Automated Cervical Cell Classification Using Vision Transformers: A Deep Learning Approach for AI in Healthcare

Yash Nareshkumar Patel

June 24, 2025

Abstract

This paper presents a deep learning-based approach to classify cervical cell images using a Vision Transformer (ViT) model. The goal is to support early detection of cervical cancer through automated image classification. The methodology includes data preprocessing, model training using transfer learning, performance evaluation, and visual explanations through Grad-CAM and t-SNE. The proposed system demonstrates the application of modern AI techniques in healthcare, particularly in digital pathology.

1 Introduction

Cervical cancer is one of the most preventable and treatable forms of cancer if detected early. Traditional manual screening methods, such as Pap smear tests, are prone to human error and inefficiencies. This research implements an AI-based image classification pipeline using Vision Transformers to assist in automated cervical cell classification, aiming to improve accuracy and efficiency in diagnostics.

2 Objective

The objective of this study is to develop a deep learning model to classify cervical cancer images into various cell types using a Vision Transformer. This model will serve as a tool to enhance diagnostic precision in clinical workflows.

3 Methodology

3.1 Data Preprocessing

We apply several data augmentation techniques, including resizing, flipping, rotation, and color jitter, to increase model generalization. Images are normalized to match ImageNet standards.

3.2 Dataset Splitting and Balancing

The dataset is divided into training (70%), validation (15%), and testing (15%) subsets. Class imbalance is handled using weighted cross-entropy loss, calculated based on class frequency.

3.3 Model Architecture

We utilize the `vit_base_patch16_224` architecture from the `timm` library, pretrained on ImageNet. The model's classification head is fine-tuned to match the number of classes in the cervical image dataset.

3.4 Training Process

We use the Adam optimizer with a learning rate of 1×10^{-4}, and a Cosine Annealing learning rate scheduler. Early stopping is employed to prevent overfitting, based on validation loss.

3.5 Evaluation Metrics

We evaluate the model using accuracy, precision, recall, F1-score, confusion matrix, and ROC-AUC. Additionally, features are visualized using t-SNE, and Grad-CAM is used to interpret predictions.

4 Results

4.1 AI for Early Cancer Detection

Early detection of cervical cancer significantly increases survival rates. AI systems can analyze large volumes of medical images more quickly and consistently than human experts, making them valuable diagnostic tools.

4.2 Benefits of Vision Transformers

Unlike CNNs, Vision Transformers use self-attention to model global relationships in the image, which can be particularly beneficial for capturing complex spatial patterns in medical images.

4.3 Explainability and Trust

Grad-CAM visualizations highlight image regions that influence the model's decision, building trust with clinicians. t-SNE visualizations show how well the model has learned to distinguish between classes in a lower-dimensional space.

5 Conclusion

This project demonstrates the feasibility and effectiveness of using Vision Transformers for automated cervical cancer diagnosis. The approach offers a robust pipeline from preprocessing to explainable AI, paving the way for practical deployment in healthcare settings.

6 Future Work

Future improvements could include:

- Expanding to multimodal data (e.g., clinical history + images)
- Integrating with hospital information systems
- Evaluating generalization on datasets from different populations or hospitals

References

[1] Dosovitskiy, A., et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *International Conference on Learning Representations (ICLR)*, 2021.

[2] Selvaraju, R. R., et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Conference on Computer Vision (ICCV)*, 2017.

[3] scikit-learn developers, "scikit-learn documentation for evaluation metrics," `https://scikit-learn.org/stable/modules/model_evaluation.html`, Accessed: 2025.