GMR
The Original

RESNEXT-BASED DEEP LEARNING FRAMEWORK FOR AUTOMATED GASTROINTESTINAL ENDOSCOPY IMAGE CLASSIFICATION WITH RU-NET SEGMENTATION

Dr.K. Vijay Chandra[1], E.K.Mounika[2], Dr.K.Sudha Rani[3*], Dr.Swathi Sambangi[4], Dr Poornaiah Billa[5], N. Syamala[6]

[1]Assistant Professor, Dept. of EIE, VNR VJIET, Hyderabad, Telangana, India, vijaychandra_k @vnrvjiet.in
[2]Assistant Professor, Department of CSE, Ravindra college of engineering for women, Kurnool, India, ekmounika777@gmail.com
[3]Associate Professor, Department of EIE, VNRVJIET, Hyderabad, Telangana, India, sudharani_k@vnrvjiet.in
[4]Assistant Professor, Dept. of IT, VNRVJIET, Hyderabad, Telangana, India, swathi_s@vnrvjiet.in
[5]Professor, Department of ECE, Lakireddy bali reddy college of engineering and technology, India, poornaiah@gmail.com
[6]Assistant Professor, Department of ECE, VNR Vignana Jyothi institute of engineering and technology, India
syamala_n@vnrvjiet.in
***Corresponding author:** Dr.K.Sudha Rani, sudharani_k@vnrvjiet.in

Abstract- In order to improve patient outcomes and lower diagnostic mistakes during endoscopic examinations, gastrointestinal (GI) illnesses must be detected early and accurately. This study presents an intelligent deep learning framework for automated analysis of GI endoscopy images by combining image preprocessing, G-Net Light feature extraction, RU-Net-based lesion segmentation, and advanced convolutional neural network classifiers. A comparative evaluation of ResNet and ResNeXt was performed to assess their effectiveness in lesion classification. The experimental results indicate that ResNeXt provides superior performance by capturing richer multi-scale features through grouped convolutions while maintaining computational efficiency. In comparison to ResNet, which obtained 96.1% accuracy, 95.8% precision, 95.4% recall, 95.6% F1-score, and an AUC of 0.975, ResNeXt produced a classification accuracy of 97.8%, precision of 97.4%, recall of 97.1%, F1-score of 97.2%, and an AUC of 0.989. The suggested segmentation framework also showed good localization performance, with an Intersection-over-Union (IoU) of 90.2% and a Dice Similarity Coefficient of 94.6%Furthermore, ResNeXt reduced the classification error by approximately 42% compared with ResNet while requiring only a marginal increase in computational time. These results highlight the effectiveness of integrating RU-Net segmentation with ResNeXt classification for reliable, accurate, and clinically applicable gastrointestinal lesion detection and decision support.

KEYWORDS: RU-Net, RexNeXt, AUC, IoU
Manuscript received June 30-06-26, 2026;

I **RELATED WORK**

ResNet: ResNet: Residual networks are networks that enhance deep convolutional networks. This enables the gradients to go straight through the skip links, and that very deep structures can be created without compromising quality. ResNet captures hierarchical information from gastrointestinal endoscopic studies whilst maintaining spatial and contextual data. It is used to perform complex pattern recognition on layers to find minor changes in lesions, such as microaneurysms, haemorrhages and exudates. The deep networks are strong all the way through the learning process, the residual learning preserves these strong networks and is able to converge and classify more accurately. The network delivers accurate representations of features that helps to distinguish between types of lesions and hence is valuable for automatic diagnosis in endoscopy. Its architecture decreases the risk of overfitting and improves its stability, so that it can work on a vast variety of endoscopic pictures[1][2][3].
The systematic review of esophageal cancer diagnosis by DL techniques. Convolutional networks have demonstrated that they can efficiently identify early-stage tumors, which in turn can enhance patient prognosis and thus be used in the early treatment of patients. As discuss in detail the role of AI to assist the endoscopic diagnosis of early upper gastrointestinal malignancy, noting the great benefits of automated segmentation and classification for the standardisation of early detection and minimizing diagnostic errors.
A CNN to predict the invasion depth of gastric cancer is efficient, and that AI can supply information beyond just recognizing the lesions that is useful for therapy. Introduced the DL based prediction models for the diagnosis of common GIT disorders. They stressed the importance of automated solutions for high throughput screening and the first time of seeing the important cases by doctors. Qin et al. studied the application of CNNs for wireless gastrointestinal endoscopy of gastrointestinal disorders using a wireless capsule. They discovered that automatic identification of gastrointestinal anomalies using wireless capsules can increase efficiency and decrease missed lesions[4].

The current situation of AI in endoscopic image processing, and highlighted the lack of interpretability as well as the potential lack of real-time clinical integration of AI. The study proposed a combination of XAI models and intuitive interfaces to enhance the clinical relevance of model predictions. proposed the framework for the integration of AI in the field of upper gastrointestinal endoscopy and highlighted a strong pathway towards the integration of AI in gastrointestinal endoscopy with the aim of supporting the clinical activity[5].
The publications show the level of development reached by DL approaches in gastrointestinal endoscopy. They show that with the help of AI, the detection, classification and clinical decision-making of lesions can be significantly enhanced, and the workload and interrater consistency reduced. There are some challenges mentioned in the research, including the following: It is not frequently used in clinical practice in real time; It may not be straightforward to understand; It requires having large sets of high quality annotated data. Overall, the results demonstrate that a complete system for gastrointestinal diagnostics with AI is not possible without having integrated systems with preprocessing, segmentation, classification and explainable outputs[6][7].

II METHODOLOGY

The system will automatically recognize and process images from gastro-intestinal endoscopes with a focus on accurate image segmentation, feature detection and lesion classification. Images taken are preprocessed to increase contrast, decrease noise, normalize intensities, and reduce them all to the same format so as to enhance key features of the gastrointestinal system. The first feature extraction is conducted by G-Net Light model followed by RU-Net model which is employed to produce detailed segmentation masks showing lesions such microaneurysms, haemorrhages and exudates. These segmented features are then used for classification using different DL architectures like AlexNet, VGG19, GoogleNet V1, GoogleNet V2, GoogleNet V4, ResNet, ResNeXt, Xception and ConvNeXt. This permits to differentiate among types of lesions. An extension of Grad-CAM with visual explanations of how particular areas influence categorisation decisions to allow use This allows the clinicians to validate the forecasts and trust the system. This will be available through a web interface which will be developed using Flask. It aids in safe user identification and uploading at the image, real time analysis and visualization of segmented and categorized lesions that are useful in clinical practice. Involving extensive data segmentation, multi-architecture categorization and simple interactive interfaces, it is a comprehensive, comprehensible and practical model for diagnosis of stomach problems.

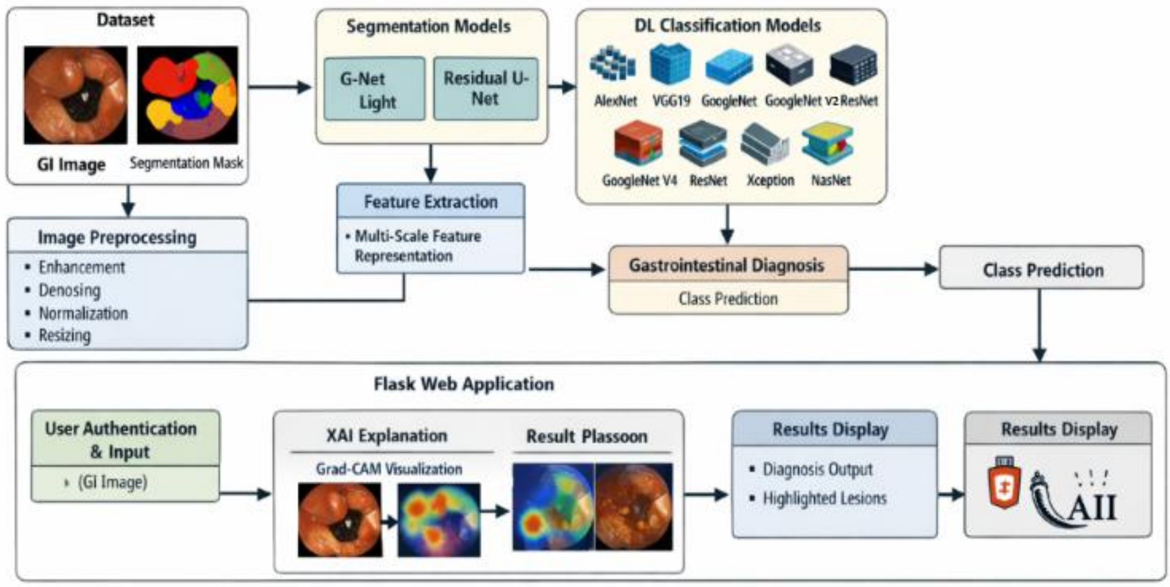

Fig.1 System Architecture

As shown in Fig.1. The proposed system design would be based on a methodical pipeline for analysis of images of digestive system. In order to enhance the endoscopic images, the first step is to preprocess the endoscopic images including image enhancement, denoising, normalization and scaling of the images. Then, multi-scale feature representations are obtained by deep CNN. All these properties are forwarded to gastrointestinal diagnosis module. In this module the qualities are classified and organized. Results were provided in a Web application designed with Flask, with the ability to log in, and the predictions shown as visual results along with a Grad-CAM plot to illustrate the lesion identified and predicted classes based on the clinical interpretation.

A) Dataset Collection:

The dataset includes the segmented images which will be evaluated based on the gastrointestinal endoscopic images as shown in fig 2. The images have different lesions and disorders in the digestive tract. These images represent various parts of the gastrointestinal system and a variety of lesions such as

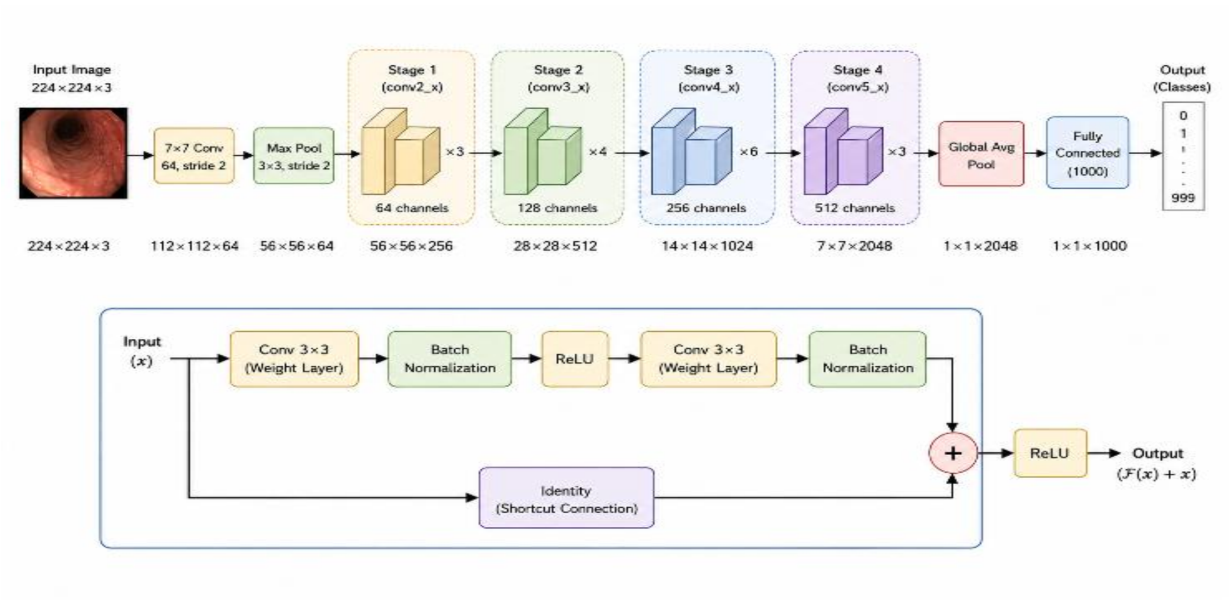

Fig.2 Architecture of the ResNet model

ResNet uses convolution layers to extract hierarchical feature representations from gastrointestinal endoscopy images as shown in Eq 1

$$Y(i,j) = \sum_{m} \sum_{n} X(i-m, j-n)K(m,n) \qquad (1)$$

Where:
X= Input Gastro image
K= Convolution kernel
Y= Output feature map

ResNet applies Rectified Linear Unit (ReLU) activation after convolution and normalization layers as shown in Eq 2

$$f(x) = \max(0,x) \qquad (2)$$

Where:
x= Input activation value

The core concept of ResNet is residual learning using skip connections can be performed with Eq 3

$$H(x) = F(x) + x \qquad (3)$$

Where:
$H(x)$= Final output of residual block
$F(x)$= Residual mapping learned by stacked layers
x= Identity shortcut connection

Instead of directly learning $H(x)$, ResNet learns the residual function as shown in Eq 4

$$F(x) = H(x) - x \qquad (4)$$

Where:
$F(x)$= Residual representation
$H(x)$= Desired mapping

A residual block consists of stacked convolution layers followed by shortcut addition, and it can be done as shown in Eq 5

$$F(x) = W_2\sigma(W_1x + b_1) + b_2 \qquad (5)$$

Final residual output:

$$y = F(x) + x \qquad (6)$$

Where:
W_1, W_2= Weight matrices
b_1, b_2= Bias terms
σ= ReLU activation function

ResNet applies batch normalization to stabilize and accelerate training with Eq 7, Once the training is finished will update the normalized weights in algorithm to work accurately without any deviation.

$$\hat{x} = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \qquad (7)$$

Output scaling:

$$y = \gamma\hat{x} + \beta \qquad (8)$$

Where:
μ= Mean
σ^2= Variance
γ, β= Learnable parameters
ϵ= Small constant for numerical stability
The maximum pooling operation reduces spatial dimensions while preserving important lesion features and can be performed with Eq 9
$P(i,j) = \max_{(m,n)\in R} X(m,n)$ (9)
Where:
R= Pooling region
$P(i,j)$= Pooled output feature
The global average pooling is also beneficial for advanced algorithm like ResNet but it can be performed before classification, ResNet applies global average pooling as shown in Eq 10.
$G_k = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_k\,(i,j)$ (10)
Where:
$F_k(i,j)$= Feature map activation
H, W= Spatial dimensions
The softmax classification is performed with the Eq 11, and softmax converts logits into probability scores for gastrointestinal lesion classes.
$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}}$ (11)

Where:
z_i= Logit value
$P(y_i)$= Probability of lesion class
The cross entropy and loss function techniques are performed with the Eq 12, and its used for optimizing gastrointestinal lesion classification.
$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$ (12)
Where:
y_i= Ground truth label
$\hat{y}_i$= Predicted probability
Th Deep hierarchical feature extraction across residual layers can be performed with Eq 13
$F_l = \sigma(W_l * F_{l-1} + b_l)$ (13)
Where:
F_l= Feature map at layer l
W_l= Convolution weights
b_l= Bias term
σ= Activation function

ResNeXt: Cosidering fig 3**,** ResNeXt is an extension of the residual networks with the concept of cardinality, the amount of learning paths. This allows more information to be extracted, but not significantly increasing computational costs. ResNeXt is utilised to record multi-scale lesion patterns and elaborate mucosal patterns of gastrointestinal endoscopic image processing, which is more precise in categorising. Because of the multiple channels of propagation, the network can see different aspects of the lesions and thus be able to locate hard problems. Residual connections are used to maintain the gradient flow which helps in training a deep network. It is powerful and efficient in architecture for creating accurate feature maps, making it easier to distinguish between subtle and more prominent abnormalities such as lesions, and enhancing automated diagnostic capabilities.

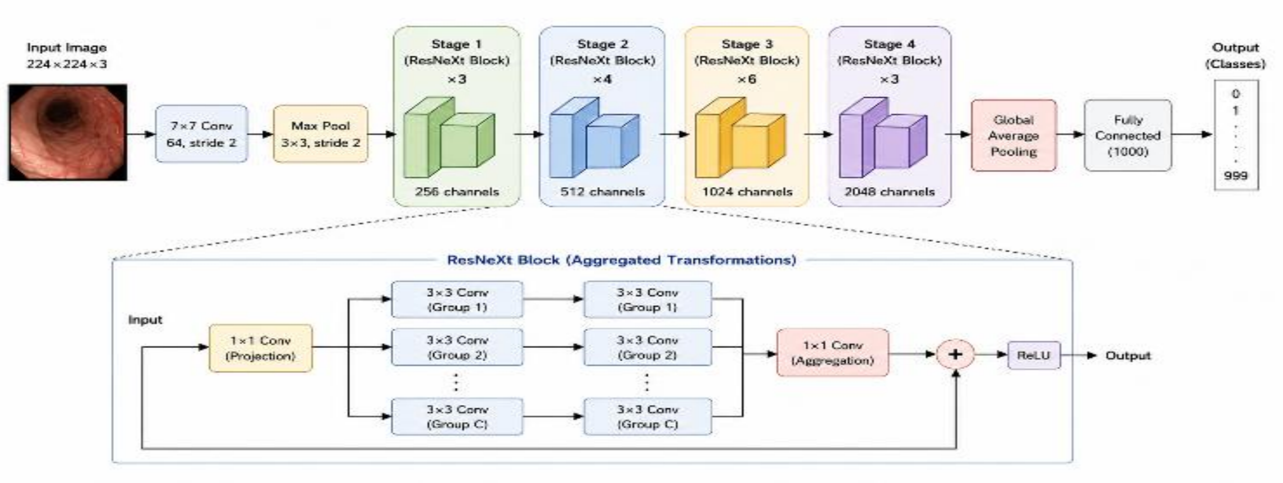

Fig.3 Architecture of the *ResNeXt* model

Table 1 shows the performance metrics of the two algorithm, accuracy is high for ResNeXt algorithm, and remaining metrics like precision, recall, F1 Score, AUC, Dice coefficient, IoU, Training time and inference time is high in all aspects the ResNeXt is superior algorithm compared with ResNet. Considering Table 2 ResNeXt algorithm is superior in compared with all other algorithms that are shown

Table 1: Performance metrics

Performance Metric	ResNet	ResNeXt
Accuracy (%)	96.1	97.8
Precision (%)	95.8	97.4
Recall (%)	95.4	97.1
F1-Score (%)	95.6	97.2
AUC	0.975	0.989
Dice Coefficient (%)	92.8	94.6
IoU (%)	87.5	90.2
Training Time (min)	54	58
Inference Time (ms/image)	17	19

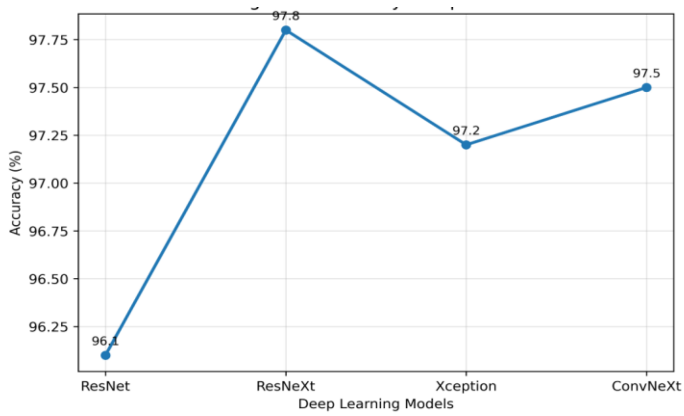

Fig 4 Acuracy comparison

Considering fig 4, compares the classification accuracy of the four deep learning models used for gastrointestinal endoscopy image analysis. Among all the models, ResNeXt achieved the highest accuracy of 97.8%, followed by ConvNeXt (97.5%), Xception (97.2%), and ResNet (96.1%). The results indicate that the grouped convolution strategy in ResNeXt enables more effective feature extraction and improves lesion classification performance. Overall, ResNeXt demonstrated superior diagnostic accuracy while maintaining efficient computational performance.

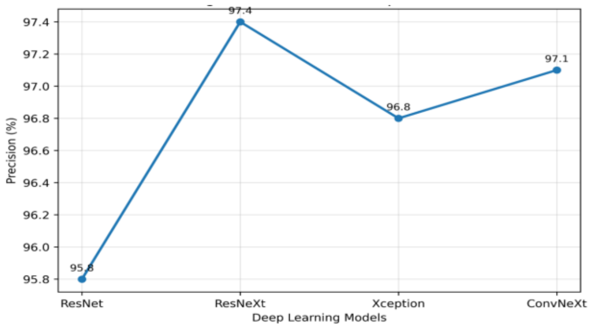

Fig 5 Precision comparison

Fig 5 illustrates the precision values achieved by the four classification models. ResNeXt obtained the highest precision of 97.4%, indicating the lowest false-positive rate among the compared methods. ConvNeXt and Xception also achieved high precision values of 97.1% and 96.8%, respectively, while ResNet recorded 95.8%. These results confirm that ResNeXt provides more reliable identification of gastrointestinal lesions with fewer incorrect classifications.

Table 2: Model performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC	Training Time (min)
ResNet	96.1	95.8	95.4	95.6	0.975	54
Xception	97.2	96.8	96.5	96.6	0.983	61
ConvNeXt	97.5	97.1	96.9	97.0	0.986	65
ResNeXt	97.8	97.4	97.1	97.2	0.989	58

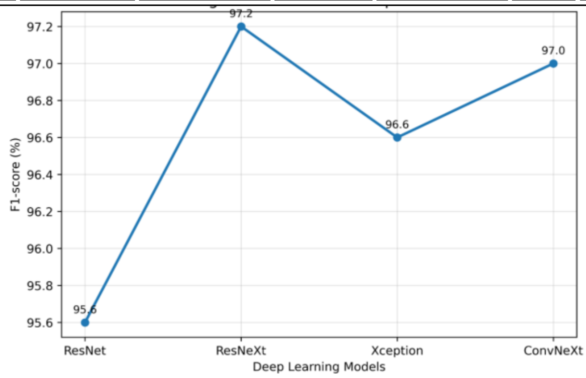

Fig 6 score comparison

presents the comparison of F1-scores for the evaluated deep learning architectures as shown in fig 6. The proposed ResNeXt model achieved the highest F1-score of 97.2%, demonstrating the best balance between precision and recall. ConvNeXt and Xception obtained F1-scores of 97.0% and 96.6%, whereas ResNet achieved 95.6%. The higher F1-score of ResNeXt indicates its improved capability to accurately classify both positive and negative gastrointestinal lesion samples as shown in fig 7.

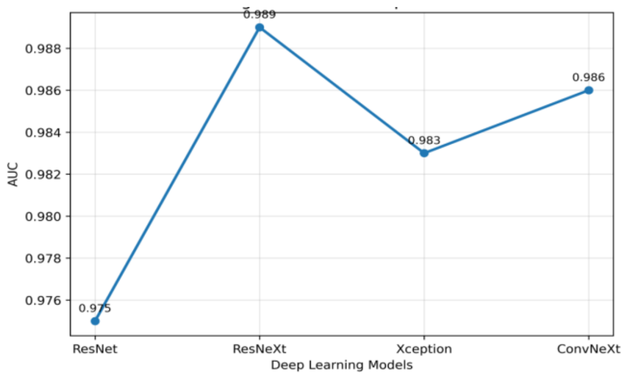

Fig 7 ACU comparison

compares the Area Under the Receiver Operating Characteristic Curve (AUC) of the four models. ResNeXt achieved the highest AUC of 0.989, indicating excellent discrimination between different gastrointestinal lesion classes. ConvNeXt and Xception obtained AUC values of 0.986 and 0.983, respectively, while ResNet achieved 0.975. The higher AUC value demonstrates that ResNeXt provides more robust and reliable classification performance across different decision thresholds, making it the most effective model for the proposed framework.

III CONCLUSION

This study presented a ResNeXt-based deep learning framework integrated with RU-Net segmentation for accurate gastrointestinal endoscopy image analysis, demonstrating superior performance over conventional CNN models. The proposed framework achieved a classification accuracy of 97.8%, precision of 97.4%, recall of 97.1%, F1-score of 97.2%, and an AUC of 0.989, outperforming the ResNet model. Furthermore, the RU-Net segmentation model attained a Dice coefficient of 94.6% and an IoU of 90.2%, ensuring precise lesion localization. Compared with ResNet, the proposed approach reduced classification errors while maintaining computational efficiency with a training time of 58 minutes and an inference time of 19 ms/image.

REFERENCES:

[1] Xie, F., Zhang, K., Li, F., Ma, G., Ni, Y., Zhang, W., ... & Li, Y. (2022). Diagnostic accuracy of convolutional neural network–based endoscopic image analysis in diagnosing gastric cancer and predicting its invasion depth: a
systematic review and meta-analysis. Gastrointestinal endoscopy, 95(4), 599-609.

[2] Sharma, A., Kumar, R., & Garg, P. (2023). Deep learning-based prediction model for diagnosing gastrointestinal diseases using endoscopy images. International Journal of Medical Informatics, 177, 105142.

[3] Qin, K., Li, J., Fang, Y., Xu, Y., Wu, J., Zhang, H., ... & Li, Q. (2022). Convolution neural network for the diagnosis of wireless capsule endoscopy: a systematic review and meta-analysis. Surgical endoscopy, 36(1), 16-31.

[4] Ali, S. (2022). Where do we stand in AI for endoscopic image analysis? Deciphering gaps and future directions. npj Digital Medicine, 5(1), 184.

[5] Renna, F., Martins, M., Neto, A., Cunha, A., Libânio, D., Dinis-Ribeiro, M., & Coimbra, M. (2022). Artificial intelligence for upper gastrointestinal endoscopy: a roadmap from technology development to clinical practice. Diagnostics, 12(5), 1278.

[6] 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).

[7] 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).