Deep Learning for Classifying Subcellular Patterns of Proteins in Microscopic Images

Rohan Sikand rsikand@stanford.edu
Stanford University
CS 279: Computational Biology

1. Introduction

The Human Protein Atlas project aims to map every human protein found in cells, tissues, and organs using a variety of technologies such as antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems biology [8]. The human cell is composed of subcelluar structures. Molecules such as proteins localize within these structures. The structure and location at which these proteins localize plays a large role in cell function. Mislocalization can lead to things like disease. Thus, it is important to study and analyze the subcellular patterns of proteins in cells. To do this, one can use microscopy to obtain high resolution micrographs of the desired molecules in the subcellular regions.

Analyzing the images that are produced by microscopy can be a time consuming and inefficient task. While the method of microscopy contains powerful traits for analyzing subcellular molecules, a microscopy imaging session could generate well over thousands of images [4]. Thus, computational techniques are required to yield meaningful analysis and results. In recent years, deep learning methods have taken the world by storm with applications from self-driving cars to biological image analysis. Deep learning models are end-to-end meaning that the analysis of the images is completely automated with no human in the loop (i.e. no feature engineering required). Here, we apply deep learning techniques to automate the process of classifying subcellular patterns of proteins in microscopic images.

2. Background

The eukaryotic cell is composed of composed of subcellular compartments (e.g. organelles). The localization of proteins to these compartments enables cellular processes to take place. Examples of these cellular processes include signalling, growth, proliferation, programmed cell death, and mutations [1]. Studying these cellular processes can unlock insights into diseases caused by these processes. Specifically, it is the mislocalization of proteins that can cause diseases. Thus, it is important to analyze the location of proteins in various subcelluar compartments. With the advent of advanced technologies, researchers and scientists are able to develop methods to

determine the localization of proteins within subcellular structures [1]. One of these these spatial proteomic technologies includes fluorescence microscopy.

Imaging the cell allows us to determine the structural components of it. The structure of molecules (such as proteins) within the cell often play a crucial role in the functionality of said molecules. As a consequence, the functionality of these molecules affect the greater cellular and biological system. One method to image the cell is via light microscopy. In this method, a light is shined on the cellular sample placed underneath the microscope. The microscope itself then measures the light that is reflected. While light microscopy is a powerful technique, the resolution of the images it produces is not great enough for viewing and analyzing subcellular structures. Even further, it is often difficult to differentiate between structures at the subcellular level. On the other hand, fluorescence microscopy aims to make the image such that the subcellular molecules appear to glow and are thus more differentiable from one another. Technically, this is done by attaching a fluorescent molecule called a fluorophore to the molecule of interest. Then, when a light is shined on it, these molecules will emit a different wavelength [2]. Confocal microscopy is another technique which uses lasers in combination with fluorescence. The images used in this project were confocal microscopy images.

3. Methods

The goal of this project is to develop and test deep learning methods for classifying subcellular patterns of proteins in microscopic images. The first step in the experiment procedure was to gather the data. From here, preprocessing and data analysis was performed. Finally, the models were built and then trained and evaluated on the data.

3.1 Data

The data used in this experiment was obtained from "Human Protein Atlas Image Classification" (link here) [5]. The data set consisted of confocal microscopy images of cells. The goal is to predict protein organelle localization labels (i.e. subcellular patterns) from a microscopy image.

Due to computing constraints, the model was trained and evaluated on only a small fraction of the data set. The subset of the data used was then further split into a train/test split where 90% of the samples were allocated for training and the other 10% were used for testing.

Figures 1 and 2 show some visualizations and statistics of the data set. Figure 3 shows image samples from the data.

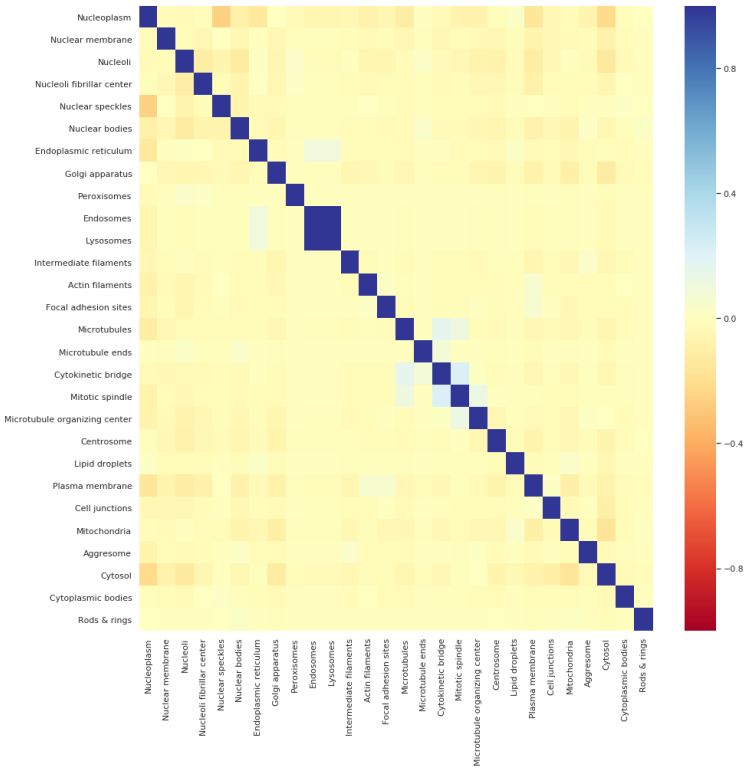

Figure 1: Heat map showing the correlation between samples of each class [5].

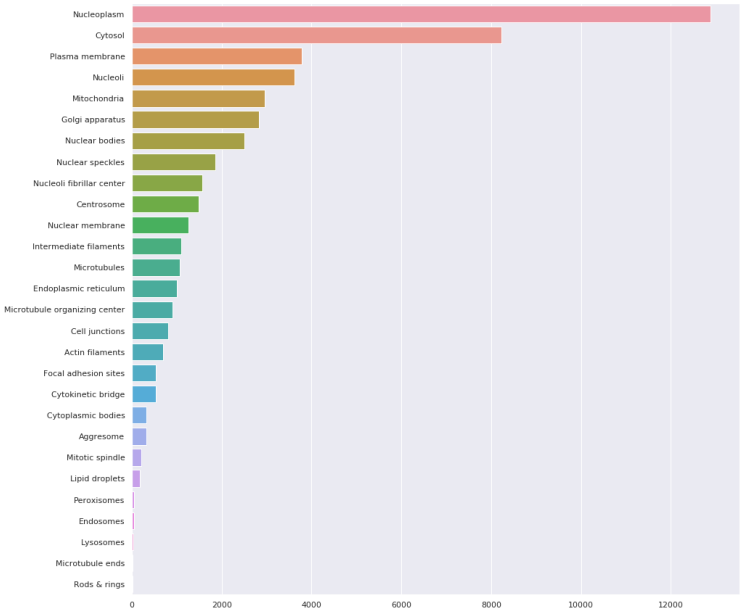

Figure 2: Distribution of the amount of samples per label. [5]

.

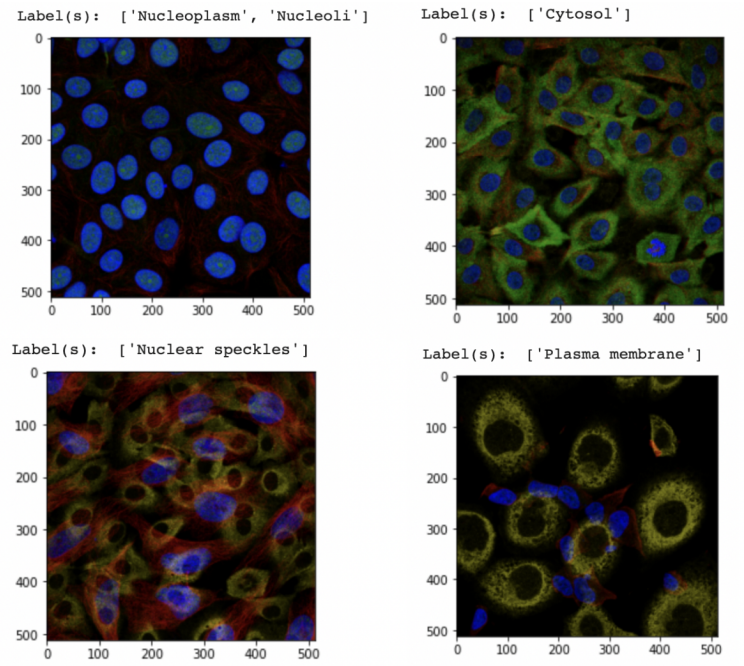

Figure 3: Samples from the data set.

3.2 Model and Training

We built and implemented a simple 6 layer fully-connected neural network (Multilayer Perceptron). The model layers are summarized in 3.2.

The model was trained for 250 epochs via backprogagation [6] using the Adam optimizer [3]. We used the binary cross-entropy loss function defined as

$$\text{loss} \ = -\frac{1}{N}\sum_{1}^{M} T_i \log\left(x_i\right)$$

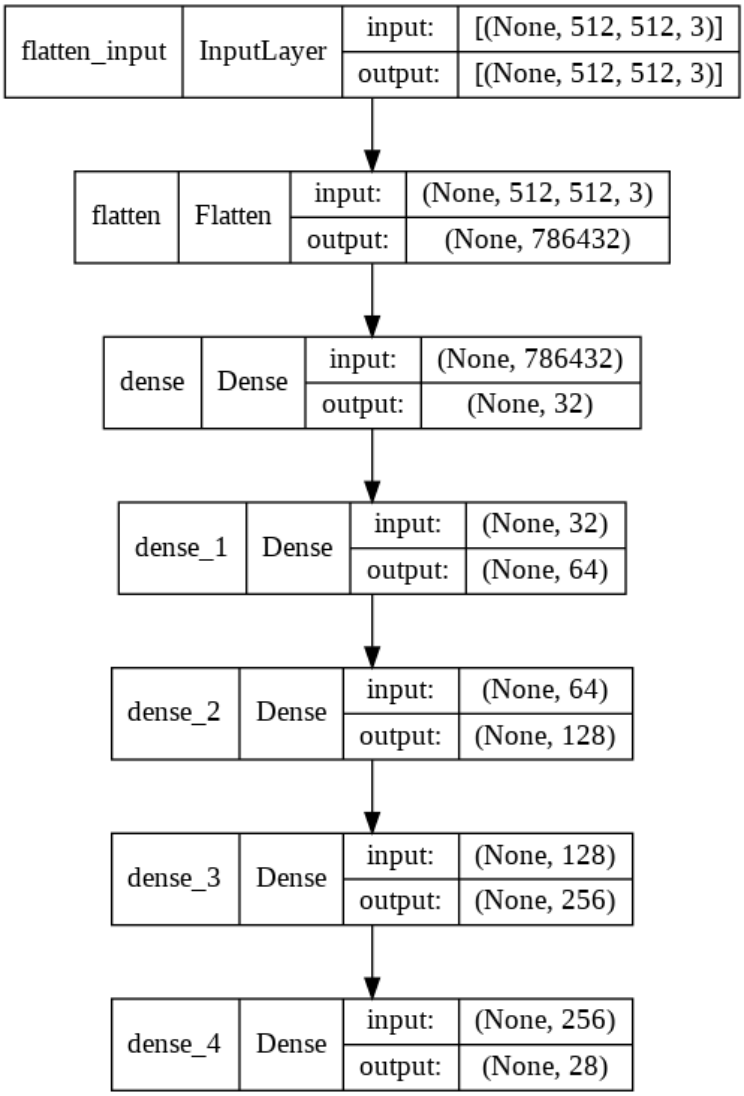

Figure 4: Diagram of the model's layers. The input is $512 \times 512 \times 3$ and the output is a 28-vector probability distribution where each index is the probability that the label represented by that index was predicted by the model as true for the given sample.

4. Results and Analysis

The training accuracy and training loss graphs are shown in figure 5.

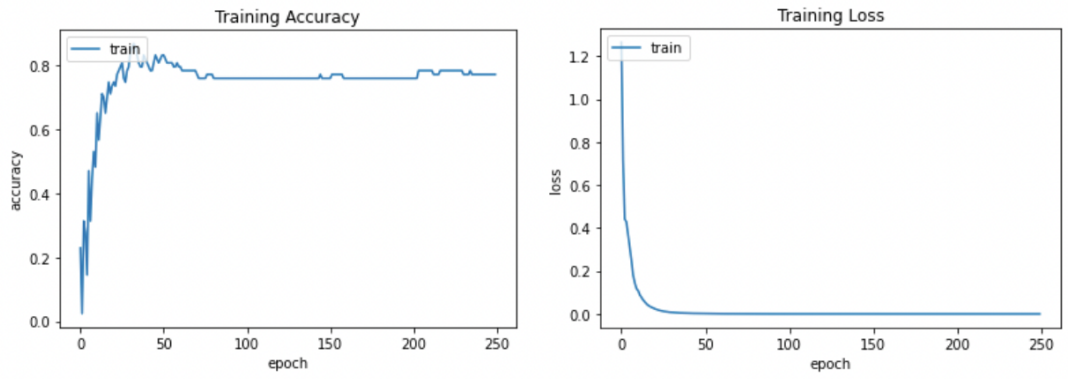

Figure 5: Plots of the training accuracy and training loss graphs.

From these graphs, it is clear that the model's accuracy is increasing while the loss is decreasing which means the model is learning and performing well on the training set. We will see how the model generalizes to new by predicting labels for samples on the test set. We show these results in section 4.1.

4.1 Evaluation

This deep learning problem is a multi-label problem meaning that each sample that is to be predicted on can be classified into one or more classes. Gauging how well a multi-label classifier performs is not as straightforward as in single-label classification problems. Thus, it is helpful to calculate several different metrics from which we can then analyze to gauge model performance. Specifically, it is difficult because some labels may have been predicted correctly while others are predicted incorrectly. To account for this partiality, we use several metrics developed for the multi-label classification problem (mostly ones defined in [7]). To adapt these metrics to the multi-label problem, we used the micro averaged versions. Specifically, we evaluate each of the labels separately and then average the results over the all of the labels. These metrics are defined as follows [7]:

$$Exact\ Match\ Accuracy = \frac{1}{n}\sum_{i=1}^{n} I\left(Y_i = Z_i\right)$$

$$Proportional\ Accuracy = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}$$

$$\text{Precision, } P_{\text{micro}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j Z_i^j}{\sum_{j=1}^{k}\sum_{i=1}^{n} Z_i^j}$$

$$\text{Recall , } R_{\text{micro}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j Z_i^j}{\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j}$$

and

$$F_{1-\text{ micro}} = \frac{2\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j Z_i^j}{\sum_{j=1}^{k}\sum_{i=1}^{n} Y_i^j + \sum_{j=1}^{k}\sum_{i=1}^{n} Z_i^j}$$

Y_i are the ground truth labels and Z_i are the predictions. Mathematically, this is defined as

$$Y_i^{\lambda} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ actually belongs to class } \lambda \\ 0 & \text{otherwise} \end{cases}$$

and

$$Z_i^{\lambda} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ is predicted to be belong to to class } \lambda \\ 0 & \text{otherwise} \end{cases}.$$

To calculate these metrics, we used the trained model to predict labels for samples in the test set. From here, these above metrics were calculated. The results of these calculations are shown in table 1.

Metric	Value
Proportional Accuracy	94.0%
Exact Match Accuracy	22.2%
Precision	0.33
Recall	0.15
F1 score	0.21

Table 1: Metric calculations

4.2 Analysis

Unfortunately, the model did not perform as well as intended and it is clear that the model has overfitted to a certain degree (since the training graphs showed positive results). Multi-label classification is a difficult problem since the model is predicting whether a particular label is present rather than a singular class of an image. This

difficulty is in part why the model performed poorly for certain scores. However, the model did actually perform quite well for proportional accuracy which is likely due to the fact that the proportional accuracy score is a forgiving metric by allowing the model to incorrectly classify some labels and not be penalized as much as in exact match accuracy.

The most probable reason for why the model did not perform well overall was because only a small portion of the data was being used for training. This was due to computational constraints. Another reason why the model may not have generalized to the test data is the large class imbalance (figure illuminates this visually). For example, nearly 40% of samples had the "nucleoplasm" label while less than 1% had the "rods & rings" label. Additionally, the neural network we used is relatively simple compared to the modern networks out today. Again, the reasoning for not using modern neural network architectures was strictly due to computational constraints. Future work would include training with more data, training for longer, and using advanced neural network architectures.

References

[1] Mathias Dreger. "Subcellular proteomics". In: *Mass spectrometry reviews* 22.1 (2003), pp. 27–56.

[2] Ron Dror. *CS 279 lecture 10 - Microscopy.*

[3] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv preprint arXiv:1412.6980* (2014).

[4] Romain F Laine et al. "Avoiding a replication crisis in deep-learning-based bioimage analysis". In: *Nature methods* 18.10 (2021), pp. 1136–1144.

[5] Wei Ouyang et al. "Analysis of the human protein atlas image classification competition". In: *Nature methods* 16.12 (2019), pp. 1254–1261.

[6] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. "Learning representations by back-propagating errors". In: *nature* 323.6088 (1986), pp. 533–536.

[7] Mohammad S Sorower. "A literature survey on algorithms for multi-label learning". In: (2010).

[8] Mathias Uhlén et al. "A human protein atlas for normal and cancer tissues based on antibody proteomics". In: *Molecular & cellular proteomics* 4.12 (2005), pp. 1920–1932.