Multi-Class Abnormality Classification for Video Capsule Endoscopy

Navneet Jha 1 [a], Rishabh Negi 2 [b],
Raj Krishna Chaudhary 3 [c], Abhinav Singh 1 [d]
[a]Amity University, Noida, [b]Amity University, Noida
[c]Amity University, Noida, [d]Amity University, Noida
[a]navkj2003@gmail.com, [b]negirishu685@gmail.com
[c]rkchy12k@gmail.com, [d]itsabhinavsingh10@gmail.com

Abstract

The increasing prevalence of Gastro-Intestinal (GI) and liver diseases worldwide is a growing public health concern, driven by factors such as industrialization, dietary changes, and heightened antibiotic use. Video Capsule Endoscopy (VCE) emerges as a promising non-invasive diagnostic tool that enables direct visualization of the GI tract, significantly enhancing the detection of anomalies, particularly in small bowel diseases like Crohn's disease and intestinal cancer. However, the potential of VCE is constrained by lengthy video frame analysis times, manual biases, and hardware-related limitations. This paper emphasizes the urgent need for innovative solutions to automate the classification of abnormalities in VCE video frames through artificial intelligence (AI) technology. By developing robust, user-friendly, and interpretable AI models for multi-class classification, the initiative aims to reduce the diagnostic burden on gastroenterologists, expedite the inspection process, and maintain high accuracy. Access to comprehensive training, validation, and test datasets will enable the creation of vendor-independent AI-based models for automatic classification across ten distinct labels: angioectasia, bleeding, erosion, erythema, foreign body, lymphangiectasia, polyp, ulcer, worms, and normal. This effort holds the potential to transform the landscape of VCE technology and improve patient outcomes in GI health.

1. Introduction

The global burden of Gastro-Intestinal (GI) and liver diseases has risen significantly, influenced by various environmental factors, including industrialization, dietary changes, and the increasing use of antibiotics. This has led to a greater demand for effective diagnostic and management techniques. One such advancement is Video Capsule Endoscopy (VCE), a non-invasive method that allows for direct visualization of the GI tract using a small capsule-shaped device equipped with an optical dome, battery, illuminator, imaging sensor, and transmitter.

VCE stands out due to its non-invasive nature, eliminating the sedation-related complications commonly associated with traditional endoscopy. It has enhanced physicians' capabilities to detect a range of anomalies within the GI tract, particularly in small bowel diseases such as Crohn's disease, Celiac disease, and intestinal cancer. Despite its advantages, the full potential of VCE is hindered by challenges related to the lengthy reading time of video frames without compromising report quality, as well as the costs associated with the capsules.

During a typical VCE procedure, which lasts between 6 to 8 hours, the device captures between 57,000 and 1,000,000 frames. Experienced gastroenterologists currently spend approximately 2 to 3 hours analyzing these video frames on a frame-by-frame basis. This manual analysis is prone to human bias and high false-positive rates, influenced by factors such as bubbles, debris, intestinal fluid, and food residues that obscure mucosal frames. Additionally, a global shortage of gastroenterologists exacerbates delays in diagnosis. Hardware-related challenges, such as capsule retention, battery limitations, and bowel obstructions, further complicate the VCE process.

The integration of artificial intelligence (AI) into VCE technology presents a promising avenue for enhancing abnormality classification. There is an urgent need to develop robust, user-friendly, and interpretable AI models capable of multi-class abnormality classification. These models can significantly alleviate the workload of gastroenterologists by reducing the time required for VCE frame inspections while maintaining diagnostic accuracy.

The aim is to foster the development, testing, and evaluation of AI models designed for the automatic classification of abnormalities captured in VCE video frames. Participants will have access to distinct training, validation, and test datasets, facilitating comprehensive model training and validation processes. This initiative promotes the creation of vendor-independent and generalized AI-based models for an automatic abnormality classification pipeline, encompassing 10 class labels: angioectasia, bleeding, erosion, erythema, foreign body, lymphangiectasia, polyp, ulcer, worms, and normal.

2. Methods

The model begins with an input layer that accepts images of size 224x224 pixels with 3 color channels (RGB). The input is then passed through a rescaling layer, which normalizes pixel values to a range between 0 and 1. This normalization helps improve model convergence during training.

The first stage of the network involves a 2D convolutional layer with 64 filters, each of size 3x3, followed by a max pooling layer. This pooling layer reduces the spatial dimensions of the input feature map by half (224x224 becomes 112x112), retaining the most prominent features while reducing computational complexity.

Next, the model goes through another convolutional layer with 64 filters (3x3), followed by another max pooling layer, further down sampling the feature map to a size of 56x56.

In the third stage, a conv2D layer with 128 filters is applied, followed by another max pooling layer, reducing the spatial dimensions to 28x28.

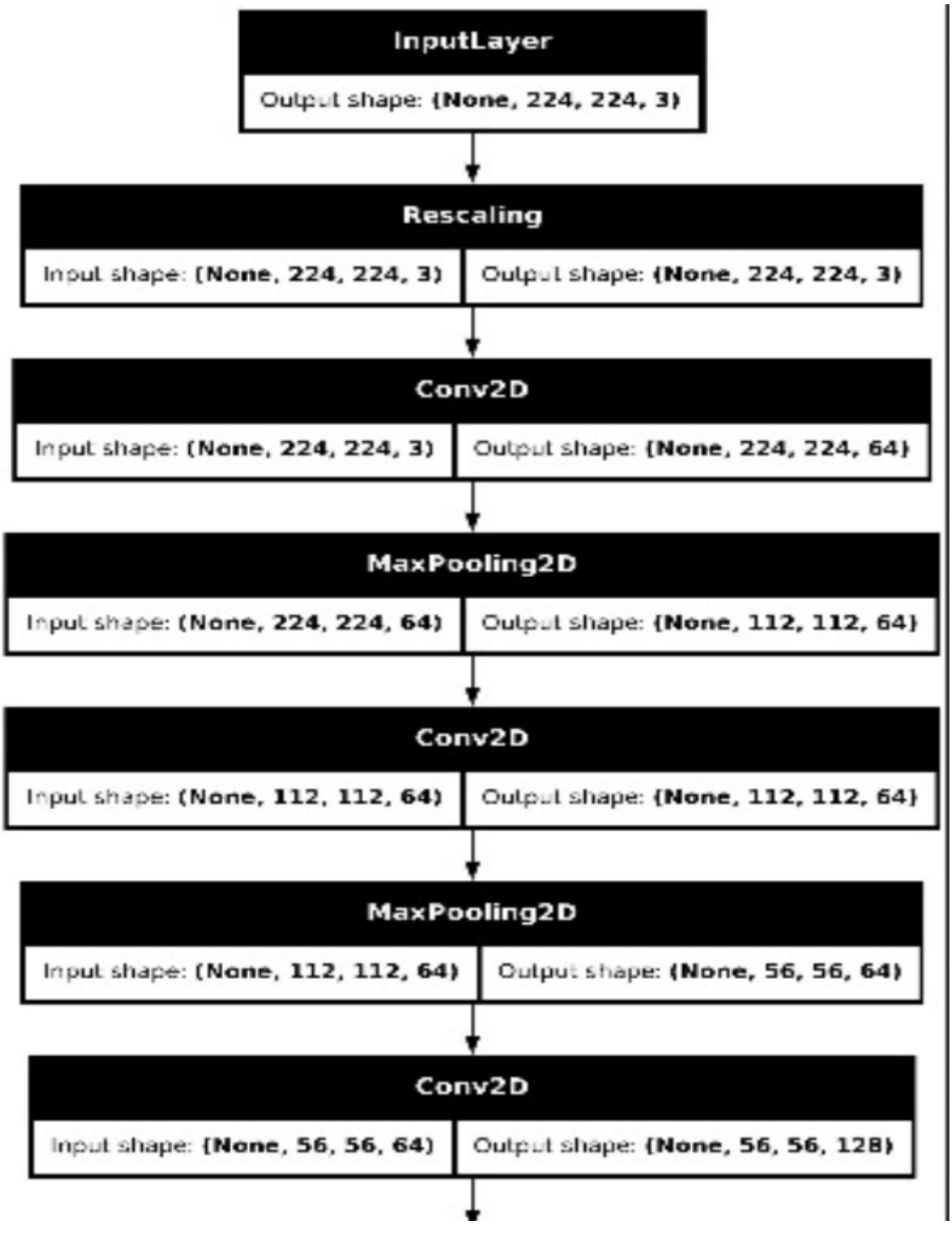
The fourth convolutional stage applies 256 filters, again followed by max pooling, which shrinks the feature map to 14x14. After this, the model uses dropout to randomly ignore 50%

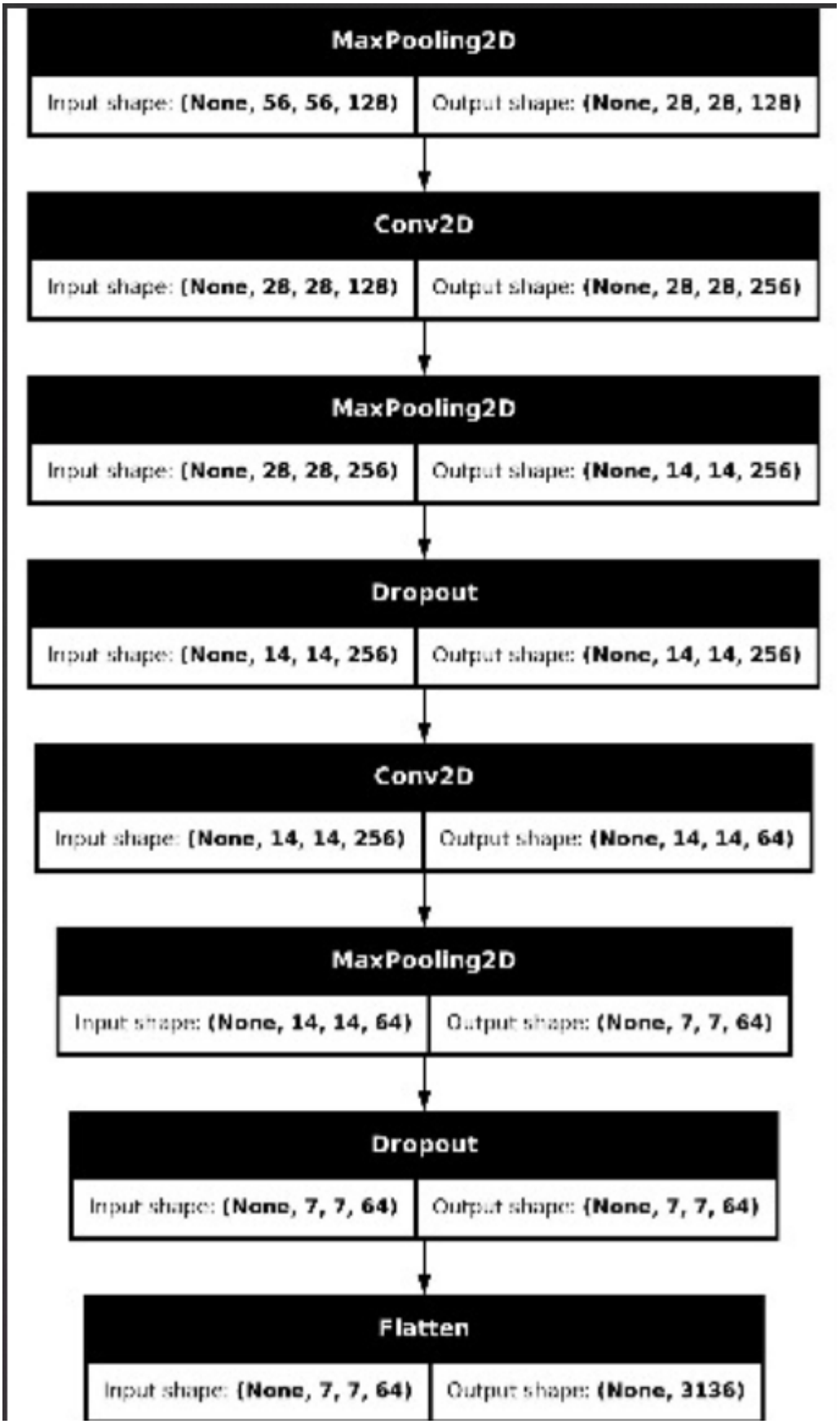
of the neurons during training, helping prevent overfitting.

A fifth conv2D layer with 64 filters follows, along with a max pooling layer, reducing the feature map to 7x7. Dropout is applied again to further regularize the model.

At this point, the model's output is flattened into a 1D vector of 3136 elements (7x7x64) via the flatten layer. The flattened vector is then passed through a dense (fully connected) layer with 100 neurons, followed by another dropout layer to prevent overfitting. Finally, the model ends with a dense output layer with 10 neurons, corresponding to the 10 possible output classes, and uses a SoftMax activation function to produce class probabilities.

This pipeline is typical of a CNN architecture used for image classification tasks, where the combination of convolutional layers, max pooling, and dropout ensures effective feature extraction and generalization.




MaxPooling2D
Input shape: (None, 56, 56, 128)
Output shape: (None, 28, 28, 128)
Conv2D
Input shape: (None, 28, 28, 128)
Output shape: (None, 28, 28, 256)
MaxPooling2D
Input shape: (None, 28, 28, 256)
Output shape: (None, 14, 14, 256)
Dropout
Input shape: (None, 14, 14, 256)
Output shape: (None, 14, 14, 256)
Conv2D
Input shape: (None, 14, 14, 256)
Output shape: (None, 14, 14, 64)
MaxPooling2D
Input shape: (None, 14, 14, 64)
Output shape: (None, 7, 7, 64)
Dropout
Input shape: (None, 7, 7, 64)
Output shape: (None, 7, 7, 64)
Flatten
Input shape: (None, 7, 7, 64)
Output shape: (None, 3136)

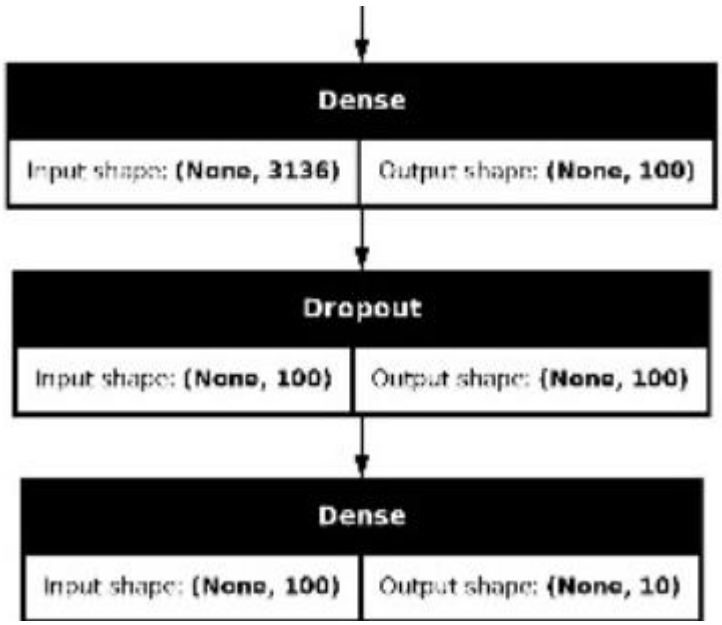


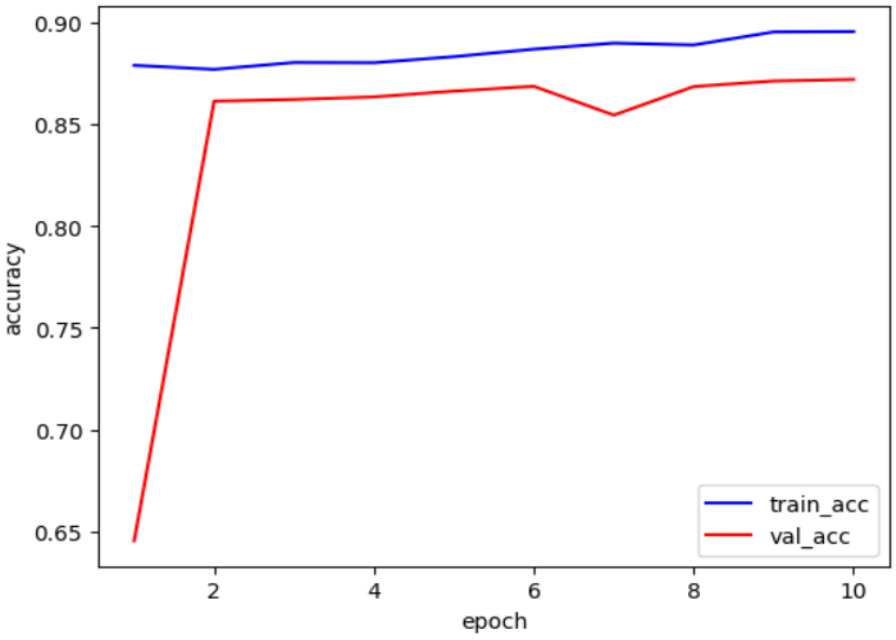
Figure 1: Block diagram of the developed pipeline.

3. Results

The convolutional neural network (CNN) model was trained over 30 epochs, with the performance particularly notable in the final epoch. By the end of training, the model achieved an accuracy of 95.34% and a loss of 0.1330.

3.1 Achieved results on the validation dataset:

The validation metrics also showed promising results, with a validation accuracy of 88.38% and a corresponding validation loss of 0.3974. Precision for validation dataset came to be 0.901 and the recall came out to be 0.8704.



These results indicate that the model effectively learned to classify abnormalities within the Video Capsule Endoscopy (VCE) video frames, demonstrating a robust ability to generalize to unseen data. The final epoch's performance highlights the model's potential to assist gastroenterologists in diagnosing gastrointestinal anomalies more efficiently, addressing the critical need for automated solutions in VCE analysis.

4. Discussion

The results obtained from the convolutional neural network (CNN) model demonstrate a significant advancement in the automated classification of abnormalities in Video Capsule Endoscopy (VCE) video frames. With an impressive training accuracy of 95.34% and a validation accuracy of 88.38%, the model has shown that it can effectively learn and generalize from the provided data, making it a promising tool for assisting gastroenterologists in their diagnostic processes.

The relatively low validation loss of 0.3974, compared to the training loss of 0.3130, indicates that the model is not overfitting, which is a common concern in deep learning applications. This suggests that the model maintains its performance even when faced with unseen data, thus enhancing its reliability in real-world applications. The findings align with previous studies that highlight the potential of AI technologies to reduce the time and effort required for manual analysis of VCE frames, which traditionally demands extensive expertise and is subject to human bias.

Furthermore, the results emphasize the importance of developing robust AI models that can operate independently of vendor-specific technologies. This approach not only promotes a broader implementation across various healthcare settings but also facilitates the integration of these models into existing clinical workflows, ultimately leading to improved patient outcomes.

5. Conclusions

In conclusion, the CNN model's performance in classifying abnormalities in VCE video frames indicates a significant step toward automating the analysis of gastrointestinal anomalies. The achieved accuracy rates and validation metrics underscore the model's capability to serve as a reliable diagnostic tool, addressing the pressing need for efficient solutions in gastroenterology. As the healthcare landscape continues to evolve, integrating AI technologies like this CNN model can substantially enhance diagnostic processes, reduce the burden on medical professionals, and improve patient care. Future work will focus on further refining the model, expanding the dataset for training, and exploring additional AI techniques to enhance classification accuracy and robustness.

6. Acknowledgement

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 Kakarl a, 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.