ECE 662 - Deep Learning with Python

Project #1 - MedMNIST

Richard Groves

November 16, 2023

1 Introduction

Deep Learning is revolutionizing the field of medical image analysis. The ability of these networks to learn features from complex image data has led to significant advancements in diagnostic tools and methodologies. This project focused on the application of a neural network (NN) to classify medical images from the TissueMNIST dataset, a subset of the MedMNIST database, which contains segmented human kidney cortex cells.

The TissueMNIST dataset presented a challenge due to its multi-class nature and the small details present in medical images. The goal of this project was to design and train a NN to classify these images into one of eight categories, based on the type of kidney cell they represent.

In this report, I describe the dataset, the design of the neural network, the training process, and the performance of the model on the test dataset. I also compare my results with existing literature [1] to understand the effectiveness of my approach in the broader context of medical image classification using deep learning.

2 Background on the TissueMNIST Dataset

The TissueMNIST dataset is a part of the larger MedMNIST collection. Originating from the Broad Bioimage Benchmark Collection (BBBC051), the TissueMNIST dataset is categorized for multi-class classification tasks. This dataset is used to advance computational pathology by offering a diverse set of medical images for algorithm development and evaluation.

2.1 Dataset Characteristics

The TissueMNIST dataset focuses on human kidney tissues, segmented into eight distinct categories. These categories include various kidney cell types such as Collecting Duct, Connecting Tubule, Distal Convoluted Tubule, Glomerular Endothelial Cells, Interstitial Endothelial Cells, Leukocytes, Podocytes, Proximal Tubule Segments, and Thick Ascending Limb. Each category represents a unique aspect of kidney tissue.

2.2 Data Preparation and Preprocessing

Initially, the BBBC051 dataset contained 236,386 cells from 3 reference tissue specimens. For the TissueMNIST dataset, these cells were organized into the categories mentioned earlier and split into training, validation, and test sets. Each gray-scale image in the dataset was originally 32×32×7 pixels, representing 7 slices. The images were reduced to 32x32x1 by taking the maximum pixel value of the 7 slices. The images were then resized to 28x28x1 before being published to the MedMNIST database.[1]

3 Approach

3.1 Model Selection and Architecture

I started the project with the selection and design of a suitable NN architecture. This architecture was iteratively modified/refined, aiming to balance the model's complexity with its performance, ensuring efficient training without compromising on accuracy.

Network Configuration	**Learning Rate**	**Batch Size**	**Epochs**	**Accuracy**	**Loss**
Conv2d-1-10-3x3, MaxPool2d-2x2, Conv2d-10-20-3x3, Linear-980-100, Linear-100-50, Linear-50-8	0.01	223	5	41.90%	0.007162
Conv2d-1-10-3x3, MaxPool2d-2x2, Conv2d-10-20-3x3, Linear-980-300, Linear-300-150, Linear-150-8	0.01	223	5	46.50%	0.006846
Conv2d-1-10-3x3, MaxPool2d-2x2, Conv2d-10-20-3x3, Linear-980-100, Linear-100-50, Linear-50-8	0.1	223	5	54.30%	0.002780
Conv2d-1-10-3x3, MaxPool2d-2x2, Linear-1960-100, Linear-100-50, Linear-50-8	0.01	223	5	47.50%	0.006591
Conv2d-1-10-3x3, MaxPool2d-2x2, Conv2d-10-20-3x3, Linear-980-100, Linear-100-50, Linear-50-8	0.001	223	5	32.10%	0.004079
Conv2d-1-10-3x3, MaxPool2d-2x2, Conv2d-10-20-3x3, Linear-980-100, Linear-100-50, Linear-50-8	0.01	223	75	58.80%	0.002534
Conv2d-1-10-3x3, MaxPool2d-2x2, Conv2d-10-20-3x3, Linear-980-100, Linear-100-50, Linear-50-8	0.01	223	150	59.00%	0.002626

Table 1: Summary of network configurations, hyperparameters and their performance

3.2 Training Process

The training of the NN involved several phases. Initially, I trained the model on a subset of the TissueMNIST dataset, monitoring its performance on both training and validation sets. The learning rate, batch size, and other hyperparameters were adjusted in an effort to optimize the performance.

3.3 Performance Evaluation

After training, I tested the model on the TissueMNIST test dataset. This step was crucial to assess the model's effectiveness in classifying data. The performance was measured in terms of accuracy and loss, and the results were compared with existing benchmarks in the literature [1] to gauge the model's efficacy.

4 Experiment

4.1 Training

4.1.1 Data

The dataset was partitioned into training, validation, and test subsets, with the training set containing 165,466 images. Each gray-scale image was 28×28. Normalization was applied to the images using a mean and standard deviation of 0.5 to ensure consistency in the data fed into the neural network.

4.1.2 Model Architecture

The neural network, designed specifically for this project, comprised convolutional, pooling, and fully connected layers. The model architecture included two convolutional layers ('conv1' and 'conv2'), each followed by a max-pooling layer. The first convolutional layer had a kernel size of 3x3, taking a single channel input and producing 10 output channels. The second convolutional layer expanded this to 20 output channels. The network architecture includes two max-pooling layers, each utilizing a 2x2 window with a stride of 2. After flattening the output from the convolutional layers, the network had three fully connected layers ('fc1', 'fc2', and 'fc3'), reducing the dimensionality to the final output of 8 units, corresponding to the number of classes in the dataset.

The architecture of the neural network is as follows:

```
Net(
  (conv1): Conv2d(1, 10, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), padding_mode=replicate)
  (pool): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)
  (conv2): Conv2d(10, 20, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1), padding_mode=replicate)
  (fc1): Linear(in_features=980, out_features=100, bias=True)
  (fc2): Linear(in_features=100, out_features=50, bias=True)
  (fc3): Linear(in_features=50, out_features=8, bias=True)
)
```

4.1.3 Training Parameters and Procedure

A learning rate of 0.01 was set, and the model was trained over 150 epochs with a batch size of 223. The batch size of 223 was chosen because it will evenly divide the training dataset into 742 batches. The optimization algorithm used was Stochastic Gradient Descent (SGD). Cross-Entropy Loss was employed as the loss function to guide the training process.

4.2 Results

4.2.1 Validation and Testing

The model's performance was evaluated on both validation and test datasets during and after training. The validation accuracy and loss were monitored after each epoch to adjust hyperparameters and prevent overfitting. Upon completion of the training, the model achieved a test accuracy of 59.0% (27895/47280) with a loss of 0.002626 on the test dataset, demonstrating its capability to classify the medical images from the TissueMNIST dataset effectively.

4.2.2 Validation Accuracy and Loss

Throughout the training process, the model's performance was evaluated based on its accuracy and loss on the validation set. The following figures illustrate the model's performance over 150 epochs, which indicate a steady improvement in validation accuracy and a consistent decrease in validation loss.

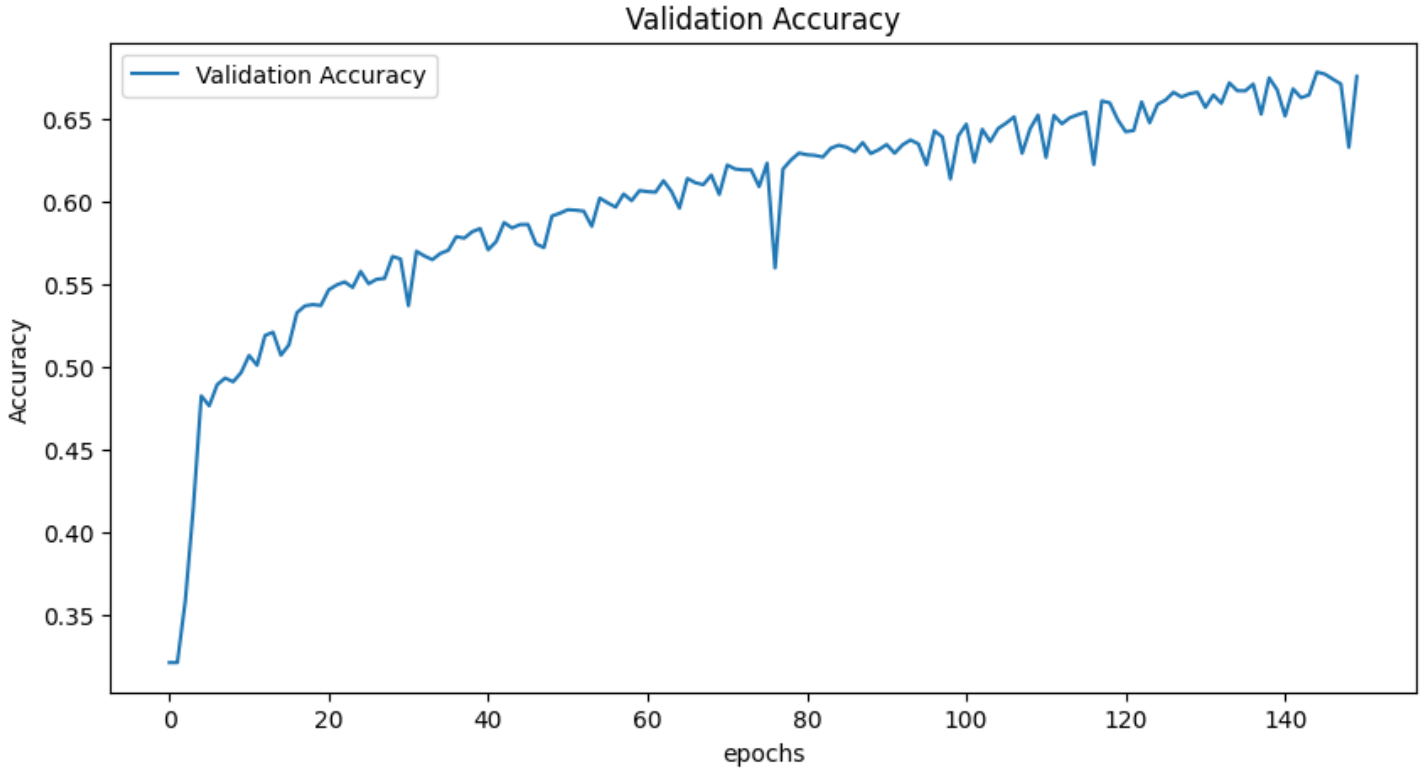

Figure 1: Validation Accuracy over 150 epochs. The model's accuracy improved steadily, indicating effective learning over time.

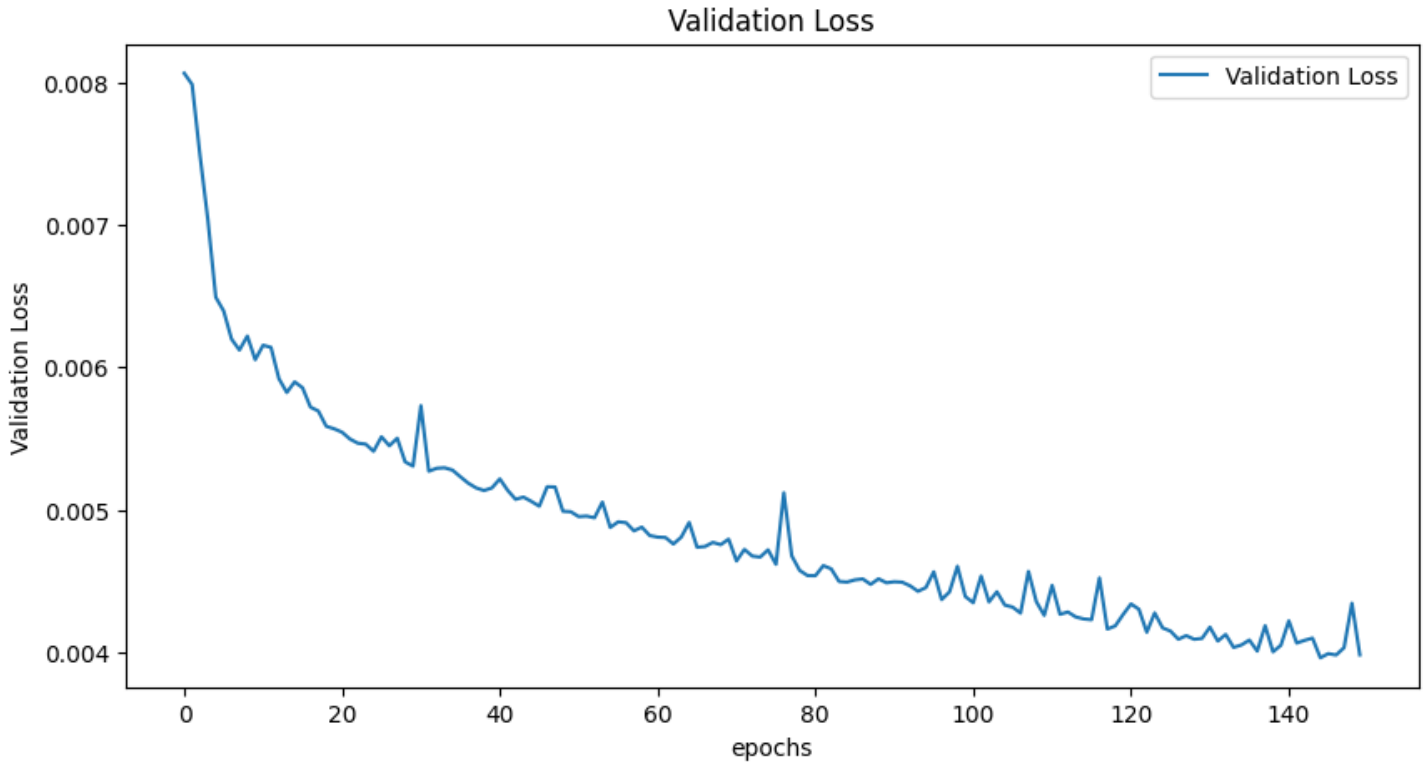

Figure 2: Validation Loss over 150 epochs. The loss decreased as the model became better at generalizing from the training data.

As shown in Figure 1, the validation accuracy increased from around 35% to just over 65%, demonstrating the model's capacity to learn and make accurate predictions as training progressed. Conversely, Figure 2 depicts a reduction in validation loss, indicating an improvement in the model's prediction capabilities. These results are encouraging when considering the model's final performance on the unseen test data.

4.3 Model Predictions and Misclassifications

During the testing phase, certain images were incorrectly classified by the model. To illustrate this, images from the predicted category, the actual category, and the incorrectly categorized instance are presented below for each class. These examples help to visualize where the model performed well and where it had difficulties.

4.3.1 Analysis

The test results were compared to existing benchmarks and literature [1] in the field of medical image classification using NNs. The accuracy achieved suggests the model's reasonable performance. Further tuning of hyperparameters, exploring different model architectures, or using more advanced training techniques might yield improved results.

5 Conclusion

In this project, a Neural Network was developed to classify images from the TissueMNIST dataset, a subset of the MedMNIST collection. The model demonstrated a test accuracy of 59.0% (27895/47280) and a loss of 0.002626 on the test dataset. When

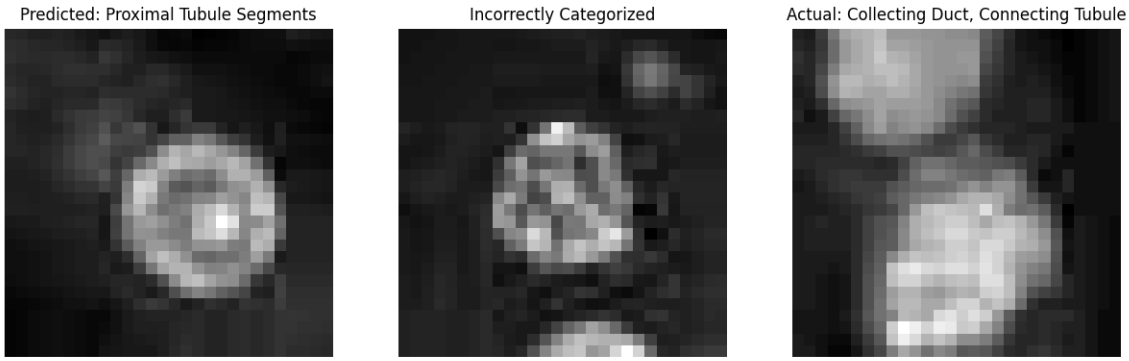

Figure 3: Misclassification example for category 0 (Collecting Duct, Connecting Tubule). Left: Image predicted as Proximal Tubule Segments. Middle: Incorrectly categorized image. Right: Actual class Collecting Duct, Connecting Tubule.

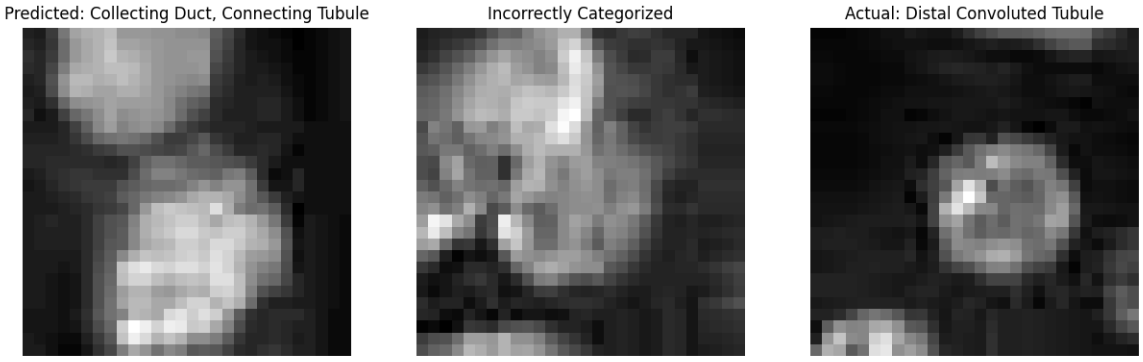

Figure 4: Misclassification example for category 1 (Distal Convoluted Tubule). Left: Image predicted as Collecting Duct, Connecting Tubule. Middle: Incorrectly categorized image. Right: Actual class Distal Convoluted Tubule.

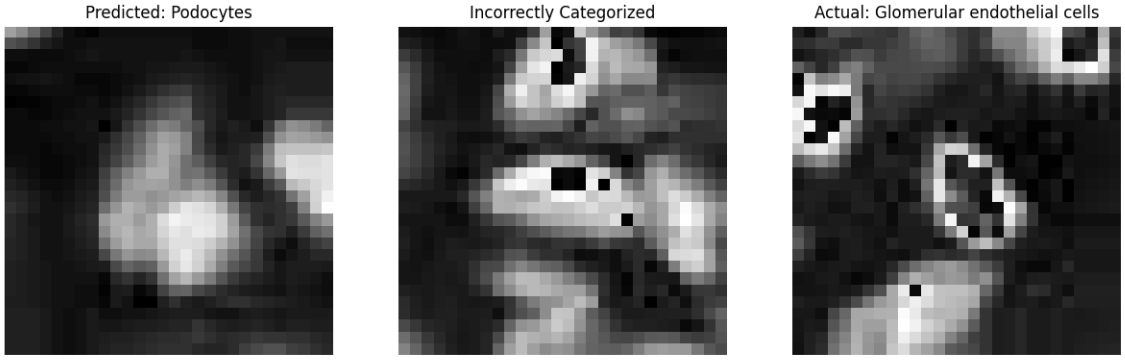

Figure 5: Misclassification example for category 2 (Glomerular Endothelial Cells). Left: Image predicted as Podocytes. Middle: Incorrectly categorized image. Right: Actual class Glomerular Endothelial Cells.

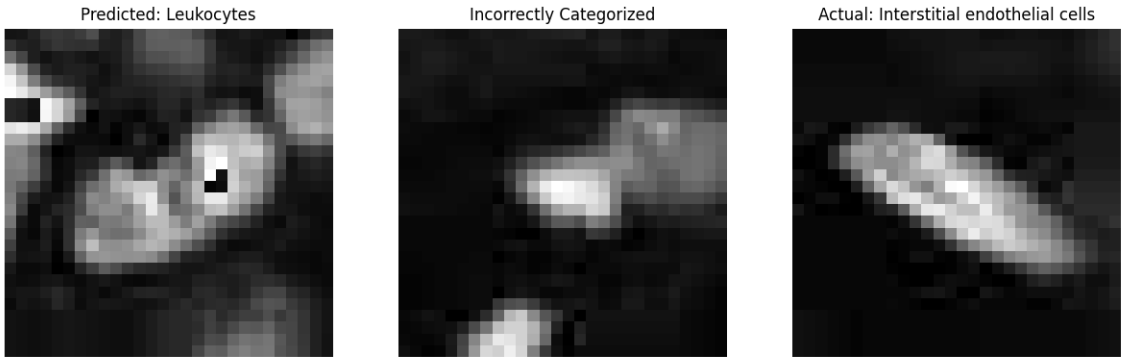

Figure 6: Misclassification example for category 3 (Interstitial Endothelial Cells). Left: Image predicted as Leukocytes. Middle: Incorrectly categorized image. Right: Actual class Interstitial Endothelial Cells.

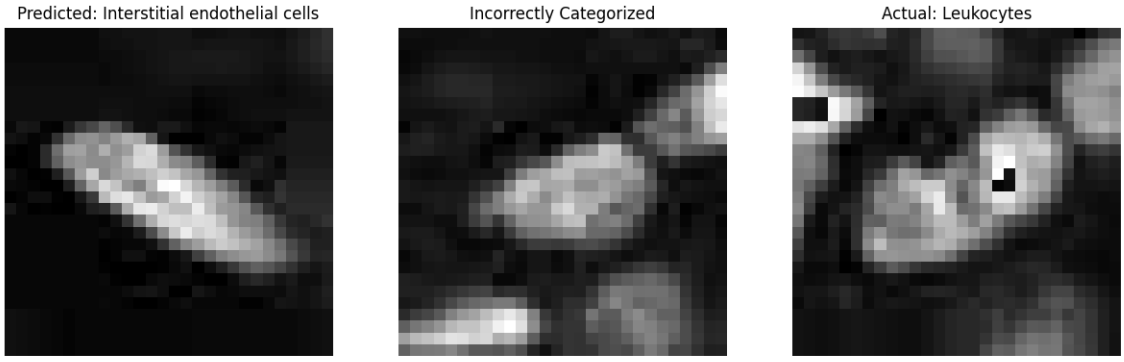

Figure 7: Misclassification example for category 4 (Leukocytes). Left: Image predicted as Interstitial Endothelial Cells. Middle: Incorrectly categorized image. Right: Actual class Leukocytes.

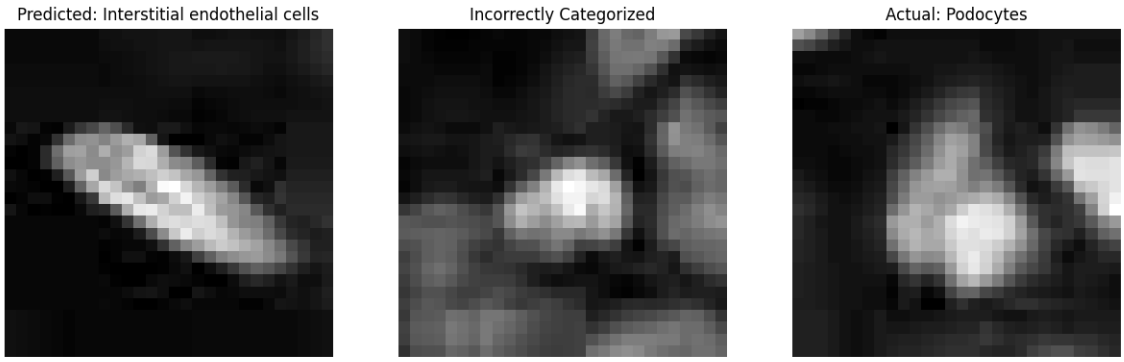

Figure 8: Misclassification example for category 5 (Podocytes). Left: Image predicted as Interstitial Endothelial Cells. Middle: Incorrectly categorized image. Right: Actual class Podocytes.

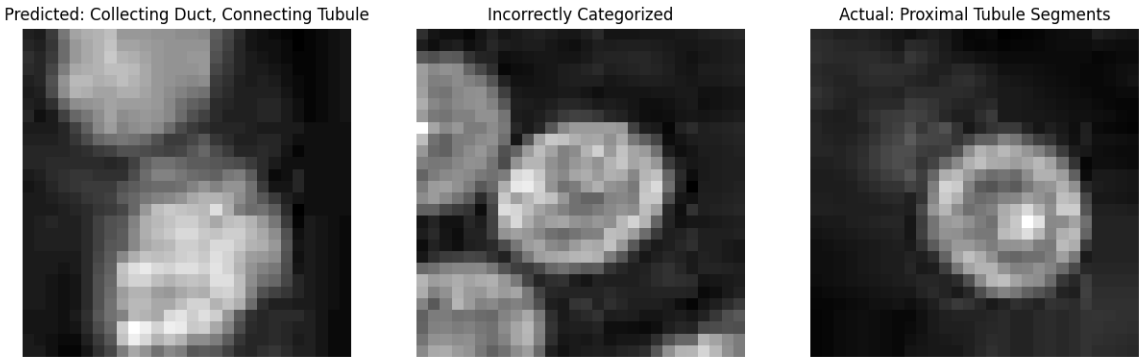

Figure 9: Misclassification example for category 6 (Proximal Tubule Segments). Left: Image predicted as Collecting Duct, Connecting Tubule. Middle: Incorrectly categorized image. Right: Actual class Proximal Tubule Segments.

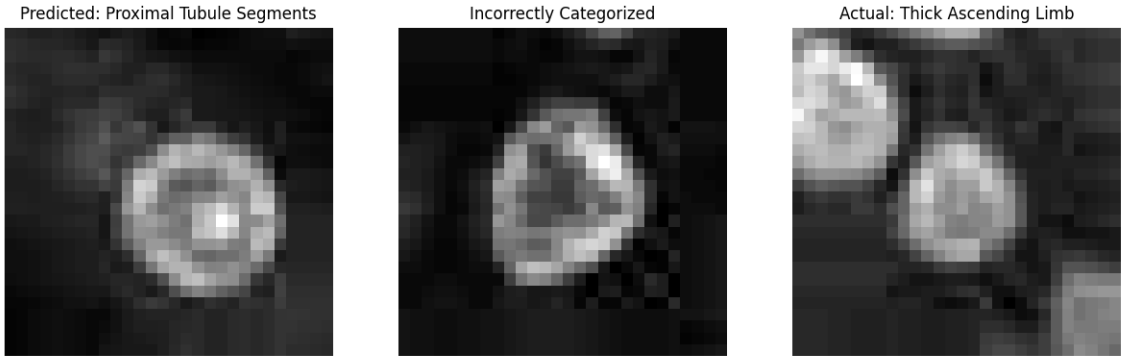

Figure 10: Misclassification example for category 7 (Thick Ascending Limb). Left: Image predicted as Proximal Tubule Segments. Middle: Incorrectly categorized image. Right: Actual class Thick Ascending Limb.

compared to the benchmarks provided by various implementations on MedMNIST, such as ResNet-18 and ResNet-50 with different image resolutions, auto-sklearn, AutoKeras, and Google AutoML Vision, the results indicate that there is room for improvement in the model's performance. [1]

Specifically, the benchmark for TissueMNIST using "AutoKeras" reported an accuracy of 70.3%, which is significantly higher than the current project's accuracy. However, the benchmarks utilized advanced techniques and possibly larger network architectures, which contributed to their higher accuracies.

Overall, the project served as a practical application of deep learning techniques in medical image analysis and provides a foundation for more sophisticated models and methods in the future.

References

[1] Jiancheng Yang et al. "MedMNIST v2 - A large-scale lightweight benchmark for 2D and 3D biomedical image classification". In: *Name of Journal (if available)* Volume Number (if available).Issue Number (if available) (Year of Publication). Shanghai Jiao Tong University; Boston College; RWTH Aachen University; Department of Endocrinology and Metabolism, Fudan Institute of Metabolic Diseases, Zhongshan Hospital, Fudan University; Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine; Harvard University, Page Range (if available). URL: `https://medmnist.com/`.