Revolutionizing Esophageal Cancer Diagnosis: A Deep Learning-Based Method in Endoscopic Images

Shincy P Kunjumon, S Felix Stephen
Department of Electronics and Instrumentation Engineering,
Noorul Islam Centre for Higher Education, Tamil Nadu, India

***Abstract*—Esophageal cancer (EC) is a severe and commonly increasing disease due to the uncontrolled growth in the esophagus. It is the sixth leading cause of cancer-related deaths worldwide. The traditional methods for the diagnosis of EC are not only time-consuming but also suffer from inconsistencies due to human factors such as experience and fatigue. This paper proposes a deep learning (DL) approach for the detection of EC from endoscopic images to improve efficiency and accuracy. The study utilizes an endoscopic image dataset of 2000 images evenly split into cancerous and non-cancerous cases. After image preprocessing and augmentation, these images are fed into the proposed Inception ResNet V2 model. The extracted features were processed by the final classification layers and produced class probabilities. The simulation results revealed that the suggested model attained 98.50% of accuracy, 97.50% of precision, 98.75% of recall and 98.00% of F1 score after fine-tuning. These results underscore the model's capability to accurately identify EC, minimizing false positives and enhancing diagnostic reliability. The proposed DL framework for automated EC detection, promising advancements in clinical workflows and patient care.**

Keywords—Deep learning; esophagus cancer; transfer learning; endoscopic images; inception ResNet V2; fine tuning

I. Introduction

Cancer involves a range of illnesses caused by the uncontrolled growth of cells, which can impact any part of the body. Over the past century, the number of new cancer cases diagnosed within a specific period of time and mortality rates have significantly increased worldwide. This rise can be attributed to several factors, including changes in lifestyle, an aging population, genetic tendencies, and environmental influences such as pollution and dietary habits. Among the many types of cancer, EC is the 6th leading cause of cancer-related deaths worldwide, highlighting its severity and significant impact on public health [1]. In less developed regions the impact of EC is significantly greater, where 80% of cases arise. About 70% of these cases are diagnosed in males, with new diagnosis and mortality rates being two to five times greater in men compared to women, increasing with age. The frequency of EC is rising due to factors such as population growth and increased life expectancy. Risk factors like smoking and excessive alcohol consumption also play a role in the increase of EC, as depicted in Fig. 1 [2]. It begins within the mucosal layer of the esophagus and gradually extends outwards, making early identification more challenging. As a result, individuals might postpone seeking medical help until the cancer has reached an advanced stage. Therefore, it's crucial to raise awareness about risk factors and encourage early screening, particularly for individuals with specific demographics, lifestyle habits, or medical conditions [3].

EC can be broadly divided into 4 categories based on the type of cells from which the cancer originates as shown in Fig. 2. Squamous cell carcinoma (SCC) develops from the thin, flat cells lining the esophagus, with risk factors including smoking, excessive alcohol consumption, and specific dietary factors. Adenocarcinoma develops from glandular cells located in the lower part of the esophagus, close to the junction with the stomach. Risk factors for this type of cancer include obesity and smoking. Sarcomas, which develops from connective tissues such as muscle or cartilage in the esophagus, are rare and consist of only a small fraction of EC cases. Lymphoma, a cancer of the lymphatic system, can occur in the esophagus but is rare compared to other types of EC.

The TNM (Tumor, Node, Metastasis) staging system is employed in clinical practice to examine the extent of EC. It categorizes tumors on the basis of three factors: Tumor (T), assessing the size and invasion of the primary tumor; Node (N), indicating lymph node involvement; and Metastasis (M), evaluating distant organ spread [4]. Combining T, N, and M categories allows clinicians to stage EC (I-IV), assisting treatment decisions and providing prediction data. Conventional EC detection and classification involve manual inspection of endoscopic images by trained professionals, which is time-consuming and subjective. This approach results in variability in diagnoses and missed detections. Human interpretation can be influenced by factors like observer experience, tiredness, and personal opinion, affecting diagnostic accuracy and consistency. Thus, there's a demand for more objective and streamlined approaches to detect and classify EC, enhancing diagnostic accuracy, enabling early intervention, and improving patient outcomes [5].

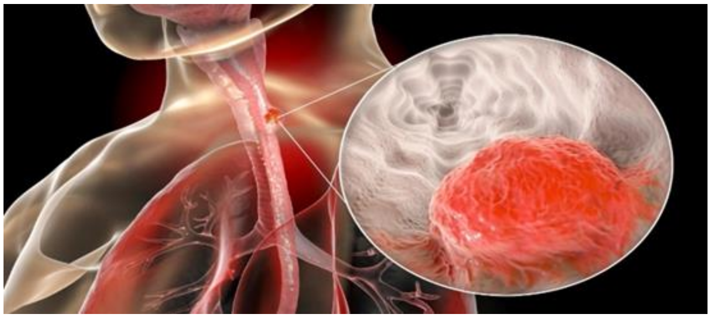


Fig. 1. Esophagus cancer.

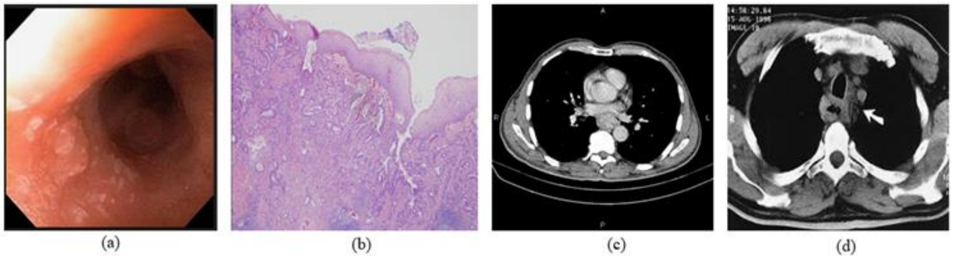


Fig. 2. Types of esophagus cancer: (a) Squamous cell carcinoma, (b) Adenocarcinoma (c) Sarcoma and (d) Lymphoma.

Deep learning, a subset of machine learning (ML) has shown impressive performances in extracting complex patterns and features from large datasets. This has resulted in notable progress in tasks like recognizing, classifying, and segmenting images. In medical imaging, such as endoscopic images for EC detection and classification, DL has shown immense potential. These algorithms can efficiently analyze vast amounts of medical images, accurately identifying diseases that are hidden. For example, DL models can distinguish between normal and abnormal tissue, detect early signs of cancerous lesions, and even predict disease progression based on imaging data [6]. Moreover, DL algorithms can be integrated into clinical workflows to help healthcare professionals in making more accurate and timely diagnoses. Automating the analysis of medical images can help lessen the burden on endoscopists and radiologists, enabling them to concentrate on cases requiring more complex interpretation. However, the use of DL in medical image analysis shows certain challenges including the need for large, high-quality labeled datasets, robust validation methods, and interpretability of model predictions. Addressing these challenges is crucial to ensure the reliability and safety of AI-assisted medical diagnosis and treatment. In this study, an effective DL approach for detecting the EC from endoscopic images is proposed. The method uses an Inception ResNet V2 model to categorize the endoscopic images accurately as “EC” or “No_EC”. The work proposed offers the following key contributions:

- To develop a DL-based model for the detection and classification of EC form endoscopic images.
- To improve the diagnosis of EC from endoscopic images and obtain an optimum accuracy.
- To compare and analyze the performance of the method suggested with the existing methods.

The paper proceeds as follows: Section II reviews previous methods relevant to the current study. Section III outlines the proposed approach. Section IV presents the experimental results and their interpretation. Finally, Section V provides the study's conclusion.

II. Literature Review

Chin et al. (2024) [7] aimed to develop a diagnostic system using DL to differentiate EC from non-contrast CT images of chests. They studied 398 people with EC and 255 healthy individuals without esophageal tumors. They employed a technique called nnU-Net for segmenting the esophagus and used a decision tree (DT) to determine the presence or absence of cancer. Their DL-based method demonstrated strong diagnostic performance, achieving 0.900 of sensitivity, 0.882 of accuracy, 0.880 of specificity, 0.890 of AUC and an 0.891 of F-score. Similarly, the study faced certain limitations, including difficulty in identifying all early-stage cancers and also the number of patients involved in this study is limited.

Li et al. (2024) [8] presented a deep-learning approach for segmenting EEC lesions. They utilized the YOHO framework, as it depends only on a single image from each patient to ensure complete patient privacy. This "one-image-one-network" learning strategy avoided the generalization issues by training the network exclusively on the input image itself, without using data from other patients. The YOHO framework was evaluated on an EEC dataset, attaining a mean Dice score of 0.888.

Yasaka et al. (2023) [9] studied the efficacy of a DL model in detecting EC on contrast-enhanced CT images. Their study comprised 252 patients with EC and 25 patients with No EC. They developed a DL model using data from patients with EC for training and validation. Then, they applied the developed model to a test dataset containing patients with and without EC, achieving AUCs of 0.98 and 0.95 for image-based and patient-based analyses, respectively. Also, the study shown certain limitations, including a training dataset of relatively small size and the restriction to patients with EC visible on CT images.

In their study, Fang et al. (2022) [10] used a semantic segmentation method to predict and label early-stage EC. They utilized a combination of ResNet and U-Net as the fundamental artificial neural network (ANN) architecture to extract the feature maps used in classifying and predicting the cancer’s location. A total of 90 narrow-band images (NBI) and 75 white-light images (WLI) were used. The research found that, on average, it took 111 ms to make predictions for each image in the test set. NBI showed 84.724% of high accuracy rate compared to WLI, which achieved 82.377%. These findings indicate that the proposed method is suitable for EC detection.

In their study, Tsai et al. (2022) [11] introduced a new method that integrate hyperspectral imaging (HSI) through band selection. They transformed WLIs into NBIs and developed a single-shot multi-box detector (SSD) model to predict the location and stage of EC, using a total of 1780 EC images. The outcomes shows that the mean average precision (mAP) for WLIs was 80%, for HSI images was 84% and for NBIs was 85%.

In their research, Zhang et al. (2022) [12] proposed an automated DL system for detecting esophageal cancer on barium esophagram. They employed five datasets derived from barium esophagram to progressively train, validate, and test the DLS. The method was evaluated and achieved a specificity, accuracy and sensitivity of 88.7%, 90.3% and 92.5%, respectively, in detecting EC. The study notes some limitations, such as the data collected only from a single medical center and the use of high-quality barium esophagram images for both testing and training purposes.

Mohammed (2022) [13] aimed to create a computer system utilizing modern image processing techniques and algorithms for the early identification of EC. The study employed the Fuzzy C-Means (FCM) algorithm for segmentation and clustering, and utilized a convolutional neural network (CNN) algorithm for detection. When tested on 100 color esophagogastroduodenoscopy (EGD) images, the proposed system achieved an accuracy of 95%. Observations indicated that combining these two algorithms enhanced the detection of EC.

Gong et al. (2022) [14] conducted a study where they developed a DL model capable of diagnosing ECs, non-neoplasms, and precursor lesions using endoscopic images. A total of 5163 (WLIs) were used to train and test the model. They utilized a no-code DL tool to build the model. It achieved an internal test accuracy of 95.6%, with precision at 78.0%, F1 Score at 85.2%, and recall at 93.9%. Furthermore, the external test accuracy reached 93.9%. However, a limitation of the study was that the established model's diagnostic performance was comparatively lower in comparison to other classes.

Chen et al. (2021) [15] introduced an EC detection model based on DL. They employed the Faster RCNN method, incorporating a technique called online hard example mining (OHEM), for detecting objects in EC images. The experiment included 1525 gastrointestinal CT images collected from 420 patients. The improved Faster RCNN's performance was examined by evaluating its mAP, F-1 measure and detection time. The experimental results indicated that the improved Faster RCNN outperformed the other two networks. The proposed method achieved a mAP of 92.15%, an F-1 measure of 95.17%, and a detection time per CT of only 5.3 seconds.

Takeuchi et al. (2021) [16] proposed a system based on AI for diagnosing EC from CT images, employing a group of 458 patients with primary EC in their study. A DL based image recognition model VGG16, was fine-tuned specifically for detecting EC. The CNN's diagnostic accuracy was examined using a test dataset comprising 46 cancerous images from CT scans and 100 non-cancerous images. The CNN-based system demonstrated an F-value of 0.742, a diagnostic accuracy of 84.2%, a specificity of 90.0% and a sensitivity of 71.7%. The study's limitations include insufficient datasets, which limits the model's performance.

Tsai et al. (2021) [17] employed an HSI and a DL model to determine the stage of EC and mark their positions. The study generated spectral data from the images using a special algorithm developed for this purpose. An SSD system was used in DL methods for the diagnosis and classification of EC. The prediction model for EC was evaluated using WLI and NBI images. The accuracy in detecting EC was 88% for WLI and 91% for NBI. Additionally, the algorithm required 19 seconds for result prediction.

Sui et al. (2021) [18] aimed to develop a DL model using the thickness of esophagus for detecting EC from unenhanced CT images. They identified 141 patients with EC and 273 without EC for the model training. A CNN model was created by collecting unenhanced CT images for diagnosing EC. Specifically, in this study, CNN utilized a VB-Net segmentation model, designed to separate the esophagus in images, measure the thickness of the mucosal layer of the esophagus and identify any lesions in the esophagus. The model's results demonstrated an average specificity of 74.33%, an average sensitivity of 77.67% and an average accuracy of 76%. The study’s limitation highlighted that the developed DL model depended only on the thickness of the mucosal layer of the esophagus and couldn’t identify the texture and other radiomic features.

There are several gaps in current research related to the detection and segmentation of esophageal tumors from unenhanced CT images. Firstly, it's challenging to identify these images and tumors specifically around the esophagogastric junction. Secondly, the detection performance is said to be weak when dealing with low-quality images. Additionally, the model's performance can heavily depend on the size of the learning rate used during training. Moreover, if the initial weight vector of a neuron is too distant from the input vector, it can lead to a decrease in the performance. Also, there's a poor prediction performance when generating depth maps. Tumors at different stages vary significantly in its shape, volume, and complexity, which affects the accuracy of automated segmentation. Finally, the use of limited and biased datasets during training may have limited the overall performance of DL-based models.

III. Materials and Methods

Detecting EC from endoscopic images is crucial for medical diagnosis. In this study, a DL model incorporating an Inception ResNet V2 is utilized for the precise detection and classification of EC. An outline of the work suggested is illustrated in Fig. 3. The model takes endoscopic images from the dataset as input. These images undergo further preprocessing and augmentation. Subsequently, the preprocessed images are given as an input to the pretrained Inception ResNet V2 model to identify the features and classifies the images into two categories: “EC” or “No_EC”.

A. Dataset Description

The dataset for detecting EC from endoscopic images was obtained from the Kaggle repository [19]. It comprises 2000 endoscopic images, with 1000 images depicting EC and 1000 images showing no EC. These images are stored in "jpg" format, ensuring ease of access and compatibility. Fig. 4 displays sample images from the endoscopic image dataset.

B. Data Preprocessing and Augmentation

In the proposed framework for detecting EC from endoscopic images, preprocessing plays a crucial role in improving image quality by reducing noise and improving the contrast. This involves resizing the images and normalization. To enhance training efficiency, OpenCV was employed to standardize all images to 224x224 pixels. Data augmentation increases dataset sizes by applying random alterations to existing images. Techniques such as rotation, flipping, shearing, and zooming create varied versions of the original images enhancing the model’s ability to generalize and recognize cancerous patterns under different conditions [20]. Following data augmentation, the dataset is split into training and testing sets with a ratio of 80:20.

C. Proposed Methodology

1) Convolutional Neural Network (CNN): A CNN network is a DL model designed specifically for processing and analyzing visual data. It comprises various layers, such as pooling layers, convolutional layers and fully connected layers. The CNNs architecture is illustrated in Fig. 5. In this architecture, features from input images are extracted by the convolutional layers by applying filters or kernels across the image. These layers capture patterns such as textures edges and shapes. The feature maps produced from convolutional layers are subsequently down-sampled by pooling layers, which lowers the spatial dimensions of the data without losing important information. Finally, the fully connected layers process the features extracted and carry out regression or classification tasks. CNNs are efficient at recognizing objects in images due to their ability to share parameters and connect nearby pixels. This helps them learn patterns at different levels, like shapes and textures. Consequently, CNNs are valuable for tasks such as object detection, image classification and segmentation [21].

2) Inception ResNet V2: Inception ResNet V2 is a deep CNN architecture that merges the principles of the Inception and ResNet models [22]. This hybrid model is employed for detecting EC. The basic architecture of the Inception ResNet V2 model, is shown in Fig. 6, which includes the inception modules, convolutional layers and residual connections.

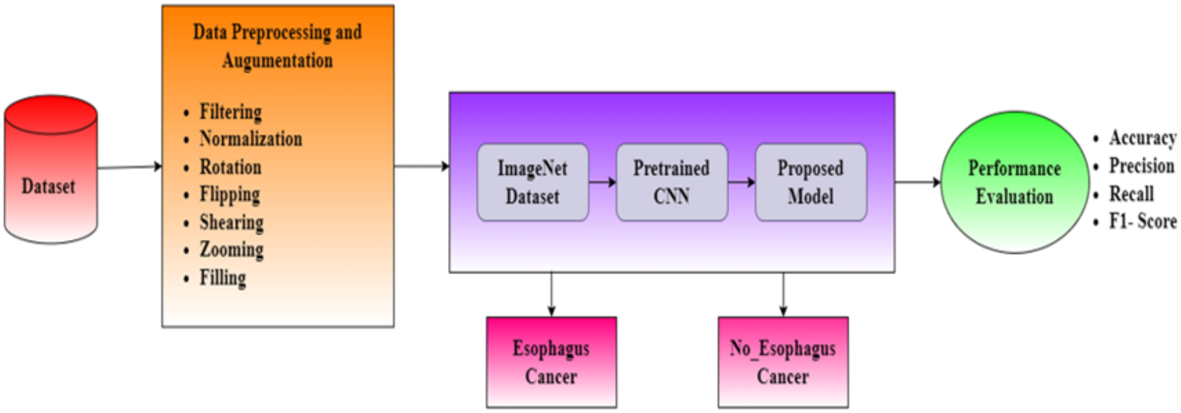


Fig. 3. Block diagram of the proposed methodology.

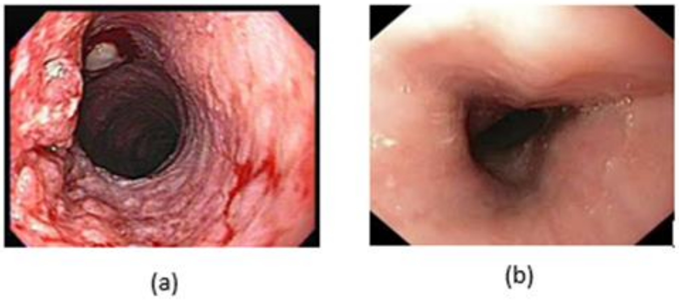


Fig. 4. Sample endoscopic images of (a) Esophagus cancer (b) Normal.

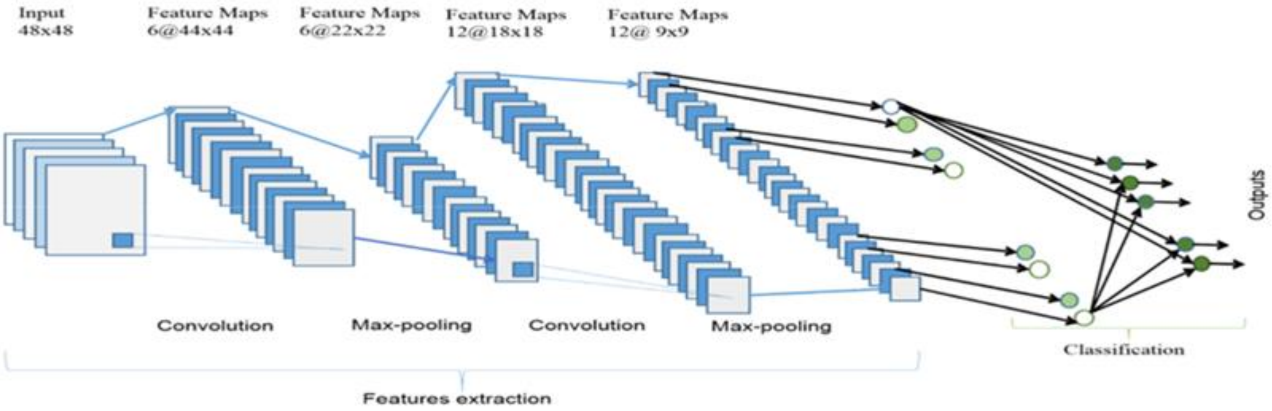


Fig. 5. Basic block diagram of CNN.

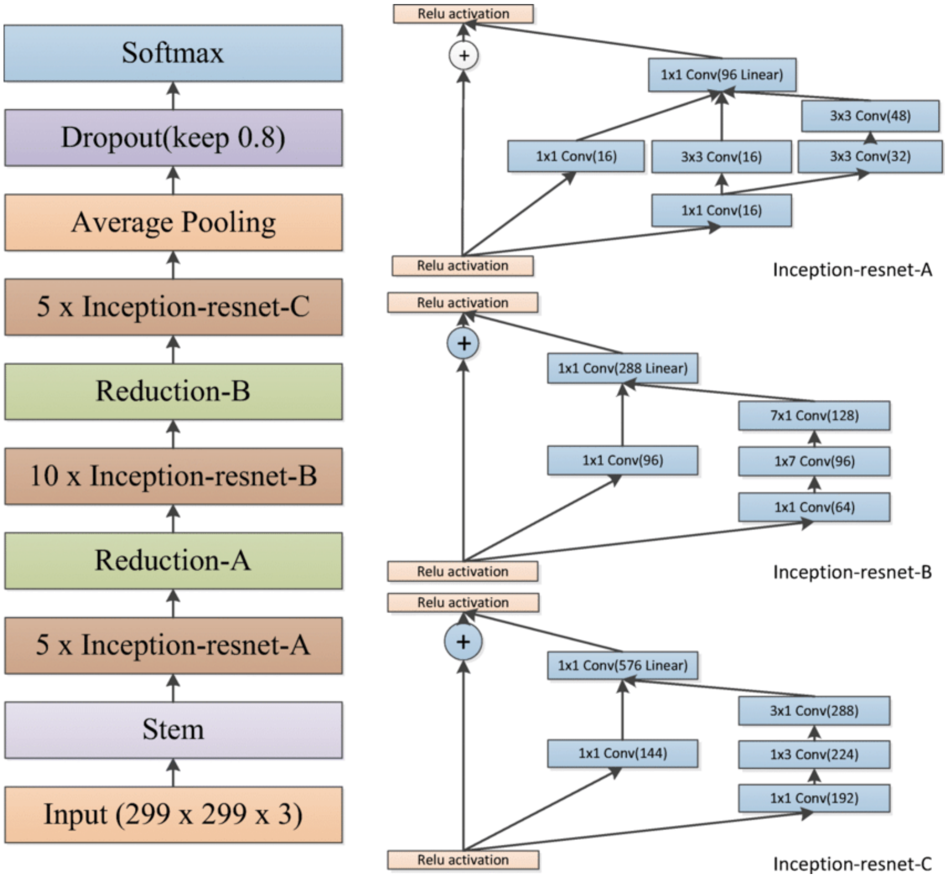


Fig. 6. Basic architecture of inception ResNet V2 model.

The endoscopic images are given as input to the Inception ResNet V2 model, functioning as a feature extractor. This allows the model to capture the features from the input image, including textures, shapes, and patterns associated with EC. The inception modules within the architecture conduct parallel convolutions at different scales, facilitating the model in capturing multi-scale features. The residual connections within each block facilitates gradient propagation. By combining the advantages of inception modules and ResNet's skip connections, Inception-ResNet V2 achieves high accuracy and computational efficiency in deep network training. In the proposed fine-tuned model, a pre-trained Inception ResNet V2 functions as a feature extractor, extracting significant features from the input endoscopic images. These features extracted are subsequently fed into dense layers comprising fully connected neural network layers for classification. The proposed framework architecture is depicted in Fig. 7.

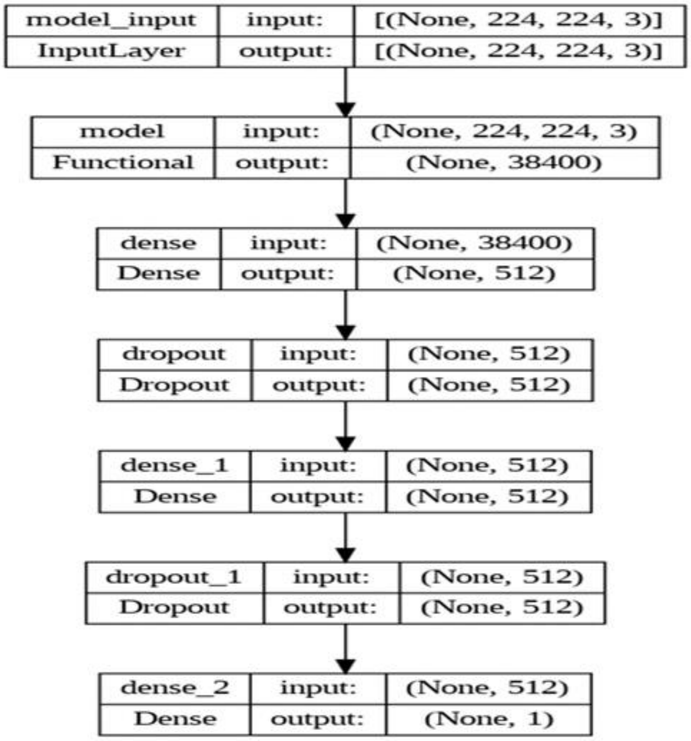

model_input | input: | [(None, 224, 224, 3)]
InputLayer | output: | [(None, 224, 224, 3)]
model | input: | (None, 224, 224, 3)
Functional | output: | (None, 38400)
dense | input: | (None, 38400)
Dense | output: | (None, 512)
dropout | input: | (None, 512)
Dropout | output: | (None, 512)
dense_1 | input: | (None, 512)
Dense | output: | (None, 512)
dropout_1 | input: | (None, 512)
Dropout | output: | (None, 512)
dense_2 | input: | (None, 512)
Dense | output: | (None, 1)

Fig. 7. Proposed model architecture.

The initial dense layer comprises 512 units and employs the ReLU activation function, introducing non-linear characteristics to the model, enables to learn more complex patterns and relationships within the data. To reduce the issue related to overfitting, a dropout layer with a dropout rate of 0.3 is applied after the initial dense layer. Following, another dense layer with 512 units and ReLU activation, similar to that of the previous layer, captures high-level representations and patterns from the data, while a dropout of 0.3 is applied again to prevent overfitting. At last, the output layer of the model consists of a single neuron with sigmoid activation. This configuration is well-suited for binary classification, effectively distinguishing between endoscopic images depicting “EC’’ or “No_EC”.

Fine-tuning is a specific approach within the TL where the pretrained model's parameters are fine-tuned using the new dataset as shown in Fig. 8 [23].

In endoscopic image-based EC detection, fine-tuning involves in adapting a pretrained deep learning model, such as Inception-ResNet V2, that has previously been trained on a large dataset. During fine-tuning, the initial layers of the pretrained model, which capture general features are kept fixed or "frozen" to preserve the knowledge gained during the original training. This ensures that the model retains its ability to recognize basic patterns and structures. Next, the model is trained on the new dataset of endoscopic images depicting EC. This model adjusts the weight of the latter layers to extract features specific to the EC detection. These later layers, starting from the 600th layer onwards in this case, are responsible for capturing more specific features relevant to the new task or dataset. Applying a low learning rate during fine-tuning allows the later layers of the model to adapt slowly to the new dataset. As the training progresses, the model learns to extract task-specific features from the endoscopic images, such as shapes, textures, and patterns associated with EC for accurate predictions. Finally, the dense layers at the end of the model are used for classification, detecting whether the endoscopic images depict EC or not. Table I provides a summary of the proposed model, both before and after fine-tuning.

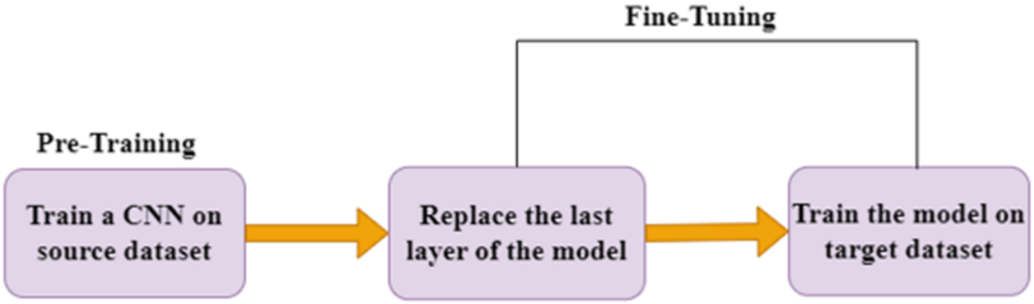

Fine-Tuning
Pre-Training
Train a CNN on source dataset
Replace the last layer of the model
Train the model on target dataset

Fig. 8. Block diagram of fine-tuning.

TABLE I. SUMMRY OF THE DESIGNED MODEL

	Before Fine-tuning	After Fine-tuning
Total Parameters	74,261,217	74,261,217
Trainable Parameters	19,924,481	46,509,569
Non-Trainable Parameters	54,336,736	27,751,648

The algorithm for the proposed model is outlined below:

```
Algorithm
Input: Endoscopic image dataset, labels determine Esophagus cancer or No_Esophagus cancer.
Output: Predictions of whether the input image contains esophagus cancer or not
Begin:
Load and preprocess data:
    1. Collect dataset: S= {($M_i, n_i$), where $M_i$ is an endoscopic image and $n_i \in \{0,1\}$ $n_i i \in \{0,1\}$ (1: No_EC, 0: EC).
    2. Preprocess:
        • Resize: $M_i \rightarrow M_i' \in R^{224\times224}$
        • Normalize: $M_i' \rightarrow \frac{M_i'-\mu}{\sigma}$
        • Data Augmentation: $M_i' \rightarrow \{M_i''\}$ (Shear, Zoom, Flipp, Rotation)

Define Base Models:
    1. Load Inception ResNet V2
    2. Input: $224 \times 224 \times 3$
           Dense (512, activation='relu')
           Dropout (0.3)
           Dense (512, activation='relu')
           Dropout (0.3)
           Dense (1, activation='sigmoid')
    3. Weight Initialization: Random initialization for new layers.

Fine tune the Model:
    1. Compile Modified Model:
       model. Compile (loss='binary_crossentropy', optimizer='Adam')
    2. Fine-tuning from the 600th Layer Onwards:
           for l≥600, layer. trainable=True
           for l<600, layer. trainable=False
    3. Train the Model with Fine-tuning Hyperparameters:
           Base_learning_rate= $\eta$
           Lower_layer_learning_rate= $\frac{\eta}{10}$
           Optimizer_higher_layers= Adam(learning_rate= $\eta$)
           Optimizer_higher_layers= Adam(learning_rate= $\frac{\eta}{10}$)
       history = model.fit (train_data, epochs=num_epochs, validation_data= (val_data, val_labels))
    4. Update the Lower Layers:

Model Evaluation:
    1. Evaluate:
           metrics=M.evaluate( $X_{test}$ , $y_{test}$ ), where metrics include accuracy, precision, recall and f1- score.
    2. Adjust Hyperparameters:
       if test_accuracy < desired_accuracy: adjust hyperparameters and retrain

Save the Model
End
```

D. Hardware and Software Setup

The method proposed for detecting EC from endoscopic images is implemented and evaluated on the Google Colaboratory platform. Two different learning rates, 0.0001 and 0.00001, are selected for the training process. The Adam optimizer is chosen for its effectiveness in optimizing DL models by adapting the learning rate during training. Additionally, the binary crossentropy loss function is employed which is commonly used for the binary classification distinguishing "EC" and "No EC". A batch size of 8 samples per iteration is utilized during training such that the model processes eight images at a time before updating its parameters. The training process is conducted over 10 and 20 epochs, with each epoch representing one complete pass through the entire training dataset. The hyperparameters of deep neural networks are determined empirically and have a notable impact on the learning process, as detailed in Table II.

TABLE II. HYPERPARAMETERS

Parameters	Value
Image Size	224*224
Batch Size	8
Optimizer	Adam
Learning rate	0.0001, 0.00001
Number of epochs	10,20
Activation function	Relu, Sigmoid
Loss	Binary crossentropy
Class mode	Binary

IV. RESULTS AND DISCUSSION

A. Evaluation Metrics

Evaluation metrics offer a quantitative assessment of performance of the model, facilitating a structured and a comprehensive evaluation of its effectiveness. Table III shows several key evaluation criteria from the proposed study.

Table IV shows the classification report of the proposed models for EC detection from endoscopic images, revealing significant performance improvements after fine-tuning. Initially, the model achieved 94.49% accuracy, which is increased to 98.50% after fine-tuning. Precision improved from 95.99% to 97.50%, while recall rise from 96.24% to 98.75%. The F1-score also increased substantially from 94.99% to 98.00%. These enhancements demonstrate that fine-tuning the model led to a more accurate and precise classification performance, effectively identifying positive instances while minimizing false positives.

TABLE III. EVALUATION METRICS

$Accuracy = (T_P + T_N)/(T_P + T_N + F_P + F_N)$	(1)
$Recall = (T_P)/(T_P + F_N)$	(2)
$Precision = (T_P)/(T_P + F_P)$	(3)
$F1 - Score = 2[(Recall * Precision)/(Recall + Precision)]$	(4)
$T_P = True\ Positive, T_N = True\ Negative, F_P = False\ Positive, F_N = False\ Negative$	

TABLE IV. CLASSIFICATION REPORT OF PROPOSED METHOD

Metrics	Before Fine-tuning	After Fine-tuning
Accuracy	94.49 %	98.50 %
Precision	95.99 %	97.50 %
Recall	96.24 %	98.75 %
F1-score	94.99 %	98.00 %

Plots, such as accuracy and loss plots, are utilized in EC detection using endoscopic images to visualize the performance of ML models during training. The accuracy plot displays how well the model performs in terms of correctly predicting the target variable over each epoch of training. Conversely, the loss plot illustrates the value of the loss function across each epoch, representing how well the predictions of the model match with the actual target values. As training progresses, the accuracy tends to increase while the loss decreases, indicating that the model is learning to make more accurate predictions.

Fig. 9 illustrates the accuracy plot and loss plot of the model before fine-tuning. Initially, in Epoch 1, the proposed model attained an accuracy of around 83.36% on the training dataset and 91.87% on validation dataset, indicating a good performance at the start of training. With each epoch, the accuracy gradually improves. By the final epoch (Epoch 10), the accuracy increases to approximately 96.02% on the training dataset and 96.25% on the validation dataset. Regarding loss of the model, it begins with a relatively high value of 0.7302 in the initial epoch and progressively decreases over subsequent epochs. By the final epoch, the loss reduces to 0.1039, indicating that the model's predictions become more accurate as training progresses.

Fig. 10 illustrates the accuracy plot and loss plot of the model after fine-tuning. Initially, in Epoch 1, the proposed model attained an accuracy of about 90.39% on the training dataset and 96.25% on the validation dataset. As training progressed, the accuracy is improved, reaching approximately 98.83% on the training dataset and 99.06% on the validation dataset at Epoch 30. Regarding the loss, the initial epoch (Epoch 10) showed high loss around 0.2356, indicating initial errors in prediction. However, as training continued, the loss slowly decreased, reaching approximately 0.0304 by the final epoch (Epoch 30). This decrease indicates that the model's predictive accuracy results in more precise classification as "EC" or "No EC".

A randomly selected image from the dataset is subjected to classification using the proposed model, accurately identifying it as either "EC" or "No EC." This successful classification, depicted in Fig. 11, underscores the model's effectiveness and reliability in accurately identifying and categorizing images within the dataset. Table V presents a comparison of the accuracy between the proposed and current techniques.

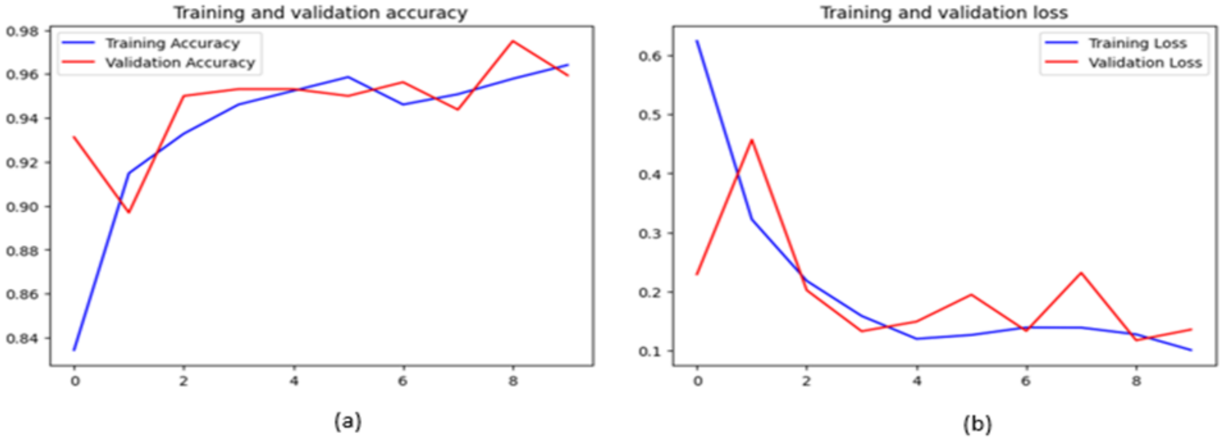


Fig. 9. (a) Accuracy plot and (b) Loss plot of the model before fine-tuning.

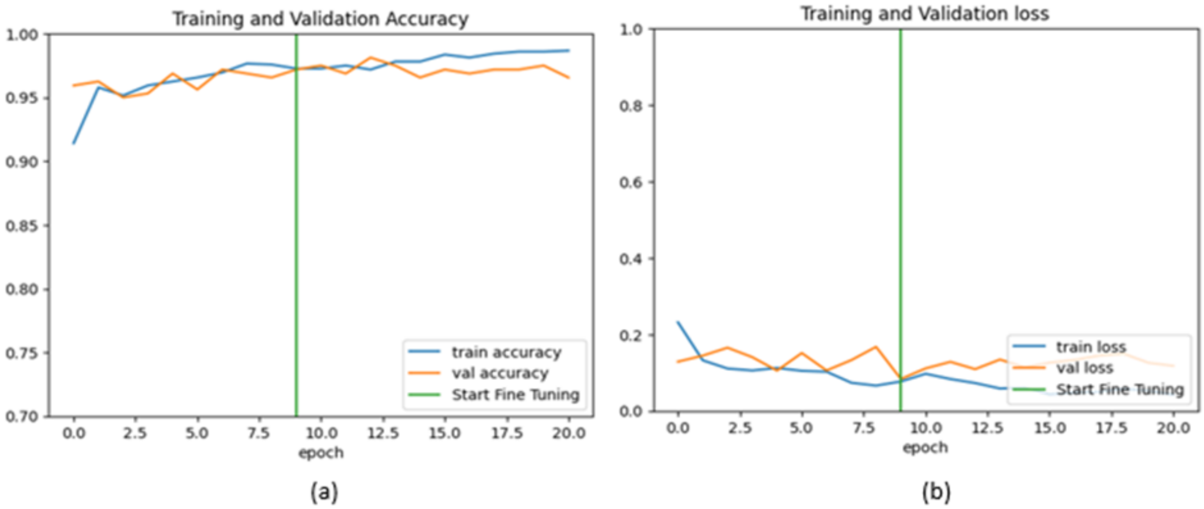
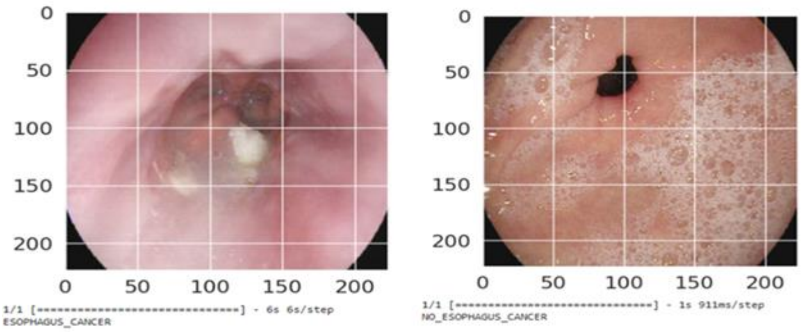


Fig. 10. (a) Accuracy plot and (b) Loss plot of the model after fine-tuning.



```
1/1 [==============================] - 6s 6s/step
ESOPHAGUS_CANCER
```

```
1/1 [==============================] - 1s 911ms/step
NO_ESOPHAGUS_CANCER
```

Fig. 11. Sample classification outputs.

TABLE V. COMPARISON BETWEEN THE PROPOSED METHOD AND EXISTING METHODS

SL. No:	Author	Methodology	Accuracy (%)
1.	Chong Lin et al. [7]	nnU-Net	88.20
2.	Fang et al. [10]	U-Net & ResNet	84.724 (NBI), 82.377 (WLI)
3.	Zhang et al. [12]	Two-stage DLS	90.3
4.	Mohammed [13]	FCM & CNN	95
5.	Gong et al. [14]	DL	95.6
6.	Chen et al. [15]	Faster RCNN	93.53
7.	Takeuchi et al. [16]	VGG16 CNN	84.2
8.	**Proposed Methodology**	**Inception ResNet V2 with Fine tuning**	**98.50**

The following table presents the comparison of the proposed method for EC diagnosis from endoscopic images with existing approaches. The proposed methodology which uses Inception ResNet V2 with fine-tuning, achieved the highest accuracy at 98.50%. This outperforms the performance of other methods, such as nnU-Net by Chong Lin et al. with 88.20% accuracy, U-Net & ResNet by Fang et al. with 84.724% (NBI) and 82.377% (WLI), and the two-stage DLS by Zhang et al. with 90.3%. Other notable methods include FCM & CNN by Mohammed at 95%, DL by Gong et al. at 95.6%, Faster RCNN by Chen et al. at 93.53%, and VGG16 CNN by Takeuchi et al. at 84.2%. The results indicate that the proposed method significantly outperforms existing approaches, highlighting its potential effectiveness in accurately diagnosing esophageal cancer from endoscopic images. Fig. 12 shows the graphical representation of the comparison of the proposed and existing approaches.

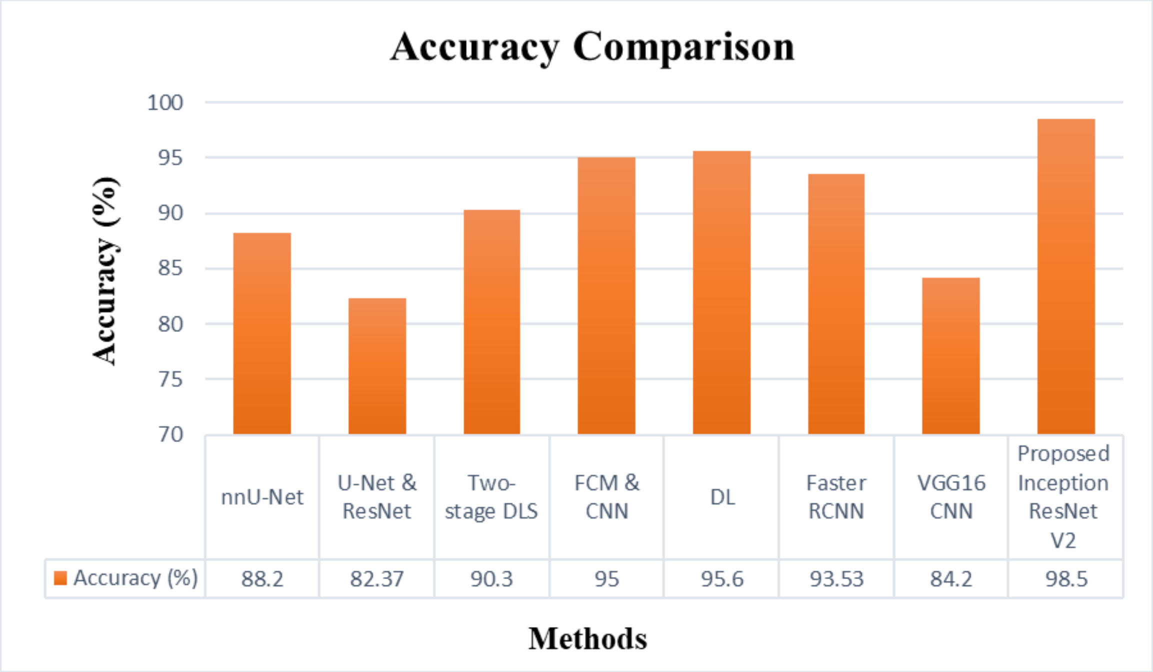


Fig. 12. Accuracy comparison of existing and proposed methods.

V. Conclusion

The early identification of EC is essential in enhancing treatment effectiveness and improving patient outcomes. This study proposes an effective DL method for detecting EC from endoscopic images. The methodology employed a deep CNN architecture, specifically the Inception ResNet V2 model. Preprocessed images are fed into the Inception ResNet V2 model, which serves as a feature extractor. TL was used to enhance the model for EC diagnosis. Through fine-tuning, the model successfully classified images depicting EC or not. The results show the efficacy of the suggested model showing significant improvements in the model exhibiting 98.50% of accuracy, 97.50% of precision, 98.75% of recall and 98.00% of F1 score. These improvements show that the model can accurately identify positive instances while minimizing the false positives which is crucial for the cancer diagnosis. Thus, the study presents a robust DL approach for EC detection from endoscopic images, providing exciting opportunities for enhancing treatment efficiency. One of the major limitations of the proposed work is its computational complexity.

Acknowledgment

I would like to express my sincere gratitude to all those who contributed to the completion of this research paper. I extend my heartfelt thanks to my supervisor, my family, my colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

References

[1] Uhlenhopp DJ, Then EO, Sunkara T, Gaduputi V. Epidemiology of esophageal. cancer: update in global trends, etiology and risk factors. Clin J Gastroenterol. (2020) 13:1010–21.

[2] Huang FL, Yu SJ. EC: risk factors, genetic association, and treatment. Asian J Surg. (2018) 41:210–5.

[3] Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. OEC. Nat Rev Dis Primers. (2017) 3:3. doi: 10.1038/nrdp.2017.48.

[4] Hong, S. J., Kim, T. J., Nam, K. B., Lee, I. S., Yang, H. C., Cho, S., ... & Lee, K. W. (2014). New TNM staging system for esophageal cancer: what chest radiologists need to know. Radiographics, 34(6), 1722-1740.

[5] Ba-Ssalamah A, Zacherl J, Noebauer-Huhmann IM, Uffmann M, Matzek WK, Pinker K, et al. Dedicated multi-detector CT of the esophagus: spectrum of diseases. Abdom Imaging. (2009) 34:3–18.

[6] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. (2015) 521:436–44. doi: 10.1038/nature14539.

[7] Lin, C., Guo, Y., Huang, X., Rao, S., & Zhou, J. (2024). EC detection via non-contrast CT and deep learning. Frontiers in Medicine, 11, 1356752.

[8] Li, H., Liu, D., Zeng, Y., Liu, S., Gan, T., Rao, N., ... & Zeng, B. (2024). Single-Image-Based Deep Learning for Segmentation of Early EC Lesions. IEEE Transactions on Image Processing.

[9] Yasaka, K., Hatano, S., Mizuki, M., Okimoto, N., Kubo, T., Shibata, E., ... & Abe, O. (2023). Effects of deep learning on radiologists' and radiology residents' performance in identifying EC on CT. The British Journal of Radiology, 96(1150), 20220685.

[10] Fang, Y. J., Mukundan, A., Tsao, Y. M., Huang, C. W., & Wang, H. C. (2022). Identification of early EC by semantic segmentation. Journal of Personalized Medicine, 12(8), 1204.

[11] Tsai, T. J., Mukundan, A., Chi, Y. S., Tsao, Y. M., Wang, Y. K., Chen, T. H., ... & Wang, H. C. (2022). Intelligent identification of early EC by band-selective hyperspectral imaging. Cancers, 14(17), 4292.

[12] Zhang, P., She, Y., Gao, J., Feng, Z., Tan, Q., Min, X., & Xu, S. (2022). Development of a Deep Learning System to Detect EC by Barium Esophagram. Frontiers in Oncology, 12, 766243.

[13] Mohammed, F. G., & Thamir, N. N. (2022). EC Detection Using Feed-Forward Neural Network. Webology, 19(1), 6121-6145.

[14] Gong, E. J., Bang, C. S., Jung, K., Kim, S. J., Kim, J. W., Seo, S. I., ... & Lee, J. J. (2022). Deep-learning for the diagnosis of ECs and Precursor Lesions in endoscopic images: a Model Establishment and Nationwide Multicenter Performance Verification Study. Journal of Personalized Medicine, 12(7), 1052.

[15] Chen, K. B., Xuan, Y., Lin, A. J., & Guo, S. H. (2021). EC detection based on classification of gastrointestinal CT images using improved

Faster RCNN. Computer Methods and Programs in Biomedicine, 207, 106172.

[16] Takeuchi, M., Seto, T., Hashimoto, M., Ichihara, N., Morimoto, Y., Kawakubo, H., ... & Sakakibara, Y. (2021). Performance of a deep learning-based identification system for EC from CT images. Esophagus, 18, 612-620.

[17] Tsai, C. L., Mukundan, A., Chung, C. S., Chen, Y. H., Wang, Y. K., Chen, T. H., ... & Wang, H. C. (2021). Hyperspectral imaging combined with artificial intelligence in the early detection of EC. Cancers, 13(18), 4593.

[18] Sui, H., Ma, R., Liu, L., Gao, Y., Zhang, W., & Mo, Z. (2021). Detection of incidental ECs on chest CT by deep learning. Frontiers in Oncology, 11, 700210.

[19] *Esophageal Endoscopy Images*. (2020, March 21). Kaggle. https://www.kaggle.com/datasets/chopinforest/esophageal-endoscopy-images/data.

[20] Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1–48.

[21] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. electronics, 8(3), 292.

[22] Wang, J., He, X., Faming, S., Lu, G., Cong, H., & Jiang, Q. (2021). A real-time bridge crack detection method based on an improved inception-resnet-v2 structure. IEEE Access, 9, 93209-93223.

[23] Chen, Z., Cen, J., & Xiong, J. (2020). Rolling bearing fault diagnosis using time-frequency analysis and deep transfer convolutional neural network. *Ieee Access*, *8*, 150248-150261.