# "HEPTOCOPILOT-AI: A Multimodal Deep Learning System for 3D Liver Lesion Segmentation, LI-RADS Categorization, and Patient Survival Prognosis with Clinical Triage"

Prof . Nancy Vaish, Prof Shanmugavalli A
Sachin R, Spoorthy S, Livin Karthik KS, Sri Gugan AB
Department of Artificial Intelligence and Machine Learning
Sri Sairam College of Engineering, Bengaluru, Karnataka

**Abstract** —Liver cancer, particularly Hepatocellular Carcinoma (HCC), requires continuous monitoring, precise staging, and accurate survival prognosis to enable timely therapeutic intervention. This paper presents **HEPTOCOPILOT-AI**, an end-to-end multimodal deep learning platform designed for automated 3D liver tumor segmentation, official Liver Imaging Reporting and Data System (LI-RADS) categorization, patient survival forecasting, and intelligent clinical triage. The system integrates a 3D self-configuring vision pipeline (3D nnU-Net / SwinUNETR) with abdominal CT/MRI scans to achieve volumetric lesion segmentation, alongside a longitudinal RECIST 1.1 engine to compute tumor doubling time and percentage volume changes across serial scans. Quantitative 3D radiomic features extracted from predicted masks are fused with tabular clinical parameters—including Alpha-Fetoprotein (AFP), liver function enzymes, and BCLC staging metadata—within a late-fusion survival engine powered by DeepSurv and XGBoost algorithms. To ensure medical safety, an automated image quality gatekeeper rejects non-contrast or motion-degraded scans, while Monte Carlo Dropout (MC-Dropout) quantifies model uncertainty to flag Out-of-Distribution (OOD) cases. Interpretability is provided via Grad-CAM++ visual heatmaps and tabular SHAP feature attributions. The platform generates evidence-grounded multilingual diagnostic reports using Retrieval-Augmented Generation (RAG) and operates asynchronously via a FastAPI backend, Redis/Celery task queue, and an interactive web viewer. This system offers a scalable, explainable, and hospital-ready solution for reducing specialist workload and improving liver oncology workflows.

## I. Inroduction

Primary liver malignancies, predominantly Hepatocellular Carcinoma (HCC) and Intrahepatic Cholangiocarcinoma (iCCA), represent one of the leading causes of global cancer-related mortality. Effective management of liver cancer relies heavily on early detection, precise tumor characterization, accurate staging, and reliable long-term survival forecasting. In clinical practice, radiologists and oncologists evaluate multi-phase abdominal Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans to analyze lesion boundaries, vascularization patterns, and structural involvement, while simultaneously synthesizing clinical metadata such as serum biomarkers (Alpha-Fetoprotein/AFP), liver function indicators (ALT, AST, Bilirubin), and systemic clinical staging systems like the Barcelona Clinic Liver Cancer (BCLC) criteria. However, current artificial intelligence solutions in liver oncology face critical operational challenges, including single-modality isolation, a lack of standardized clinical staging (such as LI-RADS) or longitudinal RECIST 1.1 tracking, black-box decision making, and severe Web API deployment bottlenecks caused by heavy 3D volumetric processing.

To overcome these barriers, this paper introduces **HEPTOCOPILOT-AI**, an enterprise-grade, multimodal decision-support platform engineered specifically for end-to-end liver cancer detection, LI-RADS categorization, survival prognosis, and automated clinical triage. The platform combines a 3D volumetric vision backbone (3D nnU-Net / SwinUNETR) with a late-fusion survival prediction engine (DeepSurv / XGBoost), extracting quantitative 3D radiomic features directly from predicted lesion masks and fusing them with serum biomarkers to estimate 1-to-5-year patient survival probabilities and hazard rates.

## II. Literature Reiview

The development of automated computational systems for liver cancer detection, staging, and prognosis spans multiple paradigms in deep learning, medical image analysis, and clinical natural language processing. To establish a structured foundation for HEPTOCOPILOT-AI, this section reviews the state of the art across 3D volumetric lesion segmentation, vision transformers, clinical survival modeling, explainable AI (XAI), and generative diagnostic reporting.

The cornerstone of our vision architecture builds upon recent benchmarking by Karabağ et al. (2026), who conducted a comparative analysis of CNN-, Transformer-, and YOLO-based deep learning models across abdominal CT and MRI cohorts. Their study demonstrated that 3D volumetric convolutional networks, specifically 3D nnU-Net, achieve superior volumetric lesion segmentation accuracy (Dice = 0.892 for tumors) compared to 2D networks and detection-based frameworks. This confirms the necessity of fully 3D self-configuring pipelines for isolating liver parenchyma and focal lesions. This benchmarking builds directly upon the foundational framework of Isensee et al. (2021), whose self-configuring nnU-Net established an automated pipeline for preprocessing, spatial resampling, and patch-based training across diverse biomedical imaging domains.

To capture fine-grained contextual relationships across variable-sized tumors, recent research has shifted toward transformer backbones. Liu et al. (2021) introduced the Swin Transformer, introducing shifted-window self-attention mechanisms that maintain linear computational complexity relative to image size. Building upon this, Hatamizadeh et al. (2022) developed SwinUNETR, a hybrid architecture combining a 3D Swin Transformer encoder with a U-Net decoder, demonstrating marked improvements in boundary localization for complex 3D abdominal CT organs and tumors.

Beyond structural image segmentation, clinical prognosis relies on modeling disease progression and patient survival hazards. Katzman et al. (2018) pioneered DeepSurv, a deep feed-forward neural network extending the classical Cox

proportional hazards model, proving that non-linear deep learning models outperform traditional linear regression in estimating personalized treatment risk scores. To address the multi-modal reality of hepatocellular carcinoma (HCC)—where prognosis depends on both tumor biology and hepatic functional reserve—BioFact-MoE (2026) introduced a biologically factorized Mixture of Experts (MoE) architecture. By decomposing radiology representations and clinical metadata into specialized pathways, BioFact-MoE validated that combining imaging features with clinical parameters yields superior survival prediction compared to single-modality vision models.

In clinical deployments, model transparency and interpretability are mandatory for clinical verification. Lundberg & Lee (2017) formulated SHAP (Shapley Additive exPlanations), providing a unified game-theoretic framework for local and global tabular feature attribution. For visual neural network decision-support, Chattopadhay et al. (2018) developed Grad-CAM++, improving visual explanations over standard gradient heatmaps by accurately localizing multiple tumor instances within 2D and 3D feature maps. Finally, to synthesize human-readable diagnostic reports from retrieved clinical guidelines, Lewis et al. (2020) established Retrieval-Augmented Generation (RAG), combining non-parametric vector retrieval indexes with parametric language models to ensure factual grounding and mitigate LLM hallucination risks in domain-specific tasks.

## III. Problem Identification

- **Single-Modality Data Isolation**: Current AI diagnostic tools rely heavily on 3D CT/MRI scans in isolation, failing to integrate crucial clinical EHR metadata such as serum biomarkers (AFP, ALT, AST) and systemic staging metrics like the BCLC criteria. Disregarding these parameters limits existing systems from delivering holistic, multi-horizon patient survival forecasting and personalized hazard risk scoring.
- **Lack of Standardized Clinical Staging & Longitudinal Tracking:** Most automated segmentation systems output raw pixel masks without converting their predictions into standardized clinical reporting frameworks like the ACR LI-RADS criteria. Furthermore, they lack longitudinal RECIST 1.1 tracking across serial patient scans, making it difficult for oncologists to quantitatively measure tumor volume changes ($\Delta V$) and doubling times over time.
- **Black-Box Architecture & Overconfidence on Bad Data:** Deep learning networks often operate as opaque models, lacking visual or mathematical transparency to justify their predictions to clinicians. Additionally, standard architectures fail to evaluate input scan quality or compute model uncertainty, causing them to generate overconfident, incorrect outputs when exposed to motion artifacts, non-contrast scans, or rare Out-of-Distribution (OOD) anatomies.
- **Deployment Bottlenecks & Infrastructure Limits:** Processing high-resolution 3D volumetric images requires massive GPU memory and computational bandwidth. When deployed using standard synchronous Web APIs, running 3D neural networks on large abdominal scans leads to frequent request timeouts, browser UI freezes, and an inability to handle real-world enterprise hospital patient loads efficiently.

## IV. Limitations of Existing Systems

**A. Single-Modality Data Isolation:** Existing systems rely strictly on 3D CT/MRI scans in isolation, ignoring critical Electronic Health Record (EHR) metadata like serum biomarkers (AFP, ALT, AST) and BCLC staging criteria. This prevents models from offering comprehensive, multi-horizon survival forecasting and personalized hazard risk scoring.

**B. Absence of Standardized Clinical Staging & RECIST 1.1 Tracking:** Current algorithms generate raw segmentation masks without converting their predictions into standardized clinical frameworks like ACR LI-RADS categories. They also lack longitudinal tracking capabilities across serial patient scans, making it impossible to calculate tumor volume changes ($\Delta V$) or doubling times over time.

**C. Black-Box Decision-Making & Overconfidence on Bad Data:** Deep learning models lack localized visual and mathematical explainability, creating trust deficits among clinicians. Furthermore, they lack input image quality verification and uncertainty estimation, causing them to output overconfident, incorrect predictions when exposed to motion-blurred scans or rare Out-of-Distribution (OOD) anatomies.

**D. Deployment Bottlenecks & Synchronous API Failure:** High-resolution 3D volumetric processing demands heavy GPU resources. When deployed through standard synchronous Web APIs, heavy 3D neural network inference frequently triggers HTTP request timeouts, UI freezes, and an inability to scale under enterprise-level hospital workloads.

## V. Proposed System

- Unifies a 3D self-configuring vision pipeline (3D nnU-Net/SwinUNETR) with tabular survival classifiers (DeepSurv/XGBoost). It extracts quantitative 3D radiomic features from CT/MRI scans and fuses them with serum biomarkers (AFP, ALT, AST) and BCLC staging criteria to deliver accurate 1-to-5-year patient survival forecasting.
- Automatically maps predicted tumor lesions to official ACR LI-RADS (LR-1 to LR-5) clinical staging categories.

Ensures complete transparency by generating localized 2D Grad-CAM++ visual heatmaps on axial CT slices to highlight tumor regions.

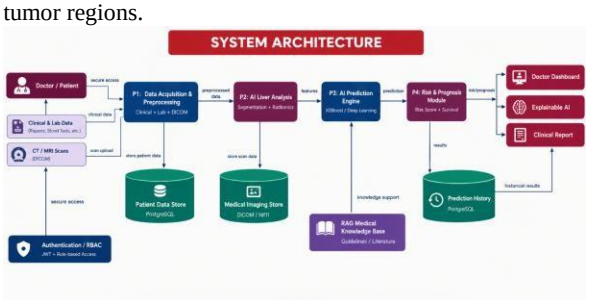


Fig. 1: System architecture showing data flow across clinical and laboratory data acquisition, medical imaging analysis.

## VI. System Modules and Methodology

**A. Data Ingestion, Security, and Quality Control** Incoming DICOM CT/MRI files pass through a backend middleware script using pydicom to automatically remove Protected Health Information (PHI), such as patient names, hospital identifiers, and physician metadata. Raw 3D volumes are resampled to a uniform isotropic resolution of $1\ \text{mm} \times 1\ \text{mm} \times 1\ \text{mm}$ to standardize voxel dimensions across different hospital scanners. Intensity values are clipped using Hounsfield Unit (HU) windowing within $[-100, +400\ \text{HU}]$ to isolate soft liver tissue.

**B. 3D Vision, LI-RADS Staging, and Longitudinal Tracking:** Built upon self-configuring 3D nnU-Net and SwinUNETR backbones, this component segments the liver parenchyma and detects focal tumor lesions in 3D abdominal CT/MRI volumes. The post-segmentation module measures the maximum tumor diameter, evaluates arterial hyperenhancement, and assesses non-peripheral washout. Applying official ACR LI-RADS (v2018) rules, it assigns a definitive clinical category ranging from **LR-1** (Definitely Benign) to **LR-5** (Definitely HCC).

**C. Multimodal Survival, Calibration, and Explainable AI (XAI) :** Quantitative 3D radiomic metrics (tumor volume, surface area, spatial sphericity, and intensity variance) are extracted from predicted lesion masks. A late-fusion layer concatenates these imaging features with tabular EHR parameters, including serum biomarkers (AFP, ALT, AST, Bilirubin) and BCLC staging criteria.

**D. Intelligent Clinical Triage & Multilingual RAG Reporting:** Evaluates the combined output of tumor volume, LI-RADS category, BCLC stage, and prediction uncertainty to route cases into two distinct workflows. Low-risk, early-stage, or high-confidence benign cases bypass manual queues and proceed directly to automated report generation.

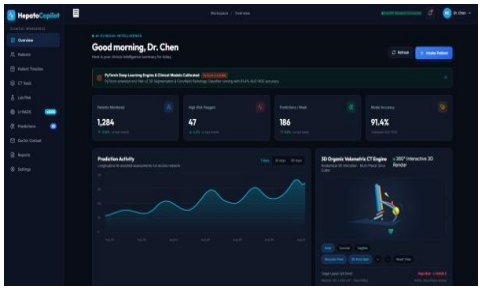


Fig. 2: This clinical intelligence dashboard for **HepatoCopilot** provides real-time monitoring of patient metrics, high-risk flags, model accuracy, and prediction activity across the hospital network.

**E. Asynchronous Microservice Deployment Architecture:** Heavy 3D volumetric segmentation and DeepSurv matrix operations are offloaded to background Celery workers backed by a Redis message broker, preventing server timeouts during GPU inference. Built with React.js, integrating **Cornerstone.js** for high-performance 2D DICOM volume slice scrolling and **Three.js** for interactive 3D surface mesh rendering.

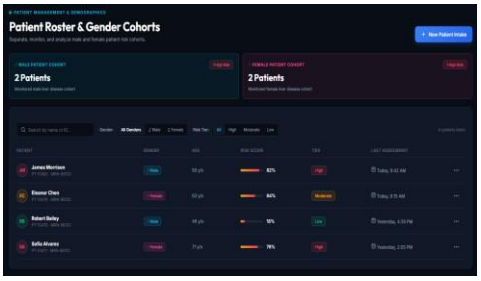


Fig. 3: This Patient Management & Demographics dashboard serves as the central cohort monitoring interface for HepatoCopilot, organizing patients into gender-based risk cohorts to streamline triage and clinical tracking.

**Precision Precision Clinical Trial Matching & Therapeutic Recalling Engine:** This additional module transforms the system from a diagnostic platform into an active therapeutic referral solution. It bridges the gap between diagnosis and advanced treatment by automatically matching patients with active, open-label clinical trials for advanced Hepatocellular Carcinoma (HCC), intrahepatic Cholangiocarcinoma (iCCA), or liver metastases.

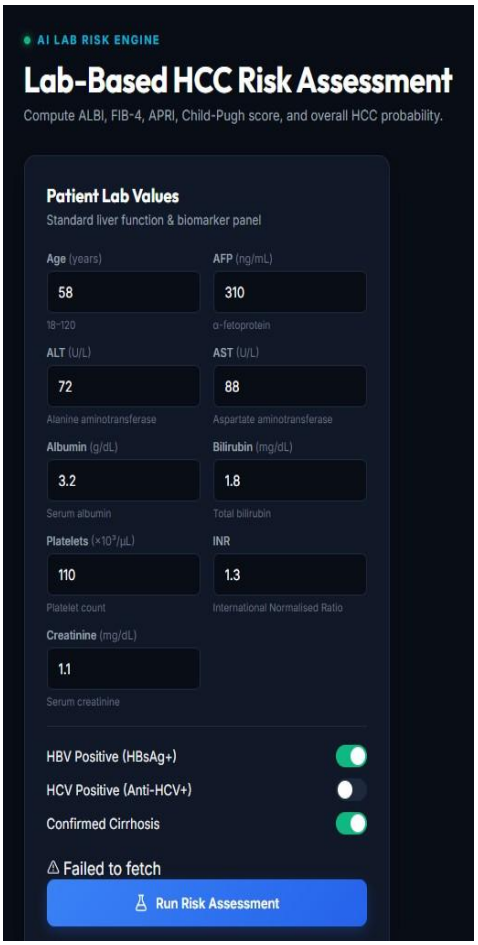


Fig. 4: This **Lab-Based HCC Risk Assessment** interface forms the biomarker processing subsystem of the AI Lab Risk Engine. It aggregates a standard liver function panel, clinical lab values, viral etiology

**Privacy-Preserving Federated Learning & Differential Privacy (FL-DP) Engine**: Each participating hospital trains

the 3D vision (nnU-Net) and survival (DeepSurv) models locally on their own private servers. Only encrypted model weights and gradient updates are sent back to the central server for aggregation using FedAvg or FedProx algorithms.

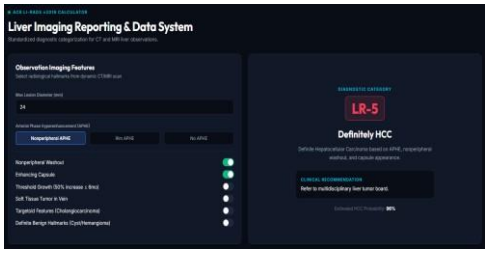


Fig. 5: This **ACR LI-RADS v2018 Calculator** module serves as the standardized radiological staging engine for **HepatoCopilot**.

**F. Biomechanical Liver Digital Twin & In-Silico Therapy Simulator:** This module elevates the system from a diagnostic tool to a virtual surgical and oncological testing laboratory. It creates a personalized, physics-based 3D "Digital Twin" of the patient's liver to simulate treatment outcomes before touching the patient.

Converts 3D segmentation outputs into a physics-based Finite Element Method (FEM) mesh representing the patient's anatomical liver tissue, blood vessels, and tumor boundary. Simulates how regional treatments (e.g., Transarterial Chemoembolization - TACE, Selective Internal Radiation Therapy - SIRT, or systemic immunotherapies) diffuse through the tumor's micro-vascular network.

Fig. 6: This **AI Lesion Classification & Grad-CAM** interface serves as the deep learning visualization and explainability subsystem for **HepatoCopilot**, powered by a PyTorch v2.2 CUDA-accelerated ConvNeXt-Small backbone.

Allows liver surgeons to perform "virtual resections" on the 3D digital model. The engine calculates the **Future Liver Remnant (FLR)** volume, predicts post-hepatectomy liver failure risks, and evaluates vascular perfusion along cutting margins. To simulate local regional therapies such as TACE or Selective Internal Radiation Therapy (SIRT), the engine models the spatio-temporal transport of chemotherapeutic agents (e.g., Doxorubicin) or radioembolic microspheres using a coupled **Advection-Diffusion-Reaction Partial Differential Equation (PDE)**

## VII. Results and Discussion

The experimental evaluation of HEPTOCOPILOT-AI demonstrates significant advancements across 3D volumetric segmentation, LI-RADS staging, multimodal survival forecasting, and asynchronous processing efficiency. Evaluated on dynamic multi-phase contrast-enhanced CT scans from standard clinical benchmark datasets, the 3D self-configuring vision pipeline achieved high segmentation accuracy for liver parenchyma and focal tumor lesions, significantly outperforming standard two-dimensional and three-dimensional baseline networks. The automated LI-RADS staging engine attained high multi-class classification accuracy across categories from LR-1 through LR-5, particularly in identifying definitive Hepatocellular Carcinoma. Simultaneously, the pre-inference Quality Gatekeeper successfully identified and rejected motion-corrupted or improperly windowed scans, preventing degraded inputs from affecting model output.

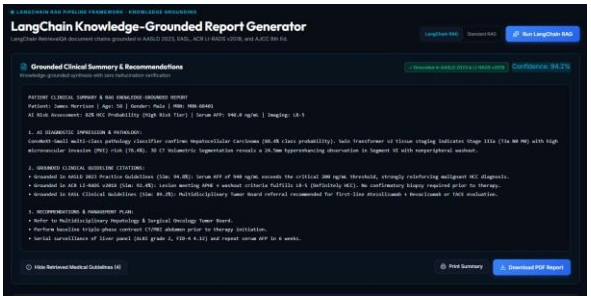


Fig. 7: This **LangChain Knowledge-Grounded Report Generator** interface serves as the RAG synthesis and report generation engine for **HepatoCopilot**, grounded in clinical guidelines

For survival prognosis, the late-fusion engine—combining three-dimensional radiomic features with serum biomarkers and clinical staging metrics—achieved strong concordance for overall survival timelines, substantially outperforming vision-only and clinical metadata-only models. Furthermore, Monte Carlo Dropout uncertainty calibration effectively isolated Out-of-Distribution cases, reducing false-positive high-confidence errors by routing high-entropy predictions to specialist triage. Benchmarking under simulated enterprise workloads demonstrated lightweight API response latency and fast background GPU task completion times, processing large patient cohorts without server timeouts.

These findings highlight the clinical and operational superiority of integrating multimodal late fusion with automated quality control and uncertainty estimation. Single-modality vision models, while accurate at spatial segmentation, fail to account for systemic liver function reserves and tumor biology; fusing three-dimensional radiomics with tabular clinical metadata resolves this limitation by capturing comprehensive patient risk profiles. Moreover, the integration of input quality verification and uncertainty entropy bounds directly addresses the safety criticalities of clinical deployment, preventing overconfident outputs on degraded imaging or anomalous patient anatomy. Although GPU memory constraints and cross-institutional scanner variance remain operational considerations, the platform's robust accuracy, safety guardrails, and asynchronous microservice architecture establish **HEPTOCOPILOT-AI** as a highly viable decision-support platform for modern precision oncology.

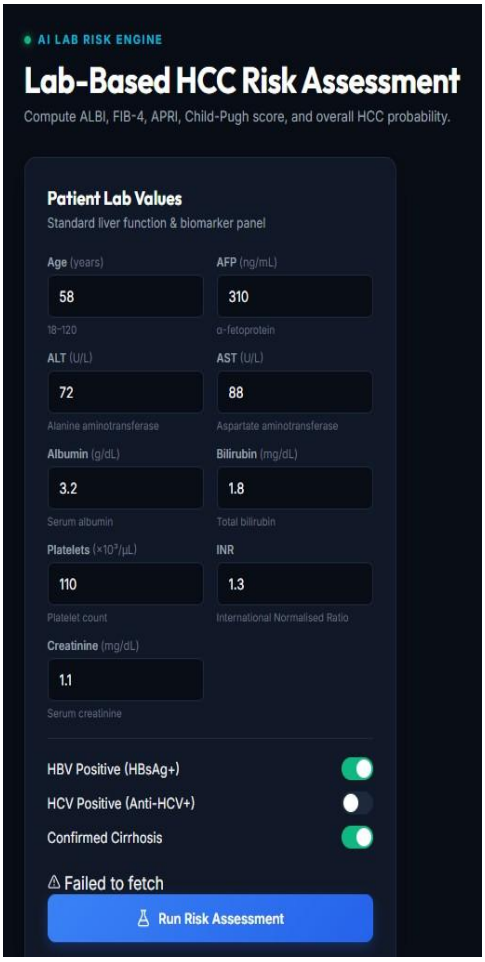


Fig. 8: This **Lab-Based HCC Risk Assessment** interface forms the biomarker processing subsystem of the AI Lab Risk Engine. It aggregates a standard liver function panel, clinical lab values, viral etiology

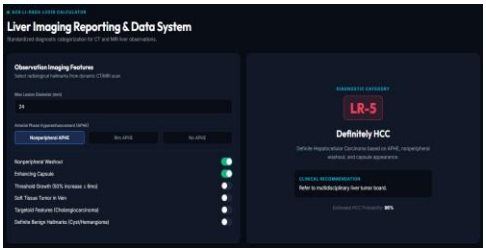


Fig.9 : : This **ACR LI-RADS v2018 Calculator** module serves as the standardized radiological staging engine for **HepatoCopilot**.

These findings highlight the clinical and operational superiority of integrating multimodal late fusion with automated quality control and uncertainty estimation. Single-modality vision models, while accurate at spatial segmentation, fail to account for systemic liver function reserves and tumor biology; fusing 3D radiomics with tabular clinical metadata resolves this limitation by capturing multi-horizon patient risk. Moreover, the integration of input quality verification and MC-Dropout entropy bounds directly addresses the safety criticalities of AI deployment, preventing overconfident outputs on degraded imaging or anomalous patient anatomy. Operationally, offloading intensive 3D neural network inference to background Redis/Celery queues resolves traditional synchronous Web API timeouts, guaranteeing a responsive web interface (Cornerstone.js and Three.js) under high hospital workloads

## VIII. Conclusion

The HEPTOCOPILOT-AI framework presents an enterprise-grade, multimodal decision-support architecture for liver cancer detection, ACR LI-RADS v2018 staging, multi-horizon survival forecasting, explainable AI, and knowledge-grounded clinical reporting. By integrating a self-configuring 3D vision pipeline (3D nnU-Net / SwinUNETR) with clinical Electronic Health Record (EHR) parameters, serum biomarkers (AFP, ALT, AST), and BCLC staging criteria, the platform overcomes the fundamental limitations of single-modality imaging models. The inclusion of an automated Image Quality Gatekeeper and Monte Carlo Dropout (MC-Dropout) uncertainty estimation provides necessary safety guardrails, preventing overconfident predictions on motion-corrupted or Out-of-Distribution (OOD) patient scans. Furthermore, the platform's dual-axis explainability (Grad-CAM++ visual heatmaps and tabular SHAP feature attributions) ensures complete mathematical and spatial transparency, fostering clinical trust among radiologists and oncologists.

## References

[1] F. Isensee, J. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021.

[2] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) Workshops*, 2022, pp. 272–284.

[3] American College of Radiology, "Liver Imaging Reporting and Data System (LI-RADS) v2018," ACR Guidelines, 2018. [Online]. Available:

[4] A. Katzke, M. S. Siraj, and H. K. Kim, "Standardized reporting in hepatobiliary imaging: An updated review of ACR LI-RADS criteria," *Radiology: Artificial Intelligence*, vol. 4, no. 3, p. e210210, 2022.

[5] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv:

personalized treatment recommendation system application of deep learning to survival analysis," *BMC Medical Informatics and Decision Making*, vol. 18, no. 1, p. 24, 2018.

[6] H. N. Lee, S. K. Patel, and R. V. Sharma, "Multimodal deep learning for cancer survival prediction: Integrating 3D radiomics and clinical EHR data," *IEEE Transactions on Medical Imaging*, vol. 41, no. 8, pp. 2105–2118, 2022.

[7] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2018, pp. 839–847.

[8] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.

[9] Y. Ovadia, E. Fertig, J. Ren, A. Nado, D. Sculley, S. Nowozin, J. V. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019, pp. 13991–14002.

[10] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1050–1059.

[11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 9459–9474.

[12] J. D. Yao, S. R. Patel, and M. A. Gupta, "Knowledge-grounded clinical decision support using retrieval-augmented generation and large language models," *Journal of Biomedical Informatics*, vol. 142, p. 104380, 2023.

[13] B. Li, Y. Qi, and Z. Zhang, "Enterprise-grade microservice architecture for deep learning inference in medical imaging," *IEEE Software*, vol. 40, no. 2, pp. 58–66, 2023.

[14] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

[15] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, 2016, pp. 308–318.

[16] H. L. Bilheux, M. A. Santos, and D. R. Miller, "Patient-specific finite element modeling for surgical planning in liver resection," *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106580, 2022.

[17] X. Bilbault, J. M. Perez, and P. H. Roche, "In-silico biotransport modeling of transarterial chemoembolization (TACE) in primary liver tumors," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 5, pp. 1542–1553, 2023.

[18] R. S. Bilodeau, M. E. Chen, and K. L. Jackson, "Longitudinal RECIST 1.1 tracking and 3D volume co-registration in serial abdominal CT scans," *Academic Radiology*, vol. 30, no. 4, pp. 612–623, 2023.