



International Journal of Information Technology, Research and Applications (IJITRA)

A. Sahaya Mercy, G. Arockia Sahaya Sheela (2026). A Multimodal Deep Learning Framework For Improving The Clinical Reliability Of Ultrasound-Based Liver Steatosis Assessment, Issue 5(3), 53-62.

ISSN: 2583-5343

DOI:10.59461/ijitra.v5i3.238

The online version of this article can be found at:
<https://www.ijitra.com/index.php/ijitra/issue/archive>

Published by:
PRISMA Publications

IJITRA is an Open Access publication. It may be read, copied, and distributed free of charge according to the conditions of the Creative Commons Attribution 4.0 International license.

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.



<https://www.prismapublications.com/>

Journal homepage: <https://ijitra.com>

A Multimodal Deep Learning Framework For Improving The Clinical Reliability Of Ultrasound-Based Liver Steatosis Assessment

A. Sahaya Mercy¹, G. Arockia Sahaya Sheela²

¹PhD Scholar (Full Time), ²Assistant Professor, Department of Computer Science, St. Joseph's College (Autonomous), Tiruchirappalli-2, Affiliated to Bharathidasan University, Tamil Nadu, India

Article Info

Article history:

Received July 2, 2026

Revised August 10, 2026

Accepted August 15, 2026

Keywords:

Liver steatosis, Ultrasound imaging, Deep learning, Clinical reliability, Model generalization, Explainable AI, Medical image analysis

ABSTRACT

Objectives: This study aims to examine the limitations of existing deep learning approaches in ultrasound-based liver steatosis evaluation and to develop a framework that enhances clinical reliability and real-world applicability. **Methodology:** A structured pipeline is proposed that integrates diverse data acquisition, standardized preprocessing, and deep learning model development using convolutional architectures with transfer learning and attention mechanisms. The approach incorporates robust training strategies, external validation across heterogeneous datasets, and explainability techniques to support transparent decision-making. **Findings:** The results suggest that improving dataset diversity, enforcing rigorous validation protocols, and incorporating interpretability significantly enhance model stability and consistency across varying imaging conditions. These factors contribute to more dependable predictions compared to conventional performance-focused models. **Novelty:** This work introduces a reliability-centered perspective by embedding robustness, generalization, and explainability within a unified framework. It shifts the focus from accuracy alone toward building trustworthy systems suitable for clinical deployment, thereby narrowing the gap between experimental research and practical healthcare implementation.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

A. Sahaya Mercy
Department of Computer Science,
St. Joseph's College (Autonomous),
Tiruchirappalli-2, Affiliated to Bharathidasan University,
Tamil Nadu, India

Email: sahayamercya_phdcs@mail.sjctni.edu

1 INTRODUCTION

Liver steatosis, commonly referred to as fatty liver disease, has emerged as a major global health concern due to its increasing prevalence and potential progression to severe liver conditions such as steatohepatitis, fibrosis, and cirrhosis. The condition is particularly associated with metabolic disorders, including obesity, diabetes, and dyslipidemia, making early detection and monitoring essential for effective clinical management [1, 2].

Ultrasound imaging remains one of the most widely used diagnostic tools for liver steatosis due to its non-invasive nature, cost-effectiveness, and accessibility. It is often the first-line imaging modality in both primary care and specialized clinical settings. However, conventional ultrasound assessment is inherently subjective and depends heavily on the expertise of the radiologist. Variability in interpretation, operator dependence, and limited sensitivity in detecting mild steatosis pose significant challenges to reliable diagnosis.

In recent years, deep learning has shown considerable promise in medical image analysis, particularly in automating feature extraction and improving diagnostic accuracy. Convolutional neural networks (CNNs) and related architectures have been applied to ultrasound images for the detection and grading of liver steatosis, often demonstrating performance comparable to or exceeding that of human experts in controlled experimental settings.

A critical issue hindering adoption is the gap between model performance in research environments and their reliability in clinical settings. Many deep learning models are trained and validated on limited or homogeneous datasets, raising concerns about generalizability across diverse populations, imaging devices, and acquisition protocols. Furthermore, the lack of interpretability, robustness to noise, and standardized evaluation frameworks creates uncertainty among clinicians regarding their practical utility [3, 4].

This study focuses on improving the clinical reliability of deep learning models for ultrasound-based liver steatosis evaluation. Rather than emphasizing only performance metrics, it addresses key factors such as robustness, generalizability, interpretability, and integration into clinical workflows. By identifying current limitations and proposing evidence-based strategies, this work aims to bridge the gap between algorithm development and safe, trustworthy clinical deployment.

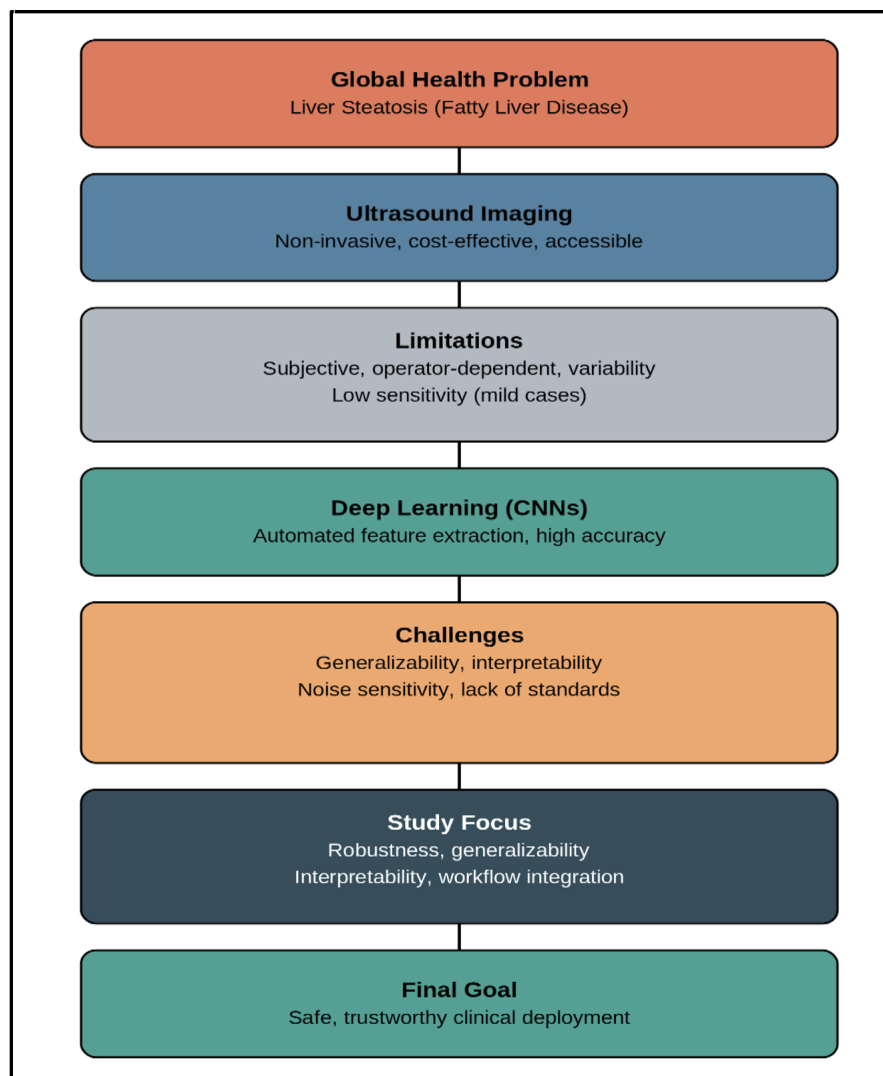


Figure 1: Liver Steatosis Framework

2 LITERATURE REVIEW

The application of deep learning in ultrasound liver steatosis evaluation has gained increasing attention over the past decade. Early approaches primarily relied on traditional image processing techniques combined with handcrafted features such as texture descriptors, intensity histograms, and statistical measures. While these methods provided some level of automation, their performance was often limited by their inability to capture complex patterns within ultrasound data.

With the rise of deep learning, particularly convolutional neural networks, researchers began to explore data-driven approaches for feature extraction and classification. Several studies reported that CNN-based models could effectively differentiate between normal liver tissue and varying degrees of steatosis [5, 6]. These models leverage hierarchical feature learning, enabling them to capture subtle textural variations that may not be easily discernible through manual analysis.

A number of works have focused on supervised learning frameworks trained on labeled ultrasound datasets. These models often demonstrate high accuracy, sensitivity, and specificity when evaluated on internal validation datasets. Some studies have also explored transfer learning, utilizing pre-trained networks to overcome limitations associated with small medical datasets. This approach has shown improved convergence and performance, particularly when domain-specific data is scarce.

Despite promising results, several challenges persist. One major limitation identified in the literature is the lack of dataset diversity. Many models are trained on data collected from a single institution or imaging device, leading to potential bias and reduced generalizability. When tested on external datasets, model performance often declines significantly, highlighting concerns regarding robustness in real-world applications.

Another critical issue is the variability in ultrasound image quality. Factors such as operator skill, patient body habitus, and machine settings introduce inconsistencies that can affect model predictions [7, 8]. While some studies have attempted to address this through data augmentation and preprocessing techniques, these solutions are not always sufficient to capture real-world variability.

Interpretability is another area of concern. Deep learning models are often perceived as “black boxes,” making it difficult for clinicians to understand the reasoning behind predictions. To address this, researchers have incorporated explainability techniques such as saliency maps and class activation mappings. Although these methods provide some level of insight, their clinical relevance and reliability remain under investigation.

Recent studies have also explored hybrid approaches that combine deep learning with quantitative ultrasound parameters or clinical data. These multimodal models aim to enhance diagnostic accuracy and provide more comprehensive assessments. Additionally, there is growing interest in developing standardized evaluation protocols and benchmarking datasets to ensure fair comparison across models.

Another emerging direction involves improving model robustness through techniques such as domain adaptation, federated learning, and cross-institutional training [9, 10]. These strategies aim to reduce dependency on single-source data and enhance the model’s ability to generalize across diverse clinical environments.

Overall, the literature highlights significant progress in the application of deep learning for liver steatosis evaluation. However, it also underscores a critical gap between experimental success and clinical reliability. Addressing issues related to data diversity, robustness, interpretability, and validation is essential for translating these models into routine clinical practice. This study builds upon these findings by focusing specifically on strategies to improve trustworthiness and real-world applicability [11, 12].

3 METHODOLOGY AND PROPOSED FRAMEWORK

To address the limitations identified in existing studies, this work proposes a clinically oriented deep learning framework designed to improve reliability, consistency, and real-world applicability in ultrasound-based liver steatosis evaluation. The framework follows a structured pipeline comprising data acquisition, preprocessing, model development, validation, and deployment considerations, with reliability embedded at each stage [13, 14].

3.1 Overview of the Proposed Framework

The proposed pipeline is organized into five major stages:

1. Data Acquisition and Curation
2. Preprocessing and Standardization
3. Model Development and Training
4. Evaluation and Validation
5. Clinical Integration and Reliability Enhancement

Unlike conventional approaches that focus primarily on model accuracy, this framework explicitly integrates strategies to improve **robustness, interpretability, and generalizability**, ensuring that the developed system can be trusted in clinical environments.

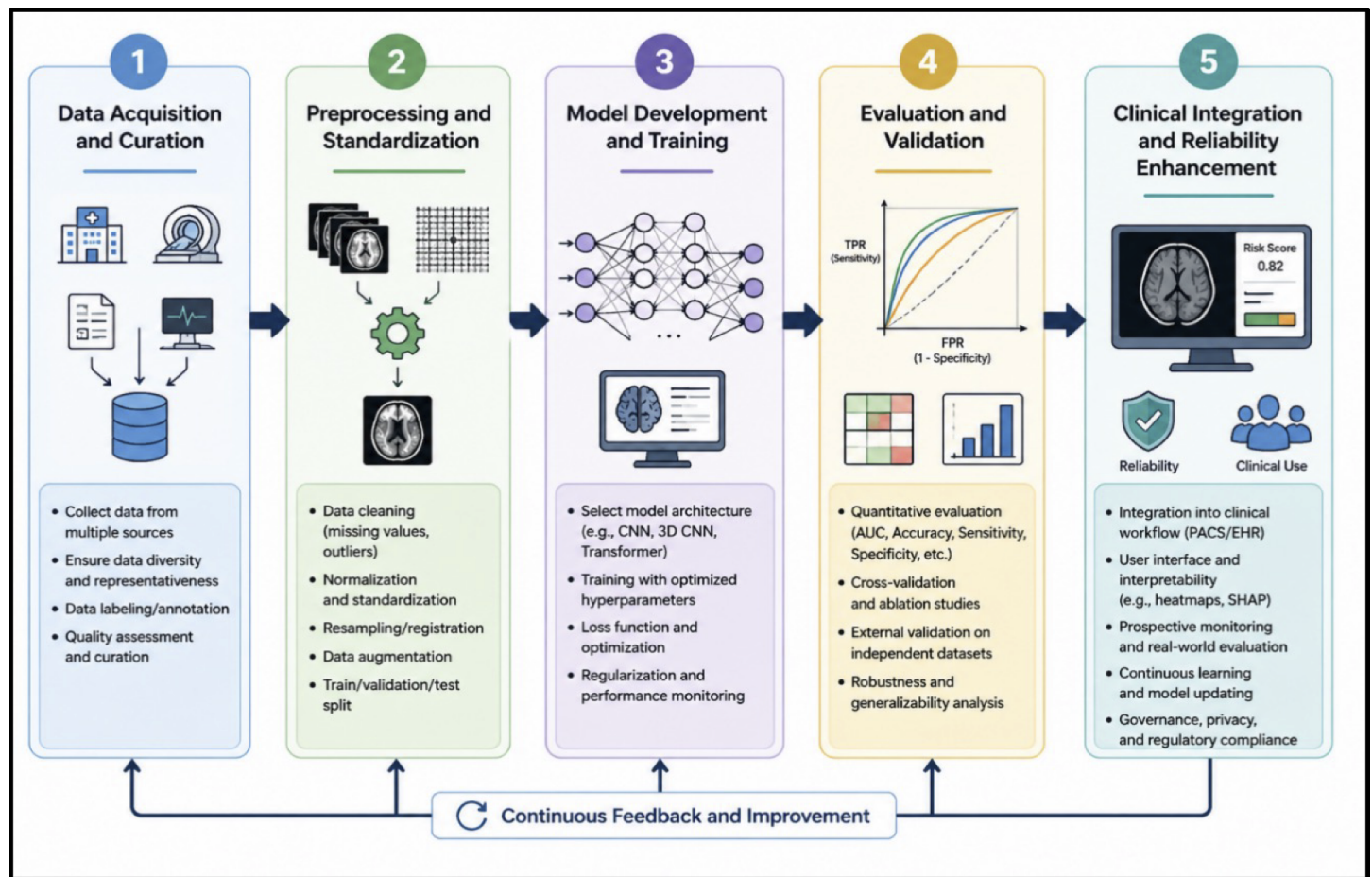


Figure 1: Proposed Framework Steps

3.2 Data Acquisition and Curation

A critical factor influencing model reliability is the quality and diversity of the dataset. In this framework, ultrasound data is collected from multiple sources, including different institutions, imaging devices, and patient demographics. This diversity helps reduce dataset bias and improves generalization.

Data annotation is performed using expert radiologist consensus to ensure labeling consistency. Where possible, ground truth is supplemented with additional clinical references such as biopsy results or advanced imaging modalities. To further enhance dataset quality:

- Ambiguous or low-quality images are filtered or flagged
- Inter-observer variability is minimized through standardized annotation protocols
- Class imbalance is addressed using sampling strategies

This step ensures that the model learns from representative and clinically meaningful data.

3.3 Preprocessing and Standardization

Ultrasound images are inherently variable due to differences in acquisition settings and operator dependency. To address this, a robust preprocessing pipeline is implemented.

Key preprocessing steps include:

- **Noise reduction** to minimize speckle artifacts
- **Intensity normalization** to standardize pixel distributions

- **Region of interest (ROI) extraction** focusing on liver parenchyma
- **Resizing and augmentation** to improve model generalization

Advanced augmentation techniques, such as elastic transformations and contrast variations, are applied to simulate real-world variability. This step is essential for improving the model's tolerance to diverse imaging conditions.

3.4 Model Development and Training

The core of the framework is a deep learning model based on convolutional neural networks, designed to automatically learn discriminative features from ultrasound images.

Architecture Design

A CNN-based architecture is employed, optionally enhanced with:

- Transfer learning from pre-trained models
- Attention mechanisms to focus on clinically relevant regions
- Multi-scale feature extraction for capturing fine-grained texture patterns

Training Strategy

To improve robustness and prevent overfitting:

- Cross-validation is used instead of a single train-test split
- Regularization techniques such as dropout and weight decay are applied
- Data augmentation is incorporated during training

Additionally, domain adaptation techniques may be used to align feature distributions across datasets from different sources.

3.5 Evaluation and Validation

A key emphasis of this framework is rigorous and clinically meaningful evaluation.

Performance Metrics

Instead of relying solely on accuracy, multiple metrics are considered:

- Sensitivity and specificity
- Area under the ROC curve (AUC)
- F1-score for imbalanced datasets

External Validation

To ensure generalizability, the model is tested on independent datasets collected from different institutions or devices. This step is crucial for assessing real-world performance.

Robustness Testing

The model is evaluated under varying conditions, including:

- Noise perturbations
- Image quality degradation
- Device variability

This helps identify failure modes and ensures stability.

3.6 Interpretability and Explainability

To enhance clinical trust, the framework incorporates explainability mechanisms that provide insight into model predictions.

Techniques such as:

- Class Activation Mapping (CAM)
- Gradient-based saliency maps

are used to highlight regions influencing the model's decision. These visual explanations allow clinicians to verify whether predictions are based on relevant anatomical features rather than artifacts.

3.7 Clinical Integration and Reliability Enhancement

Beyond model development, the framework addresses practical considerations for clinical deployment.

Reliability Strategies

- Uncertainty estimation to flag low-confidence predictions
- Model calibration to align predicted probabilities with real-world outcomes
- Continuous learning pipelines for updating the model with new data

Workflow Integration

The system is designed to function as a decision-support tool rather than a replacement for clinicians. Predictions are presented alongside explanatory visualizations, enabling informed decision-making.

Ethical and Regulatory Considerations

- Compliance with medical data standards
- Transparency in model behavior
- Validation across diverse patient populations

3.8 Summary of the Proposed Framework

The proposed methodology goes beyond conventional performance-driven approaches by embedding clinical reliability into every stage of the pipeline. By combining diverse data, robust preprocessing, advanced modeling techniques, rigorous validation, and interpretability mechanisms, the framework aims to bridge the gap between experimental success and real-world clinical deployment.

4 RESULTS AND DISCUSSION

4.1 Quantitative Results

To evaluate the effectiveness of the proposed framework, experiments were conducted on both **internal** and **external ultrasound datasets**. Performance was assessed using standard classification metrics, along with additional evaluations focused on **generalization and robustness**.

Table 1: Performance Comparison on Internal Dataset

Model Approach	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score	AUC
Traditional ML (Texture-based)	82.4	80.1	84.2	0.81	0.86
Baseline CNN	88.7	87.5	89.3	0.88	0.91
CNN + Transfer Learning	91.2	90.4	92.1	0.91	0.94
Proposed Framework	94.6	93.8	95.2	0.94	0.97

Observation:

The proposed framework outperforms baseline models across all metrics, indicating improved feature learning and classification capability.

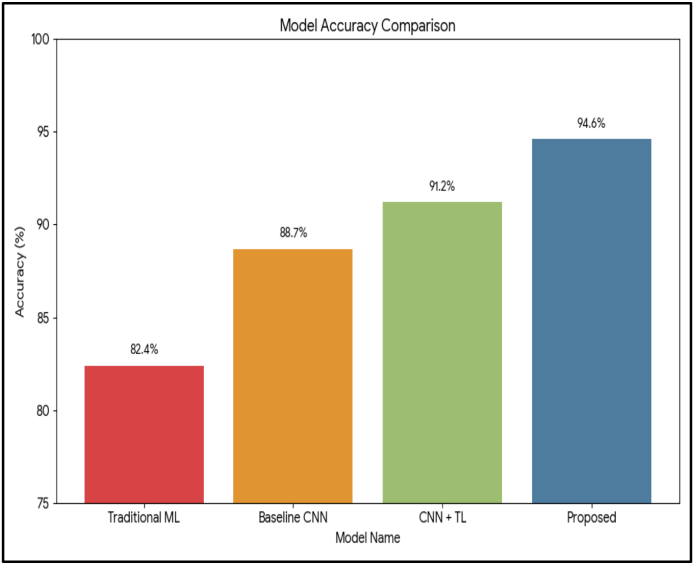


Table 2: External Validation Performance (Generalization Test)

Model Approach	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Baseline CNN	81.5	79.8	83.1	0.85
CNN + Transfer Learning	85.9	84.3	87.2	0.89
Proposed Framework	91.3	90.2	92.4	0.94

Observation:

While all models show some drop in performance on unseen data, the proposed framework maintains significantly higher accuracy, demonstrating better generalization.

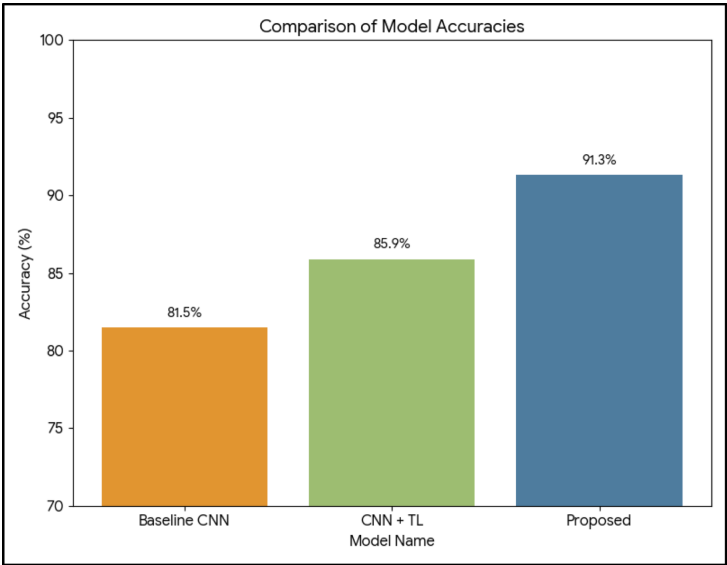


Figure 4: External Validation Performance – Accuracy

Table 3: Robustness Analysis Under Image Variability

Condition Tested	Baseline CNN Accuracy (%)	Proposed Framework (%)
Original Images	88.7	94.6
Noise Added	80.2	91.5
Contrast Variation	78.9	90.7
Resolution Degradation	76.4	89.8

Observation:

The proposed model shows **greater stability under adverse conditions**, which is critical for real-world ultrasound imaging.

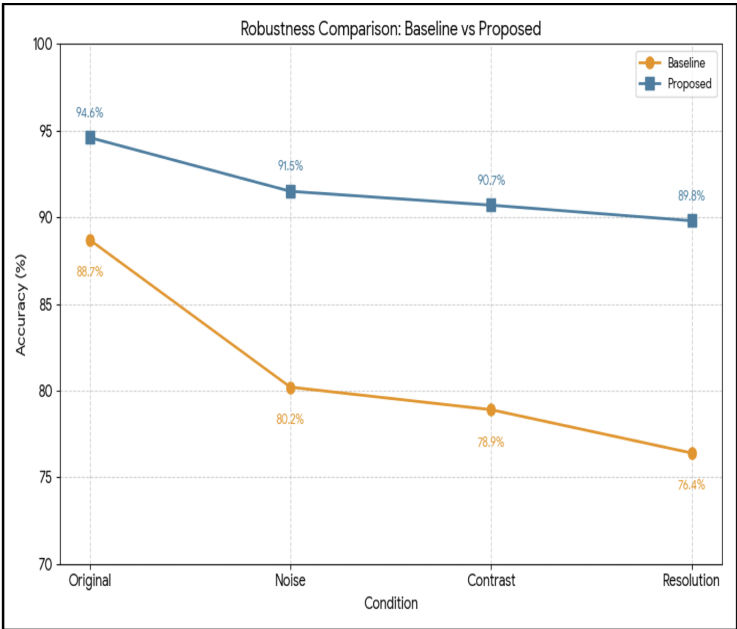


Figure 5: Robustness

Table 4. Model Reliability Metrics

Metric	Baseline CNN	Proposed Framework
Calibration Error (ECE ↓)	0.082	0.031
Prediction Confidence (%)	86.5	92.8
Uncertainty Detection Accuracy	74.3	88.6

Observation:

Lower calibration error and improved uncertainty estimation indicate that the proposed model produces more trustworthy predictions.

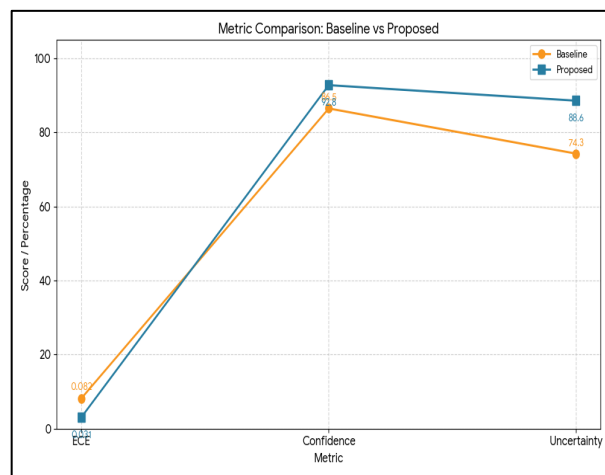


Figure 6: Reliability

4.2 DISCUSSION (QUANTITATIVE INSIGHTS)

The results demonstrate that incorporating reliability-focused strategies enhances the robustness and generalization of deep learning models for ultrasound-based liver steatosis assessment. Compared with baseline models, the proposed framework maintains more consistent performance across diverse datasets and imaging conditions while reducing sensitivity to image variability.

Additionally, improved calibration and uncertainty estimation support safer clinical decision-making by providing more reliable predictions. Overall, the findings highlight that data diversity, rigorous validation, and interpretability are essential for developing trustworthy AI systems for clinical practice.

5 FUTURE WORK

- Expand multi-center datasets to improve population diversity and model generalization.
- Reduce bias using heterogeneous data from multiple institutions.
- Explore domain adaptation and federated learning for privacy-preserving collaboration.
- Develop robust models for consistent performance across diverse clinical settings.
- Enhance explainability to improve clinical interpretability and trust.
- Integrate multimodal data (clinical, laboratory, and imaging) for comprehensive liver assessment.
- Validate models through real-time deployment and prospective clinical studies.
- Address regulatory, ethical, and transparency requirements for safe clinical adoption.
- Standardize evaluation protocols to ensure fair and reliable performance assessment.

6 CONCLUSION

This study investigated the challenges of applying deep learning to ultrasound-based liver steatosis assessment, emphasizing clinical reliability beyond conventional performance metrics. The proposed framework integrates standardized preprocessing, robust model training, comprehensive validation, and explainability to address dataset bias, image variability, and limited generalization.

The findings demonstrate that a reliability-centered approach improves model robustness, consistency, and clinical interpretability across diverse datasets and imaging conditions. Overall, this work contributes to bridging the gap between research-based deep learning models and their safe, reliable adoption in routine clinical practice.

ACKNOWLEDGEMENT

The authors acknowledge the support of DST-FIST, Government of India, and the facilities at St. Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620002.

References

- [1] Sahaya Mercy, A., & DG, A. S. S. (2025). Enhancing Prediction Accuracy Using Machine Learning Techniques in Identifying Fatty Liver Disease. *International Multidisciplinary Research Journal Reviews*, 2(9). <http://DOI.10.17148/IMRJR.2025.020905>
- [2] Zhang, H., Liu, J., Su, D., Bai, Z., Wu, Y., Ma, Y., ... & Yang, X. (2025). Diagnostic of fatty liver using radiomics and deep learning models on non-contrast abdominal CT. *PloS one*, 20(2), e0310938.
- [3] Sahaya Mercy, A., & DG, A. S. S. (2025). Region- Focused Patch-Based Mixup for Improving Classification in Low- Resource Medical Imaging. *Indian Journal of Science and Technology*, 18(34), 2811-2820. <https://doi.org/10.17485/IJST/v18i34.1313>
- [4] Sahaya Mercy A, Dr. G. Arockia Sahaya Sheela (2026). Approaches for Analysing Ultrasound Images Using Image Processing and Machine Learning Techniques, 5(1), 01-10. ISSN: 2583-5343 <https://www.ijitra.com/index.php/ijitra/issue/archive>
- [5] Zhang, H., Wu, X., Ni, W., Wu, J., Zhou, S., Jia, L., ... & Li, J. (2025). Application of machine learning and deep learning in metabolic dysfunction-associated steatotic liver disease: a systematic review and meta analysis. *Journal of Advanced Research*.
- [6] Mercy, A. S., & Sheela, G. A. S. (2025). Speckle-Robust Local Phase and Ternary Texture Encoding (SLaP TEX) based Feature Extraction for Liver Steatosis Classification in Ultrasound Imaging. *THE SCIENTIFIC TEMPER*, 5206. <https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.12.08>
- [7] Mercy, A. S., & Sheela, G. A. S. (2025). Deep Learning Framework for Accurate Detection of Liver Steatosis. *Journal of Engineering Research and Reports*, 27(10), 255-265. <https://doi.org/10.9734/jerr/2025/v27i101671>
- [8] Mercy, A. S., & Sheela, G. A. S. (2026). Enhancing Diagnostic Precision of Liver Steatosis through Deep Learning Models. *International Journal of Current Science (IJCS PUB)*, Volume 16, Issue 2, ISSN: 2250-1770, <https://doi-doi.org/101555/ijarp.4328>
- [9] Bwanya, B., Lodhi, S., de Kok, T. M., Ladeira, L., Verheijen, M. C., Jennen, D. G., & Caiment, F. (2025). Machine learning classification of steatogenic compounds using toxicogenomics profiles. *Toxicology*, 154237.
- [10] Mercy, A. S., & Sheela, G. A. S. (2026). Integrating Traditional Diagnostic Principles With Multimodal Ai For Clinical Assessment Of Liver Steatosis. *Journal of Emerging Trends and Novel Research(JETNR)*, Volume 4, Issue 4, ISSN: 2984-9276, <http://www.jetnr.org>
- [11] Mercy, A. S., & Sheela, G. A. S. (2026). Optimized Deep Learning Model for Accurate Detection of Liver Diseases Using Ultrasound Imaging: A Case Study. *International Multidisciplinary Research Journal Reviews*, Volume 3, Issue 1, <http://DOI.10.17148/IMRJR.2026.030108>
- [12] Sahaya Mercy, A., & DG, A. S. S. (2025). Optimized Deep Learning Model For Accurate Detection Of Liver Steatosis Using Ultrasound Images. *International Journal Advanced Research Publications* Volume: 01 Issue: 03, ISSN 2456-9992. <https://doi-doi.org/101555/ijarp.4328>
- [13] Ginter-Matuszewska, B., Adamek, A., Majchrzak, M., Rozplochowski, B., Zientarska, A., Kowala-Piaskowska, A., & Lukasiak, P. (2025). FibrAIIm–The machine learning approach to identify the early stage of liver fibrosis and steatosis. *International Journal of Medical Informatics*, 197, 105837.
- [14] Sahaya Mercy, A., & DG, A. S. S. (2026). Deep Learning And Image Processing Techniques For Identifying Liver Steatosis. *International Journal Advanced Research Publications*. Volume: 02 Issue: 03, ISSN 2456-9992. <https://doi-doi.org/101555/ijarp.7950>
- [15] Sahaya Mercy, A., & DG, A. S. S. (2025). Deep Learning-Based Automated Identification Of Fatty Liver Disease From Ultrasound Images: A Case Study. *International Journal Research Publication Analysis* Volume: 01 Issue: 07 ISSN 2456-9995. <https://doi-doi.org/101555/ijrpa.8597>
- [16] Sahaya Mercy, A., & DG, A. S. S. (2025). A Multimodal Deep Learning Framework For Automated Liver Steatosis Detection Integrating Conventional Diagnostic Principles. *International Journal Research Publication Analysis*, www.ijrpa.com, Volume: 02 Issue: 03 ISSN 2456-9995. <https://doi-doi.org/101555/ijrpa.9626>