Deep Learning in Medical Imaging: Discussing Video Capsule Endoscopy (PillCam)

Abstract

Medical imaging is crucial for diagnosing and studying diseases, and deep learning has significantly improved the analysis of medical images. This article explores the role of deep learning in diagnostic imaging, focusing on video capsule endoscopy (PillCam). The discussion includes a snapshot of deep learning, its application in clinical imaging, and a detailed case study on the use of convolutional neural networks (CNNs) in analyzing capsule endoscopy images.

Keyword: Deep learning, medical imaging, Video capsule endoscopy (CE), Convolutional neural network (CNN), Computer aided diagnostics (CAD)

1. Introduction to Deep Learning

Deep learning, a branch of machine learning, utilizes multiple layers of neural networks to replicate the human decision-making processes (IBM, 2023). Medical imaging involves capturing images of internal organs to diagnose and study diseases. Over the past decade, deep learning has revolutionized medical image analysis by excelling in tasks such as image segmentation, registration, extracting the features and classifying them (Halalli & Makandar, 2017). These techniques support clinicians in achieving higher diagnostic accuracy by identifying hidden patterns in images (Sistaninejhad et al., 2023).

1.1 Working of Deep Learning

Deep neural networks (DNNs) simulate the decision-making process of the human brain through multiple layers of interconnected nodes. Each layer processes input from the previous layer, refines it, and passes it on to the next layer. This process is known as forward propagation. The structure of the layers involves neurons that apply linear transformations to the input data, which are then modified by non-linear activation functions. The final output layer provides the prediction or classification result.

Backpropagation is another critical process in deep learning. It involves calculating the error in the network's prediction, propagating this error backward through the network, and adjusting the weights and biases of the neurons to minimize the error and hence increases the accuracy (IBM, 2023).

1.2 Deep Learning Frameworks and Models

Deep learning applications are typically developed using powerful frameworks such as TensorFlow, PyTorch, and JAX. These frameworks provide the tools necessary to build, train, and deploy deep neural networks efficiently (IBM, 2023). The various models used are discussed below.

1.2.1 Convolutional Neural Networks (CNNs)

CNNs are designed specifically for image processing. They employ convolutional matrix operations to identify patterns and features in input data, making them particularly well-suited for image analysis tasks.

1.2.2 Recurrent Neural Networks (RNNs)

RNNs are suitable for sequential data and are used in applications such as topic modelling, speech recognition etc.

1.2.3 Autoencoders

Autoencoders are used for unsupervised learning tasks such as dimensionality reduction and anomaly detection. They are composed of an encoder that condenses the input data and a decoder that reconstitutes it.

1.2.4 Generative Adversarial Networks (GANs)

GANs are composed of two networks, a generator and a discriminator, that operate in opposition to one another.

1.2.5 Diffusion Models

Diffusion models are a type of generative model used for tasks like image synthesis and denoising. They model the distribution of data by simulating a diffusion process.

1.2.6 Transformer Models

Transformer models, initially developed for natural language processing, have also been applied to image analysis. They utilize self-attention mechanisms to detect dependencies over long distances in the data.

1.3 Deep Learning in Computer Vision

Deep learning has significantly impacted computer vision, leading to advancements in various fields:

1.3.1 Automotive

While fully autonomous cars are still in development, deep learning technologies have been integrated into vehicles to enhance safety features. For instance, lane detection systems use computer vision to help prevent accidents by warning drivers if they veer out of their lane.

1.3.2 Healthcare

In healthcare, computer vision has been instrumental in radiology. Advanced imaging techniques powered by deep learning can detect anomalies such as cancerous tumors with high precision, aiding early diagnosis and treatment planning.

1.3.3 Marketing

Social media platforms use computer vision to enhance user experience. For example, facial recognition algorithms can suggest tags for people in photos, making it easier for users to organize and share their memories.

1.3.4 Retail

E-commerce platforms leverage visual search capabilities powered by deep learning to improve the shopping experience. Customers can upload images of products they like, and the platform suggests similar items available for purchase.

2. Deep Learning in Medical Imaging

2.1 Computer-Aided Diagnosis (CAD)

CAD is a critical research field in medical imaging, where machine learning algorithms analyze historical patient imaging data to assess conditions. These algorithms assist in identifying abnormalities, improving diagnostic accuracy, and reducing the workload on healthcare professionals. CAD systems are increasingly used in various medical imaging modalities, such as X-rays, ultrasound, MRI etc (Ritter et al., 2011).

2.2 Categories of Medical Image Analysis

The several key tasks are (Ker et al., 2018):

- **Enhancement:** Improving image quality for better visualization and interpretation.
- **Registration:** Aligning images from different sources or at different times to compare them accurately.
- **Segmentation:** Isolating specific regions of interest, such as organs or tumors, for detailed analysis.
- **Classification:** Categorizing regions based on their characteristics, such as normal vs. abnormal tissue.
- **Localization:** Identifying the precise location of abnormalities.
- **Detection:** Recognizing the presence of specific features or anomalies.

2.3 Deep Learning Architectures

Hu et al. (2018) describe the following deep learning architectures which are commonly used in medical image analysis:

- **Convolutional Neural Networks (CNNs):** Ideal for processing visual data, CNNs have been widely used in tasks such as tumor detection and organ segmentation.
- **Fully Convolutional Networks (FCNs):** These networks are an extension of CNNs, designed for pixel-wise prediction tasks, making them suitable for image segmentation.
- **Deep Belief Networks:** These generative models learn to represent complex data distributions and have been applied to medical image classification.
- **Autoencoders:** Used for unsupervised learning, autoencoders can perform

tasks such as image denoising and anomaly detection.

3. Deep Learning and PillCam

3.1 Capsule Endoscopy (CE) Overview

CE is a non-invasive diagnostic method that uses a compact wireless camera contained in a capsule to capture images of the digestive tract, with a focus on the small intestine. After being swallowed, the PillCam capsule travels through the digestive system and collects thousands of images, which are then sent to a recorder which is worn on the waist of the patient.

3.2 PillCam in Detail (Mascarenhas Saraiva et al., 2021)

3.2.1 Device Details

Figure 1 shows the structure of wireless endoscopy capsule (WCE) (Umay et al., 2017).

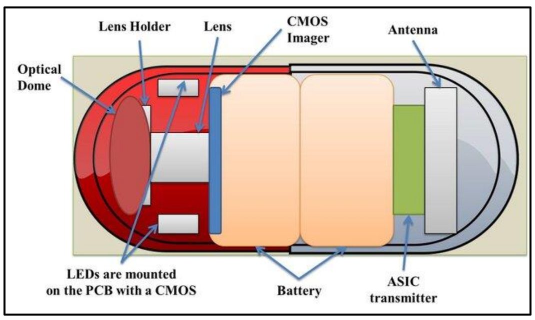

Figure 1

3.2.2 Diagnostic Capabilities

Capsule endoscopy can detect abnormalities, such as blood content, vascular lesions (e.g., angiectasia), ulcers, erosions, and protruding lesions. While other endoscopic procedures find it difficult to diagnose conditions, the CE can easily do the task, making it valuable.

3.2.3 Image Analysis Workflow

In a study by Mascarenhas Saraiva et al. (2021), a total of 53,555 images were extracted from full-length CE videos. These images were of the enteric mucosa. Of these, 18,010 images contained normal mucosa, while the remaining images included various abnormalities. The images were split into training (to train the model) and validation (to evaluate the model accuracy) datasets in the ratio of 80:20 to develop and assess the performance of a CNN. The detailed flow can be seen in figure 2.

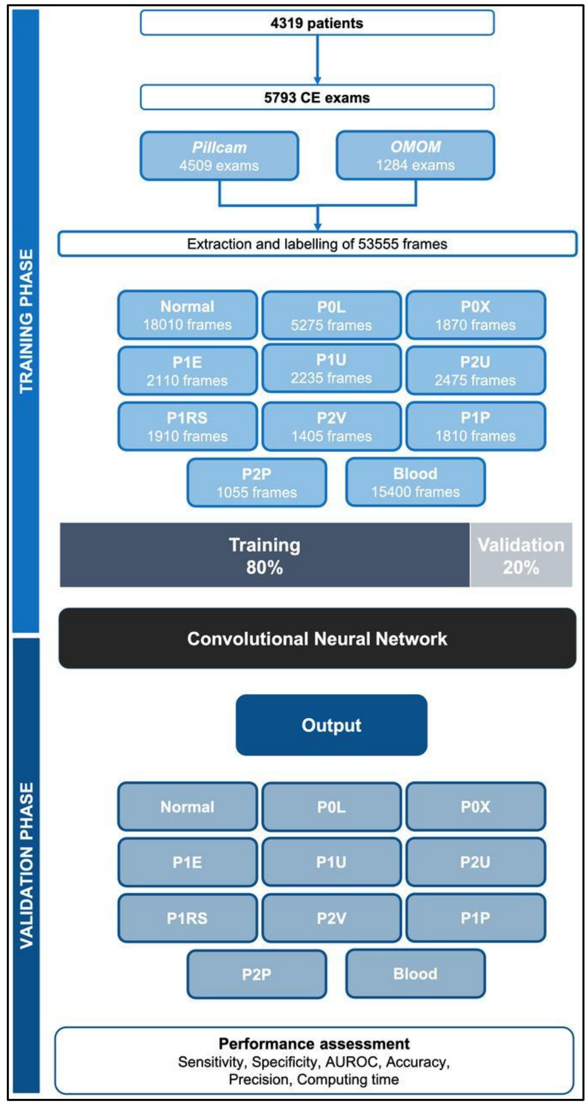

Figure 2

3.2.4 CNN Architecture

The researchers used the Xception model, pre-trained on ImageNet, as the base for their CNN. They retained the convolutional layers and replaced the fully connected layers to suit their classification task. The model was designed to classify images into three categories: P2 lesions, red spots and normal mucosa. The architecture included two blocks with fully connected layers followed by dropout layers to prevent overfitting. The training of the model was conducted with a learning rate set to 0.0001, a batch size of 32, and for a total of 100 epochs.

3.2.5 Results

The CNN's performance was evaluated using metrics such as accuracy, precision, sensitivity, and specificity. To gauge the model's capability to discriminate among the categories, Receiver Operating Characteristic (ROC) curves and the Area Under the ROC Curve (AUROC) metrics were employed. The CNN's outputs were compared with the diagnoses from specialists, which were considered the benchmark (Figure 3).

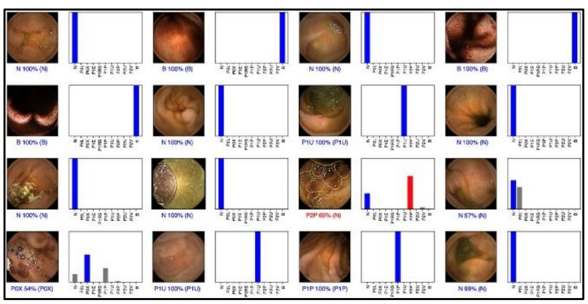

Figure 3

The model showed high accuracy in both training and validation environments (Figure 4).

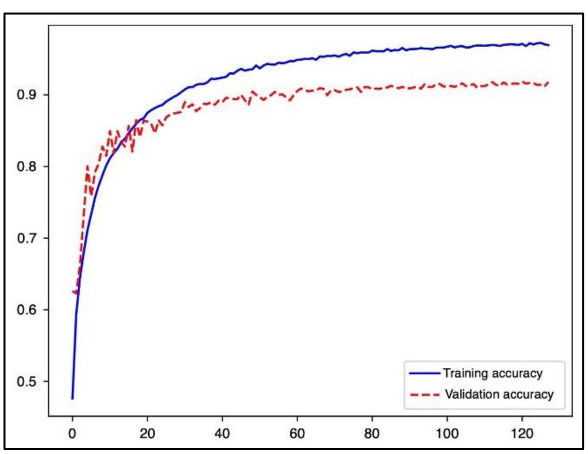

Figure 4

Heatmaps (Figure 5) generated by the software localized features predicting class probabilities, providing visual explanations for the CNN's decisions.

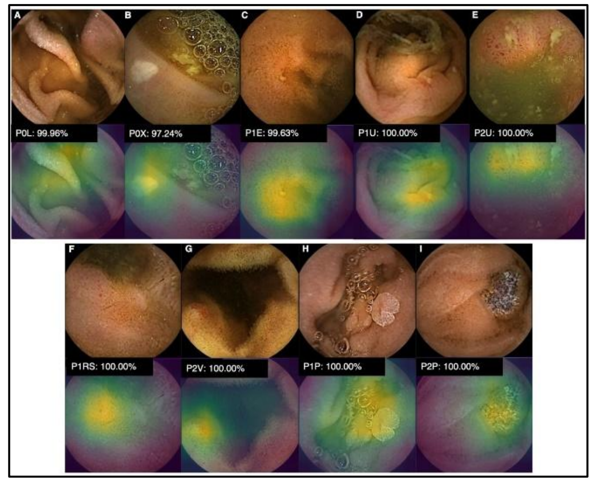

Figure 5

3.3 Significance of Automated Analysis

Reading CE examinations is time-consuming and labor-intensive, as significant lesions may be small and appear in only a few frames. The use of multilayered CNNs for automated intelligent systems for analysing CE images has yielded promising results, demonstrating the potential of deep learning (DL) to enhance medical diagnostics.

4. Discussion

The integration of deep learning in diagnostic imaging, particularly in capsule endoscopy, represents a significant advancement in healthcare technology. The ability of CNNs to analyze vast amounts of image data efficiently and accurately can lead to earlier detection of diseases, better treatment outcomes, and reduced burden on healthcare professionals. However, the full realization of deep learning's potential in medical imaging requires overcoming challenges such as maintaining data privacy, ensuring model interpretability, and managing the need for extensive annotated datasets.

5. Conclusion

Deep learning has transformed medical image analysis, offering powerful tools for enhancing diagnostic accuracy and efficiency. Capsule endoscopy, with the aid of CNNs, exemplifies the potential of deep learning in medical imaging. As technology continues to evolve, further research and development will be essential to overcome existing challenges and harness the full benefits of deep learning in healthcare.

References

1. Halalli, B., & Makandar, A. (2017). Computer Aided Diagnosis - Medical Image Analysis Techniques. *Breast Imaging*. https://doi.org/10.5772/intechopen.69792
2. Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., & Sun, Q. (2018). Deep learning for image-based cancer detection and diagnosis − A survey. *Pattern Recognition*, *83*, 134–149. https://doi.org/10.1016/j.patcog.2018.05.014
3. IBM. (2023). *What is Deep Learning?* Www.ibm.com; IBM. https://www.ibm.com/topics/deep-learning
4. Ker, J., Wang, L., Rao, J., & Lim, T. (2018). Deep Learning Applications in Medical Image Analysis. *IEEE Access*, *6*, 9375–9389. https://doi.org/10.1109/ACCESS.2017.2788044
5. Mascarenhas Saraiva, M. J., Afonso, J., Ribeiro, T., Ferreira, J., Cardoso, H., Andrade, A. P., Parente, M., Natal, R., Mascarenhas Saraiva, M., & Macedo, G. (2021). Deep learning and capsule endoscopy: automatic identification and differentiation of small bowel lesions with distinct haemorrhagic potential using a convolutional neural network. *BMJ Open Gastroenterology*, *8*(1), e000753. https://doi.org/10.1136/bmjgast-2021-000753
6. Mayo Clinic. (2019). *Capsule endoscopy - Mayo Clinic*. Mayoclinic.org. https://www.mayoclinic.org/tests-procedures/capsule-endoscopy/about/pac-20393366
7. Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., & Peitgen, H.-O. . (2011). Medical Image Analysis. *IEEE Pulse*, *2*(6), 60–70. https://doi.org/10.1109/MPUL.2011.942929
8. Umay, I., Fidan, B., & Barshan, B. (2017). Localization and Tracking of Implantable Biomedical Sensors. Sensors, 17(3), 583. https://doi.org/10.3390/s17030583
9. Sistaninejhad, B., Rasi, H., & Nayeri, P. (2023). A Review Paper about Deep Learning for Medical Image Analysis. Computational and Mathematical Methods in Medicine, 2023, e7091301. https://doi.org/10.1155/2023/7091301