Using machine learning to classify squamous cells

Omolara Dawson, Raymond Gao

Dartmouth Geisel School of Medicine

Dartmouth Health

Introduction

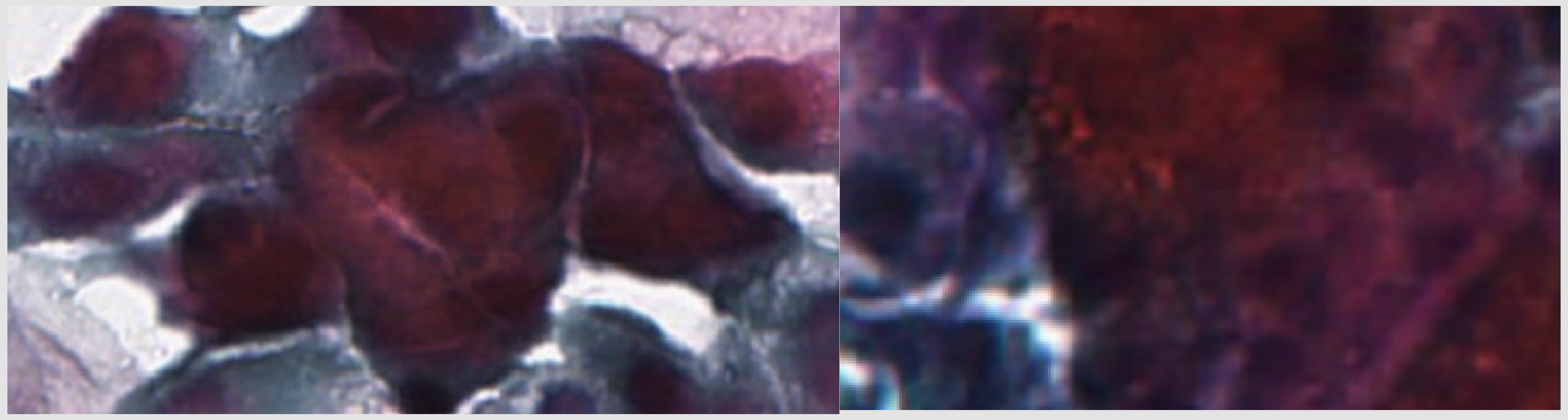
Cervical cancer, particularly squamous cell carcinoma (SCC), remains a major global health issue. Early and accurate diagnosis is essential but traditional methods, relying on manual analysis by pathologists, can be time-consuming and prone to variability. Machine learning offers a solution by automating the classification of cancerous cells, reducing human error and improving consistency.

Our project focuses on applying machine learning to classify squamous cells in cervical cancer samples. We developed a dataset of annotated whole slide images to train a model that distinguishes between normal and abnormal cells with high precision. This tool aims to enhance diagnostic accuracy, reduce diagnosis time, and ultimately improve patient care by integrating into clinical practice.

normal

some examples of abnormal and normal squamous cells

abnormal



Results

We learned how to manually annotate squamous cells and were able to distinguish between normal and abnormal squamous cells. Due to a limited amount of time, we were unable to complete certain tasks involving machine learning. The future we plan to complete the remaining steps we were unable to complete.

Methods

1. Annotated Image Dataset:
   - Whole Slide Images (WSIs): High-resolution cervical cell samples annotated by experts, including a variety of normal and abnormal cells.
   - Annotation Tools: Use tools like QuPath for detailed cell boundary and classification annotations.

2. Machine Learning Frameworks:
   - TensorFlow, PyTorch, Scikit-Learn: Libraries for building, training, and evaluating deep learning models.

3. Computational Resources:
   -GPUs: Use GPUs to speed up training for large datasets and complex models.

Discussion

Despite the advantages, challenges remain. The annotated dataset's quality and diversity are crucial for model performance, requiring accurate representation of squamous cell variability. Integrating the model into clinical workflows needs thoughtful user interface design and clinical validation. Future efforts will focus on expanding the dataset, refining model accuracy, and integrating the tool into clinical practice. Continued research will advance these goals.