**Name(s)**

**Dustin Wu**

**Project Number**

**S0833**

**Project Title**

# A Novel Fuzzy Clustering-Based Convolutional Neural Network Ensemble for Histopathologic Cancer Metastasis Detection

**Abstract**

**Objectives**

The objective of the project is to develop deep learning algorithms that can efficiently and accurately detect cancer cells in digitized lymph node slide images.

**Methods**

Used laptop that accessed a GPU through an online server, a public dataset consisting of digitized images of lymph node histopathology slides with and without cancer, Python 3, and the Keras and Scikit-learn machine learning libraries to design the architecture of and train a convolutional neural network (CNN) to determine the presence of cancer in images. An ensemble (group) CNNs that cooperate to make more accurate predictions was created by partitioning the dataset into clusters using image feature extraction and a clustering algorithm, and then training a CNN on each cluster. Compared the use of non-overlapping clusters versus overlapping (fuzzy) clusters, and the use of handcrafted features versus features automatically selected by an autoencoder.

**Results**

The first CNN created contained three convolutional layers and had an area under the ROC curve (AUC) (a performance evaluation metric ranged from 0 to 1) of only 0.7432. Through the use of data augmentation, deeper architecture, more training cycles, and other improvements, the CNN model, now with six convolutional layers, was able to achieve an AUC of 0.9796. The model performance was improved by using a non-overlapping clustering ensemble method, and the fuzzy clustering ensemble method (FCM) boosted the performance even further. Lastly, the use automatic feature extraction did not significantly impact model performance. The final FCM model improved to an AUC of 0.9862 and overall accuracy of 95% on the reserved test dataset.

**Conclusions**

In this project, I incrementally developed a novel ensemble framework to more accurately detect cancer cells in digitized lymph node histopathology images. This work will not only be useful in helping pathologists detect cancer cells in histopathology images but can also be applied to other machine learning tasks.

**Summary Statement**

I developed a novel clustering-based deep learning framework that can efficiently and accurately detect cancer cells in digitized lymph node histopathology images.

**Help Received**

I independently utilized machine learning libraries to develop and optimize various models and algorithms. I received help related to machine learning from Dr. Christopher Hart and help with understanding cancer histopathology from Dr. Jeff Engelhardt, both at Ionis Pharmaceuticals, Inc.