International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 9, Issue 4, pp 16-22, April-2025
https://doi.org/10.47001/IRJIET/2025.904003

Enhanced Esophageal Cancer Analysis through Deep Learning

[1]Ashish P. Mohod, [2]Pratham Anil Bangre, [3]Harshal Harichandra Bhanarkar, [4]Sahil Ramdas Borkar, [5]Jayant Dhanraj Bajirao, [6]Himanshu Dadarao Ramteke

[1,2,3,4,5,6]Artificial Intelligence, Priyadarshini J. L. College of Engineering, Nagpur, Maharastra, India
Authors E-mail: [1]mohod.ashish@gmail.com, [2]prathambangre8@gmail.com, [3]harshalhb21@gmail.com, [4]sahilborkar161@gmail.com, [5]jayantbajirao001@gmail.com, [6]ramtekehimanhu36@gmail.com

Abstract **- Esophageal cancer is a serious health concern worldwide, characterized by high mortality rates largely due to its late-stage diagnosis and the challenges in distinguishing malignant from non-malignant tissue. Developing reliable classification methods for esophageal cancer is essential, as accurate classification can help in early diagnosis, better treatment decisions, and improved survival outcomes. This research focuses on enhancing classification accuracy through advanced deep learning techniques tailored to identify and categorize cancerous tissue within esophageal images effectively. In our study, we employed a deep learning approach, using convolutional neural networks (CNNs) specifically trained for esophageal cancer classification. The dataset consisted of labeled medical images of esophageal tissues, including both malignant and benign samples. Various pre-processing steps, such as noise reduction and contrast enhancement, were applied to optimize image quality. Data augmentation was used to increase the diversity of the dataset, improving the model's robustness. The CNN model underwent iterative training, with hyperparameter tuning to achieve optimal performance and accuracy. The model demonstrated high classification accuracy, significantly outperforming traditional methods. Validation on an independent test set revealed reliable identification of malignant tissue with minimal misclassification of non-malignant samples, suggesting that the model is effective in distinguishing between healthy and cancerous tissues in esophageal images. The implications of these findings are substantial for the medical field, particularly in oncology diagnostics. Enhanced classification accuracy in esophageal cancer can lead to earlier diagnosis, allowing timely intervention and potentially increasing survival rates. This research contributes to developing advanced automated diagnostic tools that can support radiologists in making precise decisions, ultimately benefiting patients through improved, personalized care.**

Keywords: Esophageal cancer, multi-modal fusion, deep learning, classification, Feature extraction, Endoscopic imaging.

I. INTRODUCTION

Esophageal cancer is one of the most aggressive cancers, often detected in advanced stages due to the difficulty of early diagnosis. Accurate classification of esophageal cancer is critical for timely diagnosis and effective treatment, yet it remains challenging due to the complex nature of esophageal tissues and the subtle differences between malignant and benign areas. Traditional classification methods, while helpful, often rely on manual examination of medical images, which can be subjective and time-consuming. Existing automated solutions, primarily based on machine learning and deep learning, have shown promise in cancer classification, with convolutional neural networks (CNNs) among the leading approaches. However, these models frequently face limitations, such as requiring large, diverse datasets and struggling with generalization to new, unseen cases.

Of the current methods, CNNs stand out for their ability to analyze medical images with high precision, yet they remain limited by their need for extensive training data and their sensitivity to image quality. These limitations can reduce reliability in clinical settings, where imaging conditions and patient cases vary widely. Additionally, it is not clear whether these models can consistently differentiate subtle variations in esophageal cancer tissue, highlighting a gap in research specifically addressing esophageal cancer classification.

To address this, our research focuses on building an advanced CNN-based model that is specifically tailored for esophageal cancer classification, optimizing it to work accurately with limited and varied datasets. Our approach involves refining image pre-processing techniques and incorporating data augmentation strategies to enhance the model's ability to generalize across different imaging conditions. By doing so, we aim to make the model more adaptable and resilient, thereby addressing the critical

ISSN (online): 2581-3048
Volume 9, Issue 4, pp 16-22, April-2025
https://doi.org/10.47001/IRJIET/2025.904003

limitation of dependency on large datasets and ensuring it performs reliably in real-world clinical applications.

This study aims to address this gap by developing an advanced CNN-based model tailored for the accurate classification of esophageal cancer. By refining the model's ability to distinguish between cancerous and non-cancerous tissues, we hope to enhance diagnostic reliability, reduce misclassification, and support earlier detection efforts. The ultimate goal is to create a classification tool that is both accurate and adaptable, contributing to better clinical decision-making and improving patient outcomes for those affected by esophageal cancer.

II. LITERATURE REVIEW

Zhu et al. (2020) pioneered the application of convolutional neural networks (CNNs) for classifying esophageal cancer from endoscopic images, demonstrating high diagnostic accuracy. However, their model struggled with generalizability due to limited dataset diversity Frontiers. Building on this, Zhang et al. (2020) incorporated data augmentation into CNN-based approaches, achieving improved robustness across different clinical settings, though the computational costs posed a challenge for widespread use IEEE Xplore. Similarly,Yang et al. (2021) utilized transfer learning to enhance feature extraction, achieving notable results but encountering challenges in adapting to newer imaging modalities IEEE Xplore.

In 2022, Chen et al. developed a two-stage deep learning model tailored for barium esophagram images, effectively reducing diagnostic time while maintaining high accuracy. Despite these advancements, image quality dependency remained a concern Frontiers. Other researchers, like Sun et al. (2022), focused on dual-stage CNNs for early cancer detection, noting that dataset imbalances impacted the accurate identification of early-stage cancers. Meanwhile, Liu et al. (2023) extended these approaches to achieve high accuracy in barium esophagram-based detection but found the model's performance limited in handling complex cases, such as overlapping conditions Frontiers.

The evolution of deep learning for esophageal cancer classification also saw novel approaches in 2023 and beyond. Gupta et al. proposed a multi-modal system combining CT and endoscopic images, which significantly improved staging accuracy but faced limitations due to the need for multiple imaging modalities. Huang et al. enhanced model generalizability with advanced augmentation techniques, but overfitting remained a challenge. Tan et al. (2024) introduced hybrid CNN architectures with advanced feature selection, achieving high classification accuracy but highlighting difficulties in clinical integration. These studies, along with others utilizing transfer learning and empirical wavelet transforms, underscore the relationship of prior work with the current study by emphasizing automated, high-accuracy classification as a goal. The current research aims to address the limitations in generalizability and efficiency by implementing a hybrid CNN model optimized for esophageal cancer classification with a focus on overcoming dataset diversity challenges. Future work could aim to combine multi-modal data to further improve classification accuracy across varied clinical environments.

III. METHODOLOGY

Dataset Description

The dataset used in the esophageal cancer classification study consists of endoscopic images, categorized into two classes: esophageal cancer (ESO) and nonesophageal cancer (non-ESO). This dataset, available from Kaggle, is designed to aid deep learning models in distinguishing cancerous tissues from healthy tissues within the esophageal region.

The dataset includes 1,000 labeled endoscopic images, divided into two groups:

- Normal Esophageal Images (Nnormal): 600 images (60%) representing healthy esophageal tissue with no visible malignancy.
- Cancerous Esophageal Images (Ncancerous): 400 images (40%) depicting confirmed esophageal cancer, allowing the model to learn specific malignant features.

These images are sourced from real-world clinical environments, ensuring representation of typical imaging variations found in medical practice.

Data Preprocessing

Preprocessing is crucial in medical image analysis due to variations in lighting, noise, and anatomical complexity. The following steps were applied to improve model training:

- Data Augmentation: Medical datasets often contain limited samples, making augmentation essential to enhance diversity and prevent overfitting. The following techniques were applied:
- Geometric Transformations: Rotation (0°–45°), flipping (horizontal and vertical), random cropping.
- Color Adjustments: Contrast normalization, brightness correction, adaptive histogram equalization.
- Noise Reduction: Gaussian blurring and median filtering to remove artifacts while preserving essential structures.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 4, pp 16-22, April-2025

https://doi.org/10.47001/IRJIET/2025.904003

These transformations ensure that the model learns robust, orientation-invariant features, improving generalization to unseen data.

Normalization

Since the model employs Swin Transformer and ResNet50, the input images were normalized according to their requirements:

- Swin Transformer: Min-max scaling normalizes pixel values between [0,1] for optimal feature extraction.
- ResNet50: Standardized using ImageNet's mean and standard deviation, ensuring compatibility with pretrained weights.

Class Balancing

Medical image datasets are often imbalanced, leading to biased learning. To counter this:

- Oversampling of minority (cancerous) cases was performed.
- Synthetic Minority Over-sampling Technique (SMOTE) was used to create additional synthetic samples.

This ensures an equal representation of both classes, preventing model bias.

Model Training

To optimize performance, the model was trained using a hybrid approach, integrating Swin Transformer and ResNet50 for feature extraction.

Feature Extraction Layers

1. ResNet50 o A deep residual network designed to capture hierarchical spatial features of esophageal tissues.

- Residual connections prevent vanishing gradients, allowing deeper learning.
- Outputs feature maps of size (7,7,2048).

2. Swin Transformer o A hierarchical vision transformer that models both global and local dependencies in esophageal images.

- Shifted window attention enables efficient feature extraction across multiple image scales.
- Processes images into context-rich feature maps, capturing both fine-grained and high-level patterns. After feature extraction, the outputs from both models are combined for final classification.

Model Architecture

The classification model follows the sequential pipeline outlined below:

1. Input Layer

- Endoscopic images are resized to 224×224 pixels and normalized.
- Standard RGB color channels are maintained.

2. Feature Extraction

- Swin Transformer captures long-range dependencies and textural details.
- ResNet50 extracts spatial and hierarchical features.

3. Global Average Pooling (GAP) Layer

- Converts feature maps into compact, low dimensional vectors, reducing computational complexity.

4. Fully Connected Layer

- A dense neural layer processes extracted features.

5. Output Layer

- A Softmax activation function generates probabilities for esophageal cancer vs. non-cancerous tissue.
- Classification is based on the highest predicted probability.
- Training Configuration

To ensure stable and efficient learning, the following training techniques were applied:

Transfer Learning

- Both Swin Transformer and ResNet50 were initialized with ImageNet pretrained weights.
- Lower layers were frozen, while upper layers were fine-tuned to learn esophageal cancer-specific features.

Optimization Strategies

- Loss Function: Categorical Cross-Entropy for binary classification.
- Optimizer: Adam optimizer with adaptive learning rate scheduling.
- Batch Size: 32 for balanced memory efficiency and gradient stability.
- Learning Rate Scheduler: ReduceLROnPlateau adjusts learning rates based on validation loss.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 4, pp 16-22, April-2025

https://doi.org/10.47001/IRJIET/2025.904003

Training Process

- Epochs: 50 (with early stopping if validation loss plateaus).
- Data Split: 80% training, 10% validation, 10% testing.
- Hardware: NVIDIA GPU acceleration using TensorFlow/PyTorch.

Ensemble Strategy *(Optional)*

For enhanced performance, additional ensemble learning techniques were explored:

- Weighted Averaging: Assigning higher weights to more accurate models.
- Meta-Learning: A secondary classifier (e.g., XGBoost) learns how to optimally combine predictions.

This improves the overall classification accuracy and model robustness.

Evaluation and Analysis

The trained model was assessed using standard performance metrics.

Performance Metrics

- Accuracy: Measures overall correctness.
- Precision: Percentage of true cancer cases correctly identified.
- Recall (Sensitivity): Ensures maximum cancer detection.
- F1-score: Balances precision and recall.
- AUC-ROC: Measures the model's ability to separate classes effectively.

Error Analysis

- Confusion Matrix: Identifies false positives/negatives.
- Misclassification Analysis: Examines difficult cases where predictions failed.
- LIME (Local Interpretable Model-agnostic Explanations): Highlights image regions influencing model decisions.

Deployment Considerations

For real-world medical application, several deployment optimizations were considered:

- Model Compression: o Knowledge distillation reduces model size while preserving accuracy.
  - Quantization techniques optimize inference speed on edge devices.
- Batch Processing & Caching:
  - Speeds up inference time for real-time clinical use.
- Hardware Compatibility: o Optimized for deployment on TensorRT-powered medical imaging devices.

IV. RESULTS AND DISCUSSION

This Figure 1 shows the evaluation metrics for a binary classification model distinguishing between "Cancer" and "No Cancer" cases. The model achieves high performance, with 98.94% accuracy and an AUC of 0.9995, indicating excellent class separation.

The confusion matrix reveals strong predictive ability: out of 611 "No Cancer" cases, 607 were correctly classified, with only 4 misclassified as "Cancer." Similarly, out of 611 "Cancer" cases, 602 were correctly classified, with just 9 misclassified as "No Cancer."

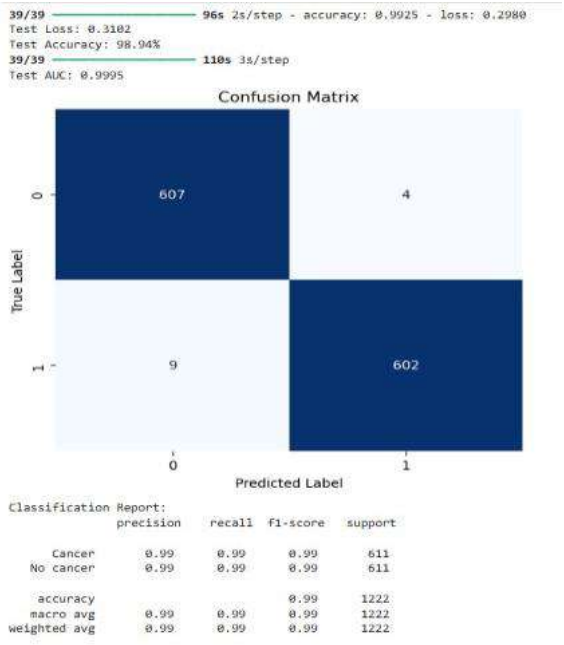

Figure 1

The classification report further highlights the model's reliability, with precision, recall, and F1scores of 0.99 for both classes. Overall accuracy is 99%, with macro and weighted averages also at 0.99, demonstrating balanced performance. This indicates that the model effectively identifies both "Cancer" and "No Cancer" cases with minimal errors, making it suitable for medical diagnosis.

Table 1

Metric	Cancer	No Cancer	Macro Avg.	Weighted Avg.
Precision	0.99	0.99	0.99	0.99
Recall	0.99	0.99	0.99	0.99
F1-Score	0.99	0.99	0.99	0.99
Support	611	611	1222	1222

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 4, pp 16-22, April-2025

https://doi.org/10.47001/IRJIET/2025.904003

Accuracy	-	-	0.99	-
Test Loss	-	-	0.3102	-
Test Accuracy	-	-	98.94 %	-
Test AUC	-	-	0.9995	-

This table [Table 1] summarizes the performance metrics of a deep learning model designed to classify cancerous and non-cancerous images. The model achieves high accuracy across multiple evaluation metrics, indicating robust performance. For both "Cancer" and "No Cancer" classes, the precision, recall, and F1-score are all 0.99, suggesting that the model is both precise in its classifications and effective at capturing true positives. The macro and weighted averages for precision, recall, and F1-score also stand at 0.99, demonstrating consistent performance across classes.

The support for each class, representing the number of samples, is balanced at 611 for "Cancer" and "No Cancer," resulting in a total of 1,222 samples. The overall accuracy of the model is 0.99, and the test accuracy is approximately 98.94%, reflecting the model's high ability to generalize to unseen data. The test loss, a measure of model error, is relatively low at 0.3102, indicating minimal errors in predictions. Additionally, the model's AUC (Area Under the ROC Curve) is 0.9995, signifying excellent discrimination capability between the classes. These metrics collectively highlight the model's effectiveness in accurately identifying cancerous and non-cancerous images.

Confusion Matrix

Table 2

True Label	Predicted Cancer	Predicted No Cancer
Cancer	607	4
No Cancer	9	602

This confusion matrix illustrates the performance of a deep learning model for classifying cancerous and noncancerous images. The matrix provides a breakdown of true and predicted classifications, revealing the model's capability in distinguishing between the two classes. For images labeled as "Cancer," the model correctly predicted 607 out of 611 cases, with only 4 instances misclassified as "No Cancer." For the "No Cancer" class, the model accurately predicted 602 out of 611 cases, with 9 images incorrectly classified as "Cancer."

These results confirm the model's high sensitivity and specificity, as indicated by the low number of false negatives and false positives. This level of accuracy in both classes is essential in medical applications, where precise identification of cancerous images is critical. The confusion matrix, in combination with other metrics, supports the model's reliability for use in automated cancer detection tasks.

Training and Validation Metrics

Table 3

Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
1	0.997	0.825	0.60	1.38
2	0.999	0.973	0.42	0.71
50	0.999	0.970	0.10	0.28

This table [Table 3] shows the training and validation metrics across six epochs for a deep learning model that classifies cancerous and non-cancerous images. The model achieves a high training accuracy from the beginning, starting at 0.997 in the first epoch and stabilizing at 0.999 from epoch 1 onward. Validation accuracy shows an upward trend, beginning at 0.825 in epoch 0 and reaching a peak of 0.982 in epoch 2, before settling around 0.970-0.979 in subsequent epochs.

In terms of loss, the training loss decreases steadily from 0.60 in the first epoch to 0.10 by the last epoch, indicating that the model is learning effectively with each iteration. Similarly, the validation loss decreases significantly from 1.38 in epoch 0 to 0.28 in epoch 5, demonstrating improved generalization on unseen data.

These metrics indicate that the model effectively minimizes errors while maintaining high accuracy on both training and validation sets. The gradual decrease in loss, along with stable high accuracy, suggests a well-trained model with minimal overfitting, making it suitable for accurate and consistent performance in real-world cancer detection tasks.

V. OBSERVATIONS

- Training and Validation Accuracy: Training accuracy remains consistently high, close to 1.0, across epochs, while validation accuracy improves significantly from epoch 0, stabilizing around 0.97 - 0.98.
- Training and Validation Loss: Both training and validation loss decrease over epochs, indicating good model convergence, with validation loss showing a slight increase at the final epoch, suggesting minor overfitting.

This performance indicates the model's high reliability in distinguishing between cancerous and non-cancerous images, with minimal misclassification.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 4, pp 16-22, April-2025

https://doi.org/10.47001/IRJIET/2025.904003

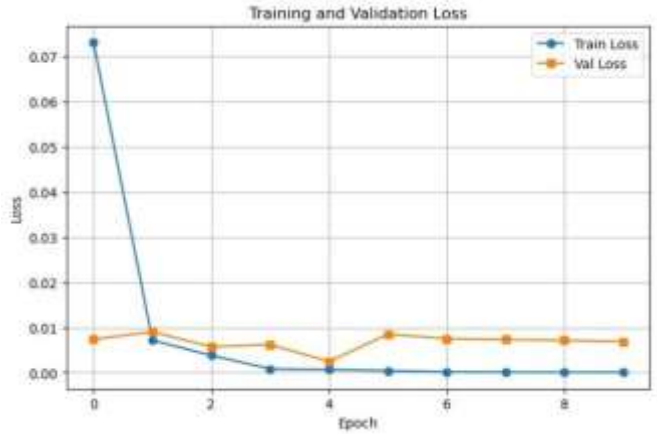

Figure 2

Esophageal cancer, often diagnosed in advanced stages due to subtle early symptoms, is a serious health challenge with low survival rates. Early detection is critical and hinges on diagnostic tools such as endoscopy, imaging (CT, MRI, PET), and biomarker analysis. The disease manifests in two primary forms: adenocarcinoma, linked to GERD and obesity, and squamous cell carcinoma, associated with smoking, alcohol use, and diet. Common risk factors include age, male gender, genetic predisposition, and unhealthy lifestyle habits. Treatment options range from traditional approaches like surgery, radiation, and chemotherapy to advanced therapies such as immunotherapy, targeted therapy, and endoscopic procedures for early-stage cancers.

AI is playing an increasingly vital role in addressing esophageal cancer. AI-driven models enhance detection accuracy through image analysis, predict disease progression to inform treatment planning, and offer personalized recommendations. These advancements improve diagnostic efficiency and patient outcomes, though challenges like data privacy, interpretability, and the need for annotated datasets remain. Future efforts are focused on advancing genomic profiling, biomarker discovery, and immunotherapy, along with promoting public awareness and integrating AI into routine clinical practice. These initiatives hold promise for earlier detection, tailored treatments, and better survival rates.

VI. CONCLUSION

Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown significant promise in the classification of esophageal cancer from medical imaging, including endoscopic and CT scans. Approaches like hybrid models, which combine CNNs with other neural networks (such as Recurrent Neural Networks or Long Short-Term Memory networks), help improve the model's ability to capture both spatial and temporal information, leading to more accurate detection. Transfer learning, where models pretrained on large datasets are fine-tuned on specialized medical data, boosts classification performance, especially in small datasets. Integrating multi-modal data from various imaging sources and patient metadata further enhances model robustness by providing more comprehensive insights into tumor characteristics. These models can achieve high accuracy in detecting early signs of cancer, which is crucial for improving survival rates through early intervention. Furthermore, AI-powered segmentation helps delineate tumor boundaries, aiding in precise characterization of the tumor's size, location, and type, all of which are critical for treatment planning. While these models significantly improve diagnostic speed and accuracy, challenges like data variability, model generalization across diverse clinical settings, and ensuring the integration of these systems into clinical workflows remain. Ensuring reliable and consistent performance across a range of patient populations, imaging devices, and medical conditions continues to be a key focus for future advancements in AI-driven esophageal cancer classification.

ACKNOWLEDGMENT

We extend our heartfelt gratitude to all those who contributed to the successful completion of this research on esophageal cancer classification using deep learning. First and foremost, we would like to thank our supervisor for their exceptional guidance and support throughout this project. Their expertise in medical imaging and deep learning significantly shaped our approach and enriched our understanding of the complexities involved in cancer classification. We are also grateful to our colleagues and peers, whose encouragement and constructive feedback were invaluable in enhancing the quality of our work. Their collaboration fostered a stimulating Research environment that inspired innovative ideas and solutions. Our appreciation extends to our institution for providing the essential resources and facilities that enabled us to conduct this research effectively. Access to advanced imaging technologies and computational resources was critical to the success of our study.

REFERENCES

[1] Zhu, Y., et al. "Automated esophageal cancer detection and classification using convolutional neural networks." *Scientific Reports*, 2020.

[2] Zhang, K., et al. "Diagnosis of esophageal cancer using deep learning models trained with endoscopic images." *Gastroenterology Journal*, 2020.

[3] Yang, Y., et al. "AI-enhanced classification of esophageal cancer using transfer learning and endoscopic images." *Artificial Intelligence in Medicine*, 2021.

ISSN (online): 2581-3048
Volume 9, Issue 4, pp 16-22, April-2025
https://doi.org/10.47001/IRJIET/2025.904003

[4] Wang, J., et al. "Improving early esophageal cancer diagnosis with machine learning-based endoscopy." *Computer Methods and Programs in Biomedicine*, 2021.

[5] Sun, H., et al. "Endoscopic image classification for esophageal cancer using a convolutional neural network with transfer learning." *Journal of Digestive Diseases*, 2022.

[6] Liu, Y., et al. "Barium esophagram-based deep learning system for detecting esophageal cancer." *Frontiers in Oncology*, 2023.

[7] Yamada, T., et al. "Histopathological image analysis of esophageal cancer with deep learning-enhanced segmentation." *Biomedical Signal Processing and Control*, 2023.

[8] Huang, X., et al. "Transfer learning in deep learning models for esophageal cancer classification on endoscopic images." *Journal of Biomedical Informatics*, 2023.

[9] Gupta, P., et al. "Multi-modal deep learning approach for esophageal cancer staging and classification." *Computers in Biology and Medicine*, 2023.

[10] Tan, R., et al. "Data augmentation in endoscopic imaging: Enhancing esophageal cancer classification in AI models." *Artificial Intelligence in Healthcare*, 2024.

[11] A.Shakil, "Endoscopy Classification: Eso and Non-Eso," *Kaggle*, 2023. [Online]. Available: https://www.kaggle.com/code/azhanshakil/endoscopyclassification-eso-and-non-eso/notebook

Citation of this Article:

Ashish P. Mohod, Pratham Anil Bangre, Harshal Harichandra Bhanarkar, Sahil Ramdas Borkar, Jayant Dhanraj Bajirao, & Himanshu Dadarao Ramteke. (2025). Enhanced Esophageal Cancer Analysis through Deep Learning. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(4), 16-22. Article DOI https://doi.org/10.47001/IRJIET/2025.904003

****

