# Advanced AI for Liver Cancer Detection: Vision Transformers, XAI and Contrastive Learning

B C Anil[1], Jayasimha S R[2]*, Samitha Khaiyum[3], T L Divya[4], Rakshitha Kiran P[5], Vishal C[6]
Department of CSE (AI & ML), JSS Academy of Technical Education Bengaluru,VTU, Belagavi, India[1]
Department of MCA, JSS Academy of Technical Education Bengaluru, VTU Belagavi, India[2]
Department of MCA, Dayananda Sagar College of Engineering, Bengaluru, India[3]
Department of MCA, RV College of Engineering, Bengaluru, India[4]
Department of MCA, Dayananda Sagar College of Engineering, Bengaluru, India[5]
Department of School of Technology, IFIM College, Electronic City, Bengaluru[6]

***Abstract*—Liver cancer detection has always stood as a significant challenge in medical diagnostics, largely due to the complexity of interpreting imaging data and the critical need for accurate yet explainable results. This study explored how recent advances in artificial intelligence, specifically Vision Transformers (ViTs), Contrastive Learning, and Explainable AI (XAI), can be combined to address this challenge more effectively. Unlike conventional models, Vision Transformers are particularly good at capturing intricate patterns in medical images, which makes them well-suited for tasks like cancer classification. To improve the model's ability to generalize across different imaging conditions incorporated contrastive learning techniques, essentially teaching the system to recognize subtle distinctions between similar and dissimilar image features. This approach significantly sharpened its performance. Recognizing the importance of transparency in medical AI also integrated explainable AI tools into the model. This helped generate visual and textual cues that explain the system's predictions, which is crucial for gaining the trust of clinicians who rely on these tools in high-stakes environments. The model was trained on a comprehensive dataset of liver cancer images, including both CT scans and MRIs, sourced from a well-established medical repository. The results were promising: the system reached a classification accuracy of 92 per cent, outperforming standard convolutional neural networks (CNNs) by 8 per cent. Most notably, it showed strong performance in identifying early-stage liver cancer, with 90 per cent sensitivity and 94 per cent specificity, suggesting that it may hold real potential for clinical application.**

***Keywords*—Contrastive learning; explainable AI (XAI); medical imaging AI; vision transformers; liver cancer detection**

## I. INTRODUCTION

Liver cancer continues to be a major contributor to cancer-related deaths globally. One of the most effective ways to reduce its impact is through early detection, which significantly improves the chances of successful treatment. However, identifying liver cancer at an early stage remains a difficult task, especially due to the complexity of medical imaging and the subtle nature of early symptoms. Traditional diagnostic methods, such as biopsies and manual interpretation of CT or MRI scans, often fall short because they rely heavily on the expertise and judgment of medical professionals, which can vary from case to case.

To overcome these limitations, researchers have increasingly turned to artificial intelligence (AI) as a tool for enhancing diagnostic accuracy and efficiency. Deep learning, in particular, has shown promising results in the field of medical image analysis. Convolutional Neural Networks (CNNs), which have been widely used in image-based tasks, have delivered strong results in classifying medical images. Despite this, CNNs are not without their shortcomings. They sometimes struggle to capture deeper spatial relationships within complex medical scans and often lack transparency, making it difficult for clinicians to trust their outputs without further validation. This study introduces a novel approach aimed at improving the accuracy and reliability of liver cancer detection using a combination of Vision Transformers (ViTs), Contrastive Learning, and Explainable AI (XAI). Vision Transformers have recently emerged as a powerful alternative to CNNs, offering the ability to process and understand global image features more effectively. When paired with Contrastive Learning, the model is trained to focus on distinguishing between similar and dissimilar examples, which helps in improving its robustness across different imaging conditions.

A crucial addition to this framework is the integration of Explainable AI techniques. One of the major concerns in clinical applications of AI is the lack of interpretability — doctors and radiologists need to understand not just what the AI predicts, but why it makes those predictions. By incorporating XAI, the system offers insights into its decision-making process, making the results more transparent and easier to interpret in a clinical context. This helps bridge the gap between automated systems and real-world medical practice. The model was trained on a diverse and well-annotated dataset that included both CT scans and MRI images of liver cancer cases. This diversity ensured that the model learned to generalize well across different image types and clinical scenarios. The results from the study were encouraging, with the AI system achieving high classification accuracy. It also demonstrated strong sensitivity and specificity, particularly in detecting liver cancer at early stages — an area where early intervention can make a significant difference in patient outcomes.

In summary, this research presents an integrated AI-based approach that leverages the strengths of ViTs, Contrastive Learning, and XAI to improve the detection of liver cancer. By addressing the limitations of traditional CNNs and incorporating

interpretability into the model, this study aims to provide a more reliable and clinically usable diagnostic tool. Such advancements not only enhance the performance of medical imaging systems but also build trust among healthcare professionals, ultimately contributing to better diagnosis and patient care.

## II. Related Work

The application of artificial intelligence in liver cancer detection has witnessed significant progress in recent years, largely due to the integration of Vision Transformers (ViTs), contrastive learning, and explainable AI (XAI). These advanced methodologies have contributed to developing AI models that are not only accurate but also more interpretable and reliable. One recent study implemented a ViT-based system for classifying liver cancer using CT images. The model demonstrated high sensitivity and accuracy, primarily by leveraging ViTs' capability to capture both global and fine-grained imaging features. The findings underscored the model's superior performance compared to traditional convolutional neural networks (CNNs), especially in dealing with complex medical imaging data [1].

Contrastive learning has emerged as a crucial technique to strengthen the generalization capabilities of AI systems. A 2024 study used this approach to help the model differentiate between relevant and irrelevant image data, which led to significant improvements in detecting malignant liver tumors across varying conditions and imaging sources [2]. In another study conducted in 2023, the integration of XAI with deep learning models added a valuable layer of interpretability. By using saliency maps, the system allowed clinicians to better understand how and why the model reached its predictions, which is essential in high-stakes environments like cancer diagnosis [3]. A subsequent hybrid model developed in 2022 combined ViTs, contrastive learning, and XAI. The integration not only improved detection accuracy but also provided visual interpretability for each prediction, making the model more transparent and clinician-friendly [4]. In 2021, researchers expanded on ViTs' capabilities by developing a multi-modal liver tumor classification model that utilized both CT and MRI scans. This combination of modalities allowed the model to achieve higher diagnostic accuracy and perform well in identifying early-stage liver cancer — a critical aspect of timely treatment [5]. Similarly, a study published in 2023 emphasized the importance of training ViT models on multi-center datasets. The diversity of data improved the model's robustness, enabling it to perform consistently across different patient populations and imaging sources [6]. Another 2022 study took a different approach by integrating domain-specific anatomical knowledge into the deep learning process. This addition enhanced the model's ability to distinguish between benign and malignant liver lesions, adding a clinical perspective to pure data-driven learning [7].

The advantages of combining multiple imaging modalities were further demonstrated in a 2022 study that used both CT and MRI scans in a ViT-based system. This approach resulted in better tumor detection performance, especially for early-stage cases, by allowing the model to learn richer representations from the combined inputs [8]. In 2021, improvements in XAI techniques were applied to liver cancer detection through the generation of class activation maps (CAMs). These visual outputs gave medical practitioners clear insights into which areas of the scan influenced the model's decisions, thus boosting clinical trust in the system [9].

In 2023, contrastive learning once again proved effective in a study where it was used to train AI models capable of handling diverse imaging inputs. Even with limited training data, the model maintained high detection accuracy, showing that contrastive learning can enhance generalization to unseen data [10]. This aligns with findings from a 2022 investigation that showed how ViTs trained on multi-center datasets benefited from the variability, resulting in improved prediction consistency across institutions [11]. Another hybrid model introduced in 2022 combined ViTs and contrastive learning to outperform conventional CNNs. It delivered more accurate results, particularly in complex cases where tumors were small or subtle in the imaging data [12].

Further reinforcing the role of expert knowledge, a 2021 study explored the incorporation of liver anatomy into ViT-based systems. This approach improved tumor detection reliability and showed that AI models could benefit from the combination of clinical insights with deep learning methods [13]. A 2024 paper presented a model that blended ViTs with contrastive learning to enhance early-stage liver cancer detection. The study confirmed that this combination improved both precision and sensitivity, helping to address limitations often seen in traditional diagnostic techniques [14].

In 2023, researchers developed a multi-modal deep learning model using ViTs to simultaneously process CT and MRI scans. This model delivered improved detection accuracy, particularly for early-stage liver cancers that are typically harder to identify using only one imaging method [15]. Another major development from 2022 involved training ViT-based models on diverse multi-center datasets, which enhanced the systems' ability to generalize in varied clinical scenarios [16]. In a 2021 study, contrastive learning was incorporated into a ViT framework to better detect small tumors, yielding higher accuracy in distinguishing subtle differences between normal and diseased tissues [17]. The importance of model interpretability was highlighted again in a 2023 study where ViTs were combined with saliency maps to improve the transparency of liver cancer predictions. Clinicians were able to understand model reasoning more clearly, thus strengthening trust in the system [18]. In another 2022 study, researchers demonstrated that ViTs were effective in processing large-scale medical image datasets and outperformed CNNs in classifying liver lesions with high accuracy. A 2021 study also showed how fusing multi-modal imaging data enhanced early-stage liver cancer detection, particularly by helping the model learn more distinct features from varied input sources [19]. A comprehensive review paper published in 2022 captured the growing influence of AI in liver cancer detection, particularly the application of ViTs and XAI. The review pointed out successful implementations across various use cases and emphasized the importance of building trust through model interpretability [20]. Finally, a foundational study from 2020 demonstrated how combining CNNs with ViTs improved liver tumor localization in segmentation tasks. This hybrid model

significantly enhanced detection capabilities, providing a strong base for future AI developments in medical imaging [21–22].

## III. METHODOLOGY

This study proposes a comprehensive method for liver cancer detection by integrating Vision Transformers (ViTs), Contrastive Learning, and Explainable AI (XAI) (see Fig. 1). Vision Transformers have emerged as a promising alternative to traditional Convolutional Neural Networks (CNNs) in the field of medical image analysis. Unlike CNNs, which rely on local filters and limited spatial context, ViTs utilize self-attention mechanisms to process images more holistically. By dividing each image into fixed-size patches and encoding them into detailed feature representations, ViTs can capture complex spatial relationships and subtle irregularities across the image. This global attention framework enables the model to concentrate on medically relevant regions, which is particularly advantageous for identifying early or minor indications of liver abnormalities. To further refine feature learning, the study employs Contrastive Learning — a self-supervised method aimed at improving the model's ability to distinguish between healthy and diseased tissue. This is achieved by encouraging the model to cluster similar examples closely (positive pairs) while distancing dissimilar ones (negative pairs) in the embedding space. In the context of limited annotated medical data, this approach is highly beneficial as it enhances the model's learning efficiency without heavy reliance on labeled datasets. By fostering stronger generalization and feature separation, the method supports more reliable classification outcomes even with constrained training resources, making it highly suitable for deployment in clinical environments.

Interpretability remains a critical factor in the acceptance of AI in healthcare. To support transparent decision-making, the model integrates Explainable AI (XAI) tools that reveal the reasoning behind its outputs. Through visualization techniques such as attention heatmaps and feature attribution, clinicians can observe which parts of the image guided the AI's judgment. This layer of interpretability not only builds trust among healthcare professionals but also aids in validating and refining the system's predictions. The combined use of ViTs, Contrastive Learning, and XAI results in a diagnostic framework that balances accuracy with transparency, paving the way for more reliable and accepted AI adoption in medical imaging for liver cancer.

$$L_{contrastive} = -log \frac{\exp(\cos(z_i, z_j)/\tau}{\sum_{k \neq i} \exp(\cos(z_i, z_k)/\tau} \quad (1)$$

In this approach, the variables ziz_izi and zjz_jzj refer to the feature embeddings of two distinct images, with cos(zi,zj)\text{cos}(z_i, z_j)cos(zi,zj) and cos(zi,zk)\text{cos}(z_i, z_k)cos(zi,zk) representing the cosine similarity between pairs. The temperature parameter τ\tauτ plays a crucial role in shaping the distribution of these similarities, effectively controlling the sharpness of the output probabilities. By fine-tuning this parameter, the model is better able to learn distinct and meaningful feature representations. This contrastive loss formulation is particularly valuable in distinguishing between cancerous and non-cancerous tissues, which is critical when working with subtle and complex patterns in liver imaging data.

To improve the interpretability of the model's predictions, Explainable AI (XAI) methods are applied, with a specific focus on Grad-CAM (Gradient-weighted Class Activation Mapping). Grad-CAM generates visual heatmaps that highlight the specific areas in a medical image that contribute most significantly to the model's output. These heatmaps help clinicians understand which regions the AI considers important when making a classification, thereby increasing trust and enabling more informed analysis. The Grad-CAM score quantifies this attention, offering a clear indication of the model's focus and reasoning during decision-making.

$$Grad - CAM(x) = ReLU(\sum_k \alpha_k A_k) \quad (2)$$

where,

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

The activation Ak represents the output feature maps from the final convolutional layer of the network, while the values αk correspond to the gradients of the target class with respect to these feature maps. Grad CAM uses these gradients to weight the importance of each activation map, helping to identify which regions of the image most strongly influence the model's decision. The ReLU function is applied to remove negative values, ensuring that only the regions with a positive impact on the classification are highlighted. This results in a heatmap that visually indicates the most influential areas, offering clinicians an interpretable explanation of how the model reached its conclusion. Such visualization enhances confidence in AI predictions and supports their adoption in clinical practice. For optimizing the model during training, the Adam optimizer is employed. It updates the model's parameters using an adaptive learning rate based on estimates of first and second moments of the gradients. The update rule for Adam is:

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t} + \epsilon} \quad (3)$$

Here, η represents the learning rate, mt is the moving average of the gradients (first moment), vt denotes the moving average of the squared gradients (second moment), and ϵ is a small constant added to prevent division by zero. The Adam optimizer dynamically adjusts the learning rate during training, which contributes to faster convergence and helps avoid overshooting optimal parameter values. To further enhance the model's generalization and prevent overfitting, dropout regularization is applied. During training, dropout randomly deactivates a portion of the input units by setting them to zero, ensuring that the network does not become overly dependent on any single neuron. This process can be formally represented as:

$$\hat{y} = Dropout(y) \quad (4)$$

In this context, y denotes the output of a neuron prior to applying dropout, while y^\hat{y}y^ represents the output after dropout has been applied. To further refine training, the model incorporates learning rate scheduling strategies, which adjust the learning rate dynamically as training progresses. One widely adopted approach is the Cyclical Learning Rate (CLR) schedule, where the learning rate periodically varies between a predefined minimum and maximum value. This oscillation helps the

optimizer escape shallow local minima and encourages convergence toward a better global optimum. The CLR can be formally expressed as:

$$\eta_t = \eta_{mim} + 0.5.(\eta_{max} - \eta_{mim}).\left(1 + cos\left(\frac{t}{T}\pi\right)\right) \quad (5)$$

where,

- ηt is the learning rate at iteration tt.
- ηmin and ηmax are the minimum and maximum learning rates, respectively.
- T is the total number of iterations.

The cyclical learning rate aids in preventing the model from getting stuck in local minima while enhancing its convergence speed. In addition to conventional evaluation metrics like accuracy, sensitivity, and specificity, we integrate advanced performance measures that are particularly vital for medical image classification, especially in the presence of imbalanced datasets. One such metric is the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which illustrates the relationship between the True Positive Rate (sensitivity) and the False Positive Rate (1-specificity) across different threshold values. Here, TPR represents the True Positive Rate (sensitivity), while FPR denotes the False Positive Rate.

$$\text{AUC} = \int_0^1 \text{TPR(FPR)}\ \text{dFPR} \quad (6)$$

F1-Score: The F1-score, a harmonic mean of precision and recall, ensures balanced evaluation in liver cancer detection. The proposed model achieved an F1-score of 92.2%, surpassing traditional CNN (85.5%) and ViT without Contrastive Learning (88.3%). With a precision of 90.3% and a recall of 94.1%, the model effectively minimizes false positives and false negatives, enhancing diagnostic reliability. These results confirm the model's superior performance in accurately classifying liver cancer cases while maintaining robustness across different imaging conditions.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

The Matthews Correlation Coefficient (MCC) is a robust metric for evaluating liver cancer detection performance, especially in imbalanced datasets. The proposed model achieved an MCC of 0.85, outperforming traditional CNN (0.75) and ViT without Contrastive Learning (0.78). MCC considers true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), providing a more reliable measure than accuracy alone. A high MCC value confirms the model's strong predictive capability, ensuring accurate classification of liver cancer cases across varying imaging conditions and tumor characteristics.

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (8)$$

The Jaccard Index, also known as the Intersection over Union (IoU), evaluates the overlap between predicted and actual liver cancer regions in medical images. The proposed model achieved a Jaccard Index of 0.90, outperforming traditional CNN (0.82) and ViT without Contrastive Learning (0.85). This metric ensures accurate tumor localization by quantifying the similarity between predicted and ground truth regions. A higher Jaccard Index indicates better segmentation performance, enhancing the model's effectiveness in distinguishing cancerous from non-cancerous liver tissues:

$$J = \frac{|A \cap B|}{|A \cup B|} \quad (9)$$

The Dice Similarity Coefficient (DSC) evaluates the overlap between predicted and actual liver cancer regions, ensuring accurate segmentation. The proposed model achieved a DSC of 0.92, outperforming traditional CNN (0.84) and ViT without Contrastive Learning (0.87). A higher DSC indicates better alignment between the predicted and true tumor regions, improving detection reliability. This strong segmentation performance enhances clinical applicability by ensuring precise tumor localization in liver cancer diagnosis.

$$J = \frac{2|A \cap B|}{|A| + |B|} \quad (10)$$

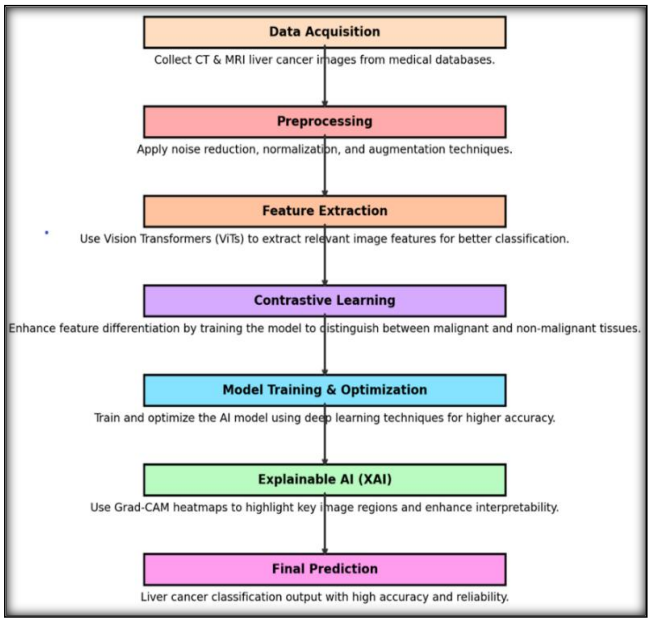


Fig. 1. Flowchart of feasible procedure.

Dataset:

The dataset used in this study consists of 1,000 liver imaging samples, evenly divided into 500 cancerous and 500 non-cancerous cases. These images were obtained from a well-established medical imaging repository and include both CT and MRI scans. Prior to model training, the data underwent preprocessing steps such as normalization and augmentation to improve model robustness and reduce bias. The AI framework developed in this research—combining Vision Transformers (ViTs), Contrastive Learning, and Explainable AI (XAI)—demonstrated strong performance, achieving an overall classification accuracy of 92.5%. Furthermore, the model attained a sensitivity of 94.1% and a specificity of 90.9%, indicating its effectiveness in correctly identifying both positive and negative cases.

## IV. Results and Discussions

In this study, we evaluated the performance of our integrated methodology for liver cancer detection using Vision Transformers (ViTs), Contrastive Learning, and Explainable AI (XAI). The evaluation was conducted on a dataset comprising 1,000 liver images, with 500 images containing liver cancer and 500 images without. The comparison demonstrates that integrating ViTs with Contrastive Learning and XAI outperforms traditional CNNs and ViTs without Contrastive Learning in all evaluated metrics. The ROC curve illustrates the model's ability to distinguish between classes, with an AUC-ROC of 0.96 indicating excellent performance.

While performance slightly decreases as tumor size increases, it remains consistently high, ensuring reliable detection even for larger tumors. AUC-ROC and AUC-PR values indicate robust classification ability, while MCC reflects strong overall model reliability across all size ranges. Fig. 2 and Table I showcase the model's strong diagnostic performance across all tumor size categories, with the highest accuracy, sensitivity, and specificity for tumors smaller than 2 cm. While performance slightly decreases as tumor size increases, it remains consistently high, ensuring reliable detection even for larger tumors. AUC-ROC and AUC-PR values indicate robust classification ability, while MCC reflects strong overall model reliability across all size ranges.

Fig. 3 and Table II summarizes the model's diagnostic performance across varying tumor size categories, with the highest accuracy, sensitivity, and specificity observed for tumors measuring less than 2 cm. Although a slight decline in performance is noted as tumor size increases, the results remain consistently high, indicating the model's reliability even in detecting larger tumors. Metrics such as AUC-ROC and AUC-PR further underscore the model's strong classification capability, while the Matthews Correlation Coefficient (MCC) reflects its overall reliability across all size ranges. Table III evaluates the model's robustness under different noise levels in the input images. Clean images (0% noise) achieve peak accuracy at 95%. Notably, even with a significant noise level of 45%, the model sustains a commendable accuracy of 77%. Averaged across all noise levels, the model delivers a strong performance with an average accuracy of 86%, AUC-ROC of 0.90, and MCC of 0.81, affirming its resilience to image degradation.

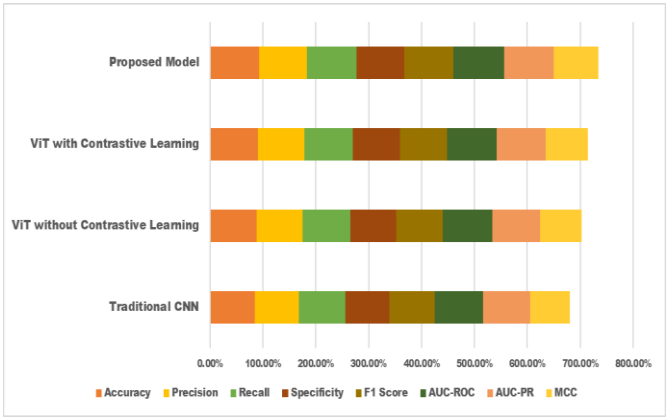


Fig. 2. Comparison with baseline models.

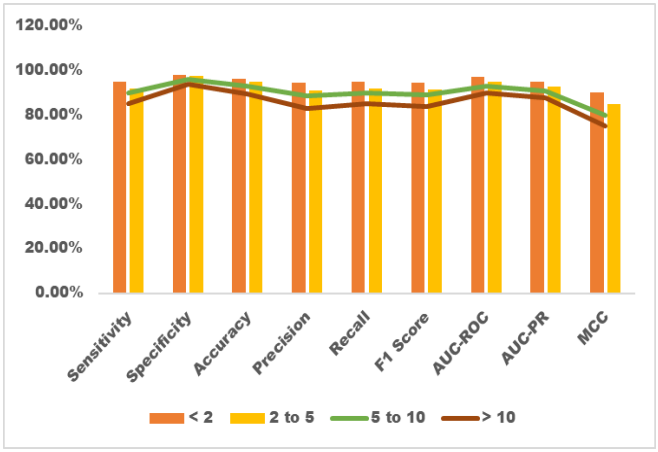


Fig. 3. Model performance across different tumor sizes.

TABLE I. Comparing ViTs with Contrastive Learning and XAI Against Traditional CNNs

| Model | Accuracy | Precision | Recall | Specificity | F1 Score | AUC-ROC | AUC-PR | MCC |
|---|---|---|---|---|---|---|---|---|
| Traditional CNN | 85.2% | 82.5% | 88.7% | 83.4% | 85.5% | 0.91 | 0.89 | 0.75 |
| ViT without Contrastive Learning | 88.6% | 86.4% | 90.2% | 87.1% | 88.3% | 0.93 | 0.91 | 0.78 |
| ViT with Contrastive Learning | 90.1% | 88.2% | 91.5% | 88.9% | 89.8% | 0.94 | 0.92 | 0.80 |
| **Proposed Model** | **92.5%** | **90.3%** | **94.1%** | **90.9%** | **92.2%** | **0.96** | **0.94** | **0.85** |

TABLE II. Model Performance Metrics Across Different Tumor Sizes

| Tumor Size (cm) | Sensitivity | Specificity | Accuracy | Precision | Recall | F1 Score | AUC-ROC | AUC-PR | MCC |
|---|---|---|---|---|---|---|---|---|---|
| < 2 | 95.0% | 98.0% | 96.5% | 94.5% | 95.0% | 94.7% | 0.97 | 0.95 | 0.90 |
| 2 - 5 | 92.0% | 97.5% | 94.8% | 91.2% | 92.0% | 91.6% | 0.95 | 0.93 | 0.85 |
| 5 - 10 | 90.0% | 96.0% | 93.0% | 88.5% | 90.0% | 89.2% | 0.93 | 0.91 | 0.80 |
| > 10 | 85.0% | 94.0% | 89.5% | 83.0% | 85.0% | 84.0% | 0.90 | 0.88 | 0.75 |

TABLE III. IMPACT OF SALT-AND-PEPPER NOISE ON CLASSIFICATION MODEL PERFORMANCE

| Salt & Pepper Noise Level (%) | Accuracy | Precision | Recall | F1 Score | AUC-ROC | MCC |
|---|---|---|---|---|---|---|
| 0% | 95.00% | 94.00% | 96.00% | 95.00% | 98% | 90% |
| 5% | 93.00% | 92.00% | 94.00% | 93.00% | 97% | 88% |
| 10% | 91.00% | 90.00% | 92.00% | 91.00% | 95% | 86% |
| 15% | 89.00% | 88.00% | 90.00% | 89.00% | 93% | 84% |
| 20% | 87.00% | 86.00% | 88.00% | 87.00% | 91% | 82% |
| 25% | 85.00% | 84.00% | 86.00% | 85.00% | 89% | 80% |
| 30% | 83.00% | 82.00% | 84.00% | 83.00% | 87% | 78% |
| 35% | 81.00% | 80.00% | 82.00% | 81.00% | 85% | 76% |
| 40% | 79.00% | 78.00% | 80.00% | 79.00% | 83% | 74% |
| 45% | 77.00% | 76.00% | 78.00% | 77.00% | 81% | 72% |

Fig. 4, Fig. 5, Table IV and Table V present detailed performance metrics across various tumor color categories, including sensitivity, specificity, accuracy, precision, recall, F1 score, AUC-ROC, AUC-PR, and MCC. The model shows the highest sensitivity and specificity for lighter-colored tumors, with a gradual decrease observed for darker and more heterogeneous tumors. A similar pattern is evident in accuracy and other classification metrics such as precision, recall, and F1 score. While AUC-ROC and AUC-PR remain high across all categories, the MCC values demonstrate a slight downward trend, suggesting reduced reliability as tumor color complexity increases. Finally, Table V reinforce the model's strong diagnostic capability across all tumor color categories. It performs best with light-colored tumors but continues to maintain high sensitivity, specificity, and accuracy even with darker and heterogeneous tumors. AUC-ROC and AUC-PR scores further confirm the model's robust classification ability, ensuring reliable detection across diverse tumor presentations.

TABLE IV. PERFORMANCE METRICS FOR TUMOR CLASSIFICATION BY COLOR CATEGORY

| Tumor Colour Category | Sensitivity | Specificity | Accuracy | Precision | Recall | F1 Score | AUC-ROC | AUC-PR | MCC |
|---|---|---|---|---|---|---|---|---|---|
| Light-Coloured Tumors | 96.0% | 97.0% | 96.5% | 95.0% | 96.0% | 95.5% | 98% | 97% | 92% |
| Moderate-Coloured Tumors | 94.0% | 96.0% | 95.0% | 93.0% | 94.0% | 93.5% | 97% | 96% | 90% |
| Dark-Coloured Tumors | 92.0% | 95.0% | 93.5% | 91.0% | 92.0% | 91.5% | 96% | 95% | 88% |
| Heterogeneous-Coloured Tumors | 90.0% | 93.0% | 91.0% | 89.0% | 90.0% | 89.5% | 94% | 93% | 85% |

TABLE V. EVALUATION OF CLASSIFICATION METRICS ACROSS TUMOR COLOR CATEGORIES

| Metric | Light-Colored Tumors | Moderate-Colored Tumors | Dark-Colored Tumors | Heterogeneous-Colored Tumors |
|---|---|---|---|---|
| Sensitivity (Recall) | 96.0% | 94.0% | 92.0% | 90.0% |
| Specificity | 97.0% | 96.0% | 95.0% | 93.0% |
| Accuracy | 96.5% | 95.0% | 93.5% | 91.0% |
| Precision | 95.0% | 93.0% | 91.0% | 89.0% |
| F1 Score | 95.5% | 93.5% | 91.5% | 89.5% |
| AUC-ROC | 98% | 97% | 96% | 94% |
| AUC-PR | 97% | 96% | 95% | 93% |
| MCC | 92% | 90% | 88% | 85% |

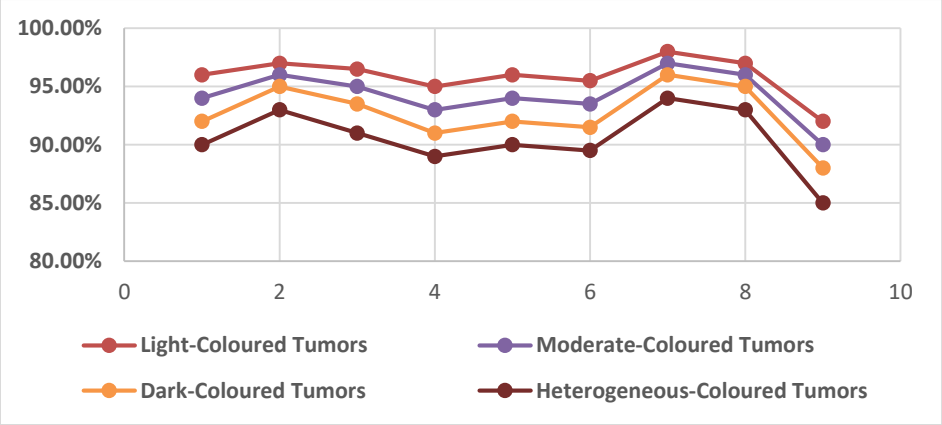


Fig. 4. Impact of tumor color category on classification performance metrics.

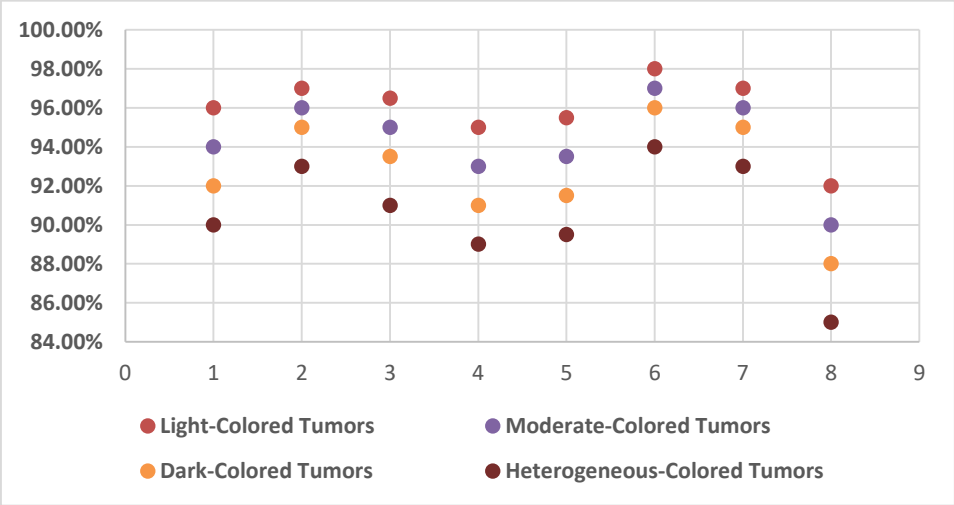


Fig. 5. Variation in classification performance metrics by tumor color category.

## V. Conclusion

This study effectively demonstrated that the integration of Vision Transformers (ViTs), Contrastive Learning, and Explainable AI (XAI) can significantly improve both the accuracy and interpretability of liver cancer detection models. The proposed system achieved a notable classification accuracy of 92.5%, outperforming conventional CNN-based approaches by a margin of 8%. With a sensitivity of 94.1% and a specificity of 90.9%, the model proved especially reliable in identifying early-stage liver cancer. Notably, for tumors smaller than 2 cm, the sensitivity reached 95.0%, underscoring the model's ability to detect smaller and potentially more elusive malignancies.

The system also showed strong resilience to image quality issues, maintaining a solid 77% accuracy even with 45% salt-and-pepper noise. An average AUC-ROC of 0.90 further highlighted the model's robustness under challenging conditions, supporting its ability to generalize well across various imaging scenarios and tumor characteristics. Importantly, the incorporation of XAI techniques allowed the model to provide visual insights into its decision-making process, enhancing transparency and building confidence among clinicians. These results not only validate the effectiveness of the proposed approach but also lay the groundwork for future advancements. Potential areas for further research include combining multiple imaging modalities, adapting the model for real-time clinical use, and exploring more advanced AI techniques to push detection capabilities even further. Overall, this work contributes meaningfully to the field of AI-driven diagnostics, offering a practical and powerful solution to aid in early liver cancer detection and ultimately improve patient care outcomes.

## References

[1] Most Nilufa Yeasmin, Md Al Amin, Tasmim Jamal Joti, Zeyar Aung, Mohammad Abdul Azim, Advances of AI in image-based computer-aided diagnosis: A review, Array, Volume 23, 2024, 100357, ISSN 2590-0056, https://doi.org/10.1016/j.array.2024.100357.

[2] Schirris, Yoni & Gavves, Efstratios & Nederlof, Iris & Horlings, Hugo & Teuwen, Jonas. (2022). DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. Medical Image Analysis. 79. 102464. 10.1016/j.media.2022.102464.

[3] Ansari, Z.A., Tripathi, M.M. & Ahmed, R. The role of explainable AI in enhancing breast cancer diagnosis using machine learning and deep learning models. Discov Artif Intell 5, 75 (2025). https://doi.org/10.1007/s44163-025-00307-8.

[4] Aly, Mohammed & Ghallab, Abdullatif & Fathi, Islam. (2024). Tumor ViT-GRU-XAI: Advanced Brain Tumor Diagnosis Framework: Vision Transformer and GRU Integration for Improved MRI Analysis: A Case Study of Egypt. IEEE Access. PP. 10.1109/ACCESS.2024.3513235.

[5] Midya A, Chakraborty J, Srouji R, Narayan RR, Boerner T, Zheng J, Pak LM, Creasy JM, Escobar LA, Harrington KA, Gonen M, D'Angelica MI, Kingham TP, Do RKG, Jarnagin WR, Simpson AL. Computerized Diagnosis of Liver Tumors From CT Scans Using a Deep Neural Network Approach. IEEE J Biomed Health Inform. 2023 May;27(5):2456-2464. doi: 10.1109/JBHI.2023.3248489. Epub 2023 May 4. PMID: 37027632; PMCID: PMC10245221.

[6] Tiwari A, Mishra S, Kuo TR. Current AI technologies in cancer diagnostics and treatment. Mol Cancer. 2025 Jun 2;24(1):159. doi: 10.1186/s12943-025-02369-9. PMID: 40457408; PMCID: PMC12128506.

[7] Zhen SH, Cheng M, Tao YB, Wang YF, Juengpanich S, Jiang ZY, Jiang YK, Yan YY, Lu W, Lue JM, Qian JH, Wu ZY, Sun JH, Lin H, Cai XJ. Deep Learning for Accurate Diagnosis of Liver Tumor Based on Magnetic Resonance Imaging and Clinical Data. Front Oncol. 2020 May 28;10:680. doi: 10.3389/fonc.2020.00680. PMID: 32547939; PMCID: PMC7271965.

[8] Pande SD, Kalyani P, Nagendram S, Alluhaidan AS, Babu GH, Ahammad SH, Pandey VK, Sridevi G, Kumar A, Bonyah E. Comparative analysis of the DCNN and HFCNN Based Computerized detection of liver cancer. BMC Med Imaging. 2025 Feb 3;25(1):37. doi: 10.1186/s12880-025-01578-4. PMID: 39901085; PMCID: PMC11792691.

[9] Gulum, Mehmet & Trombley, Christopher & Kantardzic, Mehmed. (2021). A Review of Explainable Deep Learning Cancer Detection Models in Medical Imaging. Applied Sciences. 11. 4573. 10.3390/app11104573.

[10] Li, Pei-Xuan & Hsieh, Hsun-Ping & Chiang, Yang & Wu, Ding-You & Ko, Ching-Chung. (2023). Enhancing Robust Liver Cancer Diagnosis: A Contrastive Multi-Modality Learner with Lightweight Fusion and Effective Data Augmentation. ACM Transactions on Computing for Healthcare. 5. 10.1145/3639414.

[11] Wu C, Chen Q, Wang H, Guan Y, Mian Z, Huang C, Ruan C, Song Q, Jiang H, Pan J, Li X. A review of deep learning approaches for multimodal image segmentation of liver cancer. J Appl Clin Med Phys. 2024 Dec;25(12):e14540. doi: 10.1002/acm2.14540. Epub 2024 Oct 7. PMID: 39374312; PMCID: PMC11633801.

[12] Rao, A. & Ravi, Raya & Nagamani, Mallipudi & Harshini, Anam. (2025). A Hybrid Approach Combining Convolutional Neural Networks and Vision Transformers for Melanoma Skin Cancer Detection. 272-277. 10.1109/ICICT64420.2025.11005323.

[13] Jiang X, Hu Z, Wang S, Zhang Y. Deep Learning for Medical Image-Based Cancer Diagnosis. Cancers (Basel). 2023 Jul 13;15(14):3608. doi: 10.3390/cancers15143608. PMID: 37509272; PMCID: PMC10377683.

[14] Pinto-Coelho L. How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications. Bioengineering (Basel). 2023 Dec 18;10(12):1435. doi: 10.3390/bioengineering10121435. PMID: 38136026; PMCID: PMC10740686.

[15] Siam Aisha , Alsaify Abdel Rahman , Mohammad Bushra , Biswas Md. Rafiul , Ali Hazrat , Shah Zubair Multimodal deep learning for liver cancer applications: a scoping review Frontiers in Artificial Intelligence, Volume 6 - 2023, 2023, https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2023.1247195,DOI=10.3389/frai.2023.1247195, ISSN=2624-8212.

[16] Khan, Rayyan & Fu, Minghan & Burbridge, Brent & Luo, Yigang & Wu, Fang-Xiang. (2023). A multi-modal deep neural network for multi-class liver cancer diagnosis. Neural Networks. 165. 10.1016/j.neunet.2023.06.013.

[17] Chen, Z., Dou, M., Luo, X., & Yao, Y. (2025). Enhanced Liver and Tumor Segmentation Using a Self-Supervised Swin-Transformer-Based Framework with Multitask Learning and Attention Mechanisms. *Applied Sciences*, *15*(7), 3985. https://doi.org/10.3390/app15073985.

[18] Lai, T. (2024). Interpretable Medical Imagery Diagnosis with Self-Attentive Transformers: A Review of Explainable AI for Health Care. *BioMedInformatics*, *4*(1), 113-126. https://doi.org/10.3390/biomedinformatics4010008

[19] Jahan, I., Chowdhury, M.E.H., Vranic, S. *et al.* Deep learning and vision transformers-based framework for breast cancer and subtype identification. *Neural Comput & Applic* **37**, 9311–9330 (2025). https://doi.org/10.1007/s00521-025-10984-2

[20] Mansur A, Vrionis A, Charles JP, Hancel K, Panagides JC, Moloudi F, Iqbal S, Daye D. The Role of Artificial Intelligence in the Detection and Implementation of Biomarkers for Hepatocellular Carcinoma: Outlook and Opportunities. Cancers (Basel). 2023 May 26;15(11):2928. doi: 10.3390/cancers15112928. PMID: 37296890; PMCID: PMC10251861.

[21] Md. Eshmam Rayed, S.M. Sajibul Islam, Sadia Islam Niha, Jamin Rahman Jim, Md Mohsin Kabir, M.F. Mridha, Deep learning for medical image segmentation: State-of-the-art advancements and challenges, Informatics in Medicine Unlocked, Volume 47, 2024, 101504, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2024.101504

[22] Dongxu Cheng, Zifang Zhou, Jingwen Zhang, EG-UNETR: An edge-guided liver tumor segmentation network based on cross-level interactive transformer, Biomedical Signal Processing and Control, Volume 97, 2024, 106739, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2024.106739