Attention-Driven Multi-Class Abnormality Detection in Video Capsule Endoscopy (VCE) Using Enhanced InceptionResNetV2

KAPILESH ANANDH 1 [a], SUHANI JAIN A 2 [b]

[a] Department of Computer Science, Vellore Institute of Technology, Chennai, India
[b] Department of Computer Science, Vellore Institute of Technology, Chennai, India

Email: kapilesh.da2022@vitstudent.ac.in, suhanijain.a2022@vitstudent.ac.in

Abstract

The early detection of gastrointestinal abnormalities is critical for effective patient management, particularly in conditions such as bleeding, polyps, and ulcers. This study proposes an innovative approach to multi-class abnormality detection in video capsule endoscopy by utilizing an enhanced InceptionResNetV2 model integrated with attention mechanisms. The model leverages the power of deep learning to analyze frames captured by video capsule endoscopes, focusing on ten distinct classes of abnormalities: Angioectasia, Bleeding, Erosion, Erythema, Foreign Body, Lymphangiectasia, Normal, Polyp, Ulcer, and Worms. The training dataset is enhanced using data augmentation techniques, which strengthens the model's resistance to overfitting.The proposed system is trained on a dataset provided as part of a research challenge and evaluated using various performance metrics, including confusion matrices and AUC-ROC curves. Results indicate that the attention mechanism significantly enhances the model's ability to classify abnormalities accurately, achieving a validation accuracy of 92.2%. This research advances the field of medical imaging by offering a scalable and effective solution for automated gastrointestinal abnormality detection, ultimately aiding clinicians in enhancing diagnostic precision and treatment outcomes.

1 Introduction

Artificial Intelligence (AI) has significantly impacted various fields, particularly healthcare, where deep learning methodologies have revolutionized medical image analysis. Among the innovative techniques employed, attention mechanisms stand out for their ability to enhance a model's focus on relevant features within complex datasets, leading to improved interpretability and accuracy in classification tasks. In the context of gastrointestinal (GI) diagnostics, the early detection of abnormalities is vital for effective patient management. Video capsule endoscopy is a non-invasive imaging modality that provides critical visual insights into the GI tract. However, the vast amount of data generated poses challenges for clinicians in accurately interpreting these images. To address

this, automated solutions that leverage deep learning can facilitate efficient analysis and classification of various abnormalities.

This study introduces an enhanced InceptionResNetV2 model integrated with attention mechanisms for multi-class abnormality detection in video capsule endoscopy. By focusing on key features and employing data augmentation techniques, the model is designed to tackle class imbalance and enhance its robustness against overfitting. The performance of the proposed system is assessed through multiple evaluation metrics, including balanced accuracy, mean AUC, AUC-ROC, specificity, and F1 score.

The results reveal that the integration of attention mechanisms notably improves classification accuracy, achieving a validation accuracy of 92%. This research advances the field of automated gastrointestinal abnormality detection, offering a scalable and effective tool that aids clinicians in enhancing diagnostic precision and treatment outcomes.

2 Methods

The methodology of our proposed model employs a hybrid approach that integrates an enhanced InceptionResNetV2 architecture with Attention mechanisms to optimize the classification of gastrointestinal abnormalities from video capsule endoscopy images, as illustrated in Figure 1.

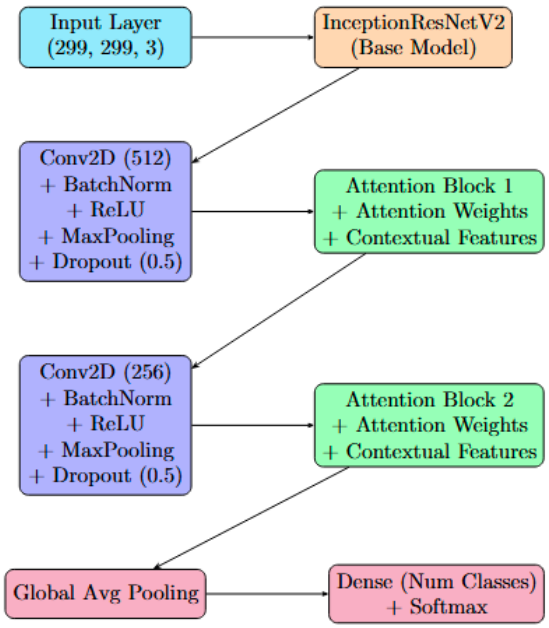

Figure 1: Attention-Based InceptionResNetV2 Architecture.

At the core of our model is the InceptionResNetV2 architecture, renowned for its ability to extract intricate features through residual connections and inception blocks. This

design captures complex patterns within the data, enhancing the model's performance in identifying subtle anomalies. The incorporation of attention blocks allows the model to focus on critical regions of the images, improving sensitivity to abnormalities that may be overlooked in standard convolutional approaches. Notably, the attention mechanism addresses class imbalance; for example, the minority class, Worms, consists of only approximately 158 images, while the majority class, Normal, contains around 28,000 images. This emphasis on significant features ensures that minority classes receive adequate representation during training, thus enhancing overall classification accuracy.

To combat the risk of overfitting, dropout layers are strategically implemented throughout the model, randomly deactivating a subset of neurons during training to promote the learning of robust feature representations. Additionally, batch normalization is employed to stabilize and accelerate the training process. The model is trained on an augmented dataset utilizing various techniques such as rotation, width and height shifts, and zooming, collectively enhancing its generalization capabilities. The use of ImageDataGenerator for data augmentation generates diverse variations of the input images, simulating a broader range of scenarios encountered in real-world applications, as depicted in the workflow shown in Figure 2.

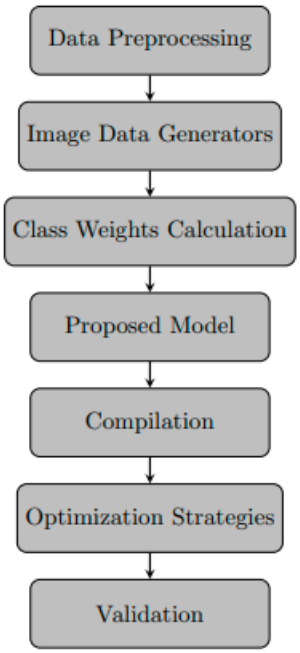

Figure 2: Block diagram of the developed pipeline.

This comprehensive methodology, integrating attention mechanisms, dropout regularization, and data augmentation, is meticulously designed to improve the model's performance in accurately identifying gastrointestinal conditions.

Figure 2 outlines the pipeline for classifying gastrointestinal abnormalities in video capsule endoscopy images. The pipeline begins with data augmentation (rotation, shifting, zooming) to enhance model generalization, followed by feeding these images into an InceptionResNetV2-based model. Key components include attention mechanisms to focus on critical regions and regularization techniques like dropout and batch normalization to ensure robust learning. The pipeline concludes with evaluation metrics, such as AUC-ROC curves and confusion matrices, to assess classification accuracy across abnormality classes.

3 Results

The results of our proposed Attention-Enhanced InceptionResNetV2 CNN model validate its effectiveness in classifying gastrointestinal abnormalities from video capsule endoscopy images. Evaluated across multiple performance metrics, the model demonstrated reliability in distinguishing between classes, particularly for minority groups. The incorporation of attention mechanisms facilitated focus on critical image regions, while dropout regularization and data augmentation enhanced the model's robustness and generalization.

3.1 Achieved results on the validation dataset

On the validation dataset, our Attention-Enhanced InceptionResNetV2 CNN model achieved key goal metrics, with a Balanced Accuracy of 84.74% and a Mean AUC of 0.99. The model also demonstrated robust performance across a comprehensive suite of metrics, including an AUC-ROC of 0.99, Mean Specificity of 0.99, Mean Average Precision of 0.88, Mean Sensitivity of 0.84, and Mean F1 Score of 0.82. Notably, the validation accuracy reached 92.2%, underscoring the model's capability to effectively handle class imbalances and accurately identify anomalies. Figure 2 presents the classification report, offering insights into performance across different classes. Additionally, Figures 3 and 4 illustrate the confusion matrix and AUC-ROC curve, respectively, visually demonstrating the model's performance in distinguishing between class

Table 1: Validation results and comparison to the baseline methods reported by the organizing team of Capsule Vision 2024 challenge.

Method	**Validation Accuracy**
Custom CNN	0.668
ResNet50	0.760
SVM	0.818
VGG16	0.716
Proposed Model	0.922

Classification Report:

	precision	recall	f1-score	support
Angioectasia	0.71	0.86	0.77	497
Bleeding	0.83	0.87	0.85	359
Erosion	0.73	0.72	0.73	1155
Erythema	0.54	0.69	0.61	297
Foreign Body	0.80	0.91	0.86	340
Lymphangiectasia	0.73	0.91	0.81	343
Normal	0.99	0.96	0.97	12287
Polyp	0.64	0.80	0.71	500
Ulcer	0.98	0.90	0.94	286
Worms	0.91	0.85	0.88	68
accuracy			0.92	16132
macro avg	0.79	0.85	0.81	16132
weighted avg	0.93	0.92	0.93	16132

Figure 3: Classification Report.

Figure 4: Confusion Matrix.

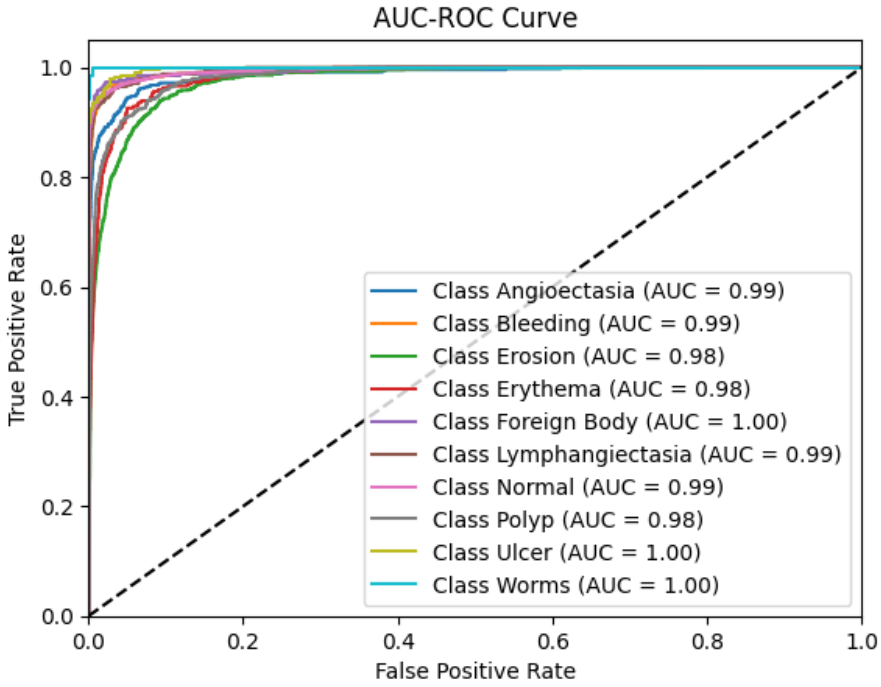

Figure 5: AUC-ROC Curve.

4 Discussion

The results from our Attention-Enhanced InceptionResNetV2 model highlight a marked improvement in automated gastrointestinal abnormality detection via video capsule endoscopy, achieving a validation accuracy of 92.2%. This outperforms baseline models, including Custom CNN, ResNet50, SVM, and VGG16, as shown in Table 1, highlighting the effectiveness of integrating attention mechanisms within the InceptionResNetV2 architecture.

Attention mechanisms significantly enhance the model's capacity to focus on critical image features, an essential aspect for medical imaging applications where subtle variances are crucial for diagnosis. By directing focus to relevant regions, the model reduces the influence of noise, improving classification accuracy for underrepresented classes, such as Worms. Data augmentation further bolsters robustness, addressing overfitting issues by introducing diverse image variations. Regularization techniques such as dropout and batch normalization aid in stable training and support the model's generalizability.

Evaluation metrics, including high AUC-ROC, specificity, and sensitivity scores, affirm the model's reliability in classifying abnormalities while minimizing false positives—an essential trait in clinical diagnostics.

5 Conclusion

This study introduces an Attention-Enhanced InceptionResNetV2 model for gastrointestinal abnormality detection, achieving a validation accuracy of 92.2% and outperforming existing baselines. Attention mechanisms contribute to improved interpretability and diagnostic accuracy by focusing on key features within endoscopy images. The model's robust performance highlights its potential in clinical applications, supporting efficient, automated diagnostics and enhancing patient management.

Future work will involve validating the model in clinical environments and expanding it to cover additional gastrointestinal conditions. Integrating this model with other diagnostic tools may offer a more comprehensive solution for patient care.

6 Acknowledgments

As participants in the Capsule Vision 2024 Challenge, we fully comply with the competition's rules as outlined in [1]. Our AI model development is based exclusively on the datasets provided in the official release in [2]. As discussed in [3], the dual-attention mechanism enhances feature extraction and classification effectiveness. Recent advancements in deep learning have enhanced abnormality detection in wireless capsule endoscopy (WCE) [4], which is akin to video capsule endoscopy (VCE). These innovations improve accuracy and efficiency in identifying gastrointestinal abnormalities. While focused on WCE, the techniques discussed can also benefit VCE image classification, aiding clinicians in diagnostic decisions. Recent advancements in class-based attention mechanisms have improved multi-label categorization in medical imaging [5]. This approach learns distinct attention masks for each class, enabling targeted localization of different pathologies within the same image. Gonçalves et al. [6] provide a comprehensive review of attention

mechanisms, emphasizing their potential in enhancing medical image analysis. The proposed architecture demonstrated enhanced classification performance on publicly available datasets. The integration of attention mechanisms in fine-tuned transfer learning models has significantly improved endoscopic image analysis for multi-class gastrointestinal disease classification [7].

References

[1] Palak Handa, Amirreza Mahbod, Florian Schwarzhans, Ramona Woitek, Nidhi Goel, Deepti Chhabra, Shreshtha Jha, Manas Dhir, Deepak Gunjan, Jagadeesh Kakarla, et al. Capsule vision 2024 challenge: Multi-class abnormality classification for video capsule endoscopy. *arXiv preprint arXiv:2408.04940*, 2024.

[2] Palak Handa, Amirreza Mahbod, Florian Schwarzhans, Ramona Woitek, Nidhi Goel, Deepti Chhabra, Shreshtha Jha, Manas Dhir, Deepak Gunjan, Jagadeesh Kakarla, and Balasubramanian Raman. Training and Validation Dataset of Capsule Vision 2024 Challenge. *Fishare*, 7 2024. doi: 10.6084/m9.figshare.26403469.v1. URL `https://figshare.com/articles/dataset/Training_and_Validation_Dataset_of_Capsule_Vision_2024_Challenge/26403469`.

[3] Haoyu Gao. Image classification based on dual-attention mechanism and multi-convolution layer. In *2022 4th International Conference on Communications, Information System and Computer Engineering (CISCE)*, pages 160–163, 2022. doi: 10.1109/CISCE55963.2022.9851102.

[4] Md. Jahin Alam, Rifat Bin Rashid, Shaikh Anowarul Fattah, and Mohammad Saquib. Rat-capsnet: A deep learning network utilizing attention and regional information for abnormality detection in wireless capsule endoscopy. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–8, 2022. doi: 10.1109/JTEHM.2022.3198819.

[5] David Sriker, Hayit Greenspan, and Jacob Goldberger. Class-based attention mechanism for chest radiograph multi-label categorization. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2022. doi: 10.1109/ISBI52829.2022.9761667.

[6] Tiago Gonçalves, Isabel Rio-Torto, Luís F. Teixeira, and Jaime S. Cardoso. A survey on attention mechanisms for medical applications: are we moving toward better algorithms? *IEEE Access*, 10:98909–98935, 2022. doi: 10.1109/ACCESS.2022.3206449.

[7] Mohamed A. Elmagzoub, Swapandeep Kaur, Sheifali Gupta, Adel Rajab, Khairan D. Rajab, Mana Saleh Al Reshan, Hani Alshahrani, and Asadullah Shaikh. Improving endoscopic image analysis: Attention mechanism integration in grid search fine-tuned transfer learning model for multi-class gastrointestinal disease classification. *IEEE Access*, 12:80345–80358, 2024. doi: 10.1109/ACCESS.2024.3408224.