Optiminds Classification Results Submission: Insights and Findings

Shubham Shashank, Rupali Choudhary

Email: optiminds00@gmail.com

Abstract

1 Introduction

In this paper, we explore various techniques for the Capsule Vision 2024 Challenge, focusing on the efficacy of different models in a multiclass classification task.

2 Methods

We adopted several approaches to tackle the classification problem, which are outlined below:

1. Two model approach (one for normal vs diseased classification, another to classify which disease).
2. Feature extraction using models like VGG16 and ResNet50, followed by classification using SVM, Random Forest, etc.
3. Fine-tuning several CNNs, including ResNet, DenseNet, and YOLO (versions 8-10).
4. Fine-tuning multiple Vision Transformers, such as DeiT, DinoV2, and Swin.

Among these approaches, only the Vision Transformer models demonstrated promising results.

3 Results

3.1 Achieved Results

4 Discussion

The results of our model indicate a promising overall accuracy of 0.9040, reflecting its effectiveness in distinguishing between various classes. Notably, classes such as Normal and Worms demonstrated high precision and recall, which suggests that the model successfully identified these categories. However, we observed that the classes Bleeding, Polyp,

Table 1: Achieved results on the validation dataset by final model

Class	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Angioectasia	0.7565	0.6194	0.7565	0.6812
Bleeding	0.3649	0.7081	0.3649	0.4816
Erosion	0.8390	0.6022	0.8390	0.7012
Erythema	0.5017	0.6564	0.5017	0.5687
Foreign Body	0.8382	0.8533	0.8382	0.8457
Lymphangiectasia	0.6968	0.8951	0.6968	0.7836
Normal	0.9205	0.9545	0.9205	0.9372
Polyp	0.8540	0.6135	0.8540	0.7140
Ulcer	0.8706	0.8646	0.8706	0.8676
Worms	1.0000	0.9714	1.0000	0.9855

Table 2: Precision, recall, and F1-score are weighted.

and Erythema were significantly more challenging to classify, with lower accuracy and F1 scores.

In our exploration of different model architectures, Vision Transformers showed the most promise. Their ability to capture complex patterns may explain their superior performance over traditional CNNs and other classifiers. Fine-tuning these models yielded valuable insights, particularly regarding hyperparameter optimization and the importance of transfer learning.

Even though class imbalance did not pose any major obstruction, it is suggested that if data had more meaningful annotations for detection or segmentation, then even CNNs could prove to be efficient.

5 Conclusion

If data is improved and annotated for detection and segmentation, then even simple models can be generalized.

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`.