International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Deep Learning Based Automated Esophageal Cancer Detection Using Medical Imaging

Manasa S R[1], Neha Naik C G[2], Sindhu B M[3]

[1,2,3]Artificial Intelligence And Machine Learning, Adichunchanagiri Institute Of Technology Chikkamagaluru India

Abstract
Early detection of esophageal cancer significantly improves patient survival rates; However,conventional diagnostic techniques are often invasive, costly, and prone to subjective interpretation by medical experts. This paper presents a deep learning based automated system for esophageal cancer detection using medical imaging modalities such as endoscopic and CT images. The proposed system employs convolutional neural networks (CNNs) to automatically extract meaningful features and perform accurate classification. Image preprocessing techniques such as normalization, noise filtering, data augmentation, and region-based segmentation are applied to enhance image quality and improve model robustness. The performance of the system is evaluated using standard metrics including accuracy, sensitivity, specificity, precision, and F1-score. Experimental results demonstrate that the proposed approach achieves reliable performance, demonstrating its effectiveness as a clinical decision- support tool for early diagnosis, reduced diagnostic workload, and improved screening efficiency in esophageal cancer detection.

I. Introduction

Early detection of esophageal cancer is essential for improving therapeutic outcomes and long-term survival. When diagnosed at an early stage, the disease can often be managed through less aggressive therapeutic interventions, significantly enhancing quality of life and long-term prognosis. However, early-stage esophageal cancer usually presents minimal or non specific symptoms such as mild swallowing difficulty or chest discomfort, which are frequently ignored or misinterpreted. Consequently ,many patients are diagnosed only after the disease has progressed to an advanced stage, where treatment options become limited and less effective. Early diagnosis of esophageal cancer is challenging due to the subtle nature of its initial clinical manifestations. Conventional screening methods are not routinely performed for high-risk populations, leading to delayed identification of malignant changes. In many cases, pathological alterations in the esophageal tissue progress silently before becoming clinically evident.

This delay in diagnosis contributes to poor survival rates and increased treatment complexity. Therefore, there is a critical need for reliable, non-invasive, and automated diagnostic approaches that can assist clinicians in detecting esophageal cancer at an early stage. Advances in medical imaging and deep learning techniques offer promising solutions by enabling accurate analysis of endoscopic images and early lesion identification.

Early detection remains a key challenge due to the lack of obvious symptoms and effective routine screening methods.

Conventional diagnostic techniques for esophageal cancer primarily include endoscopy, biopsy, and radiological imaging. While these methods are clinically effective, they are invasive, expensive, and time-consuming, requiring specialized equip- ment and highly trained medical professionals. Endoscopic procedures, in particular, may cause patient discomfort and carry certain procedural risks, making them unsuitable for routine large-scale screening. Additionally, the accuracy of diagnosis heavily depends on the expertise and experience of clinicians, leading to inter-observer variability and potential subjectivity in interpretation. These limitations hinder the widespread adoption of early screening programs, especially in resource-constrained settings and developing regions.

The increasing demand for efficient, accurate, and scalable diagnostic solutions has driven significant interest in the ap- plication of artificial intelligence (AI) within the healthcare domain. In recent years, AI-based systems have demonstrated exceptional performance in various medical image analysis tasks, including disease detection, classification, and local- ization. Among these techniques, deep learning has emerged as a powerful approach due to its ability to automatically learn complex patterns from large volumes of data. Unlike traditional machine learning methods that rely on handcrafted features, deep learning models can extract high-level and task- specific features directly from raw medical images.

Convolutional Neural Networks (CNNs), a specialized class of deep learning architectures, have proven particularly effec- tive in analyzing visual data. CNNs are capable of learning hierarchical feature representations, starting from low-level patterns such as edges and textures to high-level semantic features that represent pathological structures. This capability makes CNNs highly suitable for medical imaging applications, where subtle visual differences often indicate the presence of disease. Recent studies have demonstrated that CNN-based models can achieve performance comparable to, and in some cases exceeding, that of experienced clinicians in detecting abnormalities from medical images.

Motivated by these advancements, this work proposes a deep learning–based framework for the automated detection and classification of esophageal cancer using medical imaging modalities such as endoscopic and CT images. The proposed system aims to assist clinicians by providing an objective and reliable second opinion, thereby reducing diagnostic errors and workload. By incorporating image preprocessing techniques such as normalization, noise reduction, data augmentation, and segmentation, the system enhances image quality and improves model robustness. The ultimate goal of this re- search is to develop a clinically interpretable and scalable diagnostic tool that facilitates early detection, improves screen- ing efficiency, and contributes to better patient outcomes in esophageal cancer management.

II. Related Work

Automated detection of esophageal cancer using compu- tational techniques has gained increasing attention over the past decade due to the limitations of conventional diagnostic procedures. Early studies primarily relied on traditional ma- chine learning approaches, where handcrafted features such as texture descriptors, color histograms, edge features, and morphological characteristics were extracted from endoscopic or radiological images. These features were subsequently classified using algorithms such as Support Vector Machines (SVMs), k-Nearest Neighbors (k-NN), and Random Forest classifiers. While these approaches demonstrated initial suc- cess, their performance was highly sensitive to feature selec- tion and image quality, often resulting in limited robustness across heterogeneous datasets and clinical environments.

With the emergence of deep learning, researchers began adopting Convolutional Neural Networks (CNNs) to overcome the limitations of manual feature engineering. CNNs provide an end-to-end learning framework capable of automatically extracting hierarchical and discriminative features directly from raw images. Zhang et al. proposed one of the early CNN-based systems for detecting esophageal squamous cell carcinoma using endoscopic images. Their model achieved significantly higher sensitivity and specificity compared to traditional machine learning methods, demonstrating CNNs' ability to identify subtle mucosal and structural changes asso- ciated with early cancer stages.

A major advancement in this field was introduced by Hirasawa et al., who developed a deep CNN model trained on more than 100,000 endoscopic images. Their system achieved high diagnostic accuracy in detecting early-stage esophageal cancer and showed performance comparable to experienced endoscopists. This study highlighted the effectiveness of large- scale data-driven deep learning models and emphasized the feasibility of AI-assisted diagnosis in real clinical settings. However, such large annotated datasets are often difficult to obtain, particularly in developing regions.

To address the challenge of limited labeled medical data, transfer learning has been widely adopted in subsequent re- search. Pre-trained deep learning architectures such as VGG, Inception, DenseNet, and ResNet—originally trained on large- scale natural image datasets—are fine-tuned for medical imag- ing tasks. Luo et al. utilized a fine-tuned ResNet-50 model for classifying esophageal lesions from narrow-band imaging (NBI) endoscopic images. Their approach demonstrated im- proved classification accuracy, faster convergence, and reduced overfitting, confirming the effectiveness of transfer learning in esophageal cancer detection.

Beyond classification, several studies have focused on tumor localization and segmentation to enhance clinical usefulness. Encoder–decoder architectures such as U-Net and its variants have been extensively used to delineate cancerous regions within endoscopic and CT images. Accurate segmentation enables precise localization of lesions, which is critical for cancer staging, treatment planning, and surgical decision- making. Hybrid frameworks that integrate segmentation and classification have shown improved diagnostic performance by combining global image context with localized lesion features. Despite notable progress, existing deep learning-based esophageal cancer detection systems still face significant chal- lenges. Model generalization across different imaging devices, institutions, and patient populations remains a major concern. Additionally, many deep learning models function as "black boxes," offering limited interpretability, which reduces clin- ician trust and hinders clinical adoption. The lack of ex- plainable AI mechanisms and limited integration with hospital information systems further restrict real-world deployment. These limitations underscore the need for robust, interpretable, and scalable deep learning frameworks, motivating the devel-
opment of the proposed system.

III. PROPOSED METHODOLOGY

A. Data Acquisition

Medical images used in this study include endoscopic images, computed tomography (CT) scans, and histopatho- logical slides, collected from publicly available and annotated datasets. These imaging modalities provide complementary in- formation about esophageal tissue morphology and patholog- ical changes. To ensure model robustness and generalization, data diversity is maintained across different imaging condi- tions and patient demographics. All datasets are anonymized to preserve patient privacy and comply with ethical standards. The collected data are labeled into three classes: healthy, precancerous, and malignant.

B. Preprocessing

Preprocessing is a crucial step to enhance image quality and ensure consistency across datasets. All input images are resized to a fixed resolution of height into width so that uniform input dimensions are maintained for the convolutional neural network model. This resizing helps the model process images efficiently and ensures consistency during training and testing.

Pixel intensity normalization is applied to scale the pixel values into a standard range between zero and one. Normal- ization improves numerical stability and accelerates model convergence during training. It is performed by subtracting the minimum pixel intensity value from the original pixel value and dividing the result by the difference between the maximum and minimum pixel intensity values.

Pixel Intensity Normalization Formula:

$$\frac{I - I_{min}}{I_{max} - I_{min}}$$

a) Bilateral Filtering Formula: $I_f iltered(x) = (1/W_p) *$
$sum[I(x_i) * f_s(|x - x_i|) * f_r(|I(x) - I(x_i)|)]$

C. Feature extraction using CNN

Feature extraction is a critical stage in the proposed sys- tem, where meaningful and discriminative features are au- tomatically learned from preprocessed medical images using a Convolutional Neural Network (CNN). Unlike traditional approaches that rely on handcrafted features, CNNs learn hierarchical feature representations directly from the input data, enabling robust detection of complex patterns associated with esophageal cancer.

Step 1: Input Image Representation

Let the preprocessed image be represented as: $F_k = (W_k * X + b_k)$
where H and W denote the height and width of the image,
and C represents the number of channels (e.g., C = 3 for RGB images or C = 1 for grayscale images).

Step 2: Convolution Operation

In each convolutional layer, a set of learnable filters (ker- nels) is applied to the input image to extract spatial features such as edges, textures, and lesion patterns. The convolution operation is mathematically expressed as:

Fk = Wk * X + bk

Step 3: Activation Function

To introduce non-linearity and enable the network to learn complex patterns, an activation function is applied to the feature map. In this work, the Rectified Linear Unit (ReLU) is used:

Fk = (Wk * X + bk)

where the ReLU activation is defined as:

(z) = max(0, z)

This activation function helps in mitigating the vanishing gradient problem and accelerates convergence during training.

Step 4: Pooling Operation

To reduce spatial dimensionality and computational com- plexity, a pooling operation (typically max pooling) is applied:

P(i, j) = max(F(i, j))

Pooling improves translational invariance and helps retain the most salient features while discarding redundant informa- tion.

Step 5: Transfer Learning

To enhance learning efficiency and performance on limited medical datasets, transfer learning is employed. A pre-trained

CNN model (such as ResNet) trained on large-scale datasets like ImageNet is fine-tuned for the esophageal cancer detection task. The transferred weights serve as an effective initializa- tion:

W = Wpretrained + W

where Wpretrained represents the pre-trained weights and W denotes task-specific weight updates.

This approach reduces training time, improves convergence, and enhances feature generalization across different imaging conditions.

Step 6: Feature Vector Generation

The final convolutional feature maps are flattened and transformed into a fixed-length feature vector:

$F_k = W_k * X + b_k$

This feature vector serves as the input to the subsequent
detection and classification layers.

D.Step 4: Feature Hierarchy Formation

As the input image propagates through successive convo- lutional layers of the CNN, the network progressively learns a hierarchical representation of features, ranging from simple visual patterns to complex semantic structures. In the initial convolutional layers, the filters primarily respond to low-level features such as edges, corners, color gradients, and basic texture patterns. These features capture fundamental visual information that forms the building blocks for higher-level understanding of the image content.

In intermediate layers, the CNN combines these low-level features to identify more informative patterns such as contours, regions of interest, and localized texture variations. At this stage, the network begins to recognize structural irregularities in esophageal tissue, including abnormal mucosal patterns, thickness variations, and early lesion indicators. These mid- level features provide greater discriminatory power compared to simple edge-based representations.

In deeper convolutional layers, the learned representations become increasingly abstract and task-specific. The network integrates spatial and contextual information to extract high- level features that correspond to clinically relevant patterns such as tumor shape, lesion boundaries, texture heterogeneity, and pathological structures associated with malignant and precancerous conditions. This hierarchical feature learning enables the CNN to effectively distinguish between healthy, precancerous, and malignant tissues, making it particularly suitable for automated esophageal cancer detection.

Step E: Pooling Operation

To reduce spatial dimensionality and computational com- plexity, a pooling operation (typically max pooling) is applied:

$P(i, j) = \max(F(i, j))$

Pooling improves translational invariance and helps retain the most salient features while discarding redundant informa- tion.

Step F: Transfer Learning

To enhance learning efficiency and performance on limited medical datasets, transfer learning is employed. A pre-trained CNN model (such as ResNet) trained on large-scale datasets like ImageNet is fine-tuned for the esophageal cancer detection

Fig. 1. Esophageal cancer workflow

task. The transferred weights serve as an effective initializa- tion:

$W = Wpretrained + W$

where Wpretrained represents the pre-trained weights and W denotes task-specific weight updates.

This approach reduces training time, improves convergence, and enhances feature generalization across different imaging conditions.

Step 7: Feature Vector Generation

After the hierarchical feature extraction process, the final convolutional layers produce a set of high-level feature maps that encode rich spatial and semantic information related to esophageal tissue characteristics. To enable classification, these multidimensional feature maps must be converted into a one-dimensional representation suitable for fully connected layers. This transformation is achieved through a flattening operation, which reshapes the feature maps into a single continuous vector.

$f \quad R^{CHW}$

The generated feature vector serves as a compact and discriminative representation of the input image and is subse- quently passed to the detection and classification layers. These layers utilize the feature vector to perform class prediction by learning decision boundaries between healthy, precancerous, and malignant categories. By converting complex spatial in- formation into a structured numerical form, this step bridges the gap between feature extraction and final classification, enabling accurate and efficient automated esophageal cancer detection.

RESULT AND DISCUSSION

The overall working of the proposed Esophagus Cancer Classification system follows a structured pipeline consisting of input acquisition, CNN-based processing, prediction gener- ation, and performance analysis.

Input Image

An endoscopic image of the esophagus is uploaded through the web-based user interface. Before being processed by the CNN model, the image is resized to a fixed dimension and normalized so that all pixel values lie within a standard range. This preprocessing ensures uniform input size and improves model stability and accuracy during classification.

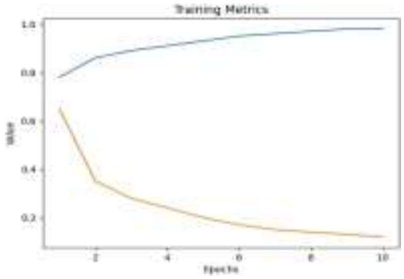

Fig. 2. Training Accuracy vs Training Loss

CNN Processing

The preprocessed image is passed through multiple convo- lutional layers, where filters extract important features such as edges, textures, color variations, and lesion patterns present in esophageal tissue. Pooling layers follow the convolution layers to reduce the spatial dimensions of the feature maps while retaining the most significant features, thereby lowering com- putational complexity and preventing overfitting. The extracted feature maps are then flattened and passed to fully connected layers, which perform high-level reasoning and classify the image based on learned representations.

Prediction Output

After processing through the CNN, the model predicts the disease category of the input image. In the given output, the predicted class is dyed-lifted-polyps. The system also displays the prediction confidence, which is shown as 100

Performance Visualization

To evaluate the effectiveness of the CNN model, several performance metrics are visualized. The training graph il- lustrates the increase in training accuracy and the decrease in training loss over epochs, confirming effective learning. The validation graph shows stable accuracy and low loss on unseen data, indicating good generalization ability. The confusion matrix provides a class-wise performance overview, where higher diagonal values represent correct classifications. Finally, the classification report summarizes precision, recall, F1-score, and overall accuracy, demonstrating the robustness and reliability of the proposed CNN-based esophagus cancer classification system. It consists of :

Training graph Validation graph Confusion matrix Classification report Training Graph :

Graph: Training Accuracy vs Training Loss

The training accuracy increases steadily from 78

The training loss decreases sharply from 0.65 to 0.12.

This indicates that the CNN model is learning effectively and converging well.

Low loss and high accuracy confirm good feature learning from esophagus images.

↑ Accuracy + ↓ Loss Effective Training Validation Graph :

Validation accuracy fluctuates between 82

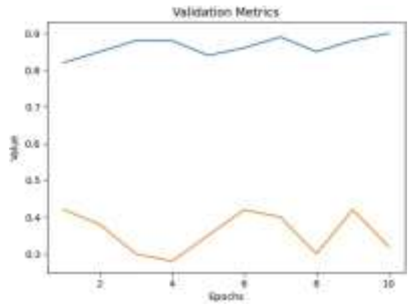

Fig. 3. Validation Accuracy vs Validation Loss

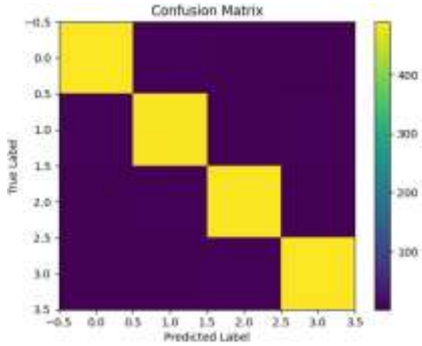

Fig. 4. confusion matrix

Validation loss remains relatively low with minor fluctua- tions.
Small variation between training and validation curves in- dicates:
No severe overfitting
Good generalization on unseen data
Training Accuracy Validation Accuracy Model Generalizes Well
Confusion Matrix :
What it Shows:
Rows → Actual class Columns → Predicted class
Most values are concentrated along the diagonal, meaning correct predictions.
Very few misclassifications between neighboring classes. Confirms high class-wise precision and recall.
Strong diagonal dominance High classification performance The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks . . .". Instead, try "R. B. G. thanks. . .". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

1. Y. Zhang, L. Zhang, H. Wang, and J. Wang, "Deep learn- ing–based detection of esophageal squamous cell carcinoma using endoscopic images," IEEE Transactions on Medical Imaging, vol. 38, no. 12, pp. 2926–2936, Dec. 2019.
2. T. Hirasawa et al., "Application of artificial intelligence using a convolutional neural network for detecting esophageal cancer in endoscopic images," Gastrointestinal Endoscopy, vol. 89, no. 3, pp. 625–632, Mar. 2019.

3. H. Luo, C. Xu, W. Jin, and J. Li, "Transfer learn- ing for esophageal lesion classification using deep convolu- tional neural networks," Computer Methods and Programs in Biomedicine, vol. 187, pp. 105–113, Apr. 2020.
4. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Con- volutional networks for biomedical image segmentation," in Proc. Int. Conf. Medical Image Computing and Computer- Assisted Intervention (MICCAI), Munich, Germany, 2015, pp. 234–241.
5. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.
6. R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Venice, Italy, 2017,
7. pp. 618–626.
8. G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, Dec. 2017.
9. J. Shin et al., "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset charac- teristics, and transfer learning," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285–1298, May 2016.
10. World Health Organization, "Cancer today: Esophageal cancer statistics," WHO, Geneva, Switzerland, 2023. [Online].
11. Available: https://www.who.int
12. S. Minaee et al., "Image segmentation using deep learning: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 7, pp. 3523–3542, July 2022.