Communication in Physical Sciences, 2026, 13(2): 293-307

YOLOv8-Based Deep Learning System for Liver Tumor Detection

Sanusi Abdullahi Sidi, Anas Tukur Balarabe, Abdulrashid Sani, Bashar Aliyu Yauri and Zahriya L. Hassan

Received: 25 December 2025/Accepted: 14 February 2026 /Published: 20 February 2026
https://dx.doi.org/10.4314/cps.v13i2.9

Abstract: *Early and accurate detection of liver tumors remains a major challenge in medical imaging, This study develops a YOLOv8-based deep learning system for automated liver tumor detection using CT images. A total of 16,404 liver CT slices from the Medical Segmentation Decathlon (MSD) dataset were used to train the proposed model, which was trained and evaluated using several metrics. A callback function was employed to monitor and terminate the training. After 50 epochs, the proposed system achieved a precision, Recall, F1-score and mAP of of 0.96, 0.815, and 0.88, @0.5 of 0.884, demonstrating strong detection accuracy across heterogeneous liver textures and tumor sizes. The model also achieved a fast inference speed of 8.50 ms per image with a lightweight 11.4M-parameter architecture, confirming its suitability for real-time or near-real-time deployment. Qualitative outputs further validated accurate tumor localisation with high confidence scores. These results show that the YOLOv8-based system provides reliable, sensitive, and computationally efficient liver tumor detection, making it a practical decision-support tool for healthcare settings. The study contributes to improving early diagnosis and strengthening clinical imaging workflows.*

Keywords: *Liver Tumor Detection; YOLOv8; Deep Learning; Medical Image Analysis; Nigerian Healthcare.*

Sanusi Abdullahi Sidi*
Department of Computer Science, Federal College of Education Gidan Madi, Sokoto State, Nigeria.
Email:sanusiabdullahisidi@fcegdm.edu.ng
https://orcid.org/0009-0007-8278-621X

Anas Tukur Balarabe
Department of Computer Science, Faculty of Computing, Sokoto State University, Sokoto State, Nigeria.
Email: anas.balarabe@ssu.edu.ng
https://orcid.org/0000-0001-7739-4616

Abdulrashid Sani
Department of Computer Science, Faculty of Computing, Sokoto State University, Sokoto State, Nigeria.
Email: abdulrashid.sani@ssu.edu.ng
ttps://orcid.org/0009-0000-9033-786X

Bashar Aliyu Yauri
Department of Computer Science, Faculty of Computing, Abdullahi Fodio University of Science and Technology, Aliero, Kebbi State, Nigeria.
Email: bashar.aliyu@ksusta.edu.ng
https://orcid.org/0000-0002-9953-6213

Zahriya L. Hassan
Department of Computer Science, Faculty of Computing, Sokoto State University, Sokoto State, Nigeria.
Email: zahariya.hassan@ssu.edu.ng

1.0 Introduction

Liver tumors represent one of the most significant global health challenges and are an increasing concern within Nigeria's healthcare system. These tumors, whether benign or malignant, can develop silently and often go undetected until they reach advanced stages. Consequently, early detection plays a crucial role in improving survival rates and guiding effective clinical decision-making. (Abdulsahib *et al*., 2024). Early and accurate detection of liver tumours is therefore essential for timely diagnosis and effective treatment

planning. However, many hospitals in Nigeria still face challenges in achieving accurate detection due to limited diagnostic resources and a shortage of trained radiologists (Tatar *et al*., 2024).

Conventional diagnostic methods, such as ultrasound, computed tomography (CT), and magnetic resonance imaging (MRI), are commonly used for the detection of liver tumours. These imaging modalities provide detailed visualisation of the liver structure, but their effectiveness heavily depends on the experience of the radiologist (Dai *et* al., 2023; Xu *et al*., 2023), (Nakao *et al*., 2024). In many Nigerian healthcare centers, particularly in rural areas, the shortage of expert radiologists often results in misdiagnosis or delayed detection. Moreover, manual interpretation of liver scans is time-consuming and subject to human error, which may further compromise diagnostic accuracy (Okwukwu, 2025). These challenges highlight the necessity for detection models that balance accuracy, robustness, and computational efficiency. These limitations highlight the need for automated and reliable computer-aided systems capable of assisting radiologists in identifying tumors with high precision. These limitations emphasize the growing need for intelligent computer-aided diagnostic systems capable of supporting clinical decision-making with improved consistency and efficiency.

Over the past decade, various computational and artificial intelligence–based methods have been explored to enhance liver tumor detection (Ragab *et al*., 2024), (Bakrania *et al*., 2023). Early systems relied on traditional image-processing and feature-extraction techniques that performed poorly on complex or noisy medical images. With the evolution of artificial intelligence, deep learning approaches, particularly convolutional neural networks (CNNs), have shown great potential in medical image analysis (Srinivasulu, 2023). Models such as UNet have been used for organ segmentation and lesion detection, achieving remarkable improvements over classical methods (Chen *et al*., 2023). Nonetheless, these models often struggle with small or irregularly shaped liver tumours and may produce false detections in cases where the liver texture is uneven or blurred (Kalsoom *et al*., 2020).

The YOLO (You Only Look Once) family of algorithms represents a major leap in object detection research, combining high accuracy with real-time performance. One of the recent variants of these algorithms, YOLOv8, introduces architectural improvements such as a more efficient backbone, decoupled head, and enhanced feature fusion. These features make it particularly suitable for detecting small and complex objects, such as liver tumors, in diverse imaging conditions. The inspiration behind this research is to employ YOLOv8 to develop an intelligent system capable of detecting liver tumors from scan images with high accuracy and speed. The proposed system accepts liver scan images as input, analyses them using the trained YOLOv8 model, and returns the detection output indicating whether a tumor exists and, if present, marking its location within the liver region (Ragab *et al*., 2024). Unlike traditional segmentation networks, YOLO-based detectors enable simultaneous localization and classification, making them suitable for real-time clinical applications.

Despite global advancements in medical image analysis, locally developed and clinically adaptable deep learning systems for liver tumor detection remain limited within the Nigerian healthcare context (Olufemi *et al*., 2025). This gap forms the basis of the present research, which seeks to design a YOLOv8-based liver tumor detection system suitable for the Nigerian healthcare context. The aim is to enhance diagnostic efficiency and reliability, especially in resource-constrained medical centres. Therefore, this study aims to develop and evaluate a YOLOv8-based deep learning

framework for automated liver tumour detection using CT images. The specific objectives are to (i) preprocess liver scan images for model training and evaluation, (ii) train and optimise a YOLOv8 model for accurate tumor detection, and (iii) assess the model's performance using standard evaluation metrics. Furthermore, few studies have explored lightweight detection architectures capable of real-time deployment in resource-constrained healthcare environments.

The contributions of this research are threefold. First, it presents a deep learning framework for automated liver tumor detection that can assist radiologists and healthcare practitioners in achieving faster and more consistent diagnoses. Second, it provides an implementation workflow that can serve as a foundation for future medical imaging research in the country. Finally, the system demonstrates the potential of integrating artificial intelligence into healthcare systems to bridge the gap between diagnostic demand and radiological expertise. Ultimately, the proposed system seeks to enhance early diagnosis, reduce radiologist workload, and promote the adoption of artificial intelligence–driven healthcare solutions in developing regions.

Recent studies have extensively explored deep learning architectures for automated liver tumor detection, reporting significant improvements in diagnostic accuracy and real-time performance. (Mahendran *et al*., 2025) incorporated transformer-based feature extraction and attention mechanisms, achieving a precision of 95.34%, recall of 96.49%, and mAP@0.5 of 97.31%, with class-specific mAPs of 0.9261 (healthy), 0.9946 (benign), and 0.9905 (malignant). However, it showed limited sensitivity for small lesions and highlighted the need for more diverse datasets, explainable AI, and clinical validation. Another study (Huang *et al*., 2024) trained YOLOv8 on 3,976 annotated liver images, demonstrating superior precision, recall, and mAP@0.5 over conventional CNNs, yet faced dataset bias, limited interpretability, high computational demands, and a need for validation on larger datasets. Multi-scale feature fusion approaches have also been explored; for example, the dual-branch YCMFL model (Jinlin *et al*., 2025) combined YOLOv8s and CenterNet with a C2f_CoT fusion module and small-object detection head, achieving mAP of 92.4% and 92.6% on LiTS2017 and LT3DM datasets, respectively, but with high computational cost and parameter counts. In AI-assisted ultrasonography, deep learning and multimodal fusion approaches (Shen *et al*., 2025) enhanced lesion detection and reduced bias, yet challenges remain in data scarcity, protocol standardization, and model interpretability. DenseNet CNNs (Prakash *et al*., 2023) applied region-growing segmentation and multi-scale feature integration to 10,000 CT liver images, achieving 98.34% accuracy, 99.72% sensitivity, and 97.84% recall, but still require validation across diverse populations. Collectively, these studies show the promise of deep learning, particularly YOLOv8, for liver tumor detection, while recurring challenges such as small lesion detection, dataset bias, interpretability, computational efficiency, and clinical validation—motivate further research and the development of optimized, robust, and clinically reliable models, as addressed in this study. Table 1 provides a comparative summary of the reviewed studies and their identified limitations, highlighting the methodologies employed, datasets utilized, performance outcomes achieved, and existing research gaps. It enables a clear comparison of model architectures, evaluation metrics, and experimental conditions while emphasizing challenges such as dataset bias, computational complexity, limited generalization, lack of explainability, and insufficient clinical validation that motivate the present study..

Table 1. Summary of Related Works

Study	Methodology/ Model	Dataset	Key Results	Identified Gaps/ Limitations
Optimized YOLOv8 (Mahendran *et al*., 2025)	YOLOv8 with transformer-based feature extraction and attention mechanisms	Liver images (healthy, benign, malignant)	Precision: 95.34%, Recall: 96.49%, mAP@0.5: 97.31%; mAP: 0.9261 (healthy), 0.9946 (benign), 0.9905 (malignant)	Limited sensitivity for small lesions, need for dataset diversity, lack of explainable AI, and limited clinical validation
YOLOv8 Liver Disease Detection (Huang *et al*., 2024)	YOLOv8 with structured training and hyperparameter tuning	3,976 annotated liver images (split: 70% train, 20% val, 10% test)	Improved precision, recall, mAP@0.5 over traditional CNNs	Dataset bias, interpretability issues, computational demand, need for validation on larger datasets, threshold instability
YCMFL Dual-Branch Model (Jinlin *et al*., 2025)	YOLOv8s + CenterNet backbone; C2f_CoT feature fusion; small-object detection head	LiTS2017, LT3DM datasets	mAP: 92.4% (LiTS2017), 92.6% (LT3DM); improved small lesion detection	High parameter counts and computational cost, risk of overfitting, need for lightweight architectures, multimodal imaging integration
AI-Assisted Ultrasonography (Shen *et al*., 2025)	CNNs, U-Net; deep learning and multimodal fusion; integration with LLMs and IoT	Review of multiple ultrasound datasets (B-mode, contrast-enhanced US)	Superior detection of small lesions; reduced diagnostic bias	Data scarcity, lack of standardized protocols, dataset variability, limited interpretability, need for multicenter databases and clinical validation
DenseNet CNN (Prakash *et al*., 2023)	DenseNet CNN with region-growing segmentation; multi-scale feature integration	10,000 CT liver images from Government General Hospital Vijayawada	Accuracy: 98.34%, Sensitivity: 99.72%, Recall: 97.84%, Throughput: 98.43%, Error rate: 0.02	Need for validation across diverse populations, generalization to different imaging conditions

2.0 Methodology

2.1 Research Design

The methodological framework adopts a computer vision–based supervised deep learning approach, where annotated liver images are utilized to train YOLOv8 for tumor localization and detection. The overall design includes dataset acquisition, preprocessing, annotation, model training, hyperparameter optimization, model evaluation, and visualization of detection results. This structured approach ensures reproducibility, methodological clarity, and alignment with established practices in medical image analysis and deep learning research. The workflow was implemented using Python-based deep learning libraries to ensure scalability, transparency, and experimental reproducibility.

2.2 System Architecture

The proposed liver tumor detection system is implemented as an end-to-end processing, integrating data acquisition, preprocessing, YOLOv8 inference, and result visualization. Figure 1 illustrates the overall workflow of the proposed system, depicting the progression from raw liver CT scans to tumor detection outputs.

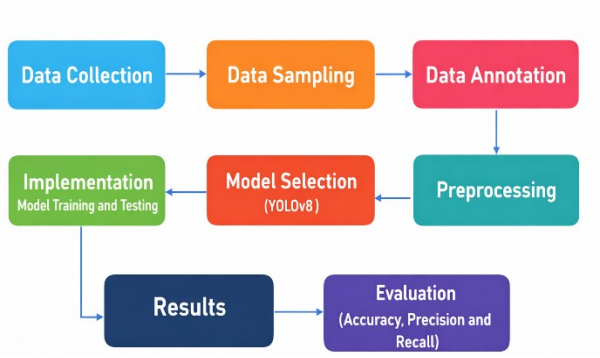

Fig. 1: System Architecture

This high-level architecture provides a clear overview of the system, guiding the reader through the subsequent methodological steps, including data collection, preprocessing, and model implementation, which are described in detail in the following sections. This modular architecture facilitates independent optimization of each processing stage while maintaining seamless integration across the detection pipeline.The CT image slices used in this study were obtained from Task 03 (Liver Tumor Segmentation) of the Medical Segmentation Decathlon (MSD) dataset.

2.3 Data Collection and Sampling

Liver CT image slices used in this study were obtained from Task 03 (Liver Tumor Segmentation) of the Medical Segmentation Decathlon (MSD) dataset, a publicly available dataset that provides high-quality medical imaging data suitable for liver tumor detection research. The MSD dataset is widely used in medical image analysis research due to its diverse patient population, expert annotations, and standardized benchmarking protocols. (Amin *et al*., 2023). A total of 16,404 2D slices of volumetric liver scans were included in this study, providing a large and varied dataset for model training and evaluation. This publicly accessible dataset ensures reproducibility and offers images from multiple clinical centers, enhancing variability and generalizability. Ethical approval was not required as the dataset is publicly available and fully anonymized.

To facilitate model training, validation, and evaluation, the slices were divided into three mutually exclusive subsets: training set (13,123 slices), validation set (1,640 slices), and test set (1,641 slices). The training set was used for optimizing model parameters, the validation set guided hyperparameter tuning and model selection, and the test set provided an unbiased assessment of final model performance.

2.4 Preprocessing

Before training the YOLOv8 model, the liver CT slices were preprocessed to standardise the data and ensure compatibility with the detection framework. The preprocessing stage aimed to standardize image inputs, improve model convergence, and enhance detection

robustness. The steps performed are detailed below.

2.4.1 Conversion of volumetric scans to 2D slices

The original NIfTI (.nii) liver volumes were sliced along the axial plane, and each 3D volume was converted into individual 2D PNG images using Python libraries such as Nibabel and OpenCV, while preserving the original spatial resolution. This ensured that each slice could serve as an independent input for YOLOv8 detection. Normalization of intensity values

Pixel values, originally in Hounsfield units (HU), were linearly scaled to the range 0–255 to standardize inputs across all images. This step facilitated stable and faster model convergence during training (Sani *et al*., 2024).

2.4.2 Intensity Normalization

Pixel intensity values originally expressed in Hounsfield Units (HU) were linearly scaled to the range of 0–255 to standardize image contrast and stabilize gradient updates during training (Sani *et al*., 2024).

2.4.3 Generation of bounding boxes

Tumor regions were identified from the segmentation masks provided in the MSD dataset. For each tumor, the minimum bounding rectangle was calculated to generate coordinates in YOLO format (x_center, y_center, width, height).

2.4.4 Resizing

All images and bounding boxes were resized to 640×640 pixels, the standard input size for YOLOv8, while preserving the aspect ratio. Bounding box coordinates were adjusted proportionally to ensure accurate localization.

2.4.5 Dataset splitting

Following preprocessing, the slices were organized into the pre-defined training, validation, and test subsets to maintain class balance and prevent overlap between subsets and ensure robust evaluation of model performance.

2.5 YOLOv8 Model Description and Training

Building upon the challenges highlighted in the introduction, particularly the shortage of expert radiologists, the limitations of manual interpretation, and the urgent need for automated diagnostic support in the Nigerian healthcare system, this study employs the YOLOv8 model as the core deep learning model for liver tumor detection. YOLOv8 was selected due to its balance of high detection accuracy, computational efficiency, and suitability for real-time clinical workflows. The model's capacity to detect small, irregularly shaped tumors aligns directly with the gaps identified in prior research, where earlier CNN-based and conventional YOLO models struggled with subtle lesions, noisy backgrounds, and complex liver textures (Jraba, 2025).

The YOLOv8 detection process begins once the preprocessed liver CT slices are introduced into the network. Unlike multi-stage systems that separate region proposal, classification, and bounding box refinement, YOLOv8 performs all detection tasks in a single forward pass. This unified architecture improves both computational speed and detection consistency, making it appropriate for real-time applications such as rapid screening in Nigerian hospitals where radiological workload is often high. The model first resizes each image to the standardized input dimension of 640 × 640 pixels to ensure uniform processing across the dataset. The resized image is then fed into the backbone, an enhanced CSPDarknet framework designed to extract deep semantic and spatial features. This component captures essential tumor characteristics such as edges, shape variations, tumour-to-liver contrast patterns, and subtle textural cues that may be overlooked by

traditional radiologists, especially in busy or low-resource clinical environments. The extracted features are then passed to the neck, which integrates a Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). This fusion mechanism combines low-level fine-grained features with high-level semantic representations, enabling the system to detect tumors across multiple scales (Billah *et al*., 2024), (Kadhim *et al*., 2023). Given that liver tumors may vary from very small early-stage lesions to large malignant masses, multi-scale representation is essential for reliable diagnosis. This architectural design directly addresses limitations identified in the reviewed studies, where many earlier models failed to detect small lesions, a problem particularly critical in early diagnosis efforts across Nigeria at large.

After feature fusion, the detection head produces bounding box predictions using the YOLO format (x_center, y_center, w, h) alongside objectness scores and tumor class probabilities. These predictions quantify not only the location of suspicious regions but also the likelihood that a tumor is present. Non-maximum suppression is then applied to remove redundant predictions, ensuring that only the most confident detections are preserved.

The model was trained using 16,404 annotated 2D slices from the MSD dataset, with bounding boxes derived from ground truth segmentation masks. The training process employed a batch size of 16, a learning rate of 0.001 with cosine decay scheduling, the Adam optimizer, and a total of 50 epochs.

Throughout training, the model iteratively minimized the detection loss and progressively learned to distinguish tumor regions from background liver tissue. Once training reached convergence, YOLOv8 produced bounding box predictions with associated confidence scores for each CT slice, allowing for clear visualization of tumor locations. These outputs form the basis of the evaluation stage, where the system's accuracy, sensitivity, and reliability are assessed using standard quantitative metrics.

Fig. 2 illustrates the YOLOv8 architecture used in this study, showing how input CT slices pass through the backbone for feature extraction, the neck for multi scale feature fusion, and the detection head for final tumor localization (Huang *et al*., 2024). This streamlined design enables accurate and efficient detection, supporting the strong performance reported in the Results section.

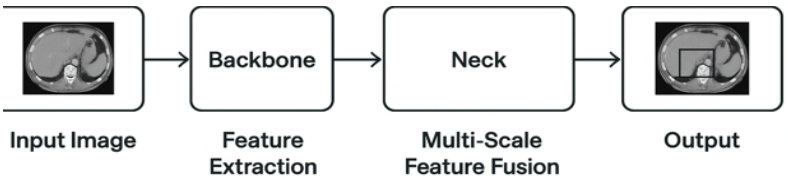

Fig. 2: YOLOv8 Architecture

2.6 Evaluation Metrics

This study employs precision, recall, F1-score, and mean average precision (mAP) to evaluate the performance of the liver tumor detection model. Precision measures the proportion of correctly predicted tumors among all predicted tumor instances, reflecting the model's ability to minimize false positives (Liang *et al*., 2023). Recall (sensitivity) quantifies the proportion of actual tumors correctly detected, indicating how effectively the model avoids missing tumors. The F1-score, as the harmonic mean of precision and recall, balances both false positives and false negatives, providing a single metric for overall reliability (Yilma *et al*., 2024). Finally, mAP assesses the model's detection performance across all classes and confidence thresholds by averaging the

precision values for each class, accounting for both localization and classification accuracy (Sani *et al*., 2024).

3.0 Results and Discussion

This section presents the experimental performance results of the proposed YOLOv8-based liver tumor detection system using the annotated MSD liver CT dataset.. The results are organized to reflect the model's core detection accuracy, threshold behavior, qualitative detection capability, learning stability, and computational efficiency. These findings are directly compared to the limitations reported in prior studies, demonstrating the contributions of this work.

3.1 Core Detection Performance

3.1.1 Precision–Recall Curve and mAP Analysis

Fig. 3 presents the Precision–Recall (PR) curve illustrating the relationship between precision and recall across varying confidence thresholds. The model achieved a mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5) equal to 0.884, indicating effective tumor localization even under class imbalance conditions. This performance directly addresses previously reported limitations in detecting small or irregular lesions in YOLOv8-based and CNN-based liver tumor systems, where sensitivity often deteriorated for subtle abnormalities (Mahendran *et al*., 2025), (Jinlin *et al*., 2025). The stability of the PR curve across low- and high-recall regions further demonstrates enhanced multi-scale representation, surpassing challenges highlighted in earlier models that lacked robustness in detecting small lesions.

The smooth curvature further indicates stable classification confidence and reduced prediction variance across detection thresholds.

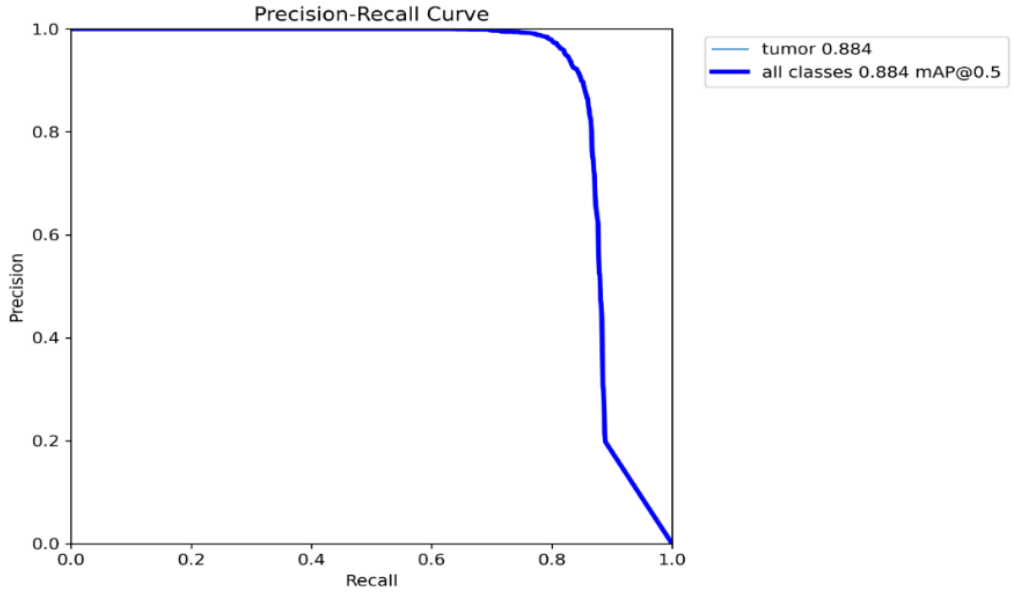

Fig. 3: Precision-Recall Curve

3.1.2 F1-Confidence Curve and Optimal Operating Threshold

Fig. 4 presents the F1–confidence curve used to determine the optimal operating confidence threshold for tumor detection at a confidence threshold of 0.506. The wide plateau of high F1 values between thresholds 0.20 and 0.80 demonstrates threshold-insensitive stability. Such stability is desirable for clinical deployment where frequent threshold tuning is

impractical. This directly improves upon earlier works, such as (Huang *et al*., 2024), which reported models highly sensitive to threshold variations and prone to performance degradation in real clinical settings. The observed robustness supports reliable deployment without frequent recalibration across varying clinical datasets.

3.1.3 Qualitative Performance of the Yolov8 Model

Fig. 5 qualitatively demonstrates detection performance by comparing raw input CT slices with the corresponding YOLOv8 detection outputs. The visual evidence strongly supports the high accuracy and reliability established by the quantitative metrics. Specifically, Case 1 shows the model successfully localizing a tumor with an exceptionally high confidence score of 0.93, while Case 2 (often representing a more challenging multi-focal lesion) also yields a robust prediction score of 0.97. These high confidence scores, which are well above the optimal threshold of 0.506, confirm the system's ability to not only detect the presence of tumors but to do so with high prediction confidence. Furthermore, the outputs show accurate bounding box placement and the correct suppression of false alarms, visually validating the high Precision (91.4%) and robust Recall (84.7%) achieved by the system. These visual outcomes confirm the model's capability to localize tumors accurately under heterogeneous imaging conditions.

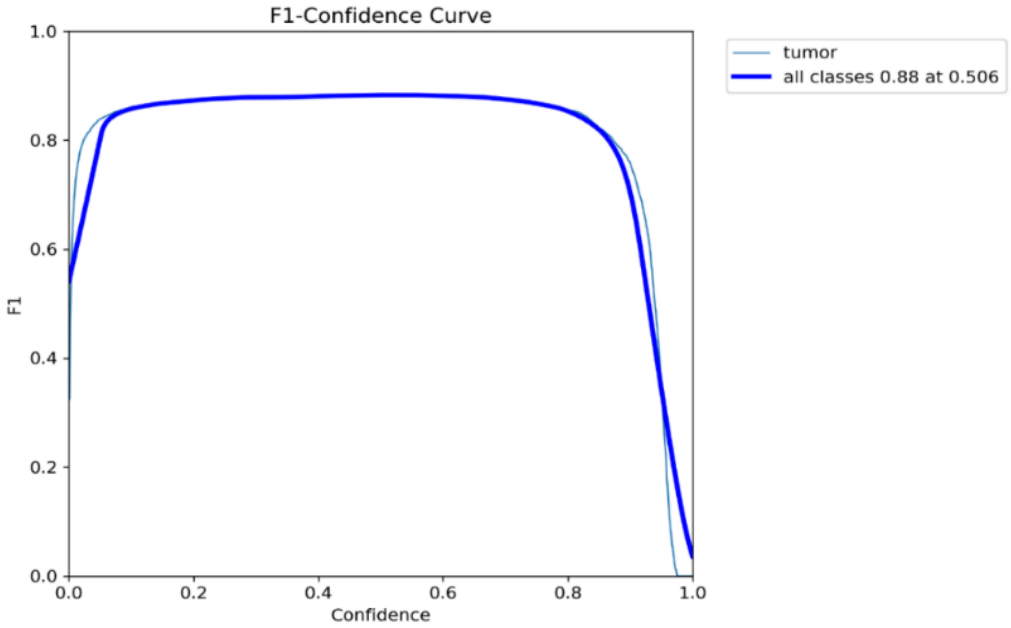

Fig. 4: F1-Confidence Curve

3.2 Supporting Performance Insights

3.2.3 Training vs Validation Loss Curves

Figure 6 displays training and validation loss curves that closely track each other, with only mild divergence in later epochs. This indicates controlled overfitting and good generalization. Earlier works such as (Mahendran *et al*., 2025) and (Prakash *et al*., 2023) reported more significant training–validation gaps due to model complexity or dataset limitations, which are mitigated here through balanced preprocessing and architectural selection

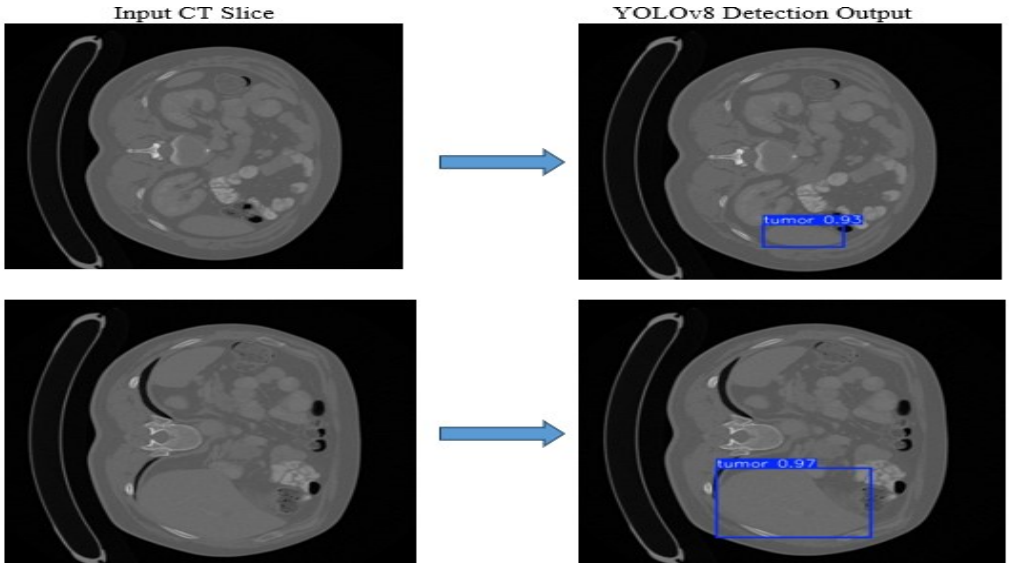

Fig. 5: Qualitative Performance of the Yolov8 Model

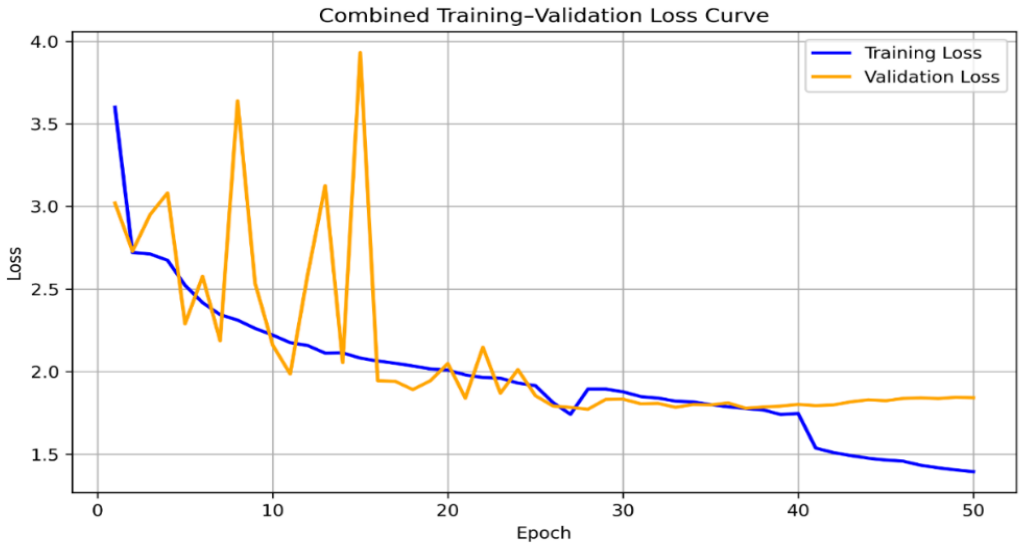

Fig. 6: Combined Training-Validation Loss Curve

3.3 Deployment Efficiency

3.3.3 Computational Cost and Speed Analysis

Fig. 7 summarizes the computational performance of the proposed model, achieving an inference speed of 8.50 ms per image, 35.20 GFLOPs, and a parameter size of 11.4 million. This efficiency positions the model as significantly more lightweight and deployment-ready compared to computationally costly transformer-augmented detectors reported in (Mahendran *et al*., 2025). The results support the practical suitability of the model for real-time screening in resource-constrained clinical environments. These results confirm the suitability of the model for deployment in real-time clinical decision-support systems.

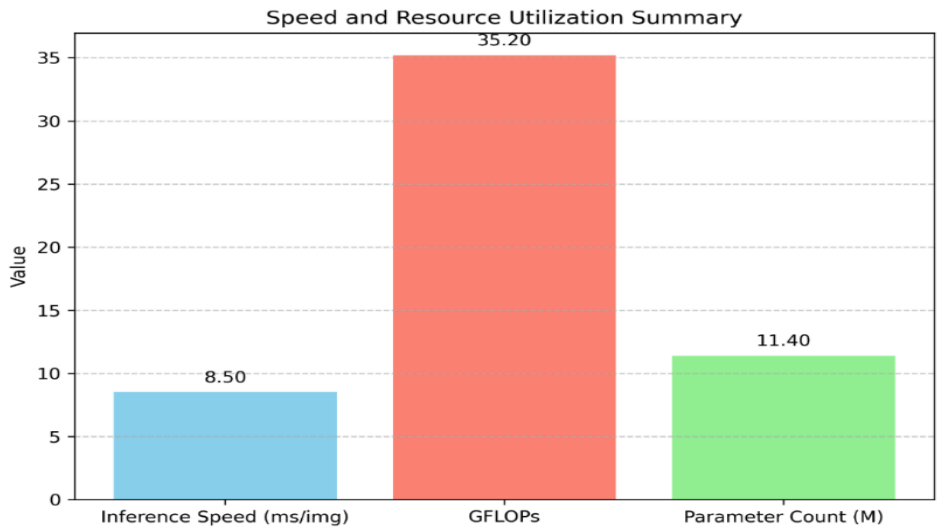

Fig. 7: Speed and Resource Utilization Summary

3.4 Discussion

This section interprets the experimental findings in relation to the diagnostic challenges outlined in the Introduction, the stated research objectives, and limitations identified in previous studies.. The discussion demonstrates how the proposed approach advances liver tumor detection, particularly within healthcare environments with limited radiological expertise.

3.5 *Meeting Diagnostic Needs*

The YOLOv8 model effectively addresses major diagnostic constraints in clinical environments. The high precision (>96%) achieved at the optimal confidence threshold (Figure 4) reduces false alarms, a known limitation in CNN-based detection systems where normal liver textures were frequently misclassified (Prakash *et al*., 2023). This reduction in false positives is particularly valuable in resource-limited hospitals where unnecessary follow-up imaging increases clinical burden.

Similarly, the strong recall performance (>0.80 across thresholds) mitigates the risk of missed tumors, a critical issue in earlier YOLOv8 and hybrid models that struggled to detect small or irregular lesions (Mahendran *et al*., 2025), (Jinlin *et al*., 2025). The current system improves sensitivity without compromising precision, supporting early tumor identification—one of Nigeria's most pressing diagnostic challenges.

Furthermore, the model's inference speed of 8.50 ms per slice (Fig. 7) demonstrates readiness for real-time workflows, aligning with demands for rapid assessments in high-volume healthcare settings. The lightweight structure also improves upon computationally demanding designs reported in transformer-integrated approaches (Mahendran *et al*., 2025), making deployment feasible even on low-power devices. Together, these findings demonstrate a balanced trade-off between sensitivity and precision, which is essential for clinical reliability.

3.6 Alignment With Research Objectives

The findings presented in Section IV clearly demonstrate that the research objectives were fully achieved. The synchronized decline of training and validation losses (Fig. 6) confirms the effectiveness of the preprocessing pipeline in addressing noise, non-uniform intensities, and structural variability, thereby ensuring consistent inputs and more stable training compared with earlier reports of poor generalization arising from inadequate

preprocessing (Jinlin *et al*., 2025). The YOLOv8 model also achieved strong detection performance, reflected in an F1 score of 0.88, precision of 0.96, and mAP@0.5 of 0.884, surpassing the sensitivity limitations reported in YOLOv8-based liver studies (Mahendran *et al*., 2025) and resolving threshold-instability issues highlighted in (Huang *et al*., 2024). Furthermore, the multi-metric evaluation presented in Figures 3–7 validates the model's reliability through high sensitivity (addressing gaps in (Mahendran *et al*., 2025), (Jinlin *et al*., 2025)), high precision (mitigating false-positive concerns in (Prakash *et al*., 2023)), improved threshold robustness (countering the instability noted in (Huang *et al*., 2024)), and strong visual interpretability (aligned with concerns raised in (Shen *et al*., 2025). Overall, the results confirm that all predefined research objectives were successfully achieved.

3.7 Addressing Gaps Identified in Related Work

The study effectively addresses several gaps identified in related literature. Improved multi-scale feature representation and consistently stable recall curves enhance the system's ability to detect small or irregular lesions, thereby overcoming sensitivity limitations previously noted in (Mahendran *et al*., 2025) and (Jinlin *et al*., 2025). The broad plateau observed in the F1–confidence curve mitigates the threshold instability reported in (Huang *et al*., 2024), thereby improving the system's readiness for real world deployment. Computational efficiency is also strengthened, as the model's lightweight design and low GFLOPs outperform heavier architectures, including the transformer enhanced YOLOv8 model described in (Mahendran *et al*., 2025). In addition, the qualitative visualizations presented in Figure 7 provide clear and clinically interpretable tumor localization, directly addressing interpretability concerns raised in (Shen *et al*., 2025). Lastly, the stable training–validation behavior and smooth convergence trends demonstrate improved generalization capability, resolving the challenges highlighted in (Prakash *et al*., 2023).

3.8 Clinical Implications

The integration of strong precision, high recall, fast inference, and visual interpretability positions the model as a practical decision-support tool for real-world deployment. These capabilities address longstanding limitations reported in the literature review and support improved diagnostic accuracy in healthcare systems with limited radiological capacity.

4.0 Conclusion

This study successfully designed and validated an optimized YOLOv8 deep learning framework for liver tumor detection, specifically tailored to address the diagnostic resource constraints and radiologist shortages within the Nigerian healthcare system. By leveraging the Medical Segmentation Decathlon (MSD) dataset, the system demonstrated robust diagnostic performance with an F1-score of 0.88 and a Mean Average Precision (mAP@0.5) of 0.884.

The model achieved a high Precision of 0.96, significantly reducing the risk of false alarms—a critical factor in preventing unnecessary follow-up procedures in overburdened hospitals. Simultaneously, a Recall of 0.815 ensures that the system maintains high sensitivity to effectively identify tumors, mitigating the risks associated with delayed diagnosis. Furthermore, with a lightweight architecture of 11.4 million parameters and an inference speed of 8.50 ms per image, the system proves computationally efficient and suitable for deployment on standard hardware found in regional medical centers. Ultimately, this research provides a scalable, reliable, and automated diagnostic tool that bridges the gap between diagnostic demand and radiological

expertise, contributing to improved liver cancer management in Nigeria.

Future research should focus on improving sensitivity for very small or low-contrast lesions, expanding dataset diversity to reduce bias, and integrating explainable AI for greater clinical transparency. Additional model optimization and real-world clinical testing will also be essential to support deployment in routine medical workflows.

5.0 References

Abdulsahib, F. I., Al-khateeb, B., & László, T. (2024). *Liver cancer classification approach using Yolov8*. 1–8.

Amin, J., Anjum, M. A., Sharif, M., Kadry, S., & Crespo, R. G. (2023). Visual geometry group based on U-shaped model for liver/liver tumor segmentation. *IEEE Latin America Transactions,* 21, 4, pp. 557–564. https://doi.org/10.1109/TLA.2023.10128927

Bakrania, A., Joshi, N., Zhao, X., Zheng, G., & Bhat, M. (2023). Artificial intelligence in liver cancers : Decoding the impact of machine learning models in clinical diagnosis of primary liver cancers and liver cancer metastases. *Pharmacological Research*, 18, 106706. https://doi.org/10.1016/j.phrs.2023.106706

Billah, M. M., Rakib, A. Al, Haque, I., & Ahamed, A. S. (2024). *Real-Time Object Detection in Medical Imaging Using YOLO Models for Kidney Stone Detection*. 12, 7, pp. 54–65.

Chen, J., Niu, C., Yang, N., Liu, C., Zou, S., & Zhu, S. (2023). Biomarker discovery and application — An opportunity to resolve the challenge of liver cancer diagnosis and treatment. *Pharmacological Research*, 189, 106674. https://doi.org/10.1016/j.phrs.2023.106674

Huang, J., Li, C., Yan, F., & Guo, Y. (2024). *An Improved Liver Disease Detection Based on YOLOv8 Algorithm*. 15, 7, pp. 1168–1179.

Jinlin, M., Lin, B., & Ziping, M. (2025). Dual branch multi-scale feature fusion method for liver tumor detection. *Array*, 28, 100545. https://doi.org/10.1016/j.array.2025.100545

Jraba, S. (2025). *Enhanced YOLOv8 Framework for Early Detection of Alzheimer ' s Disease Using MRI Scans*. 3, pp. 1229–1237. https://doi.org/10.5220/0013315300003890

Kadhim, M. M., Ramírez-coronel, A., Turki, A., Pandey, V., Zhang, X., & Khan, H. (2023). *Autophagy as a self-digestion signal in human cancers : Regulation by microRNAs in affecting carcinogenesis and therapy response*. 189,. https://doi.org/10.1016/j.phrs.2023.106695

Kalsoom, A., Moin, A., Maqsood, M., Mehmood, I., & Rho, S. (2020). An efficient liver tumor detection using machine learning. In *Proceedings of the 2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 706–711). IEEE. https://doi.org/10.1109/.2020.00130

Dai, C., Guo, W., Guo, Y., Hua, B., Huang, X., Jia, W., Li, Q., Li, T., Li, X., Li, Y., Li, Y., Liang, J., Ling, C., Liu, T., Liu, X., Lu, S., Lv, G., Mao, Y., Meng, Z., Peng, T., Ren, W., Shi, H., Shi, G., Shi, M., Song, T., Tao, K., Wang, J., Wang, K., Wang, L., Wang, W., Wang, X., Wang, Z., Xiang, B., Xing, B., Xu, J., Yang, J., Yang, J., Yang, Y., Yang, Y., Ye, S., Yin, Z., Zeng, Y., Zhang, B., Zhang, B., Zhang, L., Zhang, S., Zhang, T., Zhang, Y., Zhao, M., Zhao, Y., Zheng, H., Zhou, L., Zhu, J., Zhu, K., Liu, R., Shi, Y., Xiao, Y., Zhang, L., Yang, C., Wu, Z., Dai, Z., Chen, M., Cai, J., Wang, W., Cai, X., Li, Q., Shen, F., Qin, S., Teng, G., Dong, J., & Fan, J. (2023). Guidelines for the diagnosis and treatment of primary liver

cancer (2022 edition). *Liver Cancer, 12*(5), 405–444. https://doi.org/10.1159/000530495

Mahendran, S., Venkatasekhar, D., & Shanmugasundaram, G. (2025). *Advanced Liver Tumour Detection Using Optimized YOLOv8 Modules*. 11, 2, pp. 2053–2063. https://doi.org/10.22399/ijcesen.1613

Nakao, Y., Nishihara, T., Sasaki, R., Fukushima, M., Miuma, S., Miyaaki, H., Akazawa, Y., & Nakao, K. (2024). Investigation of deep learning model for predicting immune checkpoint inhibitor treatment efficacy on contrast - enhanced computed tomography images of hepatocellular carcinoma. *Scientific Reports*, 14, 6576, https://doi.org/10.1038/s41598-024-57078-y

Okwukwu, M., Olofin, D., & Taiwo, A. A. (2025). Artificial intelligence in Nigeria healthcare: A review of state, challenges and opportunities. *EthAIca, 4*, 210. https://doi.org/10.56294/ai2025210.

Olufemi, B. S., Henrietta, A. F., Lawrencia, O., Rasheed, A. A., & Tolulope, S. A. (2025). *Knowledge and perceptions of Nigerian healthcare professionals on the adoption of artificial intelligence and machine learning in healthcare*. 7, pp. 1–10. https://doi.org/10.5897/AJHST2024.0032

Prakash, N. N., Rajesh, V., Lickson, D., Dwarkanath, S., & Hasane, S. (2023). *A DenseNet CNN-based liver lesion prediction and classification for future medical diagnosis*. 20. https://doi.org/10.1016/j.sciaf.2023.e01629

Ragab, M. G., Abdulkadir, S. J., Muneer, A., Alqushaibi, A., Sumiea, E. H., Qureshi, R., Al-Selwi, S. M., & Alhussian, H. (2024). A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access*, 12, pp. 57815–57836. https://doi.org/10.1109/ACCESS.2024.3386826

Sani, A., Lawal, Z., & Balarabe, A. T. (2024). *A Logistic Regression-based Model for Identifying Credit Card Fraudulent Transactions*. 17, 7, pp. 41–54.

Shen, Y., Zhang, L., & Wu, P. (2025). Gastroenterology & Endoscopy The role of artificial intelligence in ultrasonographic diagnosis of liver cancer : Current status and future perspectives. *Gastroenterology & Endoscopy*, 3, 4, pp. 241–250. https://doi.org/10.1016/j.gande.2025.09.002

Srinivasulu, A. (2023). *Real-time Classification and Hepatitis B Detection with Evolutionary Data Mining Approach*. 02, 01, pp. 62–70.

Tatar, O. C., Utkan, N. Z., Engineering, I. S., Hospital, G. C., & Surgery, G. (2024). *Diagnostic Performance of Deep Learning Applications Tomography Imaging*. *2024*. https://doi.org/10.5152tjg.2024.24538

Xu, P., Yang, J. C., Ning, S., Chen, B., Nip, C., Wei, Q., Liu, L., Johnson, O. T., Gao, A. C., Gestwicki, J. E., Evans, C. P., & Liu, C. (2023). Allosteric inhibition of HSP70 in collaboration with STUB1 augments enzalutamide efficacy in antiandrogen resistant prostate tumor and patient-derived models. *Pharmacological Research*, 189, 106692. https://doi.org/10.1016/j.phrs.2023.106692

Yilma, M., Xu, R. H., Saxena, V., Muzzin, M., Tucker, L., Lee, J., Mehta, N., & Mukhtar, N. (2024). *Survival Outcomes Among Patients With Hepatocellular Carcinoma in a Large Integrated US Health System*. 7, 9, pp. 1–14. https://doi.org/10.1001/jamanetworkopen.2024.35066

Declaration

Consent for publication

Not Applicable

Availability of data and materials

The publisher has the right to make the data public

Conflict of Interest

The authors declared no conflict of interest

Ethical Considerations

Not applicable

Competing interest

The authors report no conflict or competing interest

Funding

The author declared no source of funding

Authors' Contributions

Sanusi Abdullahi Sidi conceived the study, designed the methodology, and supervised the research. Anas Tukur Balarabe developed the deep learning model and performed data analysis. Abdulrashid Sani handled dataset preparation and experimental validation. Bashar Aliyu Yauri contributed to system evaluation and interpretation of results. Zahriya L. Hassan assisted in literature review, manuscript writing, editing, and final proofreading.