Machine Learning for Automated Classification of Corneal Epithelium Tissue Simulations

Namith Manjunath Telkar [1] Shreeja Deshpande [2]

[1]ntelkar@iu.edu [2]ssdeshpa@iu.edu

Motivation

The corneal epithelium is the outmost layer of cells covering the cornea's surface, acting as a protective barrier and contributing to corneal transparency.

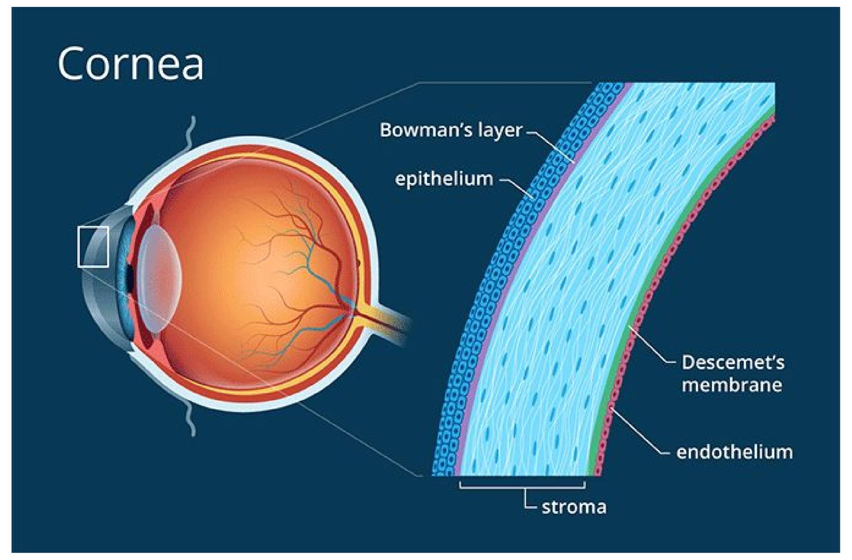


Figure 1. Image of the Cornea

The goal of the project is to perform unsupervised classification of corneal epithelium tissue images into homeostatic and non-homeostatic states, considering cell count and structural integrity. The motivation behind the project is to automate the classification process to streamline the analysis of simulation outcomes, reducing manual labor and subjectivity.

Data: CompuCell3D Generated Images

Images are generated using CompuCell3D (CC3D) simulations. CC3D simulates corneal epithelium based on diverse initial parameters. Tissue Homeostasis can be defined as the equilibrium state where cell types (BASAL, WING, STEM, SUPERFICIAL) maintain count. The images are classified as abnormal based on the following features:

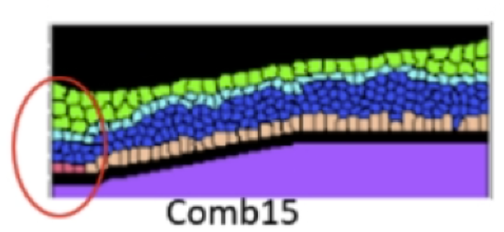


Figure 2. Stem Underproduction

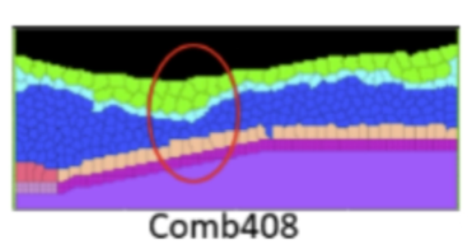


Figure 3. Middle Drop

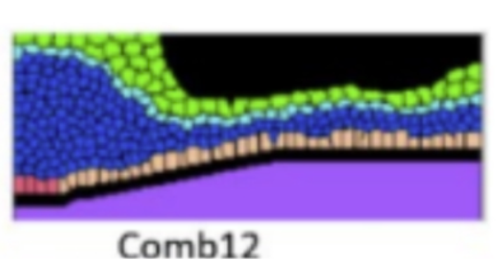


Figure 4. Stem Overproduction

Methodology

The methodology involves the following steps:

1. Train a U-Net model with the dataset of images paired with their corresponding masks.
2. Employ the trained U-Net model for predicting masks and segmenting images into binary masks to identify regions of interest.
3. Extract features related to width and continuity from the segmented images, focusing on consistent layer widths and the absence of breaks in layers.
4. Utilize the extracted features as the foundation for unsupervised classification via K-Means clustering.
5. Perform K-Means clustering to partition images into two clusters, representing normal and abnormal classes.

UNet Model for Segmentation

UNet is a popular medical image segmentation model which we will be using to learn the different layers of the image. UNet consists of encoder and decoder networks to capture context and precise localization. After learning the layers, we used KMeans clustering to cluster the different types of Corneal Epithelium images.

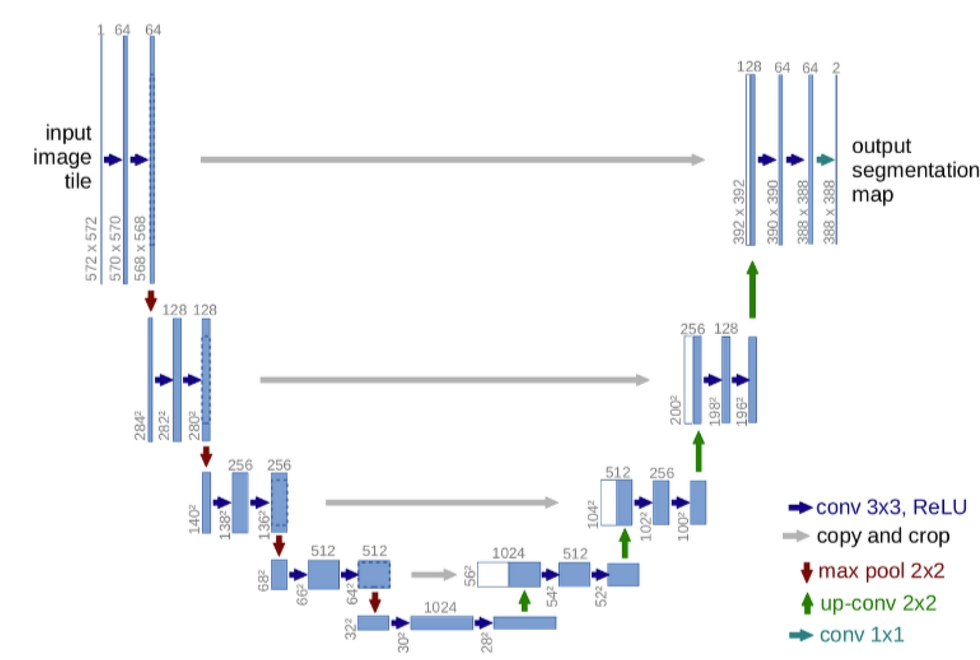


Figure 5. UNet Architecture

Experimentation

- With a dataset comprising around 2000 images, experimentation was conducted to explore various methods for feature extraction and unsupervised classification.
- One avenue of exploration involved preprocessing images with normalization and flattening, followed by training an **Autoencoder** to extract encoded representations followed by subsequent KMeans clustering-based unsupervised classification.
- **Generative Adversarial Networks (GANs)** were also considered in the initial experimentation. However, due to feature diversity constraints, the GAN approach was not pursued further in this particular study.
- Experimentation with **U-Nets** involved leveraging the dataset to construct a U-Net architecture for image segmentation, where preprocessed images were trained and encoded representations were clustered using KMeans for unsupervised classification, enhancing insights into feature extraction and classification efficacy.

Analysis and Results

Image Segmentation using UNets

The trained U-Net model achieved an impressive accuracy of approximately 98%. After loading the weights of the trained model, evaluation on the test set yielded a high accuracy score of 97.99%.

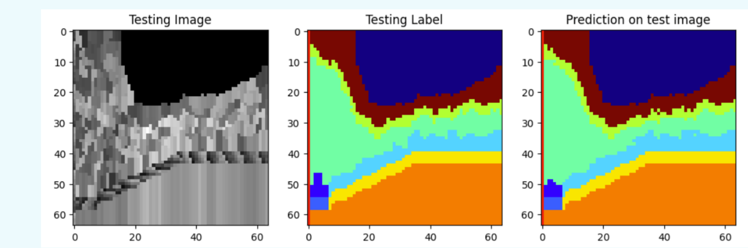
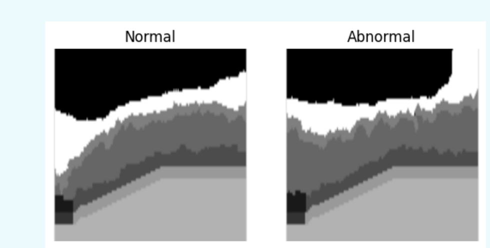


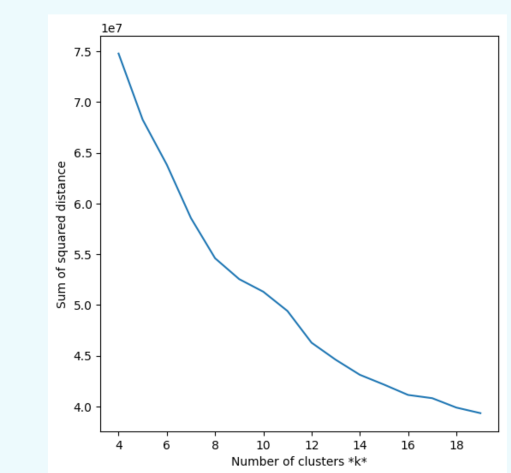
Figure 6. Predicted Masks from UNet

KMeans Clustering

After performing KMeans clustering, images were categorized into distinct clusters representing abnormal and normal classes, illuminating inherent patterns in the dataset. On further experimentation to explore different classification schemes, analysis of the features suggested an optimal number of clusters around 8, providing valuable insights into the distribution and characteristics of epithelium images.



(a) Unsupervised Classification of Images



(b) Optimal clusters

Future Work

- In future work, an exploration into identifying additional features beyond stem underproduction, middle drop, and stem overproduction is warranted to enhance the model's ability to distinguish between normal and abnormal images effectively. This expanded feature set could potentially include aspects such as texture, shape, or contextual information, which may further refine the classification accuracy.
- Additionally, considering the structural characteristics present in the epithelium images, such as variations in gender or age, offers an intriguing avenue for prediction into different classes. By integrating these factors into the model's framework, the aim can be to achieve a more nuanced understanding and classification of the images, facilitating deeper insights into the underlying patterns and characteristics of the corneal epithelium.

References

Sadanandan, S. K., Ranefall, P., Le Guyader, S., & Wählby, C. (2017). Automated Training of Deep Convolutional Neural Networks for Cell Segmentation. Scientific Reports, 7(1), 1-7. https://doi.org/10.1038/s41598-017-07599-6

Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. Nat Methods. 2019 Dec;16(12):1233-1246. doi: 10.1038/s41592-019-0403-1. Epub 2019 May 27. PMID: 31133758; PMCID: PMC8759575.

Gu Y, Zhang AC, Han Y, Li J, Chen C, Lo YH. Machine Learning Based Real-Time Image-Guided Cell Sorting and Classification. Cytometry A. 2019 May;95(5):499-509. doi: 10.1002/cyto.a.23764. Epub 2019 Apr 8. PMID: 30958640; PMCID: PMC6520141