GMR
The Original

G-NET LIGHT AND RU-NET LIGHT-BASED DEEP LEARNING FRAMEWORK FOR GASTROINTESTINAL ENDOSCOPIC IMAGE ANALYSIS

Dr.K. Vijay Chandra[1], A.Joy Pranahitha[2], Dr.K.Sudha Rani[3*], Lakkakula kavya[4], Poornaiah Billa[5]

[1]Assistant Professor, Dept. of EIE, VNR VJIET, Hyderabad, Telangana, India, vijaychandra_k @vnrvjiet.in
[2]Assistant Professor, Dept of CSE, Ravindra college of engineering for women, Joysofttech@gmail.com
[3]Associate Professor, Department of EIE, VNRVJIET, Hyderabad, Telangana, India, sudharani_k@vnrvjiet.in
[4]Assistant Professor, Dept of ECE, Sreenidhi Institute Of Science & Technology, Hyderabad, Telangana, India, kavya.l@sreenidhi.edu.in
[5]Professor, Department of ECE, Lakireddy bali reddy college of engineering and technology, India, poornaiah@gmail.com
***Corresponding author**: sudharani_k@vnrvjiet.in

ABSTRACT
Gastrointestinal endoscopy plays an important role in the early detection of inflammatory abnormalities and lesions; however, manual interpretation of endoscopic images is often affected by visual complexity and inter-observer variability. This study presents a deep learning-based framework for automated gastrointestinal endoscopic image analysis by integrating lightweight and residual segmentation networks with convolutional neural network-based classification and explainable artificial intelligence. The segmentation performance of G-Net Light and Residual U-Net (RU-Net Light) was comparatively evaluated using Dice coefficient, Intersection over Union (IoU), and pixel accuracy. In the overall segmentation analysis, G-Net Light achieved a Dice coefficient of 0.234, IoU of 0.146, and pixel accuracy of 84.7%, whereas RU-Net Light substantially improved these values to 0.726, 0.608, and 91.8%, respectively. For oesophagitis segmentation, RU-Net Light further obtained a Dice coefficient of 0.756, IoU of 0.637, and pixel accuracy of 92.5%, compared with 0.426, 0.298, and 86.7% achieved by G-Net Light. These findings demonstrate the effectiveness of residual learning and encoder–decoder skip connections in preserving lesion boundaries and extracting fine-grained mucosal features. Among the evaluated classification models, Xception achieved the highest accuracy of 93.33%, precision of 93.46%, recall of 93.09%, and F1-score of 93.14%. The integration of RU-Net Light, Xception, and Grad-CAM provides an accurate and interpretable framework for automated gastrointestinal lesion analysis and has potential to support gastroenterologists in reliable computer-assisted diagnosis.

KEYWORDS: G-Net, Residual U-Net (RU-Net Light),IoU

RELATED WORK

Gastrointestinal (GI) diseases, including oesophagitis, inflammatory disorders, polyps, ulcers, and gastrointestinal cancers, are major clinical concerns that require timely and accurate diagnosis. Endoscopy is widely used to directly examine the gastrointestinal tract and identify abnormal mucosal regions. However, the visual interpretation of endoscopic images depends considerably on the experience of the gastroenterologist and may be challenging because of variations in lesion size, shape, texture, illumination, and tissue appearance. Small or poorly defined abnormalities can be difficult to distinguish from surrounding healthy tissue, increasing the possibility of diagnostic variability. Therefore, automated computer-aided analysis of gastrointestinal endoscopic images has gained considerable attention as a supportive approach for improving lesion localisation and diagnostic consistency.

Recent advances in deep learning have significantly improved medical image segmentation and classification. Convolutional neural networks can automatically learn complex spatial and textural characteristics directly from endoscopic images without relying on manually designed features. Lightweight segmentation models are particularly attractive for real-time clinical applications because they require fewer computational resources and provide faster image processing. In this study, G-Net Light is employed as an efficient segmentation architecture with reduced filters and optimised pooling operations for extracting regions of interest from gastrointestinal images. Although its lightweight structure supports computational efficiency and scalability, the reduced feature representation may result in coarse segmentation masks and limited preservation of complex lesion boundaries.

To overcome these limitations, RU-Net Light integrates residual learning with the U-Net encoder–decoder architecture to improve gradient propagation, feature reuse, and spatial information preservation. Residual blocks allow deeper hierarchical features to be extracted, while skip connections transfer important localisation information from the encoder to the decoder for accurate reconstruction of lesion regions. The comparative results

demonstrate that RU-Net Light considerably outperforms G-Net Light, achieving an overall Dice coefficient of 0.726, IoU of 0.608, and pixel accuracy of 91.8%, compared with 0.234, 0.146, and 84.7%, respectively, for G-Net Light. In addition, multiple deep convolutional models, including VGG19, Xception, GoogleNet variants, ResNeXt, ConvNeXt-Tiny, ResNet18, and AlexNet, are evaluated for gastrointestinal image classification. The proposed framework further incorporates Grad-CAM-based visual explainability to highlight diagnostically important image regions and improve the transparency of automated predictions.

METHODOLOGY

Softmax Classification

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \quad (1)$$

Where:
z_i= Logit value
$P(y_i)$= Probability of lesion class

Cross Entropy Loss Function

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

Where:
y_i= Ground truth label
$\hat{y}_i$= Predicted probability

G-Net Light: G-Net Light is a neural network with the minimum weights to extract features efficiently and reliably without sacrificing the spatial information . It carries out the initial step of a gastrointestinal endoscopy study and has regions of interest like lesions and anomalies. It has fewer filters per layer and the pooling layers are optimised to run quickly and to use less memory. This leads to rough yet accurate segmentation masks. The reduction of the number of processes in these mask inputs demands processing networks which are designed to remove processes without the loss of the significant features. G-Net Light allows real time data processing and provides scalable data analysis. This first is the first phase in automated detection methods.

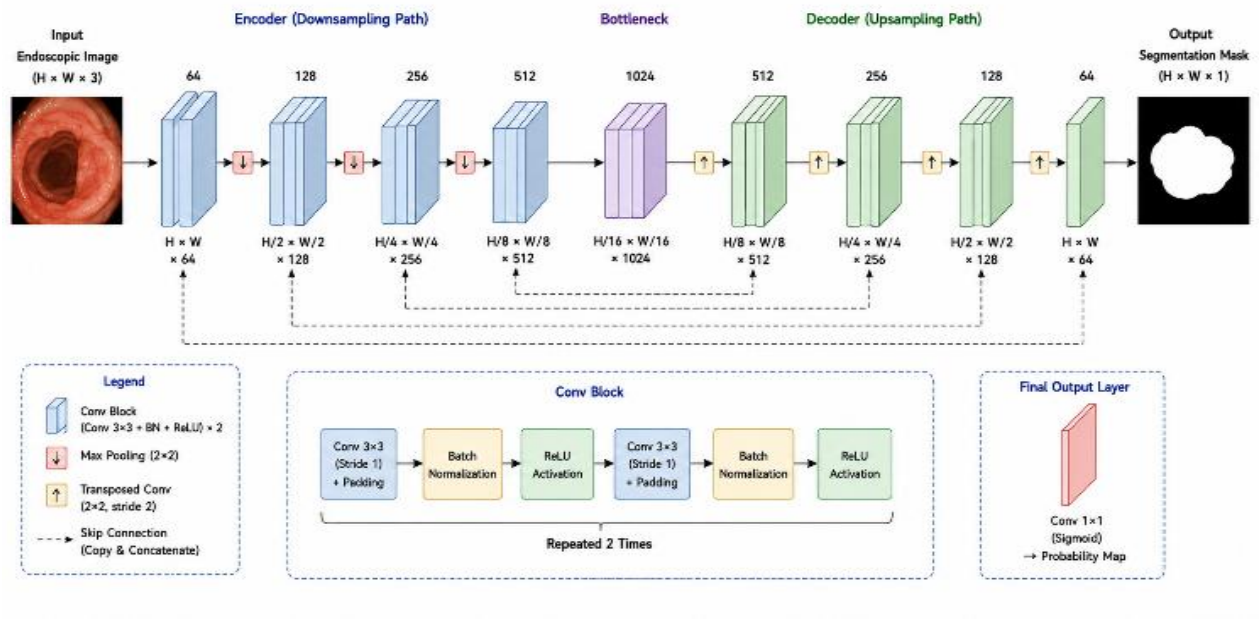

Fig.11 Architecture of the G-Net Light model

Mathematical Formulation

Convolution Operation

G-Net Light uses convolution layers to extract lesion-related gastrointestinal features.

$$Y(i,j) = \sum_{m} \sum_{n} X(i-m, j-n) K(m,n)$$

Where:
X= Input endoscopic image

K= Convolution kernel
Y= Output feature map

ReLU Activation Function

G-Net Light applies ReLU activation after convolution layers.

$$f(x) = \max(0, x)$$

Lightweight Feature Extraction Equation

Feature extraction in G-Net Light:

$$F_l = \sigma(W_l * F_{l-1} + b_l)$$

Where:
F_l= Feature map at layer l
W_l= Convolution weights
b_l= Bias
$sigma$= ReLU activation.

Pooling Operation

Pooling reduces feature dimensions while preserving important gastrointestinal lesion patterns.

$$P(i, j) = \max_{(m,n)\in R} X(m, n)$$

Where:
R= Pooling region
$P(i, j)$= Pooled output

Upsampling Equation

G-Net Light restores spatial resolution during segmentation.

$$U(x, y) = Interpolate(F(x, y))$$

Segmentation Output Equation

Pixel-wise segmentation prediction:

$$S(x, y) = \sigma(W * X + b)$$

Where:
$S(x, y)$= Segmentation probability map
W= Learnable weights
X= Input feature map
$sigma$= Sigmoid activation

Sigmoid Activation Function

Sigmoid converts segmentation outputs into probabilities.

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Dice Coefficient Equation

Dice coefficient evaluates segmentation overlap.

$$Dice = \frac{2 \mid P \cap G \mid}{\mid P \mid + \mid G \mid}$$

Where:
P= Predicted segmentation
G= Ground truth segmentation

Dice Loss Function

$$Loss_{Dice} = 1 - Dice$$

Intersection over Union (IoU)

$$IoU = \frac{\mid P \cap G \mid}{\mid P \cup G \mid}$$

Where:
$PcapG$= Intersection
$PcupG$= Union

Pixel Accuracy Equation

$$Pixel\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Binary Cross Entropy Loss

G-Net Light segmentation optimization:

$$L = -\frac{1}{N}\sum_{i=1}^{N} [\, y_i \log(\widehat{y_i}) + (1 - y_i)\log(1 - \widehat{y_i})]$$

Where:
y_i= Ground truth pixel
$\widehat{y_i}$= Predicted pixel probability

Computational Complexity Reduction
Lightweight architecture complexity:

$$Complexity \propto O(k^2 \times C \times H \times W)$$

Where:
k= Kernel size
C= Number of channels
H, W= Spatial dimensions

Residual U-Net (RU-Net): The Residual U-Net is an improvement over the original U-Net, with residual blocks added to aid in gradient flow and feature reuse. It is used for the study of gastrointestinal endoscopic images, performing fine-grained segmentation of lesions, including microaneurysms, haemorrhages and exudates. Skip connection maintains spatial data and residual blocks allow to gain more information about the image in fewer bits of data without degrading the quality. The network provides very accurate segmentation masks depicting lesion edges which enables the network to extract and classify features later. This multi scale representation, coupled with the enhanced propagation of gradients, is a useful and accurate way to identify anomalies and is one of the key elements in automated endoscopic analysis.

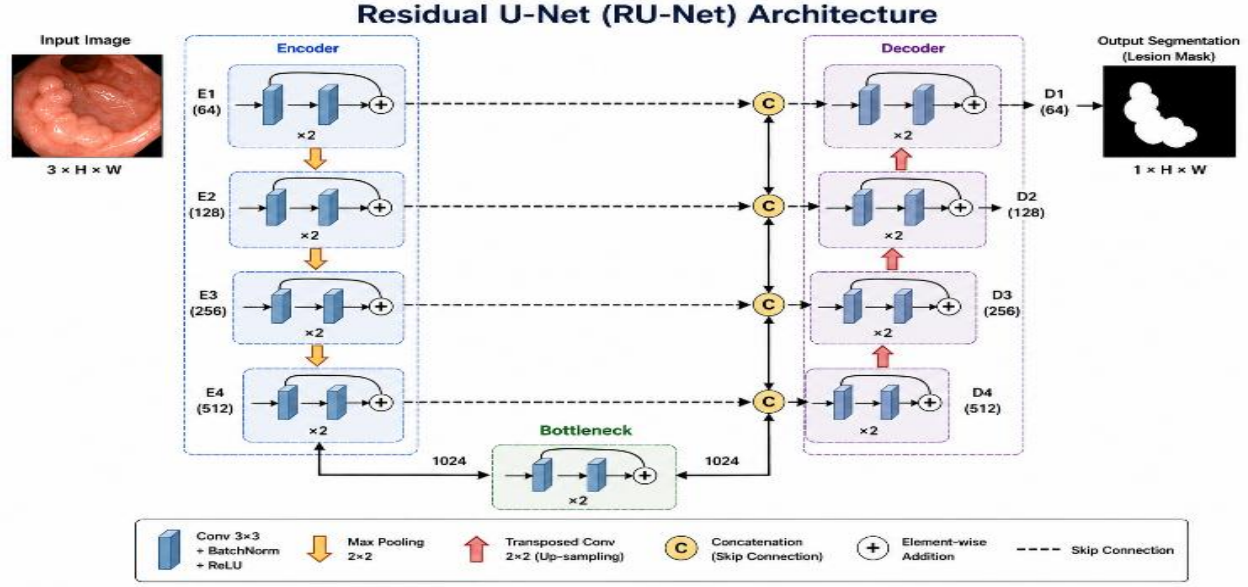

Fig.12 Architecture of the RU-Net Light model

Mathematical Formulation
Convolution Operation
RU-Net uses convolution layers for extracting gastrointestinal lesion features.

$$Y(i,j) = \sum_m \sum_n X(i-m, j-n)K(m,n)$$

Where:
X= Input endoscopic image
K= Convolution kernel
Y= Output feature map
ReLU Activation Function
RU-Net applies ReLU activation after convolution operations.

$$f(x) = \max(0, x)$$

Residual Learning Equation
Residual blocks are integrated into U-Net architecture.

$$H(x) = F(x) + x$$

Where:
$F(x)$= Residual mapping
x= Identity shortcut connection
$H(x)$= Final output
Residual Block Equation
RU-Net residual feature extraction:

$$F(x) = W_2\sigma(W_1x + b_1) + b_2$$

Final residual output:

$$y = F(x) + x$$

Where:
W_1, W_2= Convolution weights
b_1, b_2= Bias terms
$sigma$= ReLU activation

Encoder Feature Extraction Equation

Encoder path extracts hierarchical lesion features.

$$E_l = \sigma(W_l * E_{l-1} + b_l)$$

Where:
E_l= Encoder feature map
W_l= Convolution weights

Max Pooling Equation

Pooling reduces spatial dimensions.

$$P(i,j) = \max_{(m,n)\in R} X(m,n)$$

Where:
R= Pooling region

Decoder Upsampling Equation

Decoder restores segmentation resolution.

$$U(x,y) = Interpolate(F(x,y))$$

Upsampling reconstructs fine lesion boundaries.

Skip Connection Equation

RU-Net combines encoder and decoder features.

$$S_l = Concat(E_l, D_l)$$

Where:
E_l= Encoder features
D_l= Decoder features

Skip connections preserve spatial information for accurate lesion localization.

Segmentation Output Equation

Pixel-wise lesion prediction:

$$S(x,y) = \sigma(W * X + b)$$

Where:
$S(x,y)$= Segmentation mask probability
W= Learnable weights
$sigma$= Sigmoid activation

Sigmoid Activation Function

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

Maps segmentation outputs between 0 and 1.

Dice Coefficient Equation

Dice coefficient evaluates overlap between predicted and ground truth masks.

$$Dice = \frac{2 \mid P \cap G \mid}{\mid P \mid + \mid G \mid}$$

Where:
P= Predicted segmentation
G= Ground truth mask

Dice Loss Function

$$Loss_{Dice} = 1 - Dice$$

Minimizes segmentation mismatch during training.

Intersection over Union (IoU)

$$IoU = \frac{\mid P \cap G \mid}{\mid P \cup G \mid}$$

Where:
$PcapG$= Intersection
$PcupG$= Union

Binary Cross Entropy Loss

RU-Net optimization loss:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

Where:
y_i= Ground truth pixel
$\hat{y}_i$= Predicted probability

Pixel Accuracy Equation

$$Pixel\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Feature Representation Equation

Deep hierarchical feature learning:

$$F_l = \sigma(W_l * F_{l-1} + b_l)$$

Where:
F_l= Feature map
W_l= Learnable filters

Table.6 Segmentation Performance Esophagitis

Segmentation Model	Dice Coefficient	IoU	Pixel Accuracy
G-Net Light	0.426	0.298	0.867
RU-Net	0.756	0.637	0.925

Experimental results for Oesophagitis classification showed that the maximum accuracy of classification was 91.67% for Xception, which was due to its efficient depthwise separable convolution design that captures the patterns of inflammation in the mucosa and the occurrence of small anomalies. GoogleNet V4 and ResNeXt were also competitive, due to their ability to extract multi-scale contextual features. The segmentation results showed that the Dice coefficient, IoU and pixel accuracy of RU-Net were significantly higher than those of G-Net Light, indicating that RU-Net could localise the inflamed part of the oesophagus better and preserve boundary information of the gastrointestinal endoscopy lesions.

Table.7 Performance Evaluation - Classification

Model	Accuracy	Precision	Recall	F1-Score
VGG19	0.93000	0.93066	0.92858	0.92894
Xception	0.93333	0.93463	0.93087	0.93137
GoogleNetV4	0.93000	0.93079	0.92806	0.92844
ResNext	0.92833	0.92698	0.92654	0.92657
GoogleNetV2	0.92583	0.92733	0.92325	0.92376
GoogleNet	0.92500	0.92464	0.92242	0.92260
ConvNeXt-Tiny	0.92000	0.91836	0.91779	0.91765
ResNet18	0.90417	0.90731	0.90305	0.90304
AlexNet	0.90083	0.89960	0.89903	0.89903

Table.7 shows the classification accuracy, precision and recall and F1-Score of different DL models.

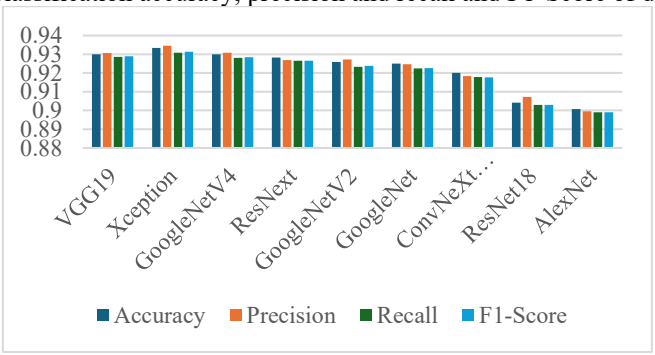

Fig.16 Comparison Graph - Classification

A comparison of the performance features of different DL models is shown in Fig. 16. Given this, it is observed that the accuracy of Xception is the highest.

Table.8 Performance Evaluation – Segmentation

Model	Dice	IoU	Pixel Acc
G-Net	0.234	0.146	0.847
RU-Net	0.726	0.608	0.918

Table.8 shows the segmentation results of G-Net and RU-Net in Dice, IoU and pixel accuracy .

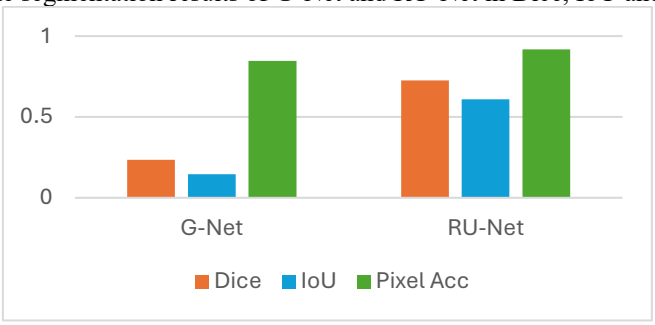

Fig.17 Comparison Graph - Segmentation

Fig. 17 shows that the Dice, IoU and pixel accuracy of RU-Net are significantly better than those of G-Net.

CONCLUSION

A combination of the DL framework of segmentation, classification and explainability algorithms has been applied for automatic interpretation of gastrointestinal endoscopic images, which was successfully done. In the segmentation task, RU-Net was able to accurately segment fine-grained margins of lesions and complicated mucosal structures. It had a Dice coefficient of 72.6, IoU of 60.8 and pixel accuracy of 91.8 which was better than the lightweight G-Net design. The benefits of segmentation are even more evident in the results of downstream diagnostic classification and even better feature representation. The analysed convolutional deep convolutational models had the best accuracy in classification which was 93.33 in Xception. This indicated that it did better than any of the other models at clarifying issues concerning the digestive system. Grad-CAM improvement: It enhances the clinical reliability by visualizing the most important parts of the image that contribute to making predictions. This has made it a lot more obvious and straightforward when to make selections. It was demonstrated that this deployment was practically viable for real-time clinical use and remote access through a web-interface and authentication of users with Flask. The global outcome will be an effective and unambiguous diagnosis system that will be scaled-up to support gastroenterologists to locate and identify lesions with precision and minimize the burden and uncertainty of diagnosis.

Future research will concentrate on improving the diagnostic ability by implementing multimodal data (e.g. the correlation between endoscopic pictures and patient metadata, histopathology report or laboratory results), to enhance the accuracy of predictions and provide a personalised diagnosis. For long-range contextual information in gastrointestinal images, one could consider using more sophisticated attention-based models like Vision Transformers or CNN-Transformer based models. Semi-supervised or self-supervised learning methods may reduce the amount of vast volumes of annotated data needed, and may make scaling to a wide range of clinical situations practicable. The real-time video endoscopic analysis, automated anomaly detection and cross-institutional validation can further improve the clinical application of the technique. More advanced XAI approaches, and more advanced implementation on the web, can enhance decision-making for gastrointestinal diagnostics, clarify and make it comprehensible.

REFERENCES:

[1] Gazzan, M., Alobaywi, B., Almutairi, M., & Sheldon, F. T. (2025). A Xia, S., Xia, Y., Liu, T., Luo, Y., & Pang, P. C. I. (2025). Application of deep learning models in gastric cancer pathology image analysis: a systematic scoping review. BMC cancer, 25(1), 1257.

[2] Rochmawati, N., Fatichah, C., Amaliah, B., Raharjo, A. B., Dumont, F., Thibaudeau, E., & Dumas, C. (2025). Deep Learning-Based Lesion Detection in Endoscopy: A Systematic Literature Review. IEEE Access.

[3] Sibomana, O., Saka, S. A., Uwizeyimana, M. G., Kihunyu, A. M., Obianke, A., Damilare, S. O., ... & Oveh, R. O. (2025). Artificial intelligence–assisted endoscopy in diagnosis of gastrointestinal tumors: a review of systematic reviews and meta-analyses. Gastro Hep Advances, 100754.

[4] Ahn, S., Hong, Y., Park, S., Cho, Y., Hwang, I., Na, J. M., ... & Kim, K. M. (2025). Development and application of deep learning-based diagnostics for pathologic diagnosis of gastric endoscopic submucosal dissection specimens. Gastric Cancer, 1-11.

[5] Kamble, A., Bandodkar, V., Dharmadhikary, S., Anand, V., Sanki, P. K., Wu, M. X., & Jana, B. (2025). Enhanced Multi-Class Classification of Gastrointestinal Endoscopic Images with Interpretable Deep Learning Model. arXiv preprint arXiv:2503.00780.

[6] Vania, M., Tama, B. A., Maulahela, H., & Lim, S. (2023). Recent advances in applying machine learning and deep learning to detect upper gastrointestinal tract lesions. IEEE Access, 11, 66544-66567.

[7] Wang, Y. Y., Liu, B., & Wang, J. H. (2025). Application of deep learning-based convolutional neural networks in gastrointestinal disease endoscopic examination. World Journal of Gastroenterology, 31(36), 111137.

[8] .Ali, H., Muzammil, M. A., Dahiya, D. S., Ali, F., Yasin, S., Hanif, W., ... & Al-Haddad, M. (2024). Artificial intelligence in gastrointestinal endoscopy: a comprehensive review. Annals of gastroenterology, 37(2), 133.
[9] Zhuang, H., Zhang, J., & Liao, F. (2023). A systematic review on application of deep learning in digestive system image processing. The Visual Computer, 39(6), 2207-2222.
[10] Lewis, J. R., Pathan, S., Kumar, P., & Dias, C. C. (2024). AI in endoscopic gastrointestinal diagnosis: A systematic review of deep learning and machine learning techniques. IEEE Access, 12, 163764-163786.
[11] Islam, M. M., Poly, T. N., Walther, B. A., Yeh, C. Y., Seyed-Abdul, S., Li, Y. C., & Lin, M. C. (2022). Deep learning for the diagnosis of esophageal cancer in endoscopic images: a systematic review and meta-analysis. Cancers, 14(23), 5996.
[12] Chandra, K.V., and Murari, B.M. Confocal corneal endothelium dystrophy's analysis using a hybrid algorithm. Journal of Engineering Science and Technology, vol. 15, no. 5, pp. 3419–3432 (2020).
[13] Chandra, K.V., and Murari, B.M. Confocal corneal endothelium dystrophy's analysis using particle filter. Journal of Engineering Science and Technology, vol. 15, no. 5, pp. 3419–3432 (2020).