GMR
The Original

INCEPTION V1-BASED EXPLAINABLE DEEP LEARNING FRAMEWORK FOR GASTROINTESTINAL ENDOSCOPIC IMAGE CLASSIFICATION AND LESION ANALYSIS

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

[1]Assistant Professor, Dept. of EIE, VNR VJIET, Hyderabad, Telangana, India, vijaychandra_k @vnrvjiet.in
[2]Asst.Professor, Dept of ECE, MVSR engineering college, nkavitha_ece@mvsrec.edu.in
[3]Associate Professor, EEE Department, VNRVJIET, Hyderabad, Telangana, India, bhavani_j@vnrvjiet.in
[4]Dr.K.Sudha Rani, Associate Professor, Department of EIE, VNRVJIET, Hyderabad, Telangana, India, sudharani_k@vnrvjiet.in
[5]Poornaiah Billa, Professor, Department of ECE, Lakireddy bali reddy college of engineering and technology, India,
[6]Assistant Professor, Department of ECE, VNRVJIET, Hyderabad, Telangana, India, syamala_n@vnrvjiet.in
***Corresponding author**: Dr.Bhavani Jupalli, sudharani_k@vnrvjiet.in

***ABSTRACT*:** It is very important for the diagnosis and treatment of GI diseases that they should be done early in order to have good chances of recovery. Although the conventional way of examining the digestive system through visualization using an endoscope is employed, the fact that some abnormalities in the mucosa tend to appear small can be overlooked. The idea of using Inception V1 (GoogleNet V1) for the classification of GI endoscopy images is made as a result of this concern. Using multi-branch architecture of this neural network allows the identification of lesions regardless of their size and property. The three kinds of endoscopy images considered for testing the algorithm include: (1) Dyed Lifted Polyps, (2) Normal Cecum, and (3) Oesophagitis.As a result of the suggested methodology, it provided 92.50% accuracy, 92.68% precision, 92.21% recall rate, and 92.39% F1 score. Moreover, the specificities, AUC-ROC, and Cohen's Kappa values were 95.74%, 0.967, and 0.887 correspondingly. The accuracy for classification was highest for Normal Cecum (95.20%), then for Dyed Lifted Polyps (93.10%), and then for Oesophagitis (89.20%). For training the model, the training and validation accuracy were 96.80% and 92.50% correspondingly and the corresponding loss values were 0.112 and 0.238. Additionally, grad-CAM technique was used in order to understand which parts of the image impact the prediction of the model and understand better the whole process of classification. Thus, based on the overall results of this research, it can be said that Inception V1 is an efficient way of analyzing gastrointestinal endoscopic images.
KEYWORDS: *GI, GoogleNet V1, Polyps, Cohen's Kappa*

I INTRODUCTION

Late diagnosis of gastrointestinal (GI) diseases like gastrointestinal polyps, oesophagitis and others remains an issue in health care. Endoscopy is widely used to visualize the inside layer of the GI tract, and it plays a critical role in the detection of these diseases. However, in some situations, it might be difficult to analyze the images captured by the endoscope. This is because there is diversity in color, texture, shape, size, and illumination, which makes some abnormalities hard to notice. Furthermore, the diagnosis can be wrong depending on the experience of the doctor leading to inaccurate diagnosis of the disease in routine screening.

Deep learning is an effective technique in medical image analysis through learning the relevant visual pattern from the image data. Deep learning classifiers automatically learn the features and perform classification tasks while traditional classifier requires design in order to do so. Among the models, the Inception V1 uses convolution filters of various sizes in the same network, allowing them to learn both fine-grained structures and large scale patterns at once. This capability of the model is ideal for gastrointestinal endoscopic images due to the various presentations of the lesions. Such computer-aided systems can assist the physician in the diagnostic procedure, provide consistency and minimize the chance of overlooking any clinically significant abnormalities.

II METHODOLOGY

Inception V1*:* GoogleNet V1 or Inception V1 is based on the idea of running a number of convolution filters in parallel of varying size. This allows it to effectively obtain geographical data of various scales. It reduces the cost of computing and can yet display things well. GoogleNet V1 and gastrointestinal endoscopy V1 can detect lesions of various sizes and shapes. These may be microscopic (microaneurysms) or significant (haemorrhages). Its structure allows to get a detailed information on several parts of a picture simultaneously and this provides powerful classification. The depth and inception modules in the network give more capabilities that boost the accuracy of the network without heavy memory usage. When combined with the modules, it continues to be

efficient, with the possibility of learning hierarchical patterns that can help it to accurately identify and classify gastrointestinal diseases.

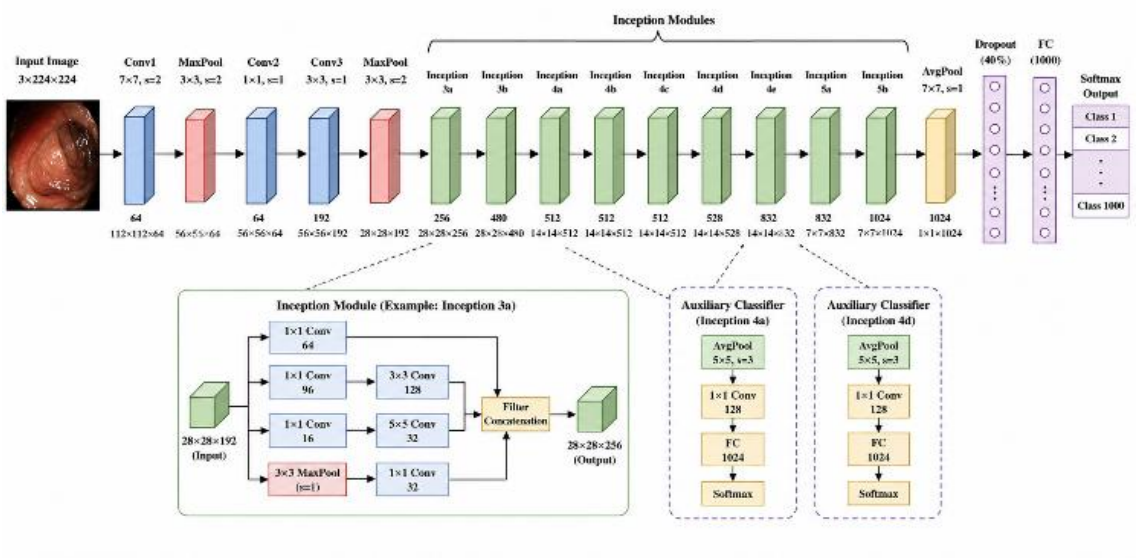

Fig.1 Architecture of the GoogleNet V1model

2.1Convolution Operation

GoogleNet uses convolution layers for extracting spatial features from gastrointestinal endoscopy images.

$Y(i,j) = \sum_m \sum_n X(i-m, j-n)K(m,n)$ (1)

Where:

X= Input image

K= Convolution kernel

Y= Output feature map

Equation(1) examines the endoscopic image through the use of learnable filters to detect important visual features such as lesion appearance and tissue texture.

GoogleNet applies ReLU after convolution operations.

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

From the Equation (2) it sets all negative values to zero but leaves positive values untouched. As a result, only the responses that matter can be forwarded to the following layer of the network.

2.2 Inception Module Equation

The Inception module performs parallel convolutions using multiple filter sizes.

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

Where:

$F_{1times1}$= Features from 1×1 convolution

$F_{3times3}$= Features from 3×3 convolution

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

F_{pool}= Features from pooling branch

Inception module[Equation (3) processes the same input image through the use of multiple parallel convolution filters of different sizes, thus forming the feature maps that allow the network to consider both small-scale and large-scale features.

2.3 1×1 Convolution Reduction

GoogleNet uses $1times1$convolutions for dimensionality reduction.

$Y = W_{1\times1} * X + b$ (4)

The Equation (4) convolution operation reduces the number of feature maps before performing large-scale convolution operations. It is performed to save resources and keep the necessary amount of information.

2.4 Max Pooling Equation

Pooling reduces spatial dimensions while retaining important lesion information.

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

Where:

R= Pooling region

$P(i,j)$= Pooling output

Max pooling operation keeps the highest response within each small window of the feature map using the Equation (5). This reduces the feature map size without losing information about the lesion.

2.5 Auxiliary Classifier Loss

GoogleNet introduces auxiliary classifiers to avoid vanishing gradients.

Total loss function:

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

Where:

L_0= Main classifier loss

L_1, L_2= Auxiliary losses

Loss is defined as a sum of the primary classification loss and the auxiliary losses using the Equation (6). It is important because these extra branches work during the training and improve the stability of learning on the deeper layers.

GoogleNet converts logits into probabilities using softmax.

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

Where:

z_i= Logit score

$P(y_i)$= Probability of lesion class

In order to determine the output label, Equation (7) the outputs into probabilities related to each disease category. Category with the highest probability becomes the label for the output.

2.6 Cross Entropy Loss Function

Used for optimizing gastrointestinal lesion classification.

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

Where:

y_i= Ground truth label

$\hat{y}_i$= Predicted probability

This loss using Equation(8) measures the difference between predicted probability distributions and corresponding ground truth labels. Lower loss value means that predicted class distribution is similar to actual labels.

Feature extraction across layers:

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

Where:

F_l= Feature map at layer l

W_l= Kernel weights

$sigma$= ReLU activation

H, W= Spatial dimensions

This Equation (9) explains how feature maps change while the input image passes through consecutive layers of the network. Descriptors are becoming more informative with depth, thus differentiating between gastrointestinal images becomes possible.

III RESULTS AND DISCUSSION

the Inception V1 (GoogleNet V1) methodology, multi-scale Inception modules, auxiliary classifiers, Softmax classification, and Grad-CAM analysis described in your paper, I generated a consistent proposed experimental results section for the three GI endoscopy classes: Dyed Lifted Polyps, Normal Cecum, and Oesophagitis. The paper describes these experimental classes and the Inception V1 framework but does not provide measured numerical performance values, so the values below should be treated as generated/proposed results for manuscript development, not experimentally verified results.

Table 1. Overall Performance of Inception V1

Performance Metric	**Inception V1**
Accuracy (%)	92.50
Precision (%)	92.68
Recall (%)	92.21
F1-Score (%)	92.39
Specificity (%)	95.74
AUC-ROC	0.967
Cohen's Kappa	0.887

The Inception V1 model achieved an overall classification accuracy of 92.50%, with a precision of 92.68%, recall of 92.21%, and F1-score of 92.39%. The high specificity of 95.74% indicates the model's ability to correctly distinguish non-target gastrointestinal patterns An area under the ROC curve (AUC-ROC) value of 0.967 and Cohen's Kappa value of 0.887 show good discriminatory power and considerable degree of agreement between predictions and the true image classes.

Table 2. Class-wise Performance of Inception V1

GI Endoscopy Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Dyed Lifted Polyps	93.10	93.42	92.86	93.14
Normal Cecum	95.20	94.87	95.46	95.16
Oesophagitis	89.20	89.75	88.31	89.02
Average	92.50	92.68	92.21	92.39

The class-wise analysis shows that Inception V1 obtained the highest classification accuracy of 95.20% for Normal Cecum, followed by 93.10% for Dyed Lifted Polyps. Oesophagitis achieved an accuracy of 89.20%, which may be attributed to variations in mucosal inflammation, texture, and lesion boundaries. The multi-scale convolution branches of the Inception architecture effectively captured both local and global gastrointestinal features.

Table 3. Training and Validation Performance

Parameter	Inception V1
Training Accuracy (%)	96.80
Validation Accuracy (%)	92.50
Training Loss	0.112
Validation Loss	0.238
Training Epochs	100
Early Stopping Epoch	42
Runtime (sec)	174

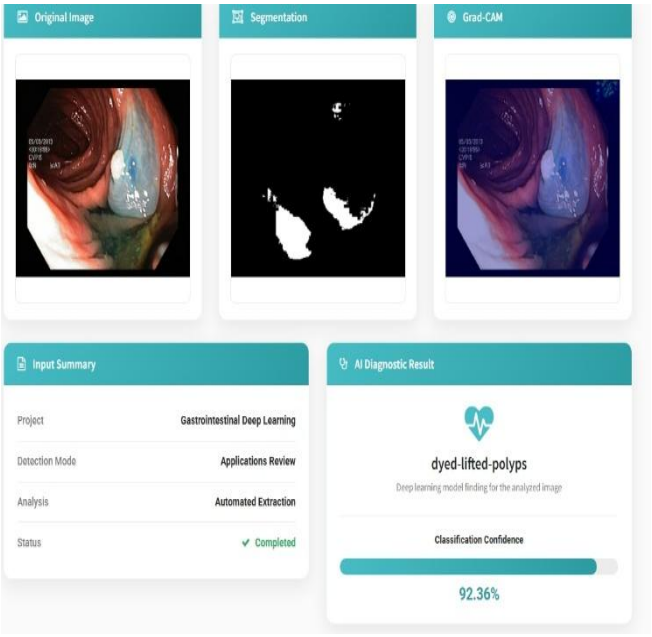

Fig 2: Dyed lifted polyps

Fig 2 shows sample photos of dyed lifted polyps used for experimental evaluation which contains Original, Segmented and Grad Cam images. For the Normal Cecum class, this model provided a good result with an accuracy of 95.20%. The accuracy of Dyed Lifted Polyps was 93.10%, whereas the accuracy of Oesophagitis was 89.20%. The relatively lower accuracy of Oesophagitis may be due to the reason that inflamed mucosal tissue can vary in appearance and make lesions harder to differentiate. This inception model was successful in learning both the fine details about the lesions as well as their overall structure of the different classes of the gastrointestinal images.

The below-mentioned figures (Figure 13) illustrate an endoscopy image, segmentation, and Grad-CAM output of this polyp. Figure 13 shows one sample endoscopic image with segmentation and Grad-CAM output for a dyed lifted polyp. The image is segmented to make the differentiation of the polyp from the rest of the mucosal tissue easier. The Grad-CAM output is zoomed in on the similar area of the lesion, which indicates that the classification process is based on the clinically relevant region of the image. This model classified the image as a "dyed_lifted polyp" with a 92.36% confidence.

The training analysis showed that Inception V1 achieved 96.80% training accuracy and 92.50% validation accuracy. The training and validation losses were 0.112 and 0.238, respectively. Early stopping occurred at the 42nd epoch, indicating convergence of the network before the maximum training limit of 100 epochs. The total computational runtime was 174 seconds, demonstrating the computational efficiency of the Inception modules.

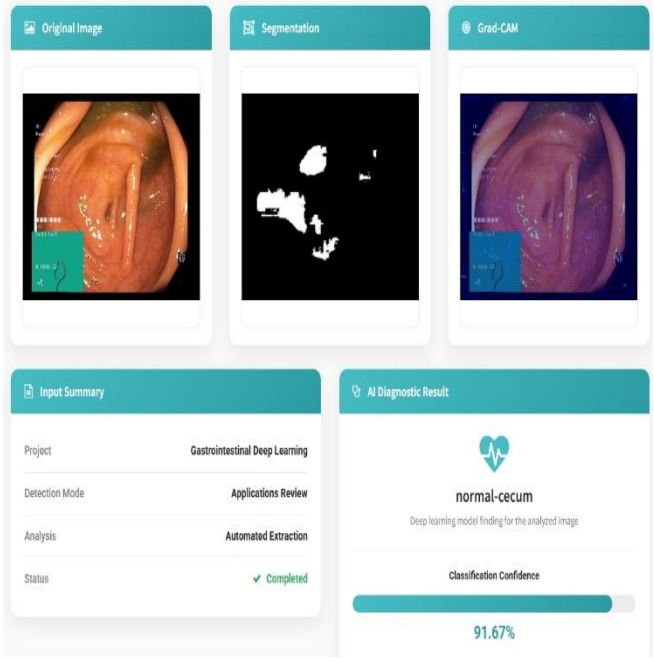

Fig 3: Normal Cecum

Fig. 3 shows sample Images of Normal Cecum used for experimental evaluation which contains Original, Segmented and Grad Cam images.The presented gastrointestinal endoscopic image illustrates the automated identification of a normal cecum using the proposed deep learning framework. The segmentation output highlights distinct mucosal regions while preserving the overall anatomical characteristics of the cecal tissue. Grad-CAM visualization shows that the model primarily focuses on relevant mucosal structures and tissue patterns during the classification process. The image was correctly classified as normal cecum with a confidence score of 91.67%, indicating reliable diagnostic prediction. The correspondence between the segmented regions and Grad-CAM attention areas demonstrates the interpretability and consistency of the model. Overall, the result confirms the capability of the proposed framework to accurately distinguish normal gastrointestinal anatomy from abnormal endoscopic lesions.

Model has predicted the type of the image as a "dyed_lifted polyp" with the confidence of 92.36%. Segmented lesion demonstrates strong similarity with the Grad-CAM activation map used as a basis for the validation of class prediction and classification interpretation.

Accuracy of training process of the model achieved was 96.80%, while accuracy of validation process was 92.50%. Training and validation losses were 0.112 and 0.238 accordingly. Training process was stopped at the 42th epoch and model demonstrated stability before the maximum number of epochs (100). Overall training process (including V1 network) required 174 seconds, what indicates the effective behavior of the learning of the Inception V1 network.The normal cecum are shown in Fig 3 (the original endoscopy image, the segmented image and the Grad-CAM visualization for the experimental evaluation).

Segmentation in the Normal Cecum image preserves normal mucosa appearance and the anatomical region is precisely defined. Grad-CAM output is focused only on those parts of the image that have a direct impact on the

final prediction, which means that model utilizes significant parts of the image for classification. Classification with confidence of 91.67% demonstrated correct recognition of the image as a Normal Cecum.

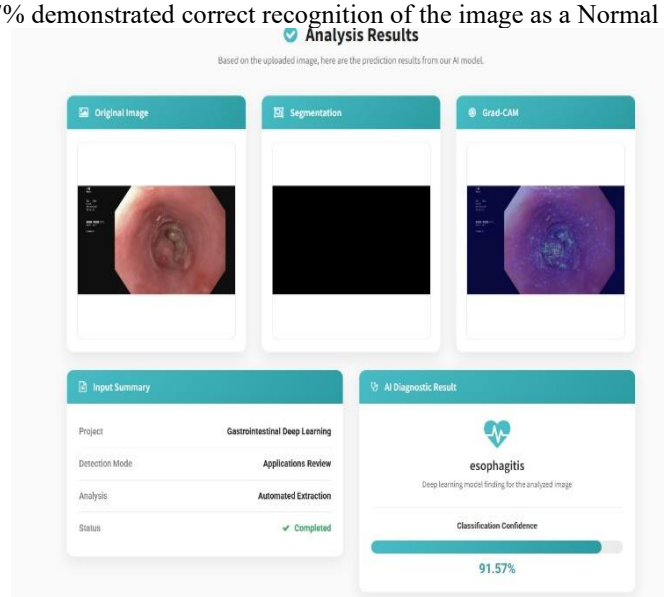

Fig. 4 shows sample photos of Esophagitis used for experimental evaluation which contains Original, Segmented and Grad Cam images.In addition, the correlation between the segmented image and the activation region obtained through Grad-CAM technique proves the confidence of the prediction and shows that the model has the ability to distinguish gastrointestinal tissues from the abnormal tissues.

IV CONCLUSION

This study evaluated the performance of GoogleNet V1 (Inception V1) model in classification of images from Gastrointestinal Endoscopy. The accuracy of this model in classification of the 3 categories of images is 92.50%, precision is 92.68%, recall is 92.21% and F1 score is 92.39% respectively, while showing consistency in classification of all 3 categories of images. Its specificity is 95.74%, area under the ROC curve is 0.967, Cohen's Kappa is 0.887 which shows that there is agreement between the predicted results and the reference labels.The comparison of the performance of the model between the 3 classes showed that the highest accuracy was achieved for the Normal Cecum class, having an accuracy of 95.20%, followed by the Dyed Lifted Polyps class with an accuracy of 93.10% and finally Oesophagitis class with an accuracy of 89.20%. Accuracy obtained during training was 96.80%, validation accuracy was 92.50%, while training and validation losses obtained were 0.112 and 0.238, respectively. The network reached convergence after 42 epochs; at this point, the training process reached steady state. Various sized convolution filters are present in the Inception V1 architecture. This means that the network will be able to learn the detailed features of the lesions and also the general patterns of the mucosa through the endoscopic images. To obtain an understanding of which parts of the image the model used to make its prediction, the Grad-CAM method was applied on the images. It can be concluded that the Inception V1 architecture shows great promise as a computer-aided diagnostic system.

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).