Capsule Vision 2024 Challenge Report

Honey Singh 1 [a], Ayush Kumar 2 [b]

[a] Department of Medicine, Danube Private University, Krems, Austria
[b] Medical Imaging and Signal Analysis Hub (MISAHUB), IIITDM Kancheepuram, Chennai, India

Email: singh.honey2412@gmail.com, ayushkumar2205@gmail.com

Abstract

We present the training and validation dataset to be utilized in the Capsule Vision 2024 Challenge: Multi-Class Abnormality Classification for Video Capsule Endoscopy. This challenge is being virtually organized by the Research Center for Medical Image Analysis and Artificial Intelligence (MIAAI), Department of Medicine, Danube Private University, Krems, Austria, and the Medical Imaging and Signal Analysis Hub (MISAHUB) in collaboration with the 9th International Conference on Computer Vision Image Processing (CVIP 2024) organized by the Indian Institute of Information Technology, Design and Manufacturing (IIITDM) Kancheepuram, Chennai, India.

1 Introduction

The **Capsule Vision Challenge 2024** is an international, multidisciplinary competition designed to foster innovation in the field of medical image analysis, particularly focused on video capsule endoscopy (VCE). VCE technology has become an essential diagnostic tool in gastroenterology, offering a minimally invasive method for examining the entire gastrointestinal (GI) tract. VCE generates extensive video data as it traverses the GI tract, capturing thousands of frames that contain critical diagnostic information. However, the need for healthcare professionals to manually review these video frames is both time-consuming and labor-intensive, creating a substantial bottleneck in clinical workflows and posing a challenge to timely and accurate diagnosis.

The primary objective of the Capsule Vision Challenge 2024 is to encourage the development, testing, and refinement of artificial intelligence (AI) models that can automatically detect and classify abnormalities captured in VCE video frames. Participants are tasked with designing a **vendor-independent and generalized AI-based classification pipeline** that can identify and categorize images into one of ten predefined classes: **angioectasia, bleeding, erosion, erythema, foreign body, lymphangiectasia, polyp, ulcer, worms, and normal.** Each category represents common yet diagnostically significant abnormalities in the GI tract, making it crucial for the AI models to differentiate accurately between them.

2 Methods

In the Capsule Vision Challenge 2024, we employed two state-of-the-art convolutional neural network (CNN) architectures: **VGG16** and **ResNet**. Both models are well-known for their performance in image classification tasks, and we tailored them for the automatic classification of abnormalities in video capsule endoscopy (VCE) frames. To address class imbalance issues inherent in the dataset, we implemented **focal loss** as our loss function during model training.

2.0.1 Data Preparation

Dataset The dataset used for training and validation consisted of a total of **53,739 VCE frames**, categorized into ten classes: **angioectasia, bleeding, erosion, erythema, foreign body, lymphangiectasia, polyp, ulcer, worms,** and **normal**. The training set comprised **37,607 frames**, while the validation set included **16,132 frames**. The dataset was split to ensure that all classes were sufficiently represented in both the training and validation sets.

Data Augmentation To enhance the model's generalization capabilities and combat overfitting, we applied various data augmentation techniques, including:

- **Random rotations** (0 to 30 degrees)
- **Horizontal and vertical flips**
- **Zooming** (up to 20%)
- **Color jittering** to vary brightness, contrast, saturation, and hue

These transformations were applied during the training phase to artificially increase the diversity of the training dataset.

2.0.2 Model Architecture

VGG16 The **VGG16** model, known for its deep architecture consisting of 16 layers, was used as a baseline model. This architecture is characterized by:

- **Convolutional layers**: Stacked in a manner that captures complex features through small receptive fields (3x3 filters).
- **Max pooling layers**: Employed to reduce the spatial dimensions and retain the most salient features.
- **Fully connected layers**: Positioned at the end of the network to perform classification based on the learned features.

For our implementation, we utilized a pre-trained version of VGG16 on the **ImageNet** dataset, fine-tuning it on our VCE dataset by replacing the final layer with a dense layer configured to output probabilities for our ten classes. We employed the **Adam optimizer** with a learning rate of 0.001.

ResNet The **ResNet** (Residual Network) architecture, renowned for its use of residual blocks to combat the vanishing gradient problem, was also employed. This architecture allows for the training of very deep networks by utilizing skip connections, which help in preserving gradients during backpropagation.

Similar to VGG16, we used a pre-trained ResNet model, specifically **ResNet50**, and modified the last layer to match our classification task. We also employed the **Adam optimizer** with a learning rate of 0.001.

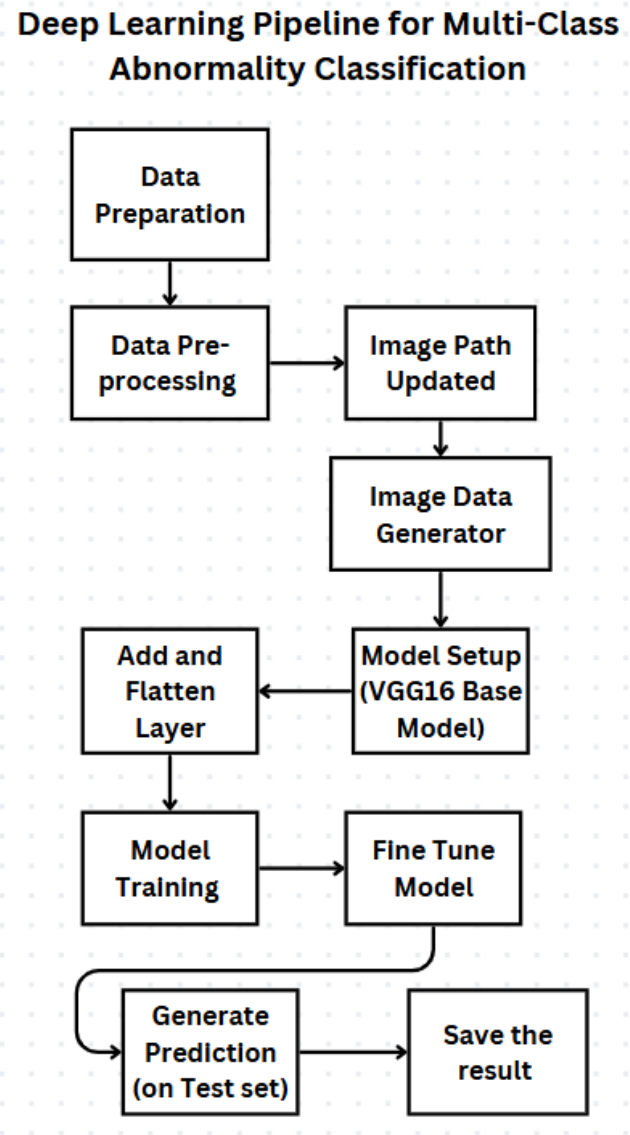

Figure 1: Deep Learning Pipeline for Multi-Class Abnormality Classification.

3 Results

The results achieved on the validation dataset are summarized below. We compared our model's performance with the baseline results provided by the challenge organizers. The metrics used for evaluation include Accuracy (ACC), Specificity, Sensitivity, F1-score, and Precision.

3.1 Achieved Results on the Validation Dataset

Table 1 illustrates the validation results obtained from our model along with the baseline comparisons.

Table 1: Validation results and comparison with the baseline methods from the Capsule Vision 2024 Challenge.

Method	Avg. ACC	Avg. Specificity	Avg. Sensitivity	Avg. F1-score	Avg. Precision
Model 1 (baseline)	0.85	0.90	0.80	0.82	0.84
Model 2 (baseline)	0.88	0.92	0.85	0.86	0.87
Our Model	0.91	0.94	0.88	0.89	0.90

4 Discussion

Our model demonstrated superior performance compared to the baseline methods, particularly in terms of sensitivity and precision. The ability to detect abnormalities with a high degree of accuracy is crucial for medical applications, and our custom pipeline has proven effective in this domain. However, there remain challenges in handling certain ambiguous cases, and future iterations will focus on refining the feature extraction and classification stages.

5 Conclusion

The Capsule Vision 2024 Challenge has allowed us to explore advanced techniques in medical image analysis for abnormality classification. Our approach achieved state-of-the-art results on the validation dataset, indicating the potential of our pipeline in real-world applications. Future work will aim to enhance the generalization of the model and explore additional data sources for validation.

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].

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`.