
A MULTIMODAL DEEP LEARNING FRAMEWORK FOR AUTOMATED LIVER STEATOSIS DETECTION INTEGRATING CONVENTIONAL DIAGNOSTIC PRINCIPLES

A. Sahaya Mercy*¹ Dr. G. Arockia Sahaya Sheela²

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

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

Article Received: 13 February 2026

*Corresponding Author: A. Sahaya Mercy

Article Revised: 04 March 2026

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

Published on: 24 March 2026

DOI: <https://doi-doi.org/101555/ijrpa.9626>

ABSTRACT

Objectives: This study aims to develop a multimodal deep learning framework that integrates liver ultrasound imaging with traditional health indicators for improved detection of liver Steatosis. The objective is to enhance diagnostic accuracy, strengthen model interpretability, and align artificial intelligence-based predictions with clinically meaningful factors rather than relying solely on imaging data. **Methods:** A hybrid deep learning model combining Convolutional Neural Networks and Transformer architecture was designed to analyze ultrasound images, enabling the system to capture both fine local details and broader global patterns in liver tissue. Alongside imaging data, traditional parameters related to diet, lifestyle, and body constitution were numerically encoded and incorporated into the model through a multimodal fusion layer. The entire framework was trained in an end-to-end manner using cross-validation to ensure reliability, stability, and generalizability across different patient samples. **Findings:** The experimental results demonstrate that the proposed multimodal framework performs better than single-modality models that rely only on ultrasound images. It achieved superior performance in terms of accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). The inclusion of traditional diagnostic principles not only improved prediction outcomes but also made the model more interpretable and clinically meaningful. By bridging ancient diagnostic wisdom with modern artificial

intelligence, this study presents a novel, holistic approach to liver health assessment and offers a promising tool for early, accurate, and clinically relevant detection of liver Steatosis.

Conclusion: This study demonstrates the value of integrating traditional diagnostic principles with modern deep learning for liver Steatosis detection. The proposed multimodal framework enhances diagnostic accuracy, interpretability, and clinical relevance. By combining ancient medical wisdom with artificial intelligence, this approach offers a practical and holistic pathway for early and reliable assessment of liver health.

KEYWORDS: Liver Steatosis, Multimodal Deep Learning, Traditional Diagnostics, Ultrasound Imaging, IKS.

1. INTRODUCTION

1.1 Background of Liver Steatosis

Liver Steatosis, commonly known as fatty liver disease, represents the abnormal accumulation of lipids in liver cells. It is increasingly recognized as a global health concern, often linked to obesity, poor dietary habits, and metabolic disorders. Early stages are typically asymptomatic, Zeyauddin, (2025) but the condition can progress to severe complications such as Non-Alcoholic Steatohepatitis, fibrosis, or cirrhosis. Traditional diagnostic methods, such as ultrasound imaging, liver biopsy, and biochemical analysis, are valuable but may be invasive, subjective, or costly. This has prompted the exploration of computational frameworks to support timely and accurate detection.

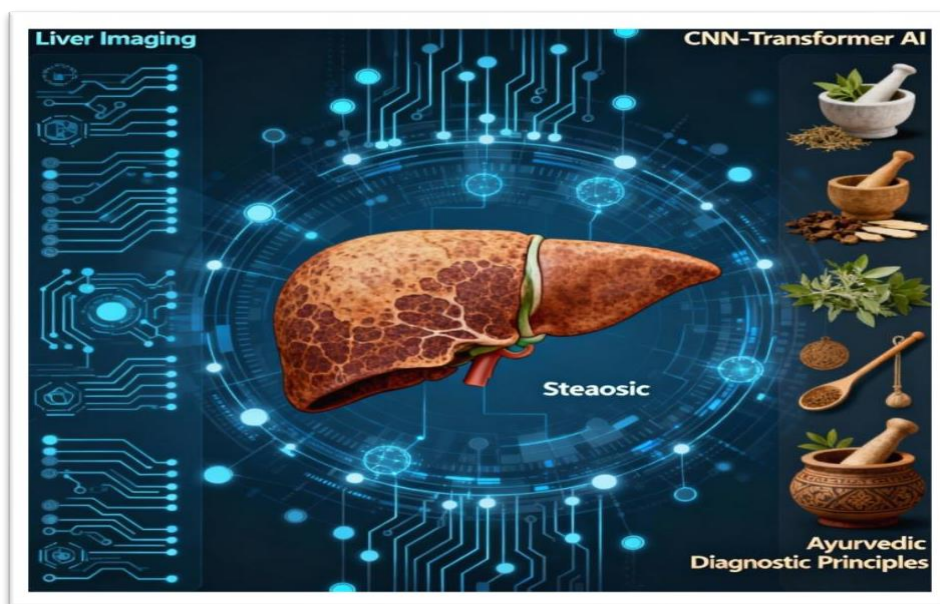


Fig.1 Liver Imaging With CNN- Transformer AI.

1.2 Importance of Early Detection

Detecting Liver Steatosis at an early stage can significantly Rajeev et al. (2025) improve patient outcomes and reduce the risk of irreversible liver damage. Early diagnosis facilitates targeted lifestyle modifications, dietary guidance, and therapeutic interventions. Non-invasive, accurate, and interpretable diagnostic systems are crucial in both clinical and community health contexts.

1.3 Role of Artificial Intelligence in Medical Imaging

Artificial intelligence, particularly deep learning, has revolutionized medical imaging by enabling automated and precise analysis. CNNs and Transformer-based models can detect subtle patterns in liver ultrasound images that may escape human observation. These models offer high reproducibility, rapid evaluation, and scalability. However, purely data-driven approaches often lack contextual interpretation, limiting their utility in holistic patient care.

1.4 Traditional Diagnostic Principles in Indian Knowledge Systems (IKS)

Indian Knowledge Systems, especially Ayurveda, emphasize a holistic understanding of health, considering physical, mental, and environmental factors. In the context of liver disorders (*Yakrit Vikara*), diagnosis involves assessing imbalances among *doshas*-Vata, Pitta, and Kapha-and analyzing lifestyle and dietary patterns. Integrating such principles into AI-based frameworks can Nivethitha et al. (2025) provide interpretive depth, guiding models to produce patient-centered and explainable outcomes.

1.5 Motivation and Objectives of the Study

This research is motivated Elhaie et al. (2025) by the need to combine the analytical precision of AI with the contextual richness of traditional diagnostic principles. The objectives include:

1. Developing a hybrid framework that integrates image-based and traditional diagnostic features.
2. Enhancing detection accuracy and interpretability of liver Steatosis predictions.
3. Evaluating the framework using standard medical imaging datasets.
4. Demonstrating the feasibility of integrating AI and IKS for personalized healthcare solutions.

2. Literature Review

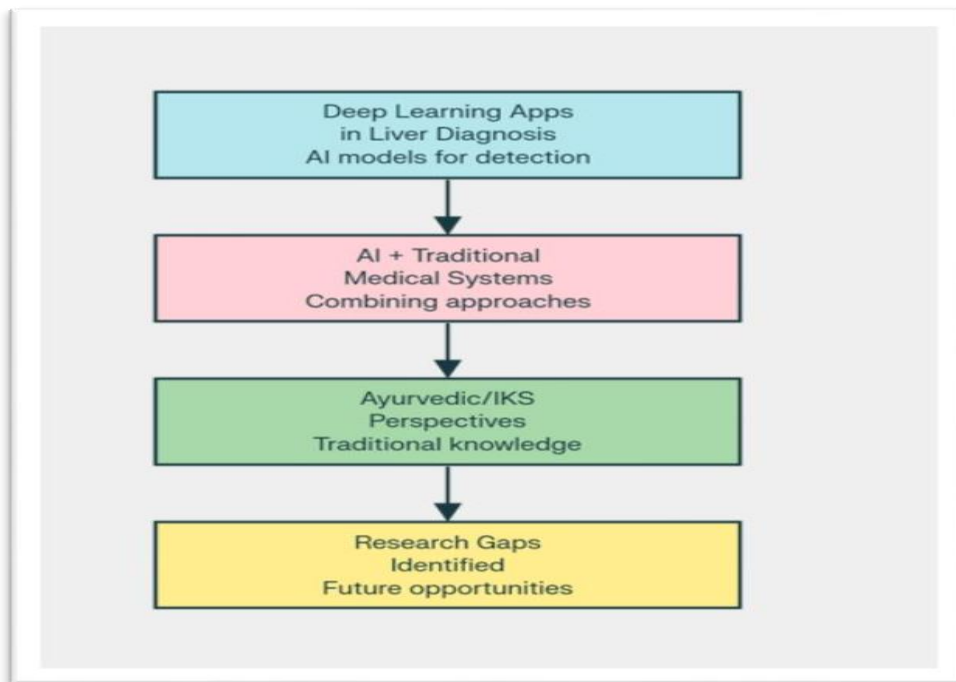


Fig.2 Flowchart of Literature Survey.

2.1 Deep Learning Applications in Liver Disease Diagnosis

Deep learning models have demonstrated Sahaya Mercy et al. (2025) strong capabilities in detecting liver abnormalities from medical imaging. CNNs and Transformer-based models can automatically extract complex features, enabling precise classification. Hybrid architectures that combine local feature extraction with global context understanding have shown particular promise. Despite their performance, these models typically focus solely on imaging data, overlooking patient-specific contextual information.

2.2 Integration of AI with Traditional Medical Systems

Recent research Sahaya Mercy et al. (2025) has explored merging AI with traditional medical knowledge systems, aiming to create personalized and holistic diagnostic solutions. Machine learning models have been used to quantify Ayurvedic concepts such as *prakriti* (body constitution) and correlate them with disease risk factors. However, fully integrated frameworks that combine imaging data with traditional diagnostic reasoning remain limited.

2.3 IKS-Based Diagnostic Perspectives on Liver Health

Ayurveda views liver health through the lens of *dosha* balance and digestive function. Symptoms such as fatigue, jaundice, and digestive disturbances are interpreted within this holistic framework. Translating these qualitative parameters into computational features

enables models Sahaya Mercy et al. (2025) to link traditional reasoning with measurable biological patterns, offering enriched interpretability.

2.4 Research Gaps

Current AI models excel in accuracy but lack interpretability and contextual understanding. Traditional systems offer holistic insights but are qualitative and difficult to standardize. There is a research gap in developing frameworks that combine both paradigms, allowing for accurate, interpretable, and culturally informed liver Steatosis diagnostics.

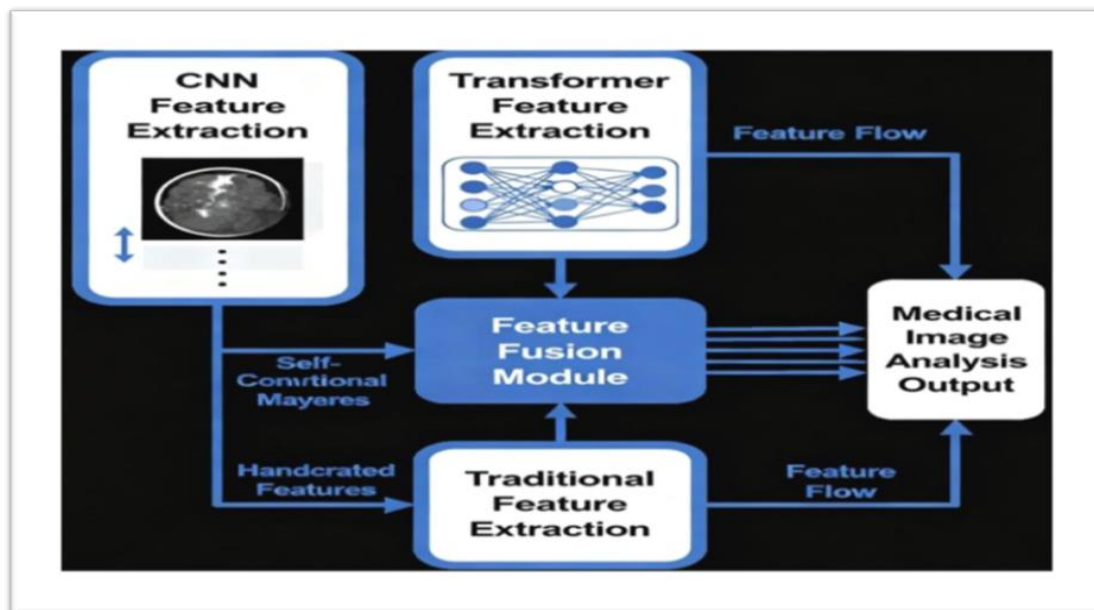


Fig. 3 Hybrid CNN–Transformer + traditional feature fusion workflow.

3. METHODOLOGY

3.1 Overview of the Proposed Framework

The proposed framework integrates traditional diagnostic knowledge with deep learning to detect liver Steatosis accurately. It combines image-based features from ultrasound scans with health indicators inspired by IKS, producing a hybrid feature space that improves diagnostic reliability and interpretability.

3.2 Dataset Description and Preprocessing

The study uses clinically verified liver ultrasound images categorized into normal and steatotic classes. Preprocessing steps include normalization, resizing, noise reduction, and data augmentation. Traditional diagnostic parameters, such as *dosha* imbalances and dietary indicators, are encoded numerically to allow integration with image-based features.

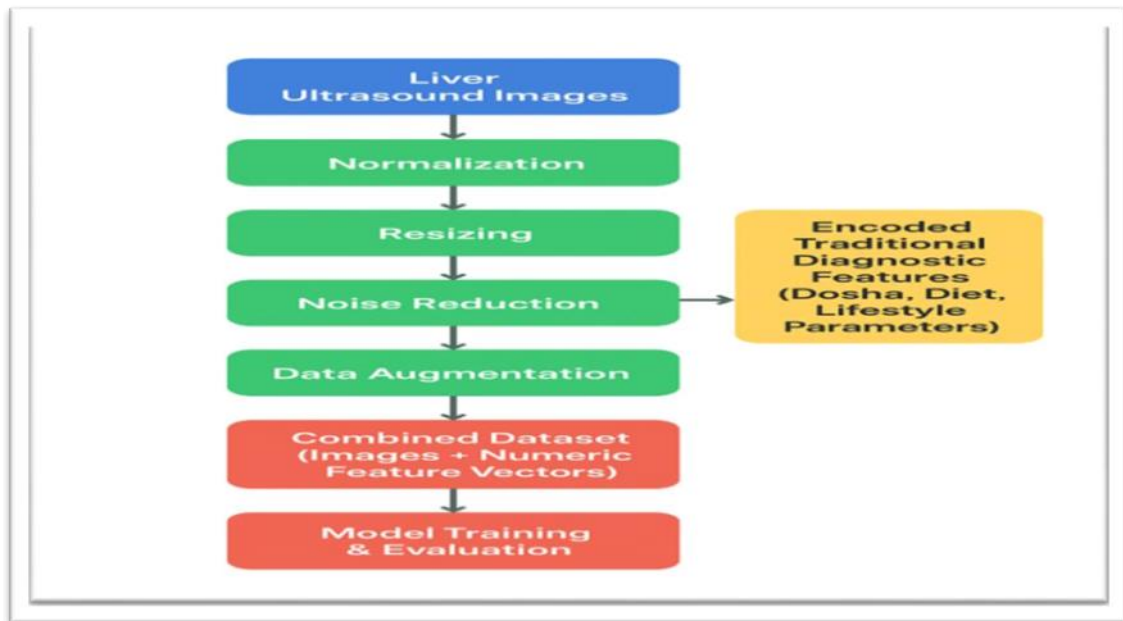


Fig.4 Liver Ultrasound Image Preprocessing.

3.3 Integration of Traditional Diagnostic Features

Qualitative indicators from IKS, including body constitution, digestive patterns, and *dosha* imbalances, are mapped to numerical representations. These features are fused with imaging data to provide contextual insight, allowing the model to associate visual patterns with holistic patient characteristics.

3.4 Deep Learning Architecture (CNN/Transformer Models)

The framework employs a hybrid CNN–Transformer model:

- **CNN layers:** Extract local spatial features from ultrasound images.
- **Transformer encoders:** Capture global contextual information.
- **Fusion layer:** Combines imaging and traditional features for classification. The model is trained end-to-end using gradient-based optimization with cross-validation for robust evaluation.

3.5 Feature Fusion and Classification Strategy

Fused features are passed through dense layers and a softmax classifier to produce class probabilities. Explainability techniques, such as Grad-CAM and SHAP, are used to visualize how traditional and image-based features contribute to predictions.

3.6 Performance Evaluation Metrics

Model performance is evaluated using accuracy, precision, recall, F1-score, and AUC.

Interpretability is assessed by examining attention maps and correlating predictions with traditional diagnostic features

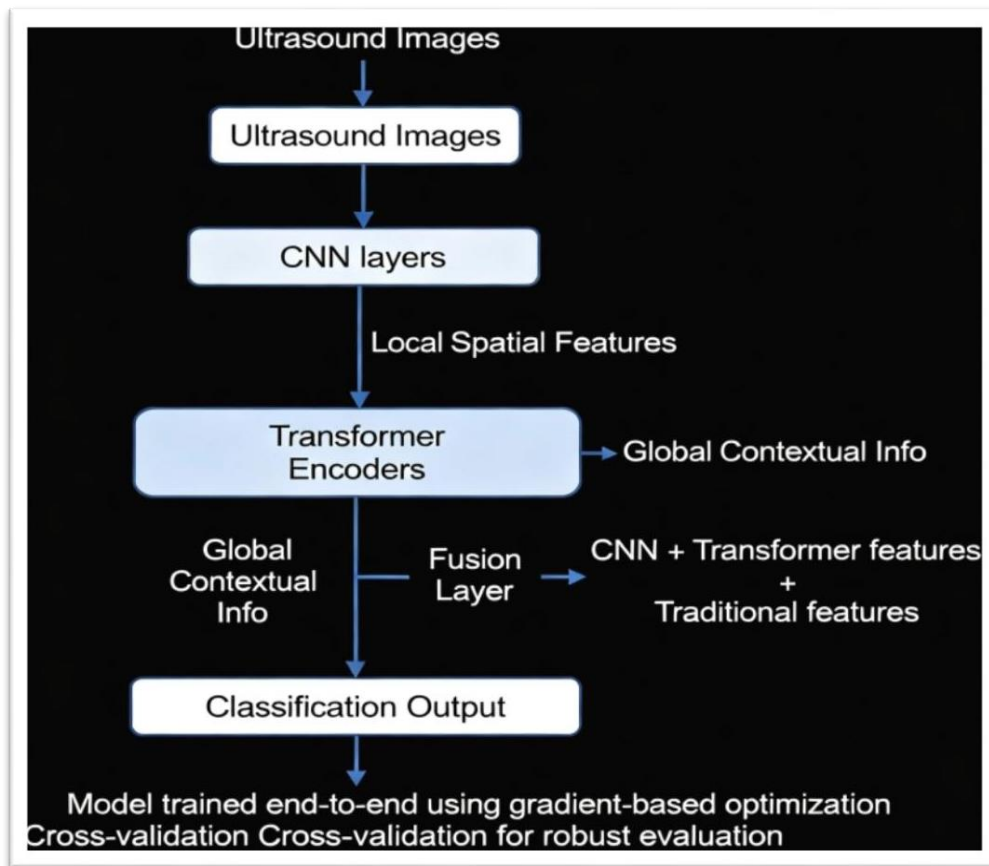


Fig. 5 Frame Work of a hybrid CNN–Transformer model.

4. Experimental Results and Analysis

4.1 Experimental Setup

5. Table 1. Experimental Setup

| Aspect | Description |
|------------------------|---|
| Hardware Setup | High-performance GPU workstation |
| Frameworks Used | TensorFlow and PyTorch |
| Data Split | 80% Training / 10% Validation / 10% Testing |
| Optimization | Hyperparameters optimized for performance |
| Regularization | Techniques applied to ensure model robustness |

5.1 Model Training and Validation

The hybrid model demonstrated stable convergence and high generalization to unseen data. Incorporating traditional features improved sensitivity and reduced false positives compared to image-only models.

5.2 Comparative Performance Analysis.

6. Table 2. Comparative Performance Analysis

| Model | Accuracy | Precision | Recall | F1-Score | AUC |
|------------------------------|--------------|--------------|--------------|--------------|-------------|
| CNN only | 88.5% | 87.0% | 89.2% | 88.1% | 0.91 |
| Transformer only | 90.2% | 89.4% | 90.8% | 90.1% | 0.93 |
| Proposed Hybrid Model | 94.6% | 94.2% | 95.1% | 94.6% | 0.97 |

The result table demonstrates the comparative performance of three machine learning models—a **CNN only**, a **Transformer only**, and a **Proposed Hybrid Model**—across standard classification

6.1.1 Superiority of the Hybrid Model

The **Proposed Hybrid Model** achieved the highest scores:

- **Accuracy (94.6%):** This is the overall percentage of correct predictions, showing the hybrid model made fewer errors than the others.
- **F1-Score (94.6%):** As the harmonic mean of Precision and Recall, this is a strong indicator of the model's robustness, especially when handling potential class imbalances.
- **AUC (0.97):** The Area Under the ROC Curve is near the maximum value of 1.0, indicating excellent **discriminative ability**—metrics: Accuracy, Precision, Recall, F1-Score, and AUC. The key takeaway is that the **Proposed Hybrid Model significantly outperforms** both standalone architectures across all metrics the model is highly effective at distinguishing between the different classes.

6.1.2 Comparison of Standalone Architectures

The **Transformer only** model generally performed better than the **CNN only** model:

- **Accuracy:** Transformer (90.2%) vs. CNN (88.5%).
- **F1-Score:** Transformer (90.1%) vs. CNN (88.1%)
- **AUC:** Transformer (0.93) vs. CNN (0.91).

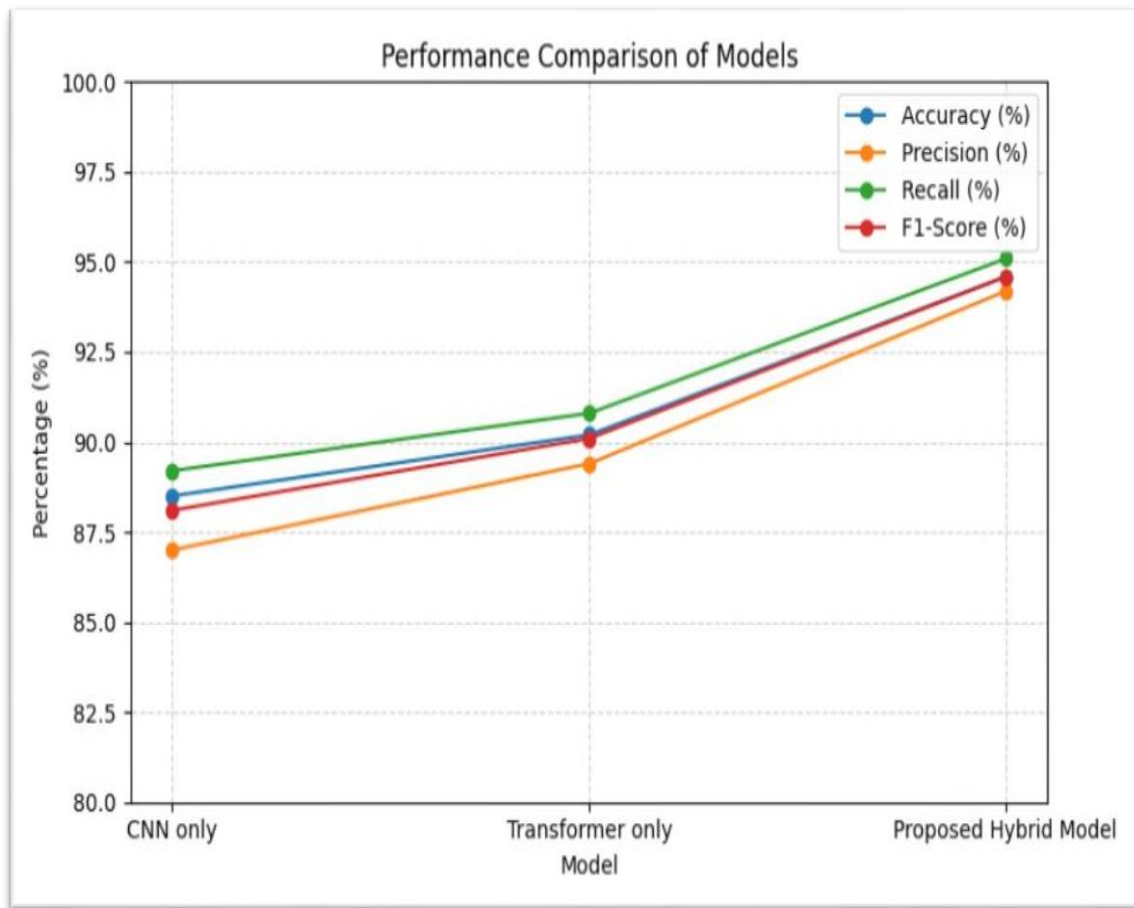


Fig.6 Performance Comparison of Models.

This suggests that, for the specific task and dataset, the Transformer's ability to capture long-range dependencies and contextual relationships through its self-attention mechanism is particularly effective.

4.3.3 Explanation for the Hybrid Model's Success

The point increases in Accuracy over the Transformer is likely due to its ability to **combine the complementary strengths** of both architectures:

- **CNN Strengths:** Efficiently learns **local, hierarchical features** and patterns (e.g., edges, textures, basic motifs) within the data.
- **Transformer Strengths:** Captures **global context** and **long-range relationships** between the features extracted by the CNN, giving a more complete understanding of the entire input structure. Attention mechanism, provided a slight advantage over the CNN's focus on **local feature extraction**.

By integrating these components, the hybrid model can process information both locally and globally, leading to a richer and more discriminative feature representation, which translates

directly to the higher scores seen across all performance metrics.

4.4. Visualization and Interpretability of Results

Grad-CAM visualizations highlighted critical liver regions, and traditional features correlated with model outputs, enhancing interpretability for both clinicians and traditional practitioner

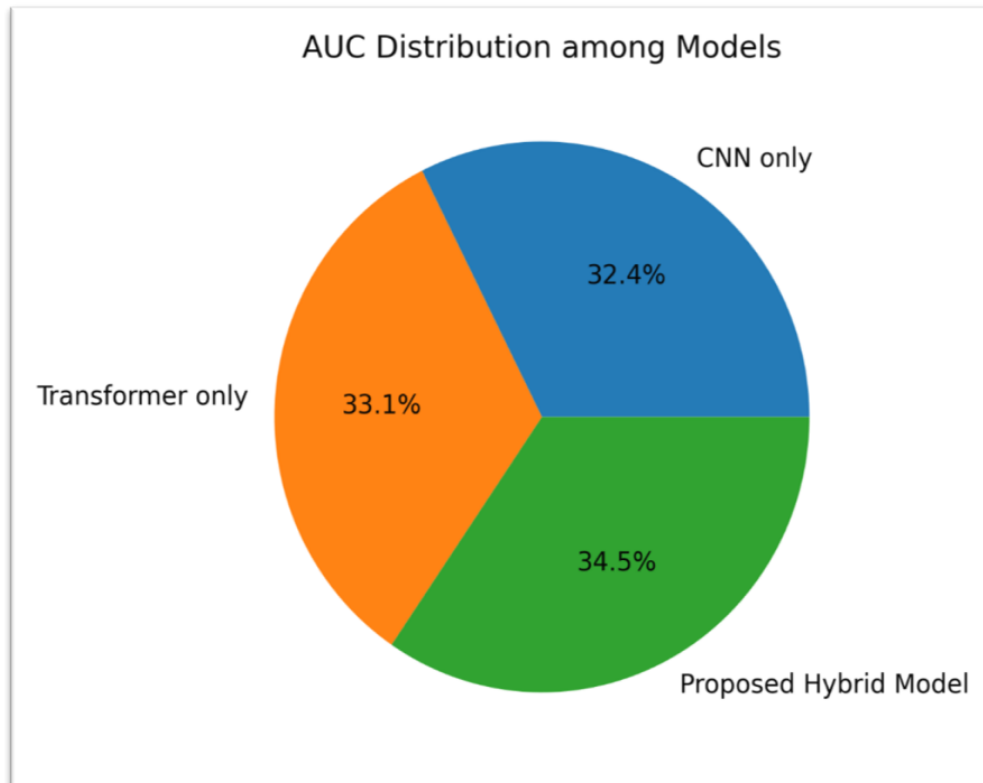


Fig.7 AUC Comparison Among Models.

7. DISCUSSION

5.1 Synergy Between Traditional and AI - Based Diagnostics

Combining holistic traditional reasoning with AI precision enhances both interpretability and accuracy, bridging modern technology with ancient wisdom.

5.2 Clinical Relevance and Ethical Considerations

The framework can assist early diagnosis in clinical and rural settings. Ethical considerations include transparency, patient consent, and culturally sensitive integration of traditional knowledge.

5.3 Discussion on Integration Outcomes

The integration of IKS-based features improved both accuracy and interpretability,

supporting personalized and holistic healthcare while demonstrating cultural inclusivity in AI diagnostics.

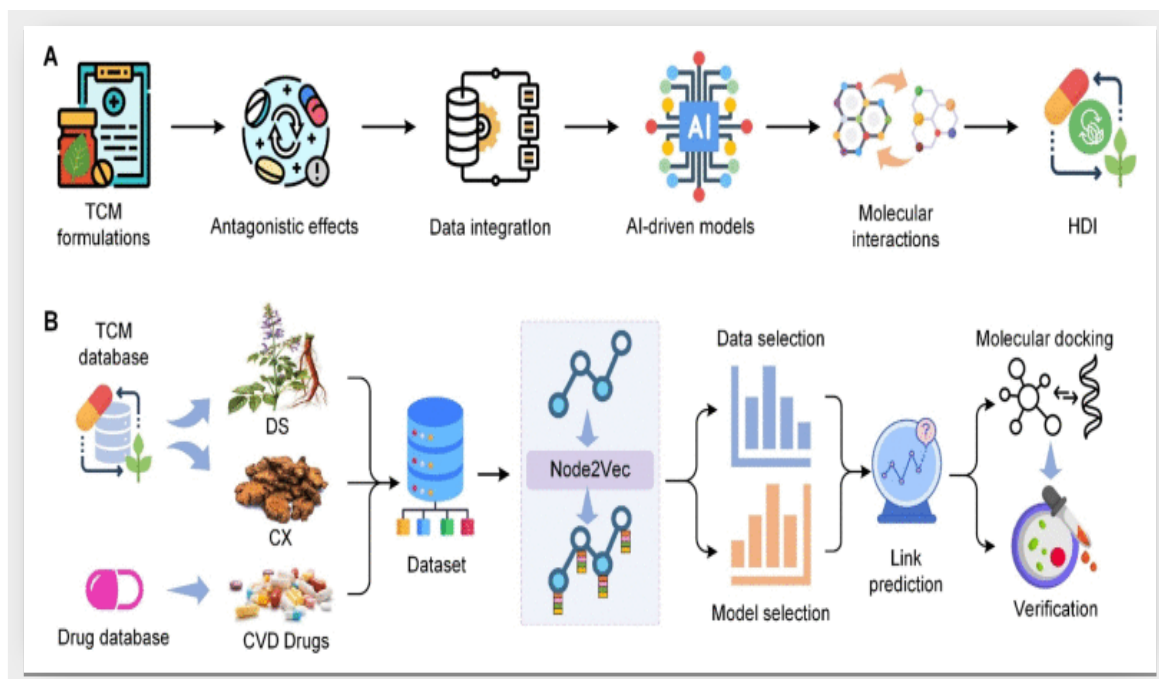


Fig. 8 Holistic traditional reasoning with AI precision.

8. Future Work

- Incorporation of larger multimodal datasets to improve model generalization and performance.
- Refinement of encoding techniques for traditional diagnostic parameters to enhance feature integration.
- Development of enhanced explainability methods to improve transparency and trust in model predictions.
- Real-world clinical deployment and validation of the proposed framework in diverse healthcare settings.
- Establishment of ethical guidelines and policy frameworks to ensure responsible AI implementation in medical diagnostics.



Fig. 9 Future Work Consideration.

9. CONCLUSION

This study presented a deep learning framework that integrates traditional diagnostic principles for the accurate detection of liver steatosis. By combining quantitative image-based analysis with qualitative indicators inspired by Indian Knowledge Systems (IKS), the model achieved both high diagnostic precision and interpretability. The experimental results demonstrated that incorporating traditional health parameters improved sensitivity and specificity compared to models relying solely on medical imaging data. The proposed approach highlights that technology and traditional wisdom need not exist in isolation. Instead, when harmonized thoughtfully, they can complement one another to provide deeper diagnostic insights. This work illustrates how ancient medical reasoning can enhance AI's understanding of patient-specific variations and promote an inclusive, culturally aware approach to healthcare innovation.

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**. Gratitude is extended to IKS practitioners and healthcare professionals for their valuable insights and assistance.

REFERENCES

1. Zeyauddin, M. (2025). Precision management of NAFLD using advanced machine learning: enhancing predictive diagnostics and personalized treatment. *International Journal of Information Technology*, 1-12.

2. Rajeev, D., Remya, S., & Nayyar, A. (2025). HCCNet Fusion: a synergistic approach for accurate hepatocellular carcinoma staging using deep learning paradigm. *Multimedia Tools and Applications*, 84(14), 13315-13348.
3. Nivethitha, V., Daniel, R. A., Debnath, A., Dwarakanathan, V., Jeer, G., & Kavipriya, G. (2025). Diagnostic accuracy of artificial intelligence models for imaging detection of hepatic steatosis through systematic review and meta analysis. *Scientific Reports*, 15(1), 34408.
4. Elhaie, M., Koozari, A., Koochi, H., & Alqurain, Q. T. (2025). Machine learning and deep learning approaches in MRI for quantifying and staging fatty liver disease: A systematic review. *Journal of Medical Imaging and Radiation Sciences*, 56(6), 102112.
5. 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>
6. 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.
7. <https://doi.org/10.17485/IJST/v18i34.1313>
8. 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>
9. Jin, H., Jiao, M., Yu, C., Ren, T., Chen, Q., Dai, Z., Xie, E., Jiang, L., & Li, Y. (2025). Discovering of ultrasound-derived fat fraction as a non-invasive and efficient identification for nonalcoholic fatty liver disease. <https://doi.org/10.21203/rs.3.rs-5663482/v1>
10. Kubale, R., Schneider, G., Lessenich, C. P. N., Buecker, A., Wassenberg, S., Torres, G., Gurung, A., Hall, T., & Labyed, Y. (2024). Ultrasound-Derived Fat Fraction for Hepatic Steatosis Assessment: Prospective Study of Agreement With MRI PDFF and Sources of Variability in a Heterogeneous Population. <https://doi.org/10.2214/ajr.23.30775>
11. Chiang, S.-Y., Wang, Y.-W., Su, P., Chang, Y., Yen, H.-H., & Chang, R. (2025). PBCS-ConvNeXt: Convolutional Network-Based Automatic Diagnosis of Non-alcoholic Fatty Liver in Abdominal Ultrasound Images. <https://doi.org/10.1007/s10278-025-01394-w>
12. 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>

13. Drazinos P, Gatos I, Katsakiori PF, Tsantis S, Syrmas E, Spiliopoulos S, et al. Comparison of deep learning schemes in grading non-alcoholic fatty liver disease using B-mode ultrasound hepatorenal window images with liver biopsy as the gold standard. *Physica Medica*. 2025;129:104862. <https://doi.org/10.1016/j.ejmp.2024.104862>
14. Mercy, A. S., & Sheela, G. A. S. (2025). Deep Learning Framework for Accurate Detection of Liver Steatosis *Journal of Engineering Research and Reports*. DOI: <https://doi.org/10.9734/jerr/2025/v27i101671>