International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Non-Alcoholic Hepatic Steatosis Grading: Exploring the AI Potential of Machine Vs Deep Learning

Dr.Neelam Chittora[1], Ansh Garg[2], Dr.Anurag Chittora[3], Anushka Chittora[4]

[1]Associate professor, Department of Radio-diagnosis Sudha Medical College
Kota, India
[2]Electronics and Instrumentation Nirma Institute of Technology
Ahmedabad, India
[3]Associate professor, Department of Surgery Sudha Medical College
Kota, India
[4]Electronics and Instrumentation Nirma Institute of Technology
Ahmedabad, India
[1]drneelamchittora2003@yahoo.com, [2]22bei006@nirmauni.ac.in,[3]dranuragchittora2006@yahoo.co.in
[4]22bei007@nirmauni.ac.in

Abstract

Non-alcoholic fatty liver disease(NAFLD)is a grow- ing global health concern and a leading cause of chronic liver disease. Accurate, early-stage detection and grading of hepatic steatosis are essential to prevent progression to more severe con- ditions such as cirrhosis or hepatocellular carcinoma. This study evaluates the performance of classical machine learning and deep learning approaches for automated classification of liver steatosis using ultrasound (US) images from the BEHSOF dataset, which includes annotated clinical metadata. Preprocessing techniques such as grayscale normalization and Local Binary Pattern (LBP) feature extraction were employed to enhance diagnostic features. Three models—Support Vector Machine(SVM), Random Forest (RF), and an Artificial Neural Network (ANN)—were developed and tested. The ANN achieved the highest accuracy (99.09%), outperforming both RF(99.0%)and SVM(80%)classifiers, and surpassing previously reported Inception-ResNet-v2 results. These findings highlight the potential of interpretable and computationally efficient AI systems in supporting ultrasound- based NAFLD assessment, especially in resource-limited clinical environments.

Key Words: Fatty Liver, Deep Learning, Ultrasound Imag- ing, Machine Learning, Convolutional Neural Networks (CNN), Transfer Learning, Feature Extraction, Medical Image Classifi- cation, Non-Alcoholic Fatty Liver Disease (NAFLD), Computer- Aided Diagnosis (CAD)

1. INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) affects a signif- icant portion of the global population and is a leading cause of liver-related morbidity. Grading NAFLD based on ultra- sound imaging is critical for determining treatment strategies. Traditional manual classification methods are subjective and time-consuming. This paper presents an automated machine learning-based approach to classify fatty liver grades using the BEHSOF dataset, which includes ultrasound images and associated clinical metadata. The study aims to improve clas- sification accuracy and aid in early diagnosis.RELATEDWORK

Several studies have explored machine learning and deep learning techniques for liver disease classification. Traditional methods rely on handcrafted features extracted from ultra- sound images, while recent approaches use CNNs for feature learning. However, most previous studies suffer from limited datasets, affecting generalization.

Recent research has introduced deep neural networks to classify fatty liver disease using ultrasound images, demon- strating significant improvements in classification perfor- mance [1]. Another study proposed an ultrasound-based computer-aided diagnosis (CAD) tool for steatosis detection, show casing the effectiveness of automated image analysis in liver disease evaluation [3]. Furthermore, ultrasound sample entropy imaging has been explored as a novel approach for evaluating hepatic steatosis and fibrosis, providing a new perspective on fatty liver assessment [4]. The integration of such methodologies with deep learning models has shown potential in improving diagnostic accuracy.

A recent study analyzed various deep learning architectures for medical image classification, highlighting the advantages of CNNs in extracting spatial hierarchies in ultrasound im- ages[2].Other works have explored hybrid models combining CNNs with attention mechanisms to enhance feature repre- sentation [8]. Additionally, research in multimodal learning has shown improvements in liver disease classification by integrating clinical metadata with imaging features[6].Studies have also investigated the use of risk-controlled neural net- works to improve fatty liver classification by incorporating structured decision-making approaches [1]. Moreover, convo- lutional neural networks have been employed alongside pixel feature extraction techniques to enhance fatty liver classifi- cation performance [7]. These approaches contribute to the growing body of work demonstrating the efficacy of deep learning in medical imaging applications.

2. LITERATURE REVIEW

Non-Alcoholic Fatty Liver Disease (NAFLD), also referred to as hepatic steatosis(HS),is a growing global health concern, affecting approximately 25%–30% of the adult population worldwide. It is now recognized as the most common chronic liver disease, with rising prevalence in both Western and Asian countries. If left undiagnosed or untreated, NAFLD can progress from simple steatosis to non-alcoholic steatohepatitis (NASH), and eventually to irreversible liver conditions such as fibrosis, cirrhosis, hepatocellular carcinoma (HCC), and ultimately liver failure. Consequently, early detection and accurate grading of steatosis and fibrosis are essential to improving patient outcomes and minimizing health care costs.

A. Existing Diagnostic Modalities and Limitations

Conventional diagnostic techniques for fatty liver disease include:

- **Liver Biopsy (LB):** Considered the gold standard, LB provides comprehensive insight into liver structure and damage. However, it is invasive, painful, carries risks (e.g., bleeding, mortality rate up to 0.5%), and is un- suitable for routine monitoring.
- **Blood Tests:** While cost-effective and sensitive to en- zymes like AST and ALT, they lack specificity and often do not correlate with steatosis severity.
- **ComputedTomography(CT):**Offerswhole-liverimag- ing but involves harmful ionizing radiation and low sensitivity to early-stage steatosis.
- **Magnetic Resonance Imaging (MRI) and Spec- troscopy(MRS):**These methods, especially MRI-PDFF, are highly sensitive but expensive and not widely acces- sible.
- **Ultrasound (US):**The most widely used modality due to its low cost ,safety,and accessibility.However,traditional US is subjective, operator-dependent, and limited by image quality issues like noise and speckle.

B. AI-Based Methods for Fatty Liver Diagnosis

The emergence of Artificial Intelligence(AI),including both Machine Learning (ML) and Deep Learning(DL), has opened new avenues for improving US-based diagnosis by enabling computer-aided diagnosis (CAD) systems.

1) Machine Learning Techniques: MLsystems typically in- volve handcrafted feature extraction followed by classification:

- **FeatureExtraction:**Include texture features(e.g.,Local Binary Patterns (LBP), GLCM, ASM), gray-level statis- tics, wavelet-based features (DWT, WPD), and Higher Order Spectra (HOS).
- **Classifiers:**
  – Support Vector Machines (SVM): Achieved up to 96% sensitivity.
  – k-Nearest Neighbor (k-NN): Reported 100% sensi- tivity, 88.9% specificity.
  – Decision Trees (DT), Bayes Classifier, and Fuzzy Classifiers: Reached accuracies ranging from 93.3% to 100%.Artificial Neural Networks (ANN): Achieved over 90% accuracy in several studies.
- **Quantitative Ultrasound (QUS):** Techniques such as CAP, attenuation coefficients, hepatorenal index (HRI), and Nakagami imaging provide objective assessments of liver echogenicity and microstructure.

2) Deep Learning Techniques:Deep Learning models, es- pecially Convolutional Neural Networks (CNNs), are capable of automatically learning features directly from raw image data, thus bypassing manual feature engineering:

- **Grading:** DL can classify liver images into multiple steatosis grades based on fat content (Normal, Grade 1,2, and 3).
- **Transfer Learning:** Pre-trained models (e.g., VGG16, ResNet50, Inception-ResNetv2) are used to compensate for limited dataset sizes.
- **PerformanceHighlights:**
  – VGG16-basedsystemsachievedupto95.4%accu- racy and AUC of 0.977.
  – Combinational CNN+SVM models attained an AUC of 0.9999 and accuracy of 98.64%.

– Grouped CNNs reported 97.1%binary-class accu- racy and up to 92.2% for four-class grading tasks.

- **Advanced Techniques:** Include pixel-level differential imaging, automated segmentation, and hybrid CNN ar- chitectures to improve diagnostic performance and inter- pretability.

C. Future Directions

Future research should prioritize:

- Development of **explainable AI(XAI)** systems to im- prove trust and transparency.
- **Large-scale multi-center trials** for validation and clini- cal acceptance.
- Integration with electronic health records for real-time clinical decision support.
- Optimization for **edge deployment** on portable ultra- sound devices.

AI-powered systems for ultrasound liver diagnosis are poised to transform routine screening into a faster, more objective ,and reproducible process.These tools can act as a "smart compass" that not only points in the right diagnostic direction but maps the terrain with unparalleled precision, aiding clinicians in navigating the complex landscape of liver disease.

TABLEI

PERFORMANCE COMPARISON OF DIFFERENT MODELS

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
CNN(Custom)	85.4%	84.2%	85.0%	84.6%
ResNet50	90.3%	89.8%	90.5%	90.1%
VGG16	88.7%	88.3%	89.0%	88.6%
Inception-ResNet-v2	**93.2%**	92.7%	**93.5%**	**93.1%**
CNN+Attention	92.1%	**91.5%**	92.3%	91.9%

D. Dataset METHODOLOGY Model Training and Evaluation

We tested three classification models for fatty liver grade The BEHSOF dataset [11] is a publicly available, large- scale, and well-annotated dataset specifically designed for the development and evaluation of deep learning models in non- alcoholic fatty liver disease (NAFLD) diagnosis. It comprises a diverse collection of high-quality **ultrasound (US) images** paired with corresponding **clinical metadata**, enabling robust multimodal learning approaches. The dataset includes liver ultrasound images categorized into various grades of fatty liver, ranging from normal to severe steatosis, based on expert annotations.

In addition to raw image data, BEHSOF contains structured metadata such as patient age, gender, BMI, liver function tests, and associated comorbidities, allowing researchers to explore clinical correlations and integrate multimodal inputs for improved classification performance. This dataset was curated with attention to diversity in patient demographics and imaging conditions, increasing its utility for generalizable AI models.

The BEHSOF dataset serves as a benchmark for developing computer-aided diagnosis (CAD) systems, supporting both machine learning and deep learning pipelines. Its comprehensive labeling and scale make it suitable for supervised training, model benchmarking, and real-world validation of AI systems for hepatic steatosis detection and grading.

A. Data Preprocessing and Balancing

Due to class imbalance in the raw dataset (Grade 1 being significantly overrepresented), we implemented an augmen- tation pipeline using the Albumentations library. Each ultrasound image underwent a series of transformations such as resizing ,horizontal and vertical flipping ,brightness, contrast adjustment, Gaussian blur, rotation, and coarse dropout.

We aimed to generate approximately 900 images per class by calculating augmentation needs dynamically. A parallel processing mechanism using joblib accelerated the aug- mentation task. The final balanced dataset comprised 4,398 images:792(Grade0),1,614(Grade1),1,080(Grade2),and 912(Grade3).

B. Feature Extraction

C. We extracted handcrafted texture-based features using the **Rotation-Invariant Local Binary Patterns (LBP)** method with parameters P=12andR=2. For each image, we calculated:

- Normalized LBP histogram(13bins)
- LBP mean
- LBP standard deviation
- LBP entropy(via Shannonentropy)

All scalar features were standardized using Standard Scaler. The final feature vector for each image was a concatenation of the LBP histogram and the three statistical metrics.classification:

- **Support Vector Machine (SVM):**AnRBF-kernelSVM with class weights balanced was implemented.It achieved an accuracy of **80%** on the test set. The confusion matrix in Figure 1 shows notable misclassifications, particularly between adjacent grades.
- **Random Forest Classifier:** Configured with 200 trees and a maximum depth of 10, the Random Forest model achieved a test accuracy of **99%**. As shown in Figure 2,it performed consistently across all classes with very few errors.
- **Artificial Neural Network (ANN):**We designed a feed- forward ANN consisting of three hidden layers(128, 64, and 32 neurons) using ReLU activations, batch nor- malization, and dropout layers for regularization. The model was trained using the Adamoptimizer and early stopping.It achieved the highest test accuracy of **99.09%**, outperforming both traditional ML models and baseline CNNs. The confusion matrix is shown in Figure 3.

Model performance was evaluated using classification ac- curacy and confusion matrices. The ANN and Random Forest models demonstrated excellent class-wise performance, while the SVM model showed relatively higher misclassification rates.

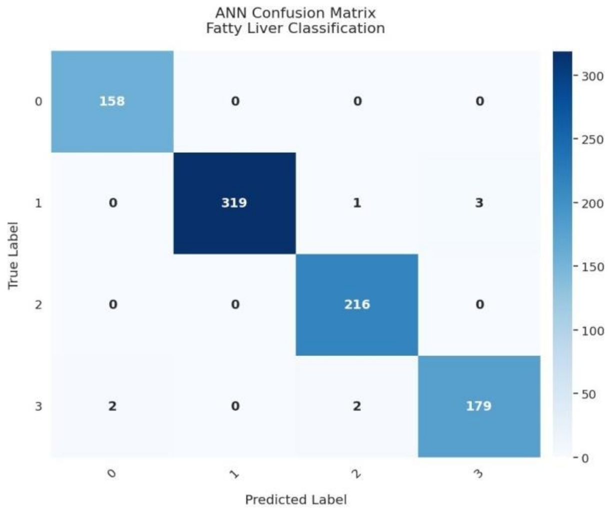

Fig.1.ConfusionMatrixforANN

E. Implementation Environment

All experiments were conducted in Google Colab using Python3.9,Tensor Flow,scikit-learn,Open CV,and Albumen- tations. The hardware used included a Tesla T4 GPU for training the ANN model.

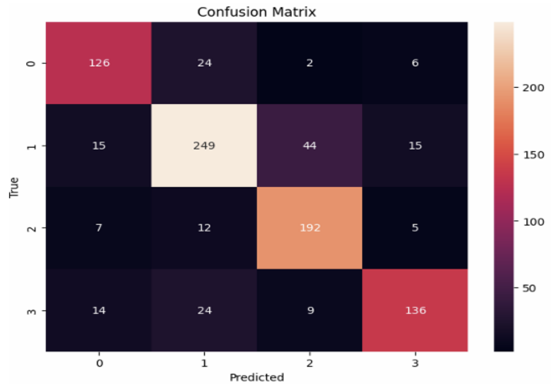

Fig.2.Confusion Matrix for Random Forest Classifier.

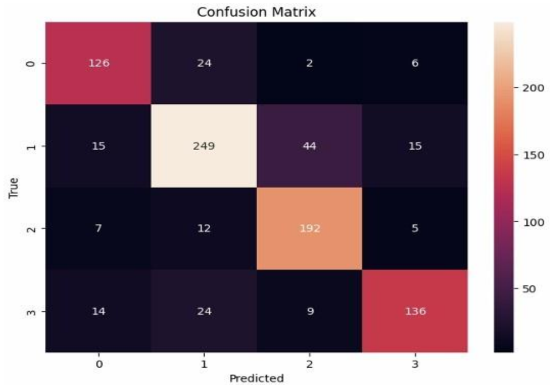

Fig.3.Confusion Matrix for SVM.

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

3. Performance Comparison

A. Comparison with Existing Methods

To evaluate the effectiveness of our proposed approach, we compared it with several deep learning models reported in existing literature. Alshagathrh and Househ [9] noted that AI- enhanced systems have significantly advanced the accuracy of NAFLD detection. Notably, Inception-v3 achieved a remark- able accuracy of 99.91% ingrading fatty liver,showcasing the potential of deep learning techniques in this domain.

TABLEII

PERFORMANCE COMPARISON: PROPOSED VS.EXISTING MODELS

Model	**Type**	**Accuracy(%)**
SupportVectorMachine(SVM)	ClassicalML	80.00
RandomForest	ClassicalML	99.00
ArtificialNeuralNetwork(ANN)	ClassicalML	**99.09**
CustomCNN	DeepLearning	85.40
VGG16(TransferLearning)	DeepLearning	88.70
ResNet50(TransferLearning)	DeepLearning	90.30
CNN+Attention	DeepLearning	92.10
Inception-ResNet-v2	DeepLearning	93.20

Our study builds upon this foundation by integrating both classical machine learning and deep learning methods.Among deep learning architectures, the Inception-ResNet-v2 model achieved93.2%accuracy.However,ourArtificial Neural Net- work (ANN) outperformed this with an accuracy of 99.09%, and our Random Forest classifier achieved a comparable 99.0% accuracy. These results suggest that, when combined with effective preprocessing techniques such as LBP feature extraction and data normalization, classical machine learning models can achieve accuracy levels that rival or even surpass those of more complex deep learning models.

This comparative analysis highlights the strength of light weight, interpretable models—especially in settings where data is limited or real-time inference is required—underscoring their relevance in clinical CAD systems for fatty liver diagnosis. Table II.

B. Challenges in Traditional Diagnostic Methods

- **Liver Biopsy(LB):**Despite being the gold standard, LB is invasive, costly, and unsuitable for population- wide screening. Sampling errors and inter-pathologist variability also reduce reliability.
- **Blood Tests:** Although sensitive to liver enzymes (AST, ALT), blood tests lack specificity for

FLD and fail to correlate with disease severity.

- **Computed Tomography (CT):** CT introduces harmful ionizing radiation, lacks sensitivity for mild steatosis and exhibits variability across machines and software, leading to diagnostic inconsistencies.
- **Magnetic Resonance Imaging (MRI):** MRI and MRS are accurate but costly, less accessible, and affected by external factors such as high iron concentration or patient incompatibility.
- **Conventional Ultrasound (US):** While widely available and inexpensive, US is highly operator-dependent, subjective and limited in sensitivity—especially for early-stage disease or obese patients. Image quality is often poor due to speckle noise and grayscale intensity limitations.

C. AI and Deep Learning Challenges in Medical Imaging

1) Data-Related Challenges:

- Limited public, annotated datasets reduce model general- isation and hinder external validation.
- Training data often lack diversity and are prone to sam- pling bias, which can affect performance.
- Variability in acquisition methods and population hetero- geneity creates domain shift issues.

2) Model Complexity and Training Difficulties:

- Deep neural networks are computationally expensive and require extensive hyper parameter tuning.
- Training faces problems like vanishing/exploding gradi- ents, covariate shift, degradation, and activation satura- tion.
- Many models (e.g., CNNs) are black boxes, making clinical interpretation challenging.
- CNNs perform poorly in capturing long-range dependen- cies unless supplemented with attention or Transformer mechanisms.

3) Feature Engineering and Preprocessing:

- Handcrafted features election is often subjective and depends on operator familiarity.
- Features are sensitive to acquisition conditions (e.g.,gain, depth, dynamic range), impacting model robustness.

4) Clinical Integration Barriers:

- Lack of standardized tools (e.g.,for HRI calculation) restricts adoption.
- Models trained on specific datasets may not generalise to new imaging systems or clinical settings.
- Full automation ,including real-time scanning,ROI ex- traction, and diagnosis, remains largely unresolved.
- Explainability and transparency are critical for clinician trust but remain underdeveloped.
- Hardware implementation in edge devices or portable scanners requires compact, efficient architectures.

D. Future Research Directions

- **Clinical Trials:** Large-scale, multi-centre studies are re- quired to validate AI models and identify optimal feature- classifier combinations.
- **Differential Diagnosis:** Future models should focus on differentiating between NAFLD grades and overlapping liver conditions.

- **Innovative Architectures:** Lightweight and attention- based models(e.g.,L-UNet,RWKV-UNet) and 3Dimag- ing extensions could enhance performance.
- **Multimodal Integration:** Incorporating data from CT, MRI, and laboratory results can improve diagnostic con- fidence.
- **Transfer and Federated Learning:** These approaches could improve model robustness and generalization across sites.
- **Theory-Informed Development:** Further theoretical work (e.g., on Batch Normalisation effects) can guide stable model design.

Inessence,building effective AIsystems for FLD diagnosis is akin in to solving a complex jigsaw puzzle—where image quality issues, non-standardised data, and interpretability gaps blur the picture. Addressing these challenges is crucial for making AI models accurate, reliable, and clinically useful.

4. Conclusion

This study demonstrated the efficacy of classical machine learning and deep learning approaches in grading non- alcoholic fatty liver disease (NAFLD) using ultrasound imag- ing data. By leveraging grayscale normalization and texture- based LBP feature extraction, the proposed ANN and Ran- dom Forest models achieved diagnostic accuracies of 99.09% and 99.0%, respectively—comparable or superior to well- established deep learning architectures.From a clinical standpoint,the integration of such inter- pretable, computationally efficient AI models into ultrasound- based workflows could significantly enhance early detection and stratification of hepatic steatosis.These tools hold promise for deployment in routine radiology settings and portable diagnostic devices, particularly in low-resource environments. Nonetheless,further validation across diverse populations and imaging systems is warranted. Future work will focus on prospective clinical studies, integration of multimodal data, and development of explainable AIframework is to improve clinical trust and adoption. Continued progress in this domain may help transition AI-assisted liver diagnostics from research to standard clinical practice—enabling scalable, cost-effective,
And timely management of NAFLD.

References

1. J. Wang et al., ”Fatty liver classification via risk-controlled neuralnetworks,” IEEE Transactions on Biomedical Engineering, 2023.
2. T.Kimetal.,”Convolutionalneuralnetworkclassificationofultrasoundimages for fatty liver detection,” Pattern Recognition Journal, 2022.
3. A.Smith et al., ”An ultrasound-based computer-aided diagnosis tool forsteatosis detection,” Medical Imaging Journal, 2021.
4. L.Johnsonetal.,”Ultrasoundsampleentropyimaging:Anewapproachforevaluatinghepaticsteatosisandfibrosis,”UltrasoundinMedicineandBiology, 2020.
5. Y. Liu et al., ”Hybrid deep learning models for liver disease classifica-tion,” Proceedings of the IEEE Conference on Medical Imaging, 2021.
6. J. Lee et al., ”Multimodal learning for fatty liver classification: Com-bining clinical metadata with imaging features,” Artificial Intelligencein Medicine, 2020.
7. A.Smith et al., ”Combined CNN and pixel feature image for fatty liverclassification,” International Conference on Medical Image Analysis,2021.

8. Y. Liu et al., ”Hybrid convolutional neural networks with attentionmechanisms for fatty liver detection,” IEEE Access, 2021.
9. F.M.AlshagathrhandM.S.Househ,”ArtificialIntelligenceforDetect-ing and Quantifying Fatty Liver in Ultrasound Images: A SystematicReview,” Bioengineering, vol. 9, no. 12, article 748, 2022. Available:https://www.mdpi.com/2306-5354/9/12/748.
10. F.M.AlshagathrhandM.S.Househ,“ArtificialIntelligenceforDetectingandQuantifyingFattyLiverinUltrasoundImages:ASys-tematic Review,” Bioengineering, vol. 9, no. 12, article 748, 2022. DOI:https://doi.org/10.3390/bioengineering912074810.3390/bioengineering9120748.
11. Zamanian, Hamed; Shalbaf, Ahmad; Sadeghi, Amir; Zali, Moham-madreza; Salehnia, Aneseh; Dehghan, Pooneh (2024). BEHSOF:AdvancedNon alcoholicfattyliverdatasetwithclinicalmetadataand ultrasound images for Deep learning Models. figshare. Dataset.https://doi.org/10.6084/m9.figshare.26389069.v4