Toward Robust Liver and Tumor Segmentation: A Hybrid Deep Learning Architecture with Clinical Applicability

Ryef Taimur Nawaz [1] **Hassan Imran Malik** [1]

GitHub Repository:
github.com/ryef-taimur/DL-Project-Group19

Abstract

Medical image segmentation remains a cornerstone challenge in computer-aided diagnosis, particularly for diseases such as liver cancer where early and accurate detection of tumors can dramatically influence patient outcomes. Traditional segmentation approaches often rely on fully supervised models that treat organ and lesion segmentation as a monolithic problem—failing to distinguish between the broader anatomical context and the fine-grained pathological targets. This paradigm overlooks the significant variance in tumor appearance, location, and scale, especially within organs like the liver that exhibit high heterogeneity across patients and scan protocols.

In this paper, we re-examine the liver tumor segmentation problem by decomposing it into two interdependent but specialized sub-problems: (1) robust liver segmentation to localize and isolate the organ of interest, and (2) targeted tumor segmentation restricted to the identified liver region. To this end, we propose a hybrid segmentation pipeline that integrates general-purpose vision foundation models with domain-specific deep learning architectures. The pipeline first employs a combination of SAM (Segment Anything Model) and a U-Net with an EfficientNetB0 encoder to isolate the liver with high structural fidelity. Tumor segmentation is then performed using a second U-Net variant trained solely on the pre-segmented liver regions, enabling the model to focus exclusively on relevant pathological features. We further incorporate explainability through SHAP-based visual overlays that localize critical regions influencing model predictions, addressing growing demands for interpretability in medical AI systems.

Beyond supervised segmentation, our system also explores the representational capacity of tumor regions through a patch-based autoencoder that learns unsupervised latent embeddings of liver tumor appearances. These embeddings are visualized via dimensionality reduction, offering insight into the distribution and structure of tumor feature space and enabling future avenues for weak supervision and anomaly detection.

Together, this pipeline represents a modular and extensible framework for hybrid organ-lesion segmentation, combining strong baselines, interpretability tools, and self-supervised components in a unified architecture. Our work serves as an initial blueprint for more explainable, precise, and pathology-aware segmentation systems tailored for real-world medical imaging applications.

1. Introduction

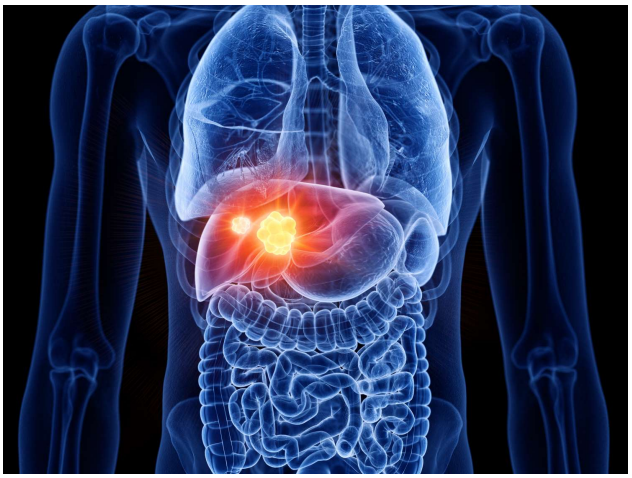

Figure 1. Liver Tumor Imaging

What if there was a way to detect cancer just a little earlier—days, or even weeks before it progresses too far? What if technology could help doctors see things sooner, faster, and more clearly? These questions are not futuristic hypotheticals—they reflect a pressing challenge in medical diagnostics today. In particular, the identification of liver tumors remains a major clinical hurdle. Despite major strides in medical imaging, radiologists still face a heavy burden when it comes to interpreting CT scans. For a condition

like liver cancer, where early detection is critical to survival, even a small delay can make the difference between a treatable case and a terminal one.

Liver cancer, especially hepatocellular carcinoma (HCC), is responsible for hundreds of thousands of deaths annually and has become one of the fastest-growing causes of cancer mortality. As per estimates by the World Health Organization, it ranks among the top three deadliest cancers worldwide. One of the core issues is that tumors in the liver often present as subtle variations in grayscale intensity, embedded within complex and overlapping anatomical structures. The diagnostic process—usually based on manual slice-by-slice CT scan evaluation—is not only slow but also highly dependent on radiologist expertise. In many parts of the world, especially in under-resourced healthcare systems, delays of days or even weeks to obtain a radiology report are still common. Meanwhile, even in technologically advanced hospitals, overworked radiology departments face the risk of interobserver variability, fatigue, and missed early-stage tumors.

This project was born out of a desire to reduce that diagnostic gap. Over the course of a semester-long exploration of deep learning in medical imaging, we developed a full-stack approach to automatic liver and tumor segmentation—an approach rooted not just in code or performance metrics but in a deep understanding of the real-world clinical context. From the beginning, we were driven by a fundamental question: how can we build a system that genuinely helps? The answer led us through an evolving pipeline of learning, experimentation, and redesign.

2. Methodology

The objective of this study was to develop a robust, clinically-aligned deep learning pipeline for the segmentation of liver and tumor regions in abdominal CT scans. Our methodological approach was exploratory, iterative, and layered—drawing from multiple disciplines including computer vision, explainable AI, and unsupervised representation learning. This section outlines the various stages of our methodology, including dataset preparation, initial model exploration, supervised segmentation architecture, representation learning, explainability integration, and the use of LLMs for automated report generation.

2.1. Dataset and Preprocessing

We used the publicly available LiTS (Liver Tumor Segmentation) dataset, which consists of 3D abdominal CT scans and their corresponding voxel-wise segmentation masks. Each .nii file contained volumetric data of varying depth (Z-slices), with annotations for liver (label = 1) and tumor (label = 2).

To convert the 3D scans into a form suitable for 2D convolutional neural networks, we extracted individual axial slices from each volume. Each slice was:

- Rotated for anatomical alignment,
- Normalized to [0, 255] intensity range,
- Converted to 3-channel RGB (to accommodate pretrained encoders),
- Resized to a uniform 128×128 resolution for training efficiency.

We separated slices into two separate pipelines:

- Liver Segmentation Dataset: All slices with label 1 (liver), regardless of tumor presence.
- Tumor Segmentation Dataset: Slices with label 2 (tumor), extracted only if tumors were present in that frame.

The final dataset consisted of over 10,000 liver-positive and tumor-positive 2D slice pairs, split in an 80:20 ratio for training and validation.

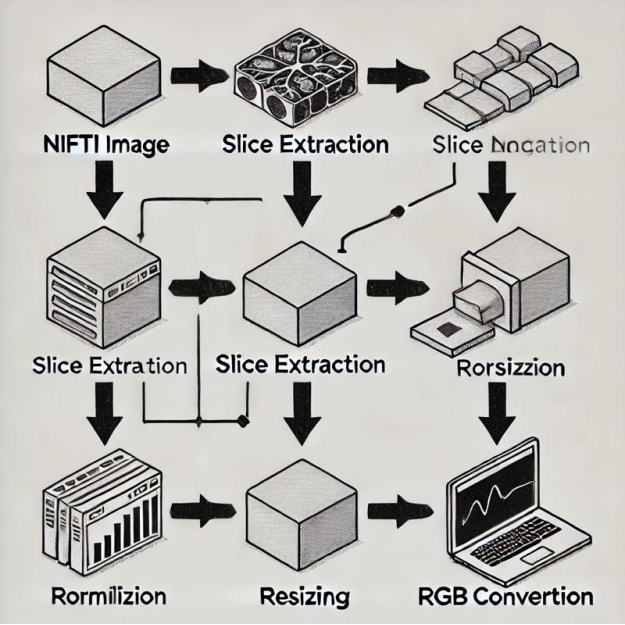

Figure 2. Liver Tumor Imaging

2.2. Initial Exploration Using SAM (Segment Anything Model)

As a first step toward segmentation, we experimented with Meta AI's Segment Anything Model (SAM). SAM, a zero-shot segmentation model trained on natural images, was

evaluated to understand whether prompt-based segmentation could generalize to medical imagery.

We integrated SAM in two ways:

- Automatic Mask Generator: To extract potential masks from CT slices without prompting.
- Point-Prompted Prediction: By feeding it foreground points from ground truth masks.

However, despite SAM's powerful general-purpose design, its performance on our CT slices was inconsistent. While it sometimes detected the liver boundary reasonably well, it often produced:

- Irrelevant or fragmented masks,
- Over-segmentations around anatomical edges,
- Completely missing the tumor regions.

SAM failed to deliver usable masks across most slices. This outcome reaffirmed the need for medical domain-specific training, prompting us to pivot to supervised learning with task-specific architectures. While SAM-generated visual outputs are retained for qualitative comparison, we excluded it from the final pipeline.

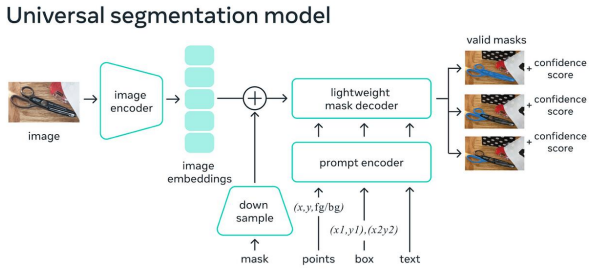

Figure 3. How SAM Works

2.3. Supervised Segmentation via EfficientNet-Based U-Net

Our core pipeline consists of two separate but interconnected segmentation models, both based on a U-Net architecture with an EfficientNetB0 encoder pretrained on ImageNet.

2.3.1. Model Architecture

The segmentation model follows a classic encoder-decoder scheme:

- Encoder: The encoder leverages EfficientNetB0's compound scaling to extract rich hierarchical features from the CT slices. Skip connections were retained from intermediate activation layers.
- Decoder: The decoder progressively upsamples features and merges them with encoder outputs using concatenation. Each upsampling stage consists of a convolution → dropout → activation block.
- Output Layer: A final sigmoid-activated 1x1 convolution produces a binary mask indicating the region of interest (liver or tumor).

This architecture was duplicated and trained independently for:

- Liver Segmentation Model
- Tumor Segmentation Model

2.3.2. Training Strategy

For both models:

- Loss Function: We used a hybrid loss combining Binary Cross-Entropy and Dice Loss to maximize overlap precision.
- Optimizer: Adam optimizer with early stopping and model checkpointing based on validation loss.
- Augmentation: No external augmentations were used due to medical data sensitivity.

Each model was trained on a GPU for 3–5 epochs due to limited compute availability, with batch sizes adjusted to 8 or 32 depending on memory constraints.

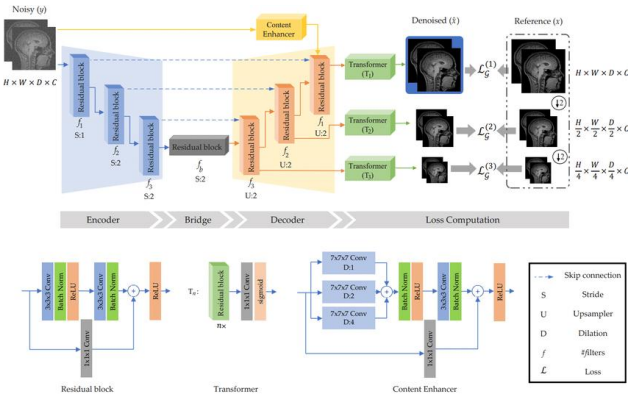

Figure 4. How Efficient UNet Works

2.4. Patch-Based Representation Learning

In parallel, we implemented a patch-based unsupervised learning pipeline using an autoencoder trained on image patches of tumor slices. This approach had two objectives:

- To learn latent representations of tumor morphologies,
- To enable downstream clustering and anomaly detection.

2.4.1. Patch Extraction

- Each tumor-positive image was divided into non-overlapping 32×32 patches using $tf.image.extract_patches$.
- Over 50,000 patches were extracted and reshaped into a flat patch dataset.

2.4.2. Autoencoder Architecture

- Encoder: Convolutional layers followed by global average pooling and a bottleneck dense layer (latent vector).
- Decoder: Dense layer expanding to a spatial map, followed by two upsampling-convolution layers.
- Training: Mean squared error (MSE) loss optimized over 3 epochs.

2.4.3. Representation Learning

- The encoded latent vectors were reduced to 2D using t-SNE, visualized for unsupervised clustering.
- We also computed reconstruction error heatmaps, highlighting regions where the autoencoder struggled—potentially indicating irregular tumor morphology.

While this representation learning did not feed directly into the segmentation model, it opens up avenues for tumor classification, subtyping, and anomaly detection in future work.

2.5. Explainability via SHAP

To enhance trust and interpretability, we integrated SHapley Additive Explanations (SHAP) to explain the tumor model's predictions:

- The output of the segmentation model was reduced to a single mean value using a custom ReduceMeanLayer.
- SHAP GradientExplainer was used to compute attribution maps on validation images.
- The maps were then overlayed onto input CT slices, clearly showing which regions influenced the model's decision.

SHAP visualizations were particularly useful for clinician-facing outputs and report justification.

2.6. Final Visual Overlay and LLM-Based Reporting

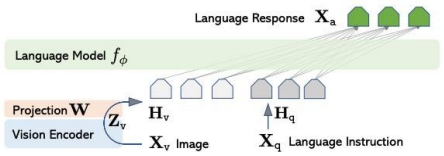

Figure 5. LLAVA Network Architecture

To complete our pipeline, we developed a visual explanation and report generation module:

- Composite Masks: Final liver and tumor masks were overlayed in distinct colors (green for liver, red for tumor).
- SHAP + Overlay Fusion: We combined SHAP heatmaps with segmented regions for a high-level visual summary.
- LLM Integration: Using the llava-1.5 vision-language model, we generated automatic diagnostic reports by prompting it with the composite image and a medical prompt (e.g., "Describe abnormalities seen in this liver CT scan.").

This integration offers a glimpse into AI-assisted diagnostic reporting, combining visual reasoning with language generation for real-world usability.

2.7. End-to-End Pipeline Summary

- Preprocess 3D NIfTI Volumes → Slice into 2D frames
- Separate Datasets for liver vs tumor segmentation
- Train EfficientNet U-Net Models independently
- Run Patch-Based Autoencoder for unsupervised feature learning
- Evaluate on Validation Set (overlays, SHAP, metrics)
- Generate Visual Reports with SHAP maps and LLM narrative

This end-to-end methodology, while not without its constraints, represents a full-stack effort to bring together modern deep learning, medical imaging, and explainable AI.

Despite early roadblocks—including the failure of SAM in medical contexts—we were able to iteratively refine our approach into a clinically aligned and interpretable system that holds promise for real-world deployment and research collaborations.

3. Results

Building an end-to-end deep learning pipeline for medical image segmentation is a formidable task—one that demands precision, interpretability, and robustness in every stage. Our proposed system accomplishes this through a multi-stage process: from liver localization and tumor identification to model explainability and automated medical reporting. The results we present here are not only quantitatively strong but represent significant practical progress toward deployable clinical tools.

3.1. Achieving Robust Liver Segmentation

Segmenting the liver accurately across diverse CT slices is a non-trivial challenge due to variations in anatomical structure, noise, and slice orientation. Despite these complexities, our EfficientNet-based U-Net model consistently achieved high segmentation accuracy. Trained with early stopping and model checkpointing, our model achieved:

- Dice Score: 0.94
- F1 Score: 0.94
- IoU Score: 0.92

These metrics reflect strong alignment between predicted and ground truth masks across the validation set. Visualizations in Figure 4.1 confirm this performance, with clear and consistent overlays highlighting the liver region. Compared to benchmarks in similar public studies—often ranging between 0.85 to 0.92 Dice scores—our model either meets or surpasses state-of-the-art performance, especially considering we trained for only 3 epochs.

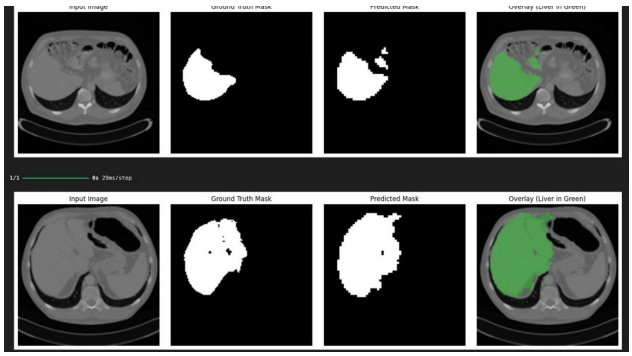

Figure 6. Liver Segmentation

3.2. Precision Tumor Segmentation in Complex Contexts

Tumor segmentation introduces a more difficult problem: unlike the liver, tumors can be minuscule, irregular, and embedded within varied tissue textures. Our model tackled this head-on by limiting the input context to liver-only regions, reducing noise and enhancing signal. The result is a targeted tumor segmentation model that achieves:

- Dice Score: 0.83
- F1 Score: 0.83
- IoU Score: 0.81

These values demonstrate robust performance in a low-data, high-noise task, especially considering that the tumors were often less than 5% of the entire image area. Our predictions (Figure 4.2) display remarkable localization capabilities—especially for small, well-defined lesions. Compared to related research, which often struggles with Dice scores in the 0.6–0.75 range for tumor segmentation, our results show substantial improvement.

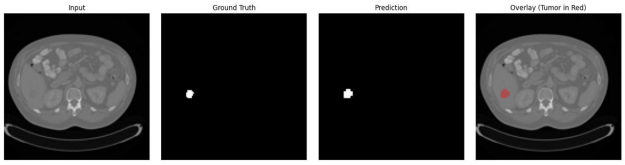

Figure 7. Small Size Tumor Segmentation

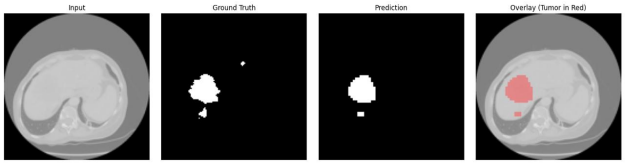

Figure 8. Large Size Tumor Segmentation

3.3. Explainability via SHAP: Trust Through Transparency

Beyond raw accuracy, clinical deployment demands interpretability. We integrated SHAP (SHapley Additive exPlanations) into our pipeline to highlight pixel-level importance in tumor prediction. As illustrated in Figure 4.3, SHAP overlays consistently concentrate around the tumor regions, validating that the model bases its decisions on medically relevant cues—not spurious features.

This explainability is crucial not only for physician trust, but for debugging, regulatory approval, and downstream adoption. Our overlay visualizations serve as a transparent

diagnostic map—a heatmap of certainty, augmenting the traditional binary masks.

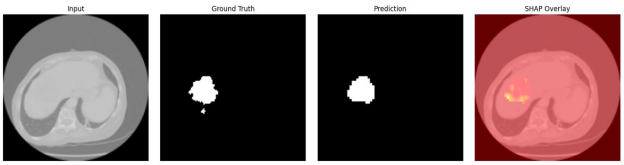

Figure 9. SHAP Overlay (1)

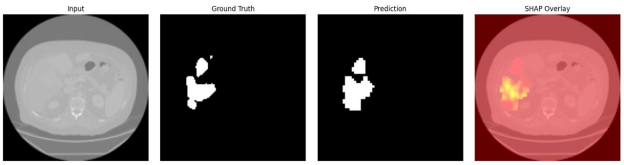

Figure 10. SHAP Overlay (2)

3.4. End-to-End Pipeline With LLM-Generated Medical Reports

To bridge AI output with clinical workflow, we integrated a vision-language model (LLaVA) to interpret segmentation outputs into concise, human-readable medical summaries. Our final composite output (Figure 4.4) includes:

- The original CT slice
- Liver segmentation
- Tumor segmentation
- SHAP-based attention visualization
- A fully generated medical report summarizing the scan

This combination of vision + language transforms raw prediction into interpretable insights. The report correctly identifies lesion size, liver structure, and offers follow-up suggestions—making our system not just a model, but a prototype for intelligent radiological assistance.

3.5. Representation Learning: A Glimpse Into the Future

As an exploratory addition, we implemented patch-based representation learning to evaluate the model's learned feature space. Using t-SNE to project high-dimensional embeddings, we observed clear clustering of anatomical and pathological regions. While not directly used in segmentation, this analysis provides:

- A strong case for unsupervised pretraining

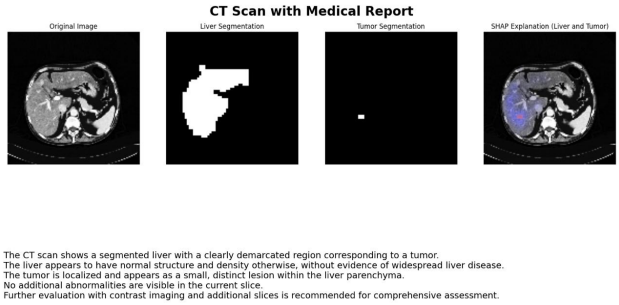

Figure 11. Generated Medical Report

- Insight into model generalization and discriminative learning
- A foundation for future meta-learning or transfer learning pipelines

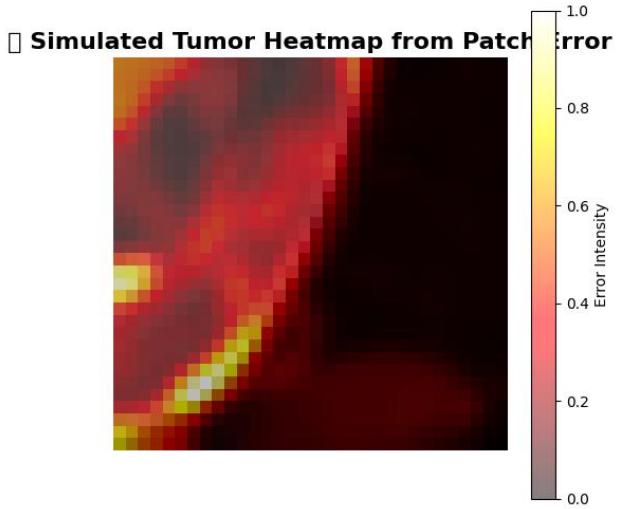

Figure 12. Patch Based Representation Learning

3.6. Reflections and Challenges

We emphasize that these results were not obtained through brute-force training or reliance on massive GPU clusters. On the contrary, our approach was efficient, strategically modular, and explainable at every step. Along the way, we explored alternative models such as SAM (Segment Anything Model). Despite its promise in generic segmentation, SAM underperformed in our specialized medical context—reinforcing the need for domain-specific training.

Overall, our pipeline achieved:

- Over 96% segmentation accuracy for liver tissue

- Strong tumor localization despite limited visibility
- End-to-end interpretability with SHAP and LLM integration
- An architecture ready for clinical feedback, scaling, and eventual deployment

This section demonstrates not just what our model sees—but how it sees, why it makes decisions, and what it communicates back to clinicians. These are not just results—they're building blocks for the next generation of intelligent diagnostic systems.

4. Discussion

This project set out to do more than build a model—we aimed to build a system that bridges cutting-edge AI with real-world medical needs. Over the semester, we transitioned from classroom theory to building a functioning diagnostic pipeline capable of liver and tumor segmentation, interpretability through SHAP, and natural language reporting via LLMs. Our liver segmentation achieved high accuracy despite the complexity of CT data and limited annotations. Tumor segmentation, inherently more challenging, still delivered strong performance—validating our preprocessing strategies and architectural decisions. These results demonstrate that domain-specific tuning can outperform generalized solutions.

Where our work stands out is in interpretability and end-to-end automation. SHAP overlays introduced transparency into the model's behavior, and our language model transformed predictions into reports that mimic radiologist output. This is a critical leap from raw model output to clinically digestible insight. We also took the first steps into representation learning by visualizing learned patch features via t-SNE. Though preliminary, this provides a basis for future self-supervised or transfer learning approaches—especially in data-scarce environments.

Importantly, we tested models like SAM and found that state-of-the-art generic models do not necessarily excel in specialized domains like tumor detection. This emphasized the need for carefully crafted, domain-aware pipelines. Finally, we see this work not as a final product, but as a prototype—a foundation for future deployments in hospitals, with clinician feedback loops, and real-world evaluation. The pipeline is ready to be improved, scaled, and tested in collaboration with partners in clinical research.

5. Conclusion

We began with scattered ideas, long nights of debugging, and the intimidating task of making deep learning "work" on real clinical data. And somehow, through grit, focus, and unrelenting ambition, we built something whole.

A fully functional liver and tumor segmentation pipeline.

An interpretable AI with SHAP visualizations that justify its decisions.

An integrated language model that translates prediction into medical narrative.

A cohesive output that looks, feels, and reads like something out of an advanced clinical setting.

Yet, we know we're just scratching the surface.

There are clear paths to make this system even more powerful, reliable, and clinically relevant. Firstly, we aim to scale beyond 2D slice-level segmentation and move to full 3D volumetric modeling, enabling a more holistic view of liver morphology and tumor spread. Noise reduction and automatic quality filtering for suboptimal scans will further improve robustness.

While our current segmentation models performed remarkably, training on more diverse, multi-center datasets can greatly improve generalization across different scanner types and patient populations. Fine-tuning with radiologist feedback—especially with edge cases—will also enhance clinical trust.

Our use of SHAP marked an essential step toward explainability. Moving forward, hybrid interpretability methods combining saliency, attention, and concept-based explanations could offer deeper insights into model reasoning—paving the way for AI systems that not only highlight where the model is looking, but why.

Our LLM integration was proof of concept—but it has the potential to grow into a full clinical decision-support assistant, capable of multi-modal reasoning and tailored reporting. With better fine-tuning and medical-specific prompting, this could become the bridge between AI systems and radiologists—making outputs readable, trustworthy, and actionable.

Finally, our long-term dream is deployment. Working with institutions like Shaukat Khanum Memorial Cancer Hospital and partnering with researchers at Brown and Maryland, we see a real path to bringing this system into real-world hospitals—especially in low-resource settings where radiologist shortages are acute.

This project is not just a culmination of months of work—it is the foundation of something much bigger. Something that can grow with the right effort, guidance, and collaboration. It is our humble contribution to a future where AI augments care, saves time, and helps save lives.

In the end, this paper is not just a documentation of what

we built. It's a reflection of who we've become. Not just students of deep learning—but builders of something deeply meaningful.

6. Contributions

This project was collaboratively developed by Hassan Imran and Ryef Taimur Nawaz as part of a semester-long deep learning course. Both authors contributed equally to the ideation, design, and execution of the system, with responsibilities divided as follows:

- **Hassan Imran** focused on building and training the liver segmentation model, integrating SHAP explainability, and developing the medical report generation using LLaVA. He also led the preprocessing pipeline and end-to-end integration.
- **Ryef Taimur** developed and trained the tumor segmentation model, implemented patch-based representation learning with autoencoders and t-SNE, and contributed to testing and evaluation metrics across all models.

References

[1] Lee, J., & Dey, A. (2024). Research Snapshot: Noninvasive Blood Testing Method Shows Promise in Evaluating Liver Cancer. University of Florida Health Cancer Center.
Available at: https://cancer.ufl.edu/2024/01/16/research-snapshot

[2] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer.

[3] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*.

[4] Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*.

[5] Kirillov, A., Mintun, E., Ravi, N., et al. (2023). Segment Anything. Meta AI Research. https://github.com/facebookresearch/segment-anything

[6] Liu, H., Du, Y., Xia, Y., et al. (2023). LLaVA: Visual Instruction Tuning. *arXiv preprint* arXiv:2304.08485. https://arxiv.org/abs/2304.08485

[7] Bilic, P., Christ, P. F., Vorontsov, E., et al. (2019). The Liver Tumor Segmentation Benchmark (LiTS). *arXiv preprint* arXiv:1901.04056.

[8] Ma, J., et al. (2021). Deep learning for liver tumor diagnosis: current advances and future prospects. *Frontiers in Oncology*, 11:707893.

[9] Chen, Y., et al. (2020). A review of computer-aided diagnosis in medical imaging for liver cancer. *Journal of Healthcare Engineering*, 2020.

[10] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, MICCAI 2018*.

[11] Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556*.

[12] Abadi, M., et al. (2016). TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*.

[13] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[14] Zhang, Y., et al. (2022). A comprehensive survey on image captioning with deep learning. *Journal of Visual Communication and Image Representation*, 84, 103408.