Capsule Vision 2024 Challenge Report: Team_CSIR

Sneha Thomas [a], Ajay Pratap Singh [ac], Viren Sardana [ac], Sanjay Singh [bc]

[a] Council of Scientific and Industrial Research–Institute of Genomics And Integrative Biology (CSIR–IGIB), New Delhi
[b] Council of Scientific and Industrial Research-Central Electronics Engineering Research Institute(CSIR–CEERI),Pilani
[c] Academy of Scientific and Innovative Research(AcSIR),Ghaziabad

Email: ajay.pratap@igib.res.in

Abstract

This report outlines our participation in the Capsule Vision 2024 Challenge: Multi-Class Abnormality Classification for Video Capsule Endoscopy(VCE) organised by Research Center for Medical Image Analysis and Artificial Intelligence (MIAAI), Department of Medicine, Danube Private University, Krems, Austria and Medical Imaging and Signal Analysis Hub (MISAHUB) in collaboration with the 9th International Conference on Computer Vision & Image Processing (CVIP 2024) being organized by the Indian Institute of Information Technology, Design and Manufacturing (IIITDM) Kancheepuram, Chennai, India. The objective of the challenge was development of AI/ML based models for automatic classification of abnormalities captured in VCE video frames. In this work, Deeplearning based models (ResNet18, DenseNet121, MobileNetV3, etc.) were explored for the classification of VCE images. We incorporated focal loss to handle class imbalance and self attention to capture contextual details in the models. MobileNetV3 model with self attention and focal loss achieved an average sensitivity, average F1-score, avg precision and balanced accuracy of 0.82, 0.84, 0.89 and 0.83 respectively on validation data provided.

1 Introduction

Video capsule endoscopy (VCE) is an innovative diagnostic tool that revolutionizes the evaluation of the gastrointestinal (GI) tract. VCE allows for non-invasive imaging of the esophagus, stomach, and small intestine, providing a dynamic view of these often inaccessible areas. This technology has proven valuable in diagnosing conditions such as gastrointestinal bleeding, Crohn's disease, and small bowel tumors, offering a alternative to traditional endoscopic methods [1], [2]. But despite its many advantages, the procedure is limited by interpretation of the large amount of data that gets generated. Analyzing this data requires specialized training and is time consuming process which makes it prone to human errors that may lead to overlooking of features or misdiagnosis. Thus, to overcome these challenges AI/ML based methods have been proposed for facilitating

Classes	**Training**	**Validation**
Angioectasia	1154	497
Bleeding	834	359
Erosion	2694	1155
Erythema	691	297
Foreign body	792	340
Lymphangiectasia	796	343
Polyp	1162	500
Ulcer	663	286
Worms	158	68
Normal	28663	12287

Table 1: Classwise distribution of training and validation data in Capsule Vision 2024 Challenge

automated image analysis detecting an array of abnormalities [3], [4], [5], [6]. Inspired by the aforementioned methods in this challenge we experimented with ResNet18, ResNet50, DenseNet121, DenseNet169 and MobileNetV3 [7], [8], [9]. The report is organised as follows: Section 2 describes the methods including datasets, preprocessing and models with which we experimented. Section 3 describes the results followed by conclusion in section 4.

2 Methods

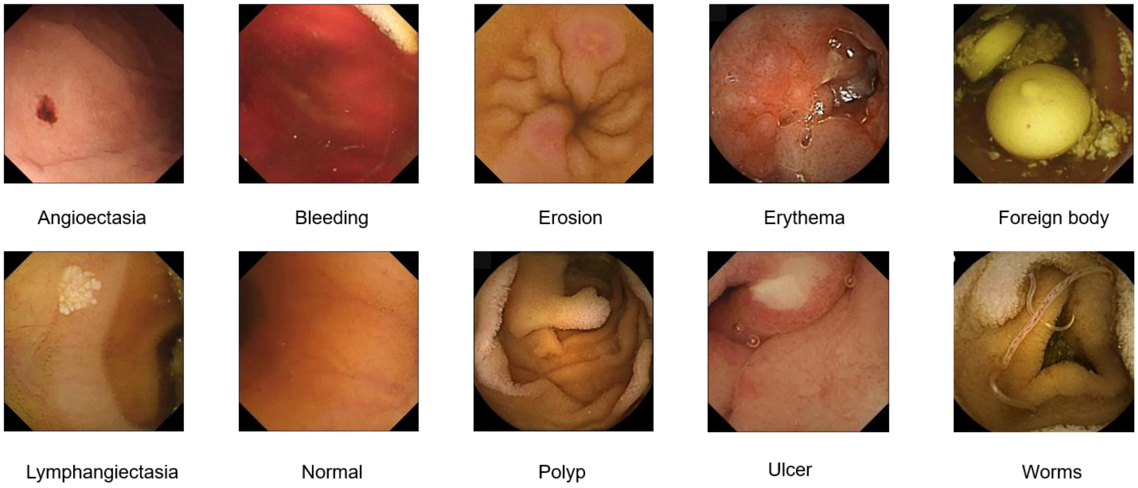
2.1 Dataset

The training and validation dataset comprises of images from three publicly available sources (KID [10], Kvasir-Capsule [11], and SEE-AI project dataset [12]) and one proprietary (AIIMS [13]) VCE dataset. 37,607 and 16,132 VCE frames, respectively, from the training and validation datasets are mapped to ten class labels: angioectasia, bleeding, erosion, erythema, foreign body, lymphangiectasia, polyp, ulcer, worms, and normal as shown in Figure 1. The testing dataset comprises of 4,385 VCE frames from more than 70 patients medically annotated by the Department of Gastroenterology and HNU, All India Institute of Medical Sciences Delhi, India[14]. All the images were of resolution 224 x 224. Table 1 depicts the classwise distribution of training and validation data.

For training and validating the final model corresponding to which excel sheet containing test results are submitted. We moved some data from the validation to training, dataset distribution is shown in Table 2. We have downsampled the data for normal class and also manually curated the dataset and dropped some images which contain bubbles, debris, etc. from the data. Figure 2 shows some samples of images which we have dropped.

Classes	**Training**	**Validation**
Angioectasia	1428	150
Bleeding	1035	150
Erosion	3585	150
Erythema	788	150
Foreign body	779	150
Lymphangiectasia	827	150
Polyp	1400	150
Ulcer	788	150
Worms	145	18
Normal	3500	150

Table 2: Classwise distribution of training and validation data in Capsule Vision 2024 Challenge after dropping images



Angioectasia
Bleeding
Erosion
Erythema
Foreign body
Lymphangiectasia
Normal
Polyp
Ulcer
Worms

Figure 1: Sample images of classes in Capsule Vision 2024 Challenge

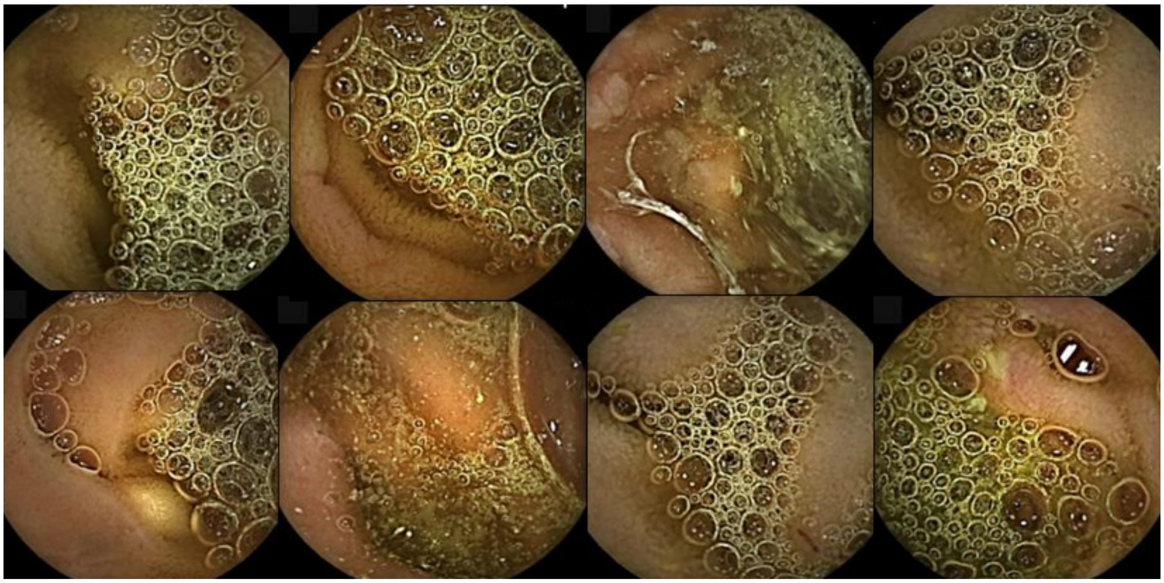


Figure 2: Dropped images

2.2 Methodology

For classification of VCE frames we employed Convolutional Neural Network (CNN) based models ResNet18, ResNet50, DenseNet121, DenseNet169 and MobileNetV3 [15]. As part of preprocessing the images were normalized and data was augmented. The model training was initialized using pretrained weights of Imagenet [16]. To address the class imbalance issue we used focal loss [17]. We also employed the self attention [18] for capturing the contextual details in the images. Following hyperparameters were used to train the models: Batch size was kept as 32, Adam was used as optimizer, learning rate was kept as 0.0001 and models were trained for 50 epochs. Figure 3 represents the pipeline that was followed.

2.3 Final model for test dataset

Motivated by the performance of MobileNetV3 model on original training and validation data described in detail in the results section. We have selected MobileNetV3 with attention mechanism and focal loss as our final model for testing data. This model is trained and validated with data shown in Table 2. Following hyperparameters were used to train the model: Batch size was kept as 32, Adam was used as optimizer, learning rate was kept as 0.0001 and models were trained for 50 epochs. Also we have incorporated gradcam to the final model to show the explainability of our model [19].

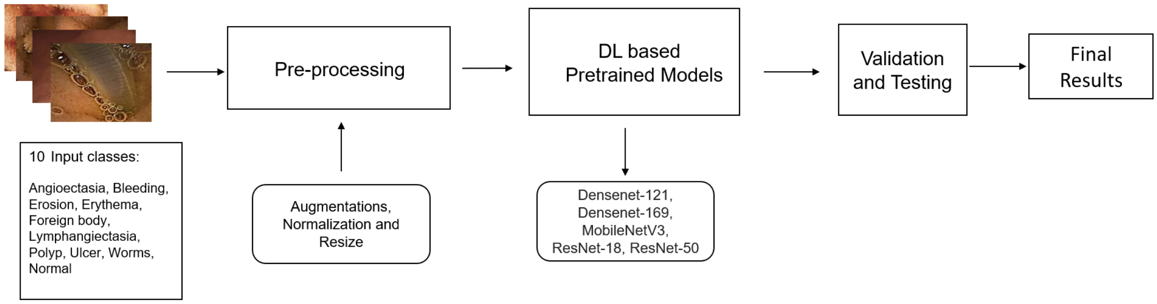
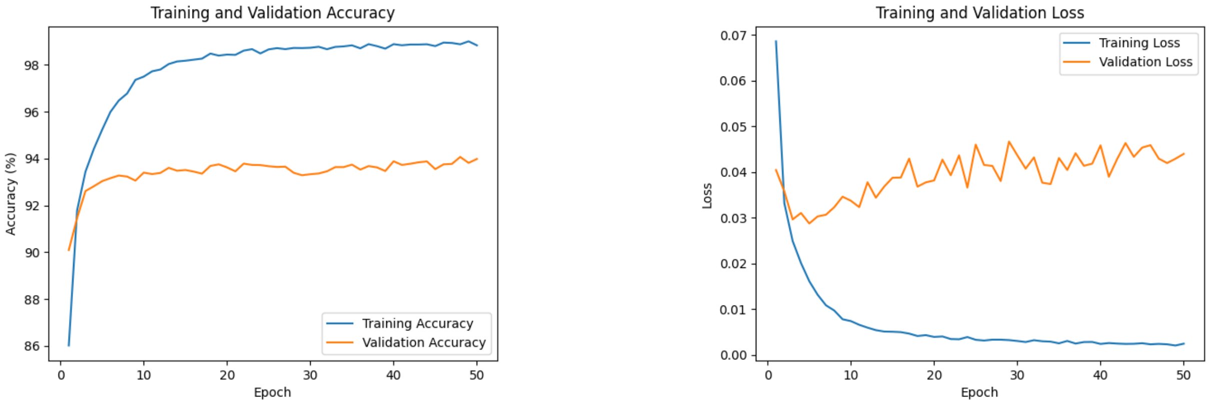


Figure 3: Block diagram of the developed pipeline.

3 Results

3.1 Results on the validation dataset

Table 3 and Table 4 depict the average AUC, average precision, average sensitivity, average specificity, average F1-score and balanced accuracy metrics for each aforementioned model. The performance of MobileNetV3 on the validation dataset as compared to the baseline model provided by the organizers is described in Table 3. Table 4 describes the performance comparison of our experimented models on the same validation dataset provided by the challenge organizers[20]. MobileNetV3 model with self attention and focal loss achieved an average sensitivity, average F1-score, avg precision and balanced accuracy of 0.82, 0.84, 0.89 and 0.83 respectively on validation data provided. Figure 4 shows the MobileNetV3 performance across 50 epochs with accuracy and loss graphs on the validation dataset. Figure 5 depicts the ROC curves of each class on the validation dataset. Figure 6 shows the confusion matrix obtained by MobileNetV3 model.



(a) Accuracy graph

(b) Loss graph

Figure 4: Accuracy graph and loss graph for MobileNetV3 model.

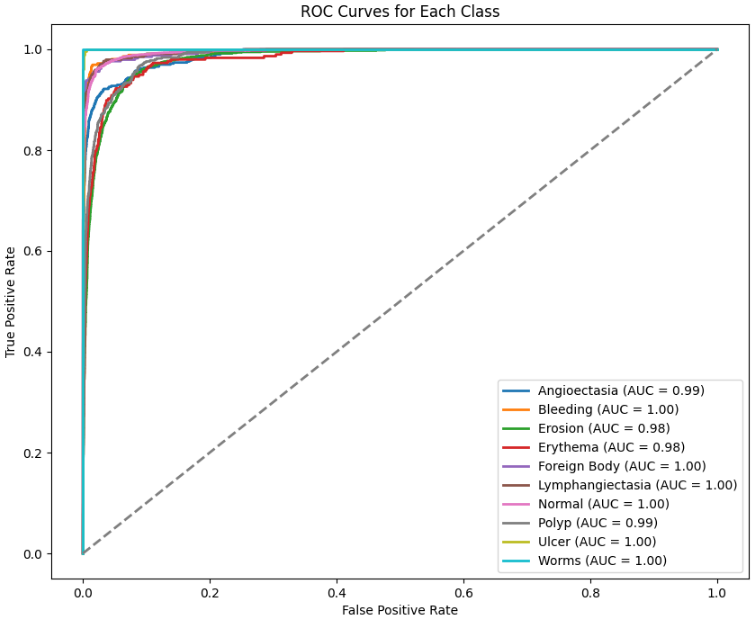


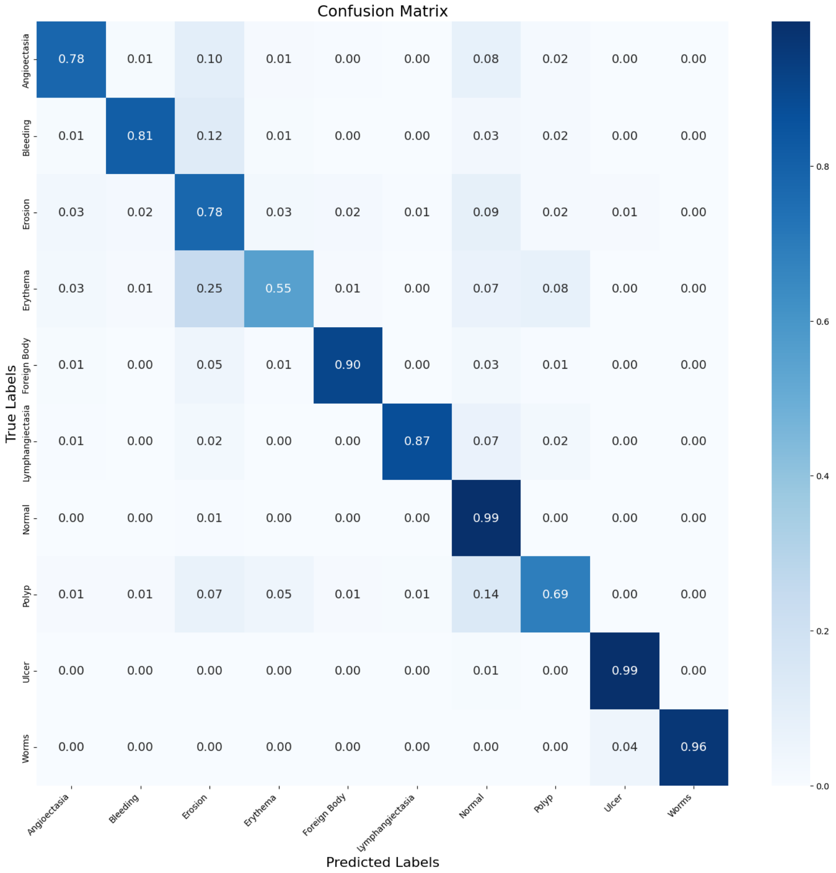
Figure 5: ROC Curve of each class for MobileNetV3 model.

Table 3: Validation results and comparison to the baseline methods reported by the organizing team of Capsule Vision 2024 challenge.

Method	**Avg. AUC**	**Avg. Specificity**	**Avg. Sensitivity**	**Avg. F1-score**	**Avg. Precision**	**Bal. Acc.**
Model 1 (VGG16)	0.92	0.97	0.54	0.48	0.52	0.57
Model 2 (SVM)	-	-	0.41	0.49	0.83	-
MobileNetV3	**0.99**	**0.99**	**0.82**	**0.84**	**0.89**	**0.83**

Table 4: Validation results and comparison to the models that we have experimented.

Method	**Avg. AUC**	**Avg. Specificity**	**Avg. Sensitivity**	**Avg. F1-score**	**Avg. Precision**	**Bal. Acc.**
Model 1 (DenseNet169)	0.99	0.99	0.82	0.84	0.88	0.83
Model 2 (DenseNet121)	0.99	0.99	0.82	0.84	0.88	0.83
Model 3 (ResNet18)	0.99	0.99	0.81	0.82	0.87	0.82
Model 4 (ResNet50)	0.99	0.99	0.81	0.84	0.88	0.83
Model 5 (MobileNetV3)	**0.99**	**0.99**	**0.82**	**0.84**	**0.89**	**0.83**

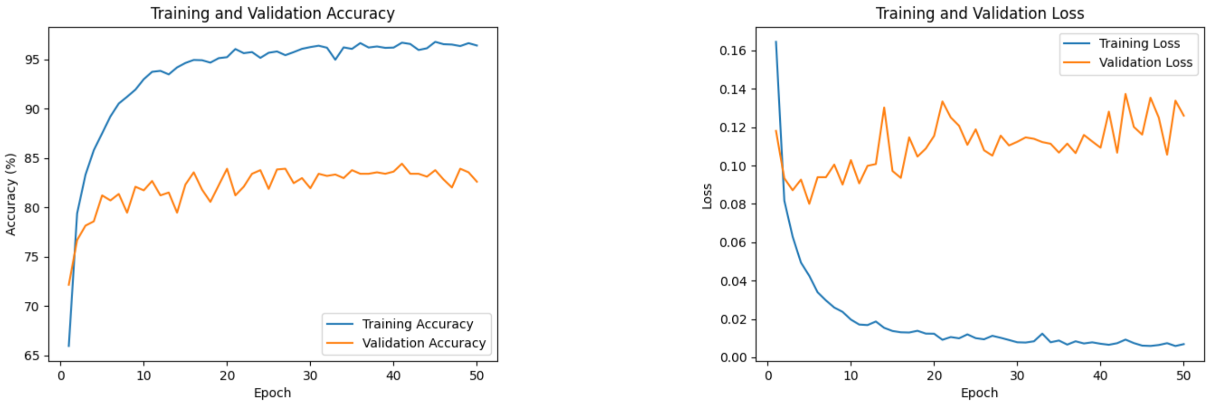


	Angioectasia	Bleeding	Erosion	Erythema	Foreign Body	Lymphangiectasia	Normal	Polyp	Ulcer	Worms
Angioectasia	0.78	0.01	0.10	0.01	0.00	0.00	0.08	0.02	0.00	0.00
Bleeding	0.01	0.81	0.12	0.01	0.00	0.00	0.03	0.02	0.00	0.00
Erosion	0.03	0.02	0.78	0.03	0.02	0.01	0.09	0.02	0.01	0.00
Erythema	0.03	0.01	0.25	0.55	0.01	0.00	0.07	0.08	0.00	0.00
Foreign Body	0.01	0.00	0.05	0.01	0.90	0.00	0.03	0.01	0.00	0.00
Lymphangiectasia	0.01	0.00	0.02	0.00	0.00	0.87	0.07	0.02	0.00	0.00
Normal	0.00	0.00	0.01	0.00	0.00	0.00	0.99	0.00	0.00	0.00
Polyp	0.01	0.01	0.07	0.05	0.01	0.01	0.14	0.69	0.00	0.00
Ulcer	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.99	0.00
Worms	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.96

Figure 6: Confusion Matrix obtained for MobileNetV3 model

3.2 Results on the test dataset

Figure 7 shows the accuracy and loss plots for the MobileNetV3 model trained and validated on the dataset described in Table 2. Figure 8 shows the samples of the predicted results along with GradCAM for the explainability on the test dataset provided by the organizers.



(a) Accuracy graph

(b) Loss graph

Figure 7: Accuracy graph and loss graph for the MobileNetV3 model trained and validated on the dataset described in Table 2.

4 Conclusion

It was indeed great for us to be a part of the Capsule Vision 2024 Challenge. This challenge involves the development and evaluation of a CNN-based models for the classification of anomalies on VCE frames. MobileNetV3 model with focal loss and self-attention obtained the best performance as compared to baseline models provided by the organizers as well as other experimented models. The CNN based deep learning models and techniques like Grad-CAM, provide powerful insights into their decision-making processes by visualizing key regions of focus, enabling a better understanding of complex features seen on VCE frames.

5 Acknowledgments

We would like to thank the organizers of the Capsule Vision 2024 Challenge for providing the dataset and platform for this competition[21]. Additionally, we acknowledge the contributions of our teammates and the support from our research institutions. As participants in the Capsule Vision 2024 Challenge, we fully comply with the competition's rules as outlined by the organizers. Our AI model development is based exclusively on the datasets provided in the official release in [20].

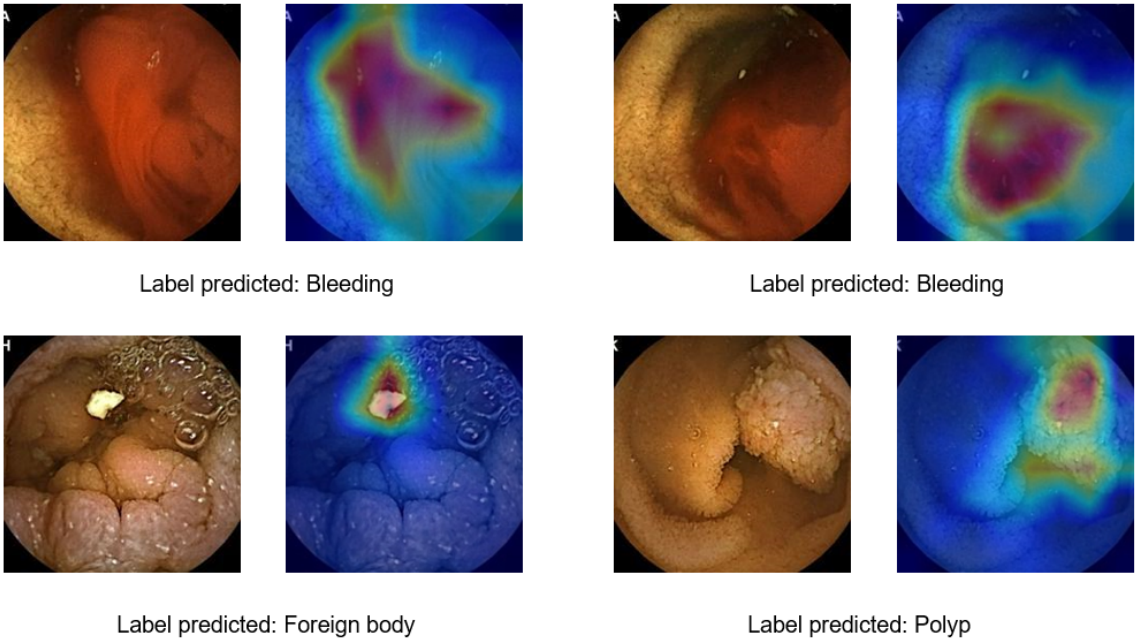

Label predicted: Bleeding
Label predicted: Bleeding
Label predicted: Foreign body
Label predicted: Polyp

Figure 8: Gradcam images generated during testing

References

[1] Frank Phillips and Sabina Beg. Video capsule endoscopy: pushing the boundaries with software technology. *Translational Gastroenterology and Hepatology*, 6, 2021.

[2] Sharib Ali. Where do we stand in ai for endoscopic image analysis? deciphering gaps and future directions. *npj Digital Medicine*, 5(1):184, 2022.

[3] AWYRC Sahafi, Y Wang, CLM Rasmussen, P Bollen, G Baatrup, V Blanes-Vidal, J Herp, and ES Nadimi. Edge artificial intelligence wireless video capsule endoscopy. *Scientific reports*, 12(1):13723, 2022.

[4] Dara Varam, Rohan Mitra, Meriam Mkadmi, Radi Aman Riyas, Diaa Addeen Abuhani, Salam Dhou, and Ayman Alzaatreh. Wireless capsule endoscopy image classification: an explainable ai approach. *IEEE Access*, 11:105262–105280, 2023.

[5] Hassaan Malik, Ahmad Naeem, Abolghasem Sadeghi-Niaraki, Rizwan Ali Naqvi, and Seung-Won Lee. Multi-classification deep learning models for detection of ulcerative colitis, polyps, and dyed-lifted polyps using wireless capsule endoscopy images. *Complex & Intelligent Systems*, 10(2):2477–2497, 2024.

[6] Prabhananthakumar Muruganantham and Senthil Murugan Balakrishnan. A survey on deep learning models for wireless capsule endoscopy image analysis. *International Journal of Cognitive Computing in Engineering*, 2:83–92, 2021.

[7] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching

for mobilenetv3. *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1314–1324, 2019.

[8] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Anastasios Koulaouzidis, Dimitris K. Iakovidis, Diana E. Yung, Emanuele Rondonotti, Uri Kopylov, John N. Plevris, Ervin Toth, Abraham Eliakim, Gabrielle Wurm Johansson, Wojciech Marlicz, Georgios Mavrogenis, et al. Kid project: an internet-based digital video atlas of capsule endoscopy for research purposes. *Endoscopy international open, vol. 5, no. 06, pp. E477–E483*, 2017.

[11] Pia H. Smedsrud, Vajira Thambawita, Steven A. Hicks, Henrik Gjestang, Oda Olsen Nedrejord, Espen Næss, Hanna Borgli, Debesh Jha, Tor Jan Derek Berstad, Sigrun L. Eskeland, Mathias Lux, et al. Kvasir-capsule, a video capsule endoscopy dataset. *cientific Data, vol. 8, no. 1, p. 142*, 2021.

[12] Akihito Yokote, Junji Umeno, Keisuke Kawasaki, Shin Fujioka, Yuta Fuyuno, Yuichi Matsuno, Yuichiro Yoshida, Noriyuki Imazu, Satoshi Miyazono, and Tomohiko Moriyama. Small bowel capsule endoscopy examination and open access database with artificial intelligence: The see-artificial intelligence project. *DEN open, vol. 4, no. 1, p. e258*, 2024.

[13] Nidhi Goel, Samarjeet Kaur, Deepak Gunjan, and S. J. Mahapatra. Dilated cnn for abnormality detection in wireless capsule endoscopy images. *Soft Computing, pp. 1–17*, 2022.

[14] Palak Handa, Amirreza Mahbod, Florian Schwarzhans, Ramona Woitek, Nidhi Goel, Deepti Chhabra, Shreshtha Jha, Manas Dhir, Pallavi Sharma, Dr. Deepak Gunjan, Jagadeesh Kakarla, and Balasubramanian Ramanathan. Testing Dataset of Capsule Vision 2024 Challenge. 10 2024. doi: 10.6084/m9.figshare.27200664.v1. URL `https://figshare.com/articles/dataset/Testing_Dataset_of_Capsule_Vision_2024_Challenge27200664`.

[15] https://pytorch.org/vision/stable/models.html.

[16] Jia GDeng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition, 248-255*, 2009.

[17] Lin T He K Ross G, Goyal P and Dollár P. Focal loss for dense object detection. *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2980–2988, 2017.

[18] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

[19] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[20] 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`.

[21] Palak Handa, Amirreza Mahbod, Florian Schwarzhans, Ramona Woitek, Nidhi Gooel, Deepti Chhabra, Shreshtha Jha, Manas Dhir, Deepak Gunjan, Jagadeesh Kakarla, and Balasubramanian Raman. Capsule vision 2024 challenge: Multi-class abnormality classification for video capsule endoscopy. 08 2024. doi: 10.48550/arXiv.2408.04940.