INTRODUCTION

Colorectal cancer incidence is on the rise globally. The disease can be challenging to treat because it is a heterogeneous cancer, meaning it consists of diverse subtypes with distinct molecular signatures and clinical behaviors. The promise of precision medicine offers hope for treatment. In this project, students will automate the classification of colorectal tissue types, which can streamline the histopathological analysis process, reduce diagnostic variability, and improve the reproducibility of results.

DATASETS AND TISSUE TYPES

This project uses a labeled histopathology image dataset consisting of microscopic tissue samples from colorectal cancer cases. Each image represents a small patch of tissue extracted from a larger whole-slide image. The dataset supports supervised machine learning by pairing each image with a label assigned by expert pathologists. These labels indicate the type of tissue present and serve as ground truth for training and evaluation.

The dataset includes eight tissue categories: tumor, stroma, complex glands, lymphocytes, adipose tissue, mucosa, debris, and empty regions. Each tissue type provides important clinical information. For example, tumor tissue indicates cancer growth, lymphocytes reflect immune response, and stroma represents the tumor's supporting environment.

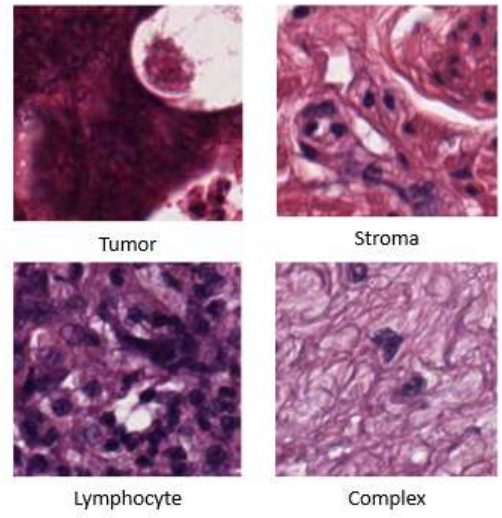

Figure 1 Different Tissue Types in the Dataset

Image data is stored as numerical pixel values and can be processed by convolutional neural networks. Tissue labels are categorical and must be converted into numerical format using one-hot encoding. This method transforms each label into a binary vector. For example, "tumor" may be represented as [1, 0, 0, 0, 0, 0, 0, 0]. By combining labeled image data with encoded labels, the dataset supports effective supervised learning and accurate tissue classification.
The dataset contains 1,024 images and was split into training and testing sets in a 3:1 ratio. Because medical datasets are often small, data augmentation techniques such as flipping, rotation, and gamma-correction were applied. These methods help the model learn meaningful features rather than superficial variations. By training on these augmented images, the model learns to focus on meaningful pathological features rather than superficial variations in orientation or scale.

THE MODEL : AI PATHOLOGIST

AI offers a promising approach for analyzing tissue samples and identifying cancer cells, immune responses, and tissue structures. Convolutional neural networks can recognize visual patterns and learn complex features automatically. When applied to histology slides, these models can detect subtle structures that may be difficult for humans to observe.

This study compares three models: a baseline CNN, a ResNet50-based transfer learning model, and an EfficientNetB3-based transfer learning model. All models were developed using Keras, which builds on TensorFlow and simplifies deep learning implementation.

Convolutional Neural Network

A convolutional neural network (CNN) is designed to recognize visual patterns in images, similar to how pathologists interpret tissue slides. It consists of several specialized layers [Figure 2].

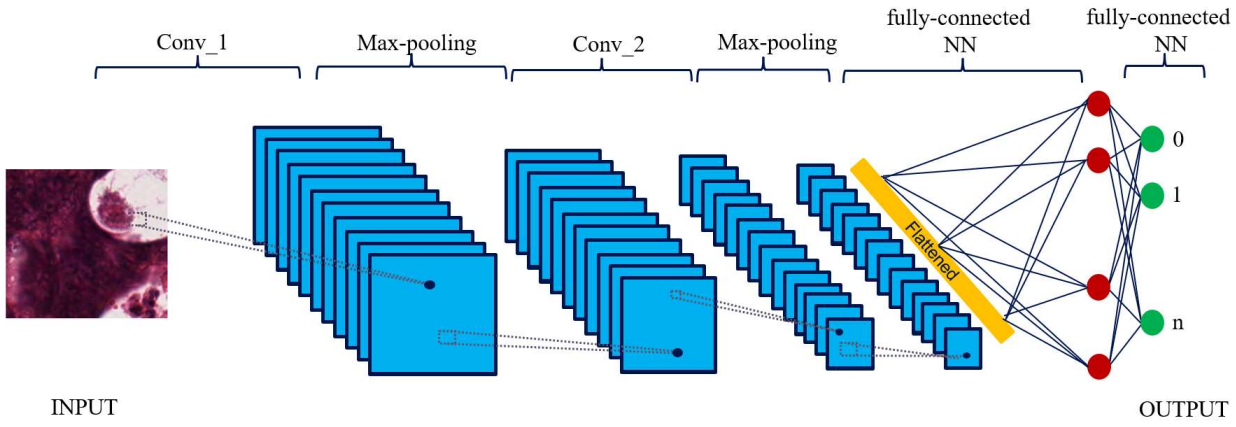

Figure 1 Basic CNN Model

Conv2D (Convolutional) Layers detect basic features such as edges and textures. As more layers are added, the model learns more complex structures, such as nuclei shapes and tissue patterns

MaxPooling2D Layers retain only the most prominent ones in each region, filtering out minor or inconsistent variations.

Dropout Layers ignore a fraction of the features the model has learned, forcing it to rely on general patterns that work across many examples. This technique prevents the model from simply memorizing the training images also called overfitting

Dense Layers are the final layers act as the model's decision-maker. They combine all information from the earlier layers and learn how to weigh that information to make a final prediction.

The CNN model was built using the Sequential API, which stacks layers in order [Figure 3]. The input layer was defined using the exact image dimensions.

Transfer Learning Model

Certain visual features, such as edges and textures, are important across many image classification tasks. CNNs learn these features in early layers. Transfer learning allows these learned patterns to be reused for new tasks.

Models such as ResNet, AlexNet, and MobileNet were trained on ImageNet, which contains millions of images. Although ImageNet does not include medical images, the learned features are still useful for histology.

In transfer learning [Figure 4], a pre-trained model is downloaded, and most layers are frozen. The final layer is modified to match the number of tissue classes. Most early layers are “frozen” so their weights do not change, while later layers are allowed to adapt to the new task. Input images are resized to meet model requirements.

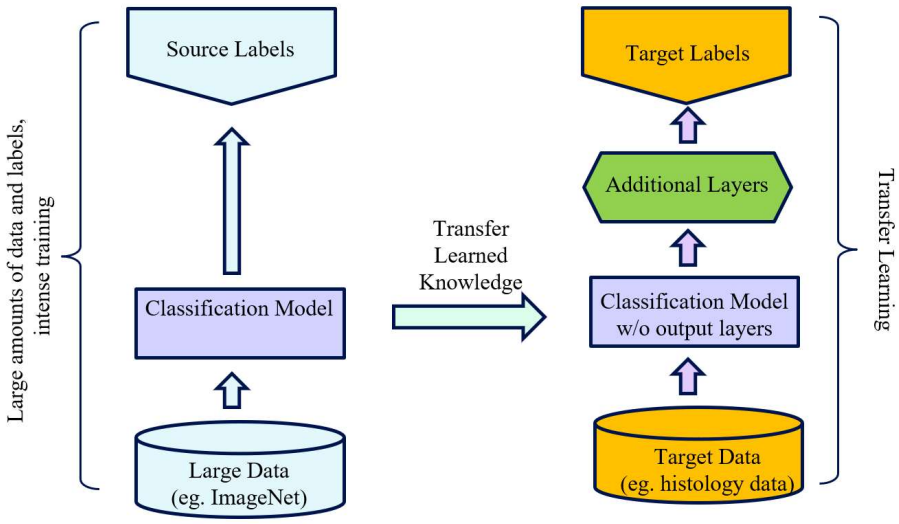

Figure 3 Transfer Learning Model

The model is then fine-tuned on histology data. This approach improves performance when working with limited datasets. In this study, ResNet50 [Figure 5] and EfficientNetB3 [Figure 6] are compared.

These models were built using the Functional API, which allows flexible connections between layers.

ResNet50 is a deep convolutional neural network with 50 layers that uses "residual connections" to prevent training problems in very deep models. These shortcut connections help preserve important features, improve accuracy, and enable effective transfer learning for medical image classification.

EfficientNetB3 balances depth, width, and resolution to achieve strong performance with fewer parameters. It is well suited for medical imaging tasks with limited data.

TRAINING THE MODELS

During training, the models analyzed labeled tissue samples, made predictions, compared them with true labels, and adjusted internal parameters. All models were trained for 20 epochs using the Adam optimizer and categorical cross-entropy loss function. Data augmentation techniques such as flipping, rotation, and gamma correction were applied.

RESULTS

In this section, we compare the training results of the three models: the baseline CNN, transfer learning using ResNet50, and transfer learning using EfficientNetB3. Training accuracy (shown by the blue line in the plots) measures how well a model performs on the data it was trained on. It represents the percentage of training examples that the model classifies correctly. High training accuracy indicates that the model is learning patterns from the training dataset. Validation accuracy (shown by the orange line) measures how well the model performs on new, unseen data that was not used during training. It is calculated using a separate validation dataset and reflects the model's ability to generalize beyond the training data.

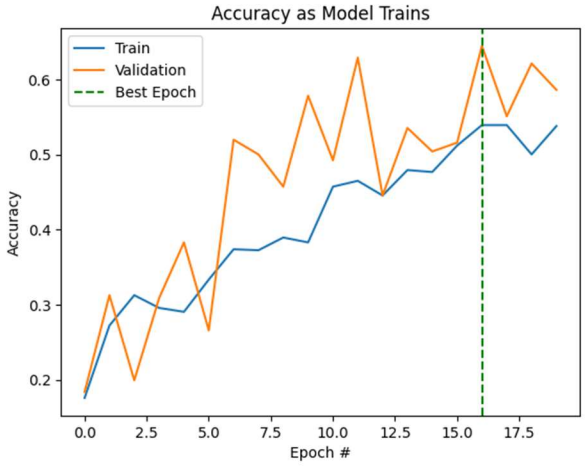

Figure 4 Accuracy for Basic CNN Model with Data Augmentation

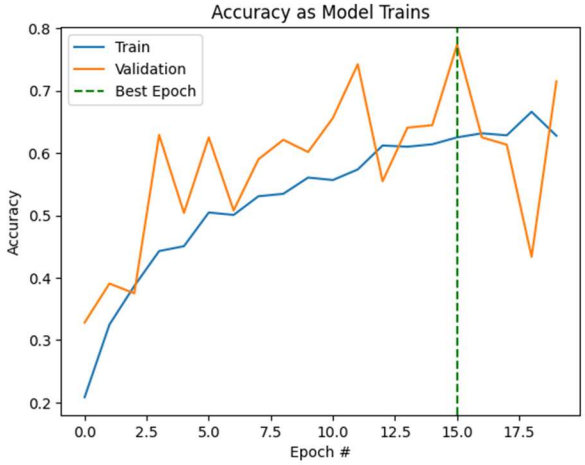

Figure 5 Accuracy for Basic CNN Model with Data Augmentation

Comparing Figures 4 and 5, training accuracy increases to very high levels, indicating that the model becomes highly effective at classifying training images, including augmented samples. However, validation accuracy remains unstable and does not show consistent improvement. Its peak performance is only slightly higher than that of the baseline model and remains highly variable. Although data augmentation increased the amount of training data, the simple CNN architecture was not powerful enough to capture robust and complex patterns in histology images.

As shown in Figures 6 and 7, the training accuracy of the transfer learning models rises quickly to high values, while validation accuracy plateaus earlier and at lower levels. This gap suggests that the models begin to memorize training samples rather than fully generalizing to new data. The validation loss reaches its minimum during mid-training and then stagnates or increases, indicating that peak performance occurs before training is complete.

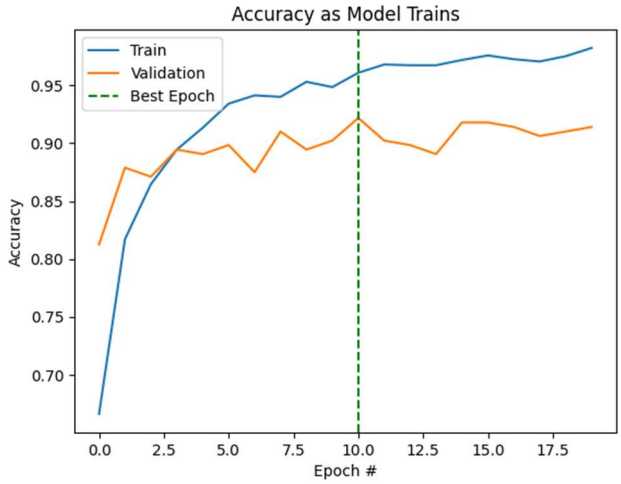

Figure 6 Accuracy for Transfer Learning from Resnet50

For the transfer learning models, Figures 8 and 9 present the classification reports, which summarize key performance metrics for each class, including precision, recall, F1-score, and support. Model effectiveness varies across tissue types. Visually distinct classes, such as adipose and empty, are classified with very high accuracy. In contrast, the complex class consistently shows lower F1-scores, indicating that it is more ambiguous and visually similar to other tissues.

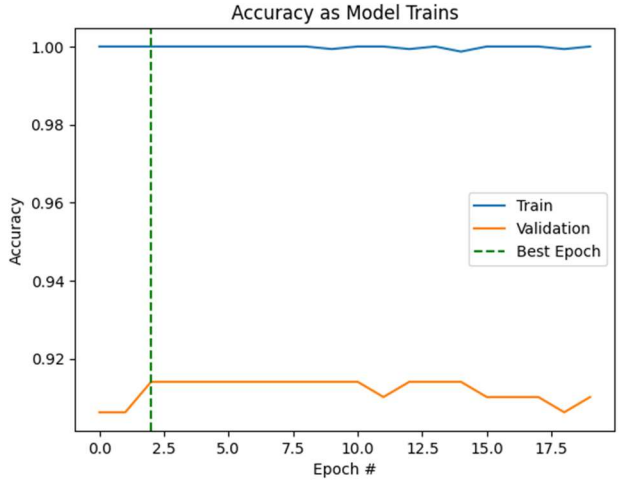

Figure 7 Accuracy for Transfer Learning from EfficientNetB3

	precision	recall	f1-score	support
adipose	0.98	1.00	0.99	43
complex	0.82	0.79	0.81	29
debris	0.91	0.87	0.89	23
empty	1.00	0.97	0.99	39
lympho	0.86	0.90	0.88	20
mucosa	0.81	0.97	0.88	31
stroma	0.91	0.88	0.90	34
tumor	0.97	0.86	0.91	37
accuracy			0.91	256
macro avg	0.91	0.91	0.91	256
weighted avg	0.92	0.91	0.91	256

Figure 8 ResNet50 Transfer Learning Classification Report

	precision	recall	f1-score	support
adipose	1.00	1.00	1.00	43
complex	0.84	0.72	0.78	29
debris	0.92	1.00	0.96	23
empty	1.00	0.97	0.99	39
lympho	0.68	0.85	0.76	20
mucosa	0.93	0.84	0.88	31
stroma	0.91	0.91	0.91	34
tumor	0.89	0.92	0.91	37
accuracy			0.91	256
macro avg	0.90	0.90	0.90	256
weighted avg	0.91	0.91	0.91	256

Figure 9 EfficientNetB13 Transfer Learning Classification Report

For the critical tumor class, the ResNet50 transfer model achieves high precision, meaning that its tumor predictions are usually correct. However, it shows noticeably lower recall, which is a major clinical concern. This indicates that the model fails to identify a significant number of actual tumors, resulting in false negatives. These values may vary across different training runs.

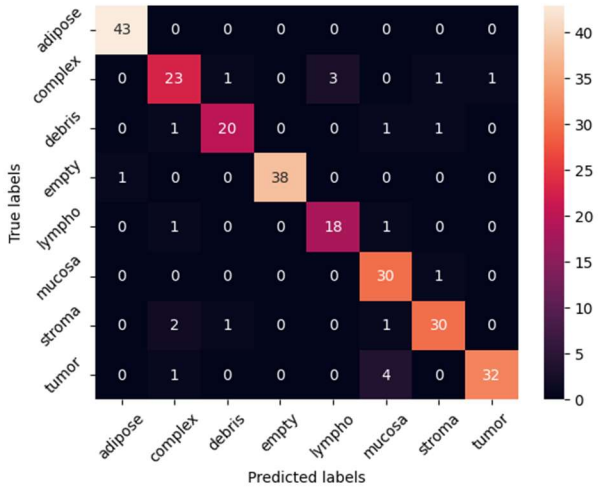

Figure 20 Confusion Matrix of the ResNet50 Transfer Learning

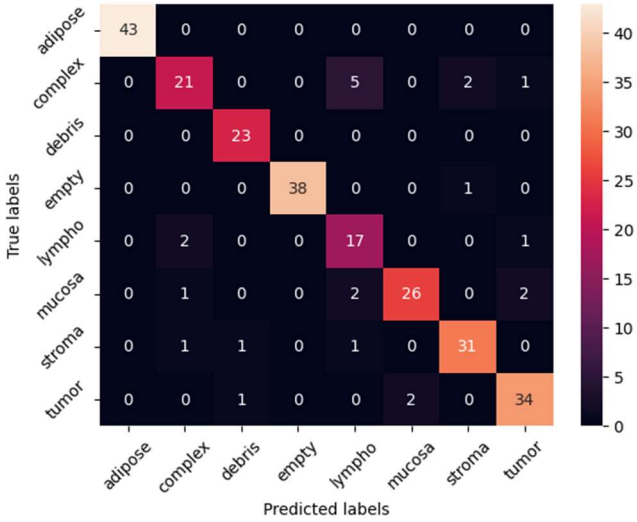

Figure 31 Confusion Matrix of EfficientNetB13 Transfer Learning