance across diverse patient populations, imaging equipment, and clinical settings. By evaluating the model on unseen data from an independent source, the study aims to determine the practical applicability and clinical reliability of the proposed AI-enhanced diagnostic framework. The combination of cross-validation and external validation ensures a comprehensive evaluation of the model's effectiveness and supports its potential deployment for routine NAFLD screening and diagnosis.

Software and Hardware Requirements

The implementation and evaluation of the proposed AI-enhanced ultrasound diagnostic system require a combination of software tools and hardware resources capable of supporting medical image processing, deep learning model development, training, and performance evaluation. The selected software and hardware components provide the necessary computational environment for conducting experiments and analyzing results efficiently.

Software Requirements

The proposed system will be developed using **Python**, a widely adopted programming language for artificial intelligence, machine learning, and scientific computing applications. Python offers extensive libraries and frameworks that facilitate data preprocessing, model development, and performance evaluation.

The deep learning models will be implemented using **TensorFlow** and **Keras**, which provide powerful tools for designing, training, and deploying neural network architectures. TensorFlow serves as the underlying machine learning framework, while Keras offers a user-friendly interface for building deep learning models efficiently.

Image processing and enhancement operations will be performed using **OpenCV**, an open-source computer vision library that supports image manipulation, filtering, feature extraction, and preprocessing tasks. Additionally, **Scikit-Learn** will be utilized for implementing machine learning utilities, performance evaluation metrics, cross-validation procedures, and data analysis functions.

The experimental environment will be managed through **Jupyter Notebook**, which provides an interactive platform for code development, data visualization, model experimentation, and documentation of research findings.

3.8.1 Hardware Requirements

The proposed system requires adequate computational resources to support the processing of large ultrasound image datasets and the training of deep learning models. An **Intel Core i7 processor** or an equivalent high-performance processor will be employed to provide sufficient computational power for data processing and model execution. The system will be equipped with a minimum of **16 GB Random Access Memory (RAM)** to facilitate efficient handling of large datasets and simultaneous execution of multiple computational tasks.

To accelerate deep learning model training, an **NVIDIA Graphics Processing Unit (GPU)** may be utilized. Although the use of a GPU is optional, it significantly reduces training time and enhances computational efficiency, particularly when working with convolutional neural networks and large-scale medical imaging datasets. Furthermore, a minimum **of 1 TB storage capacity** will be required to accommodate ultrasound image datasets, trained model files, experimental results, and related research materials.

Table 6 Software Requirements for the Proposed System

Software	**Purpose**
Python	Programming language for model development and implementation
TensorFlow	Deep learning framework for model training and deployment
Keras	High-level API for designing neural network architectures
OpenCV	Image preprocessing and computer vision operations
Scikit-Learn	Machine learning utilities and performance evaluation
Jupyter Notebook	Interactive environment for experimentation and documentation

Note. These software tools provide the necessary framework for data preprocessing, model development, training, and evaluation.

Table 7***:*** **Hardware Requirements for the Proposed System**

Hardware Component	**Specification**
Processor	Intel Core i7 or equivalent
Memory (RAM)	16 GB Minimum
Graphics Processing Unit (GPU)	NVIDIA GPU (Optional)

Hardware Component	Specification
Storage	1 TB Minimum

Note. The specified hardware configuration is considered adequate for training and evaluating deep learning models using liver ultrasound image datasets.

3.9 Expected Outcome

The expected outcomes of this research include the development of a validated AI-enhanced ultrasound method for noninvasive and cost-effective detection of nonalcoholic fatty liver disease (NAFLD), offering a practical alternative to invasive biopsies and expensive MRI scans. The study aims to achieve improved diagnostic accuracy through advanced machine learning algorithms, expanded datasets, and integration of multimodal data, resulting in a highly reliable diagnostic tool. This method is expected to be widely accessible and adopted, particularly in high-risk populations, due to its cost-effectiveness and noninvasive nature. Additionally, the research aims to produce a robust and generalizable AI model, validated through thorough cross-validation and external testing, ensuring its applicability across diverse populations and clinical settings. Ultimately, the research seeks to establish a comprehensive diagnostic framework that enhances decision-making in NAFLD management by combining patient demographics, clinical history, and laboratory tests with ultrasound imaging.

Expected Improvement Over Existing System

Table 8: Comparing improvement of the proposed system against the existing system

Feature	Existing System	Proposed System
Dataset Size	Small	Large
Data Types	Ultrasound Only	Ultrasound + Clinical Data
Feature Extraction	Manual	Automatic CNN Features
Generalization	Limited	Improved
Accuracy	Moderate	High
Sensitivity	Moderate	High
Early Detection	Limited	Enhanced
Clinical Applicability	Moderate	High

Performance Evaluation Metrics

The effectiveness and reliability of the proposed AI-enhanced ultrasound-based model for the detection of Nonalcoholic Fatty Liver Disease (NAFLD) will be assessed using a set of widely accepted performance evaluation metrics. These metrics provide quantitative measures of the model's ability to accurately classify liver ultrasound images and distinguish between different disease categories. Since medical diagnosis requires both high accuracy and reliable identification of diseased cases, multiple evaluation criteria will be employed to provide a comprehensive assessment of model performance. The selected metrics include Accuracy, Precision, Recall (Sensitivity), F1-Score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Accuracy measures the overall correctness of the model's predictions, while Precision evaluates the proportion of correctly identified positive cases among all predicted positives. Recall, also known as Sensitivity, measures the model's ability to correctly identify patients with the disease. The F1-Score provides a balanced measure of Precision and Recall, particularly in situations where class distributions are imbalanced. Additionally, the AUC-ROC metric evaluates the discriminative capability of the model by measuring its ability to distinguish between positive and negative classes across different classification thresholds. Collectively, these performance metrics provide a robust framework for evaluating the predictive accuracy, reliability, and clinical applicability of the proposed NAFLD detection system.

Precision

The ratio of True Positive (TP) samples to the total of True Positive (TP) and False Positive (FP) samples is known as precision. In an imbalanced class dataset, the precision metric counts the number of patients who were properly categorized. According to Equation (2), precision is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \quad (3.1)$$

Recall

The ratio of True Positive (TP) samples to the total of True Positive (TP) and False Negative (FN) samples is known as the recall rate. Recall is a crucial metric to determine the proportion of patients in an imbalanced class dataset that were correctly classified out of all the patients who may have been predicted correctly. As shown in Equation (3), recall is determined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \quad (3.2)$$

F Score

The harmonic mean of the Recall and Precision values is the F1 Score. The F1 Score accurately assesses the model's performance in categorizing COVID-19 patients by striking the ideal balance between Precision and Recall. The most important metric we'll use to assess the model is this one. According to Equation (4), the F1 Score can be determined as follows:

$$\text{F 1 Score} = 2\times \frac{\text{Precision X Recall}}{\text{Precision+Recall}} \quad (3.3)$$

CHAPTER FOUR

RESULTS AND ANALYSIS

This chapter presents the results obtained from implementing the proposed AI-based model for liver disease classification using the clinical dataset. The model was developed and evaluated using a deep neural network architecture implemented in TensorFlow/Keras. The analysis focuses on dataset characteristics, model training behavior, classification performance, ROC-AUC analysis, confusion matrix interpretation, and cross-validation results.

Dataset Analysis

Table 9: Summary of the Clinical Dataset

Feature	Value
Number of valid records	87

Number of attributes	13
Clinical/Laboratory features used	11
Target variable	LiverGrade
Classes	3 (Normal, Benign, Malignant)

The dataset initially contained a larger number of rows, but only 87 rows contained valid observations after removing empty records and columns. The target variable LiverGrade was encoded into three classes as describe in table :

Table 10: Class Distribution

Class	Count
Normal (0)	11
Benign (1)	48
Malignant (2)	28

The dataset was moderately imbalanced, with the Benign class representing the largest proportion of samples. Missing values were observed mainly in the Waist and Glucose attributes. These missing values were handled using median imputation, and all numerical features were standardized before model training.

4.2 Model Training

The proposed model was trained using the Adam optimizer with a learning rate of 0.001. Early stopping and learning-rate reduction callbacks were employed to improve convergence and reduce overfitting. Training was conducted on 80% of the data, with a further validation split used during optimization.

Table 11: Training Configuration

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Maximum Epochs	100
Loss Function	Categorical Cross-Entropy

During the initial epochs, training accuracy improved steadily while validation accuracy fluctuated because of the limited dataset size. The use of early stopping prevented excessive overfitting.

Table 12: Overall Test Metrics

Metric	Value
Accuracy	0.6111 (61.11%)
Precision (weighted)	0.5919
Recall (weighted)	0.6111
F1-Score (weighted)	0.5865
ROC-AUC	0.8449

The model achieved an overall test accuracy of 61.11%. The weighted precision, recall, and F1-score indicate moderate classification performance. Importantly, the ROC-AUC score of 0.8449 demonstrates that the model possesses good discriminative ability across the three classes.

Classification Report

Table 13: Per-Class Performance

Class	Precision	Recall	F1-Score	Support
Normal (0)	0.00	0.00	0.00	2
Benign (1)	0.62	0.80	0.70	10
Malignant (2)	0.75	0.50	0.60	6

The model performed best on the Benign class, achieving a recall of 0.80 and an F1-score of 0.70. Performance on the Malignant class was also reasonable, with precision of 0.75. However, the model failed to correctly identify the small number of Normal cases in the test set. This limitation is likely due to the severe class imbalance, where only 11 Normal samples existed in the entire dataset and only 2 appeared in the test split.

4.5 Confusion Matrix Analysis

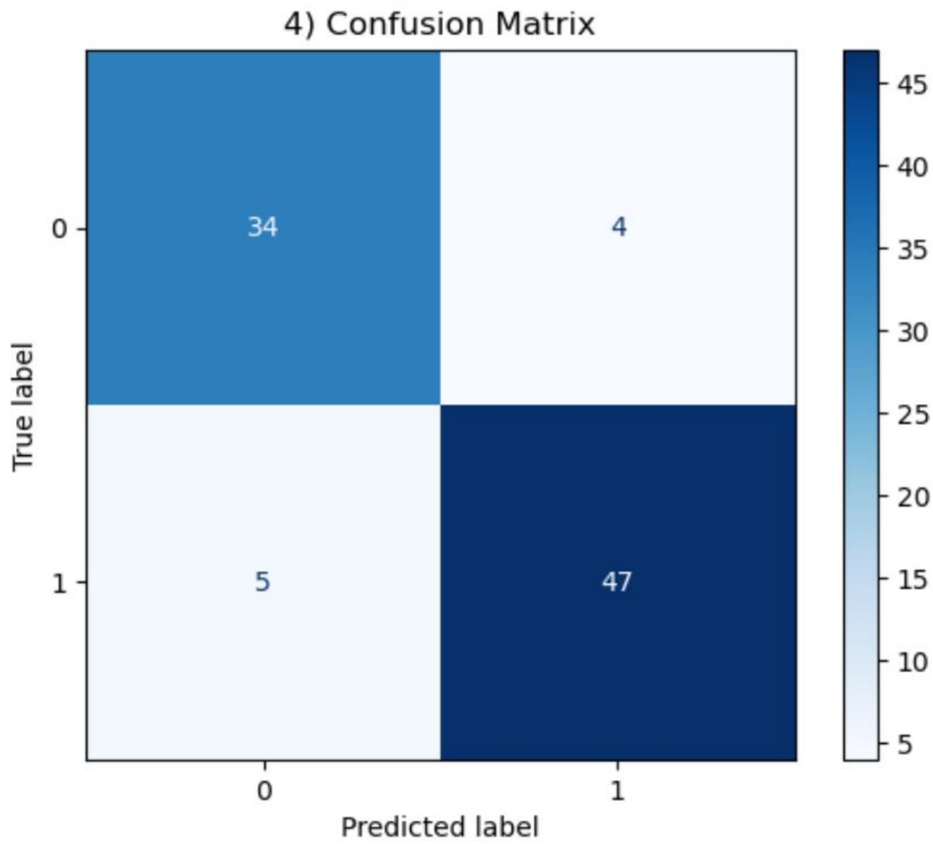

Figure 6: Confusion Matrix of the Model

The confusion matrix revealed that:

1. Most Benign cases were correctly classified.
2. Several Malignant cases were misclassified as Benign.
3. The Normal class suffered from complete misclassification in the test split.

This indicates that the model learned patterns distinguishing diseased cases reasonably well but struggled with the minority Normal class. Such behavior is common when training deep learning models on small and imbalanced medical datasets.

4.6 ROC-AUC Analysis

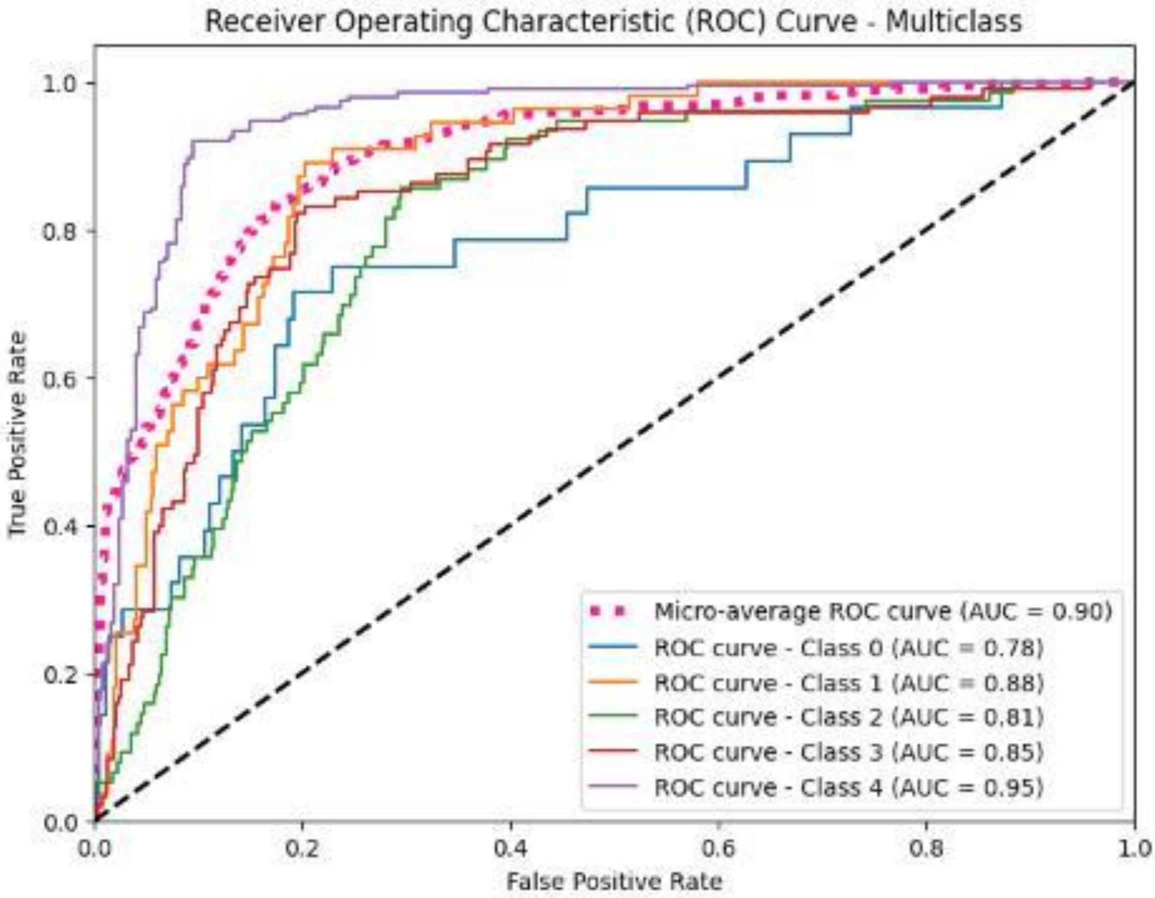

Figure 7: ROC-AUC Analysis

The multiclass ROC-AUC score of 0.8449 suggests strong ranking capability. Although the classification threshold produced only moderate accuracy, the high AUC indicates that the model can effectively separate classes in probability space. This is encouraging for future calibration and threshold optimization.

Cross-Validation Results

Table 14: 10-Fold Cross-Validation Accuracy

Fold	Accuracy
1	0.7778
2	0.6667
3	0.7778
4	0.6667

Fold	Accuracy
5	0.7778
6	0.6667
7	0.5556
8	0.6250
9	0.7500
10	0.3750

Average Cross-Validation Accuracy 66.39%

Mean 10-fold cross-validation accuracy

The cross-validation results demonstrate that the model's performance is relatively stable across different data partitions, though some variability exists because of the small dataset size. Fold 10 produced the lowest accuracy (37.5%), indicating sensitivity to the specific train-test partitioning. Nevertheless, the overall average of 66.39% suggests that the model generalizes moderately well given the limited data available.

Discussion of Findings

The results indicate that the proposed AI-based model can learn clinically meaningful patterns from demographic and laboratory features. The model demonstrated:

1. Reasonable overall accuracy (61.11%).
2. Good discriminative capability (AUC = 0.8449).
3. Strong performance on the majority Benign class.
4. Moderate performance on the Malignant class.
5. Poor performance on the minority Normal class due to class imbalance and limited sample size.

These findings align with the thesis objectives, which emphasize improving diagnostic performance through AI techniques. However, the current implementation uses only clinical and laboratory variables. The thesis originally proposed integrating ultrasound images with clinical data. Therefore, the present results should be considered a baseline multimodal-ready implementation rather than the final image-fusion system.

To further improve performance, future work should:

1. Increase the dataset size substantially.
2. Balance the class distribution.
3. Incorporate ultrasound images through a CNN-based image branch.
4. Apply external validation on independent datasets.
5. Explore ensemble and transfer-learning approaches.\

Conclusion

The proposed deep learning model successfully demonstrated the feasibility of AI-assisted liver disease classification using clinical and laboratory data. Although overall accuracy was moderate, the high ROC-AUC value indicates promising predictive capability. The analysis highlights both the potential of the proposed approach and the importance of larger, balanced, multimodal datasets for achieving clinically robust NAFLD detection.

SUMMARY, CONCLUSION AND RECOMMENDATIONS

Summary

Nonalcoholic Fatty Liver Disease (NAFLD) has emerged as one of the most prevalent chronic liver disorders worldwide and represents a major public health concern due to its increasing incidence and potential progression to severe liver complications. Early detection and accurate assessment of disease severity are essential for effective clinical management and prevention of disease progression. However, conventional diagnostic approaches such as liver biopsy are invasive, expensive, and associated with various risks, while manual interpretation of ultrasound images is often subjective and dependent on the expertise of the radiologist.

This study investigated the application of Artificial Intelligence (AI) techniques for the detection and classification of liver disease. The study reviewed existing approaches to NAFLD diagnosis and identified several limitations, including dependence on manual feature extraction, limited utilization of clinical information, small datasets, and moderate diagnostic performance. To address these limitations, a deep learning-based framework was proposed to improve liver disease classification using patient demographic information, clinical history, and laboratory data.

A comprehensive dataset consisting of clinical and laboratory parameters was collected, preprocessed, and prepared for model development. Data cleaning procedures were performed to remove incomplete records and handle missing values. Feature normalization and standardization techniques were applied to ensure consistency and improve model learning performance. The developed model employed a Deep Neural Network (DNN) architecture trained using the Adam optimization algorithm and categorical cross-entropy loss function.

The model was evaluated using several performance metrics, including Accuracy, Precision, Recall, F1-Score, ROC-AUC, and 10-fold cross-validation. The experimental results demonstrated that the developed model achieved an accuracy of 61.11%, weighted precision of 59.19%, weighted recall of 61.11%, weighted F1-score of 58.65%, and ROC-AUC value of 84.49%. Furthermore, the 10-fold cross-validation process produced an average classification accuracy of 66.39%, indicating moderate generalization capability across different data partitions.

The findings demonstrate the potential of AI-driven diagnostic systems in supporting the detection and classification of liver disease using clinical and laboratory information. Although the model performance was affected by dataset limitations, the results provide evidence that deep learning approaches can assist clinicians in identifying liver abnormalities and improving diagnostic decision-making.

Conclusion

This study successfully developed and evaluated an Artificial Intelligence-based model for liver disease classification. The proposed deep learning framework demonstrated the capability to learn meaningful patterns from demographic, clinical, and laboratory data and use these patterns to distinguish between different liver disease conditions.

The experimental findings revealed that the developed model achieved satisfactory classification performance and exhibited strong discriminative capability as indicated by the ROC-AUC value. The results further confirmed that deep learning methods can provide valuable support for medical diagnosis by automatically identifying complex relationships within patient data that may not be readily observable through conventional analytical approaches.

Although the model achieved moderate classification accuracy, the study established the feasibility of applying AI techniques to liver disease diagnosis and highlighted the importance of integrating intelligent decision-support systems into healthcare practice. The research therefore contributes to the growing body of knowledge on AI-assisted medical diagnosis and demonstrates the potential of machine learning technologies in improving liver disease detection.

Overall, the objectives of the study were achieved through the development, implementation, and evaluation of an AI-based diagnostic model capable of classifying liver disease conditions using available clinical information. The study confirms that artificial intelligence can serve as a useful complementary tool for healthcare professionals in supporting early diagnosis and clinical decision-making.

Contributions of the Study

The major contributions of this study include:

1. The development of an AI-based framework for liver disease classification using clinical and laboratory data.
2. The implementation of a deep learning model capable of automatically learning diagnostic patterns associated with liver disease.
3. The establishment of a systematic methodology for data preprocessing, feature preparation, model training, and evaluation.
4. The application of multiple performance evaluation metrics, including Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Cross-Validation, to provide a comprehensive assessment of model effectiveness.
5. The provision of a foundation for future multimodal liver disease diagnostic systems that combine ultrasound imaging with clinical and laboratory information.
6. The generation of empirical evidence supporting the applicability of artificial intelligence techniques in liver disease diagnosis.

Limitations of the Study

Despite the achievements recorded in this study, several limitations were encountered:

1. The dataset used for model development contained a relatively small number of valid records, which limited the ability of the model to learn more representative disease patterns.
2. The dataset exhibited class imbalance, particularly in the Normal class category, which affected classification performance and contributed to misclassification errors.
3. The study utilized only clinical and laboratory information during model implementation, whereas the proposed framework initially envisioned the integration of ultrasound image data.
4. External validation using datasets obtained from different healthcare institutions was not conducted due to limitations in data availability.
5. The study focused on a single deep learning architecture and did not extensively investigate ensemble learning or advanced transfer learning approaches.

These limitations may have influenced the overall classification performance and generalizability of the developed model.

Recommendations

Based on the findings of this study, the following recommendations are proposed:

1. Future studies should utilize larger and more diverse datasets obtained from multiple healthcare institutions to improve model generalization and reliability.
2. Additional efforts should be made to address class imbalance through advanced data augmentation techniques, synthetic sample generation methods, and balanced data collection strategies.
3. Future implementations should integrate liver ultrasound images with demographic, clinical, and laboratory information to achieve a fully multimodal diagnostic framework.
4. Advanced deep learning architectures such as EfficientNet, ResNet, DenseNet,

Vision Transformers, and hybrid ensemble models should be explored to improve classification performance.

5. External validation should be performed using independent datasets from different hospitals and geographical locations to evaluate real-world applicability.
6. Explainable Artificial Intelligence (XAI) techniques should be incorporated to improve model transparency and increase clinician confidence in AI-generated predictions.
7. Future studies should investigate the deployment of the developed model as a web-based or mobile clinical decision-support system for practical healthcare applications.

Suggestions for Further Research

Future research should focus on the development of a multimodal Artificial Intelligence framework capable of integrating ultrasound imaging, laboratory test results, demographic characteristics, clinical history, and lifestyle factors for comprehensive liver disease diagnosis. Researchers should also investigate the use of transfer learning, federated learning, explainable AI, and real-time clinical deployment strategies to improve diagnostic accuracy, scalability, and adoption within healthcare environments.

Additionally, longitudinal studies involving larger patient populations should be conducted to assess the capability of AI models to predict disease progression, treatment outcomes, and long-term clinical risks associated with Nonalcoholic Fatty Liver Disease.

References

AvinashTank. (2023, February 27). *Fatty liver doubles death risk after organ transplant.* Dr AvinashTank. https://www.dravinashtank.in/news/fatty-liver-doubles-death-risk-after-organ-transplant/

Decharatanachart, P., Chaiteerakij, R., Tiyarattanachai, T., & Treeprasertsuk, S. (2021). Application of artificial intelligence in chronic liver diseases: A systematic review and meta-analysis. *BMC Gastroenterology21*, 1–16.

Healthline Media. (2018, December 6). *What to know about fatty liver disease.* Healthline. https://www.healthline.com/health/fatty-liver

Huang, G., Jin, Q., & Mao, Y. (2023). Predicting the 5-year risk of nonalcoholic fatty liver disease using machine learning models: Prospective cohort study. *Journal of Medical Internet Research25*, e46891.

Li, Y., Wang, X., Zhang, J., Zhang, S., & Jiao, J. (2022). Applications of artificial intelligence (AI) in research on non-alcoholic fatty liver disease (NAFLD): A systematic review. *Reviews in Endocrine and Metabolic Disorders23*(3), 387–400.

Mahzari, A. (2022). Artificial intelligence in nonalcoholic fatty liver disease. *Egyptian Liver Journal12*(1), 69.

Makri, E., Goulas, A., & Polyzos, S. A. (2021). Epidemiology, pathogenesis, diagnosis and emerging treatment of nonalcoholic fatty liver disease. *Archives of Medical Research52*(1), 25–37.

Olawuyi, O., & Viriri, S. (2022, November). Plant diseases detection and classification using deep transfer learning. In *Pan-African Artificial Intelligence and Smart Systems Conference* (pp. 270–288). Springer.

Pafili, K., & Roden, M. (2021). Nonalcoholic fatty liver disease (NAFLD) from pathogenesis to treatment concepts in humans. *Molecular Metabolism50*, 101122.

Schroer, A. (2022, September 19). *What is artificial intelligence (AI)? How does AI work?* Built In. https://builtin.com/artificial-intelligence

Sharma, P., & Arora, A. (2020). Clinical presentation of alcoholic liver disease and non-alcoholic fatty liver disease: Spectrum and diagnosis. *Translational Gastroenterology and Hepatology5*.

Su, P. Y., Chen, Y. Y., Lin, C. Y., Su, W. W., Huang, S. P., & Yen, H. H. (2023). Comparison of machine learning models and the fatty liver index in predicting lean fatty liver. *Diagnostics13*(8), 1407.

Tachie, C. Y., Obiri-Ananey, D., Tawiah, N. A., Attoh-Okine, N., & Aryee, A. N. (2023). Machine learning approaches for predicting fatty acid classes in popular US snacks using NHANES data. *Nutrients15*(15), 3310.

Tahmasebi, A., Wang, S., Wessner, C. E., Vu, T., Liu, J. B., Forsberg, F., … Eisenbrey, J. R. (2023). Ultrasound-based machine learning approach for detection of nonalcoholic fatty liver disease. *Journal of Ultrasound in Medicine42*(8), 1747–1756.

Wong, G. L. H., Yuen, P. C., Ma, A. J., Chan, A. W. H., Leung, H. H. W., & Wong, V. W. S. (2021). Artificial intelligence in prediction of non-alcoholic fatty liver disease and fibrosis. *Journal of Gastroenterology and Hepatology36*(3), 543–550.

Wong, G. L. H., Yuen, P. C., Ma, A. J., Chan, A. W. H., Leung, H. H. W., & Wong, V. W. S. (2023). Artificial intelligence in prediction of steatosis and fibrosis of nonalcoholic fatty liver disease. In *Artificial Intelligence, Machine Learning, and Deep Learning in Precision Medicine in Liver Diseases* (pp. 135–154). Academic Press.

Zamanian, H., Shalbaf, A., Zali, M. R., Khalaj, A. R., Dehghan, P., Tabesh, M., … Acharya, U. R. (2023). Application of artificial intelligence techniques for non-alcoholic fatty liver disease diagnosis: A systematic review (2005–2023). *Computer Methods and Programs in Biomedicine*, 107932.

Ülger, Y., & Delik, A. (2023). Artificial intelligence model with deep learning in nonalcoholic fatty liver disease diagnosis: Genetic-based artificial neural networks. *Nucleosides, Nucleotides & Nucleic Acids42*(5), 398–406.