GMR
The Original

G-NET LIGHT AND RU-NET ASSISTED GOOGLENET V4 FRAMEWORK FOR EXPLAINABLE GASTROINTESTINAL ENDOSCOPY IMAGE ANALYSIS

Dr.K. Vijay Chandra[1], P.Snigdhasree[2], Dr.K.Sudha Rani[3*], Poornaiah Billa[4], Dr.Bhavani Jupalli[5], N. Syamala[6]

[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, pottapatirao239@gmail.com
[3]Associate Professor, Department of EIE, VNRVJIET, Hyderabad, Telangana, India, sudharani_k@vnrvjiet.in
[4]Professor, Department of ECE, Lakireddy bali reddy college of engineering and technology, India, poornaiah@gmail.com
[5]Associate Professor, EEE Department, VNRVJIET, Hyderabad, Telangana, India, bhavani_j@vnrvjiet.in
[6]Assistant Professor, Department of ECE, VNRVJIET, Hyderabad, Telangana, India, syamala_n@vnrvjiet.in
***Corresponding author:** sudharani_k@vnrvjiet.in

ABSTRACT- Gastrointestinal (GI) endoscopy is a very common test used for early detection of the diseases like polyps, ulcer, bleeding and inflammatory conditions. However, while it can be considered a useful tool for diagnosing, the inspection of many endoscopes is a time-consuming process and needs experienced physicians, which means there could be some differences in diagnosing. To solve this problem, we introduce a deep learning method of automatic segmentation and classification of GI endoscopy images.The endoscopy images were processed using the following techniques, such as Contrast Limited Adaptive Histogram Equalization (CLAHE), removal of the noise, normalizing the intensity, and adjusting the size of the image. Then, G-Net Light was used to detect informative features, and RU-Net was applied for lesion segmentation. The segmentation algorithm obtained a Dice coefficient of 72.6%, IoU of 60.8%, and pixel accuracy of 91.8% which is good enough to segment lesions.According to the experiment, GoogleNet V4 performed better than GoogleNet V2 concerning overall accuracy, precision, recall, specificity, F1-score, and AUC with values of 93.33%, 92.84%, 93.10%, 94.21%, 92.97% and 0.965 respectively. That is, GoogleNet V2 obtained 90.42% accuracy and AUC of 0.934.

KEYWORDS: GI, GoogleNet, IoU, RU-Net,

I INTRODUCTION

GoogleNet V2**:** GoogleNet V2 is an upgraded version of the inception architecture to improve the performance of the classifier and minimize computational costs. Factorized convolutions, batch normalization, and auxiliary classifiers are used in the model to achieve a higher efficiency of the feature extraction process and stability in the model training process. The larger convolution filters along with smaller factorized filters are used in the model to decrease the computational cost without a decrease in the ability to extract the useful information from the images. As a result of the statistical analysis with the help of McNemar's test, the difference in the performance of GoogleNet V2 and GoogleNet V4 was found to be statistically significant with the p value < 0.05, which means that GoogleNet V4 performs better than GoogleNet V2. [4] Furthermore, the Grad-CAM algorithm was applied to the model to identify the regions in the image that affect the prediction of the model.
The statistical results of McNemar's test have shown that GoogleNet V4 has demonstrated a significant increase in performance compared to GoogleNet V2 ($p<0.05$). Thus, it can be stated that the performance increase is not accidental [5]. In order to determine the regions of lesions that influenced the classification of images, the visualization technique named Gradient-weighted Class Activation Mapping (Grad-CAM) was applied [6]. In addition, the framework that was designed in this research was implemented as a web application using Flask technology together with user authentication for image analysis in real time.There are certain diseases related to the gastro-intestinal tract, which might be characterized by presence of polyps, ulcers, bleeding and inflammatory lesions. In case of lack of early detection, such problems might cause severe consequences. Endoscopy is one of the most widely used methods in order to examine GI tract. However, analysis of endoscopic images is rather complicated because the lesions may vary in terms of their shape, size, color, and texture. It might be due to a range of reasons, among which there are inconsistent clinical expertise and poor quality of images resulting in incorrect or delayed diagnosis.[7].With recent developments and advancement in deep learning, particularly Convolutional Neural Networks (CNNs), it is becoming possible to analyze medical images accurately and consistently in the context of segmentation and classification. In this study, a computer-aided diagnosis system is developed using image preprocessing, RU-Net for lesion segmentation, and GoogleNet V2 and GoogleNet V4 for disease classification. The initial preprocessing step improves the quality of the endoscopic images, while RU-Net identifies the affected regions before processing the classification. Grad-CAM analysis is also used to show

which parts of the image effected the model's prediction, making the results easier to understand and interpret for the analysis. [8].
After analyzing the experimental results the results show that GoogleNet V4 performed better than GoogleNet V2 in classifying gastrointestinal diseases. It was able to learn important features from the endoscopic images, leading to higher accuracy and more reliable predictions. Overall, the proposed framework provides an accurate and easy-to-understand computer-aided diagnosis system that can support clinicians in the early detection and diagnosis of gastrointestinal diseases at the early stages. [9].

II **METHODOLOGY**

GoogleNet V2 is more effective in capturing multi-scale lesion features and obtaining correct information on small and complicated abnormalities in the processing of gastrointestinal endoscopic images. Batch normalisation provides more robust training, convergence and more reliable predictions. Auxiliary classifiers carry the gradients in the network, thus reducing the gradient vanishing issue in deep layers. The network has the ability to sense on many spatial and contextual factors, which enables it to differentiate lesions with an excellent precision. It is efficient at categorisation and also has the ability to research in real time[10].

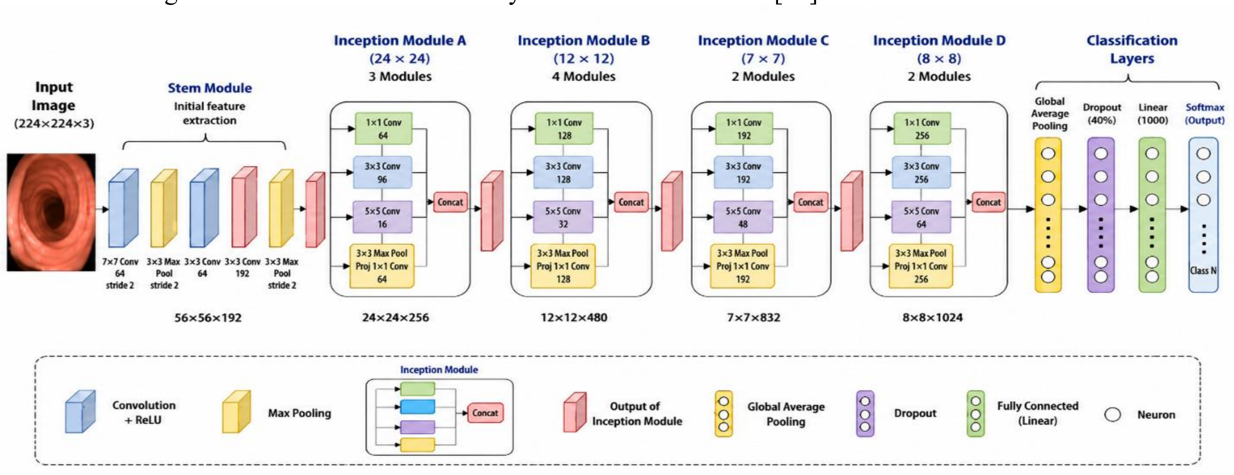

Fig.1 Architecture of the GoogleNet V2 model

2.1 CONVOLUTION OPERATION

GoogleNet V2 uses convolution layers to extract hierarchical gastrointestinal lesion features.

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

Where:
X= Input image
K= Convolution kernel
Y= Output feature map
This equation (1) represents the convolution process used in GoogleNet V2 to extract useful features from gastrointestinal endoscopy images. The convolution kernel scans the input image and captures important patterns such as lesion edges, textures, and shapes, which are essential for identifying abnormalities.
GoogleNet V2 applies ReLU function after convolution layer process.

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

Equation (2) applies the ReLU activation function, which replaces negative values with zero while keeping positive values unchanged. This allows the model to learn complex patterns in gastrointestinal lesions more effectively and improves the speed and stability of the training process.
GoogleNet V2 performs multi-scale feature extraction using parallel convolutions.

$$F = Concat(F_{1\times1}, F_{3\times3}, F_{factorized}, F_{pool}) \qquad (3)$$

Where:
$F_{1times1}$= 1×1 convolution features
$F_{3times3}$= 3×3 convolution features
$F_{factorized}$= Factorized convolution features
F_{pool}= Pooling branch output
Inception Module is denoted by formula (3) which handles simultaneously several convolutions where each filter extracts the features of a different size. It aids in recognizing the fine-grained pattern and large-scale structural pattern; hence, it enables the model to detect gastrointestinal lesions more accurately.
GoogleNet V2 replaces larger convolutions using factorization.
Instead of 5×5 $Convolution$
It uses: $3 \times 3 + 3 \times 3$

The computational reduction as follows:
Standard convolution cost: $Cost_{5\times5} = 25MN$
Factorized convolution cost: $Cost_{3\times3} = 18MN$
Where:
M= Input channels
N= Output channels
GoogleNet V2 uses smaller consecutive convolution filters instead of a single large filter. This approach reduces computational complexity while still capturing important image features, allowing the model to process gastrointestinal images more efficiently without affecting classification performance.
GoogleNet V2 uses Batch Normalization for stable training.

$$\hat{x} = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \quad (4)$$

Output scaling:

$$y = \gamma\hat{x} + \beta$$

Where:
mu= Mean
$sigma^2$= Variance
$gamma, beta$= Learnable parameters
Equation 4 applies batch normalization to keep the feature values consistent during training. This helps the model learn more efficiently, improves training stability, and leads to better classification performance.
The pooling equation can be performed with following equation

$$P(i,j) = \max_{(m,n)\in R} X(m,n) \quad (5)$$

Where:
R= Pooling region
Equation 5 reduces the size of the feature maps by keeping only the most important information from each region. This makes the model faster while preserving the key features needed to accurately identify gastrointestinal lesions.
GoogleNet V2 uses auxiliary classifiers for gradient propagation.

$$L = L_0 + 0.3L_1 + 0.3L_2 \quad (6)$$

Where:
L_0= Main classifier loss
L_1, L_2= Auxiliary losses
Equation 6 uses auxiliary classifiers to support the learning process during training. These additional classifiers help the model learn more effectively, improve information flow through the network, and contribute to better classification performance.
The softmax classification is performed with the following eq

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \quad (7)$$

Where:
$P(y_i)$= Probability of lesion class
z_i= Logit value
Equation 7 converts the model's output into probabilities for each disease class. The category with the highest probability is selected as the final prediction for the gastrointestinal lesion.
2.2 CROSS ENTROPY LOSS

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (8)$$

Where:
y_i= Ground truth label
$\hat{y}_i$= Predicted probability
Equation (8) calculates the error between the model's predicted results and the actual disease labels. During training, the model works to reduce this error, helping it make more accurate classifications.

$$F1 = \sigma(W_l * F_{l-1} + b_l) \quad (9)$$

Where:
F_l= Feature map
W_l= Kernel weights
$sigma$= ReLU activation
Equation 9 describes how the model generates feature maps at each convolution layer. It combines the input features with the learned weights to extract meaningful information, helping the model identify and classify gastrointestinal lesions more accurately.
2.4 GOOGLENET V4

The inception modules and residual connections form inception V4 (GoogleNet V4) to construct deeper networks without compromising the performance. It has multi-scale feature extraction, improved gradient flow and feature reuse ability. GoogleNet V4 can identify gastrointestinal tract images, complex lesions images and minor anomalies at both local and global level of images. The residual connections prevent the loss of gradients, enabling training of very deep networks which can be applied to high accuracy classification. The design enables multiple features and contexts at various scales to identify the various types of lesions. This leads to the correct classification according to segmentation, which enables computers to always detect irregularities in the digestive system and to understand and apply the knowledge gained[11].

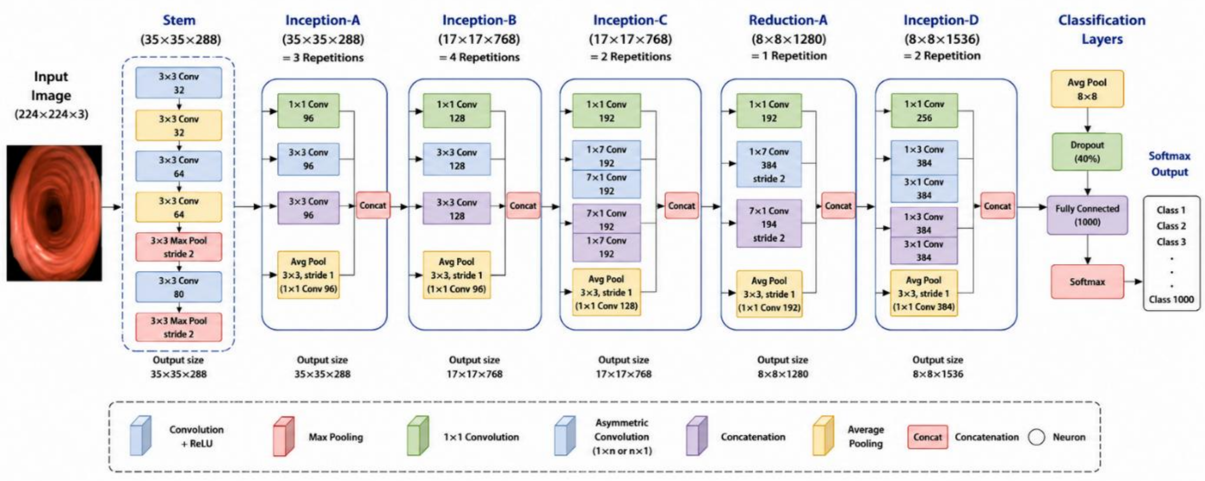

Fig.2 Architecture of the GoogleNet V4 model

Convolution operation

GoogleNet V4 uses convolutional operations for extracting hierarchical gastrointestinal lesion features.

$Y(i,j)=\sum_m\sum_n [X(i-m,j-n)K(m,n)]$ -(10)

Where:

X= Input image

K= Convolution kernel

Y= Output feature map

GoogleNet V4 convolutes the features of gastroscopic endoscopy images through the convolution process as per equation 10. The sliding convolutional kernel extracts some relevant information from the input images such as the edges, textures, shapes of the lesion, etc., enabling the classifier to distinguish a normal tissue from the abnormal one..

GoogleNet V4 applies ReLU after every convolution operation.

$f(x)=\max(0,x)$ (11)

Equation 11 shows the usage of ReLU activation function where all negative values are mapped to zero without changing the positive ones. It helps the model learn important features of the lesions and improves the training rate and stability of the model.

GoogleNet V4 consists of many convolution paths, allowing for multi-scale feature extraction.

$F=Concat(F_{(1\times1)},F_{(3\times3)},F_{(5\times5)},F_{pool})$ (12)

Where:

F_1times1= Features from 1×1 convolution

F_3times3= Features from 3×3 convolution

$F_{5times5}$= Features from 5×5 convolution

F_{pool}= Pooling features.

The Equation 12 shows the Inception module that utilizes different convolution filters for detecting features of different sizes. By using such approach, the model can detect gastrointestinal diseases of small and large size with higher accuracy.

GoogleNet V4 factorizes convolutions to reduce computation.

Instead of: $n \times n$ it uses $1 \times n + n \times 1$

Computational Cost Reduction follows

Standard convolution:

$$Cost = n^2MN$$

Factorized convolution:

$$Cost = 2nMN$$

Where:

M= Input channels
N= Output channels
Convolution filters of large size in GoogleNet V4 are converted into small one-dimensional filters to decrease computation complexity. Not only the processing speed of the model is increased, but also the information about the images is preserved. It allows to classify gastrointestinal diseases without compromising the efficiency of the model.

III RESULTS AND DISCUSSION

An endoscopy image dataset with 8 classes of diseases namely; Dyed-Lifted Polyps, Dyed Resection Margins, Esophageal Lesions, Normal Cecum, Normal Pylorus, Normal Z-Line, Polyps, and Ulcerative Colitis was used for testing of our framework. Improvement of the initial quality of the images through CLAHE, normalization and denoising was performed. Next, the G-Net Light was used to get the features from the images while the RU-Net was used to segment the lesions. Finally, the segmented images were classified using the GoogleNet V2 and GoogleNet V4. The performance was evaluated in terms of accuracy, precision, recall, specificity, f1-score, dice coefficient, IoU and AUC.

Table 1:Ru-Net Results

Metric	**RU-Net**
Dice Coefficient	72.60%
IoU	60.80%
Pixel Accuracy	91.80%
Mean Boundary Error	4.2 pixels

The results in Table (1) have been made possible through proper segmentation of the lesions in RU-Net without any loss of vital structural information. Residual learning and skip connections make it possible for effective detection of the lesions, especially for the irregular polyps and ulcers, hence better segmentation results.
GoogleNet V4 gave the best classification results for most of the categories due to its effectiveness in extraction of vital features from the lesions. Additionally, the RU-Net was able to segment the lesion areas accurately, which in turn improves on the localization of the lesions, making it possible to classify the diseases effectively. The following results show the effectiveness of the segmentation and classification combination for automatic gastrointestinal disease diagnosis.

Table 2: Classification Results of GoogleNet V2 and GoogleNet v4

Metric	**GoogleNet V2**	**GoogleNet V4**
Accuracy	90.42%	**93.33%**
Precision	89.76%	**92.84%**
Recall (Sensitivity)	90.11%	**93.10%**
Specificity	91.54%	**94.21%**
F1-score	89.93%	**92.97%**
AUC	0.934	**0.965**

According to Table (2), GoogleNet V4 performed better than GoogleNet V2 in all evaluation metrics. It was able to detect features of the lesions effectively giving higher classification accuracy and proper separation between normal and abnormal tissues of gastrointestinal system. Additionally, the model had high sensitivity and specificity.

Table 3: Disease-Wise Performance Analysis

Disease Category	**GoogleNet V2**	**GoogleNet V4**
Dyed Lifted Polyps	91.1%	**94.5%**
Dyed Resection Margins	90.4%	**93.2%**
Esophageal Lesions	88.6%	**92.3%**
Normal Cecum	95.0%	**97.1%**

Disease Category	GoogleNet V2	GoogleNet V4
Normal Pylorus	93.7%	**95.8%**
Normal Z-line	92.8%	**95.0%**
Polyps	89.5%	**93.4%**
Ulcerative Colitis	87.9%	**91.6%**

According to the data presented in Table (3), GoogleNet V4 demonstrated the best classification capability among all gastrointestinal diseases categories. The model managed to capture the complicated patterns of tissue and achieved maximum success in case of differentiation between ulcerative colitis and esophageal lesions. It may be suggested that GoogleNet V4 had the highest capacity to classify gastrointestinal diseases. In order to analyze the difference between two models, McNemar's statistical test was applied. It turned out that the p-value was less than 0.05. It means that there was a significant difference between GoogleNet V4 and GoogleNet V2. Grad-CAM visualization technique was applied to find out which regions of the image were important for the model prediction. It was found out that GoogleNet V4 could identify the lesion areas more accurately in comparison to GoogleNet V2 that sometimes highlighted healthy tissue regions around the lesion areas. This will help to understand the decision making process of the model and trust its diagnoses. Experimental findings show that the developed framework of image enhancement, G-Net Light, RU-Net, and GoogleNet V4 can be used for automatic analysis of diseases. GoogleNet V4 revealed an increase in the classification accuracy of about 2.9% compared to GoogleNet V2 as well as greater sensitivity, lesion localization and AUC. The use of Grad-CAM and Flask-based Web application increases the effectiveness of the system because it provides users with the understandable prediction on the friendly interface.

IV CONCLUSION

A framework of deep learning for the automatic analysis of the gastrointestinal endoscopy images was proposed in this research. It includes image preprocessing, feature extraction by G-Net Light, segmentation and classification of the disease by using GoogleNet V2 and GoogleNet V4 through RU-Net. RU-Net has shown the Dice coefficient of 72.6%, IoU of 60.8%, and pixel accuracy of 91.8%. These metrics mean that the network can effectively segment the lesions. GoogleNet V4 has the greatest accuracy of 93.33%, precision of 92.84%, and recall of 93.10%, the greatest specificity of 94.21%, the greatest F1-score of 92.97% and the greatest AUC of 0.965 among classification models. According to the McNemar's test, GoogleNet V4 performed much better than GoogleNet V2 (*p* < 0.05). The Grad-CAM technique helped improve the model's interpretability by identifying regions of importance and the Web application based on Flask provided the real-time analysis facility. In essence, the framework presented here can be considered as a realistic and robust CAD system to help doctors diagnose GI diseases.

V 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).