Machine Learning for Automated Classification of Corneal Epithelium Tissue Simulations

Namith Manjunath Telkar
Indiana University Bloomington
Bloomington, Indiana
ntelkar@iu.edu

Shreeja Deshpande
Indiana University Bloomington
Bloomington, Indiana
ssdeshpa@iu.edu

Abstract

The corneal epithelium, as the outermost layer of the cornea, plays a critical role in maintaining corneal transparency and providing a protective barrier. In this paper, we present a novel approach for automating the classification of corneal epithelium tissue images into homeostatic and non-homeostatic states, based on cell count and structural integrity. Using images generated by CompuCell3D simulations, we train a U-Net model for image segmentation and extract features related to width and continuity from segmented images. These features serve as the basis for unsupervised classification using K-Means clustering. Experimentation with various methods, including Autoencoders and U-Nets, demonstrates promising results, with the trained U-Net model achieving an accuracy of approximately 98% in image segmentation with a mean IOU of 94.88%. K-Means clustering reveals inherent patterns in the dataset and suggests an optimal number of clusters for effective classification. This research contributes to streamlining the analysis of corneal epithelium images, reducing manual labor, and providing valuable insights into tissue homeostasis.

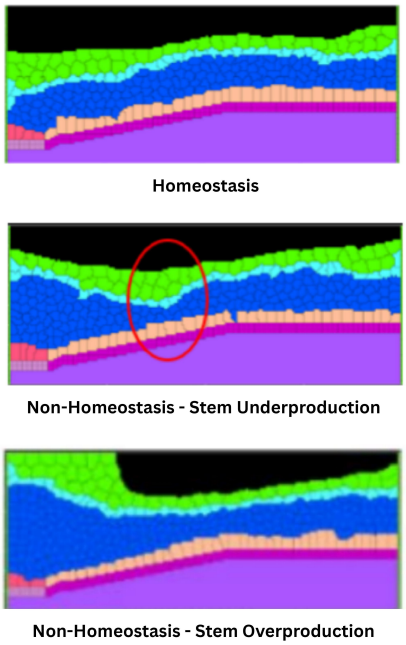

Figure 1. Image showing different states of the corneal epithelium simulations with one in Homeostatic state and two in Non-Homeostatic state.

1. Introduction

The corneal epithelium, as the outermost layer of cells covering the surface of the cornea, plays a pivotal role in maintaining ocular health by serving as a protective barrier and contributing to corneal transparency. Understanding the dynamic states of corneal epithelium tissue is essential for assessing ocular health and diagnosing various ocular conditions. However, manual classification of these tissue states based on cell count and structural integrity is labor-intensive and subjective.

To address this challenge, we propose an automated approach for classifying different corneal epithelium tissue states using machine learning techniques. By analyzing images generated by CompuCell3D (CC3D) simulations, which simulate corneal epithelium based on diverse initial parameters, we aim to classify tissue states into homeostatic and non-homeostatic categories. Tissue homeostasis, defined as the equilibrium state where cell types (BASAL, WING, STEM, SUPERFICIAL) maintain initial counts, serves as a crucial reference point for classification.

Preserving the structural integrity of tissue shape is imperative for maintaining ocular health, making it a key consideration in our classification methodology. Our goal is to perform unsupervised classification of corneal epithelium

tissue images, considering both cell count and structural integrity, to automate the classification process. By streamlining the analysis of simulation outcomes, our approach aims to reduce manual labor and subjectivity associated with traditional classification methods.

In this paper, we present our methodology for automating the classification of corneal epithelium tissue states, leveraging machine learning techniques and image analysis. We discuss the previous related work and outline our approach for unsupervised classification. Furthermore, we highlight the significance of our research in advancing ocular health assessment and diagnosis.

2. Background and Related Work

The field of automated cell segmentation and classification has undergone significant advancements driven by the integration of machine learning techniques with biological image analysis.

Sadanandan et al. (2017) [4] introduced an automated training framework for deep convolutional neural networks (CNNs) tailored specifically for cell segmentation tasks. By leveraging large-scale datasets and novel training strategies, their approach demonstrated remarkable accuracy in segmenting cells from microscopy images. This work highlighted the potential of deep learning models in automating labor-intensive tasks such as cell segmentation, paving the way for more efficient and accurate analysis pipelines.

Vasan et al. (2020) [5] delved into the applications and challenges of machine learning in enabling realistic cellular simulations. Their study emphasized the importance of incorporating machine learning algorithms into computational models to capture the complexity of cellular behavior accurately. By integrating machine learning with cellular simulations, researchers can gain deeper insights into biological processes and phenomena, ultimately advancing our understanding of cellular dynamics.

In the realm of ocular health assessment, Wang et al. (2023) [6] developed an automated evaluation system for corneal sodium fluorescein staining (CSFS) using deep learning techniques. Their system enabled accurate and efficient assessment of corneal health based on CSFS images, demonstrating the potential of deep learning models in clinical applications for diagnosing ocular conditions. This work exemplifies the translation of machine learning advancements into practical tools for improving patient care in ophthalmology.

Moen et al. (2019) [2] conducted a comprehensive review of deep learning applications in cellular image analysis. Their work highlighted the versatility of deep learning methods across various microscopy modalities and biological contexts, ranging from single-cell analysis to tissue imaging. By leveraging deep learning algorithms, researchers can extract rich information from complex biological images, facilitating advances in biomedical research and drug discovery.

Gu et al. (2019) [1] proposed a machine learning-based system for real-time image-guided cell sorting and classification. By integrating machine learning algorithms with imaging techniques, their system enabled rapid and accurate cell sorting based on predefined criteria, opening avenues for high-throughput cellular analysis and manipulation.

Collectively, these studies underscore the transformative potential of machine learning in automating cellular image analysis tasks, including segmentation, classification, and evaluation. Our research builds upon these foundations, aiming to automate the classification of corneal epithelium tissue states using machine learning techniques, thereby advancing the field of ocular health assessment and diagnosis.

3. Proposed Methodology

Our proposed methodology aims to automate the classification of corneal epithelium tissue states by leveraging machine learning techniques, specifically deep learning for image segmentation and clustering algorithms for unsupervised classification. In this section, we outline our approach, starting with the dataset used for training and evaluation, followed by a detailed description of the U-Net [3] architecture for image segmentation and K-Means clustering for classification.

3.1. Dataset

The dataset utilized in our study comprises 2000 images and corresponding masks generated by CompuCell3D (CC3D) simulations. These images simulate corneal epithelium based on diverse initial parameters, capturing variations in tissue states such as homeostasis and non-homeostasis. Each image is paired with a mask delineating regions of interest, facilitating supervised training for image segmentation tasks as seen in Figure 2.

3.2. Image Segmentation

The segmentation of corneal epithelium tissue images is a critical step in our methodology, enabling the extraction of meaningful features for subsequent classification tasks. To accomplish this, we employed the U-Net architecture, a well-established convolutional neural network (CNN) specifically tailored for biomedical image segmentation. The U-Net architecture consists of an encoder-decoder network structure, allowing it to capture both global context and precise localization information. In the encoder part, convolutional layers extract high-level features from the input image, while in the decoder part, upsampling layers reconstruct segmented masks with pixel-level accuracy. Additionally, skip connections between corresponding encoder and decoder layers facilitate the fusion of low-level and

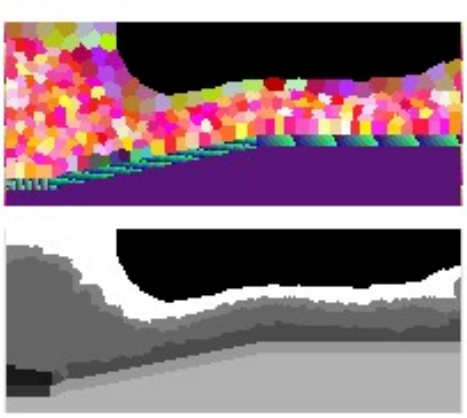

Figure 2. An example of an input image and its corresponding mask

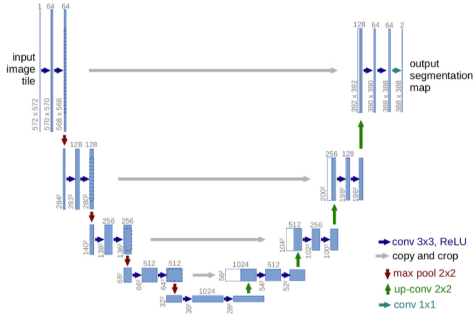

Figure 3. Architecture of the U-Net model

high-level features, enhancing segmentation performance. The U-Net architecture can be observed in figure 2.

In adapting the U-Net model for our dataset, which entails multi-class labeling with ten distinct classes representing different cell types within the corneal epithelium, several modifications were implemented. Firstly, the output layer of the U-Net architecture was adjusted to accommodate ten channels, each corresponding to a separate class label. This modification allowed the model to predict the presence of each cell type at the pixel level, facilitating fine-grained segmentation. Moreover, to optimize model performance for multi-class segmentation, we utilized the categorical cross-entropy loss function during training. This loss function enabled the network to effectively differentiate between the various cell types present in the masks, guiding the segmentation process towards accurate delineation of cell boundaries.

Through these adaptations and training procedures, the U-Net model was successfully trained to segment corneal epithelium tissue images into ten distinct classes, laying the groundwork for subsequent classification tasks. This segmentation capability is essential for extracting relevant features related to cell count and structural integrity, facilitating the automated classification of tissue states.

3.3. Unsupervised Clustering

After segmenting corneal epithelium tissue images using the trained U-Net models, the next step in our methodology involves unsupervised clustering of the extracted features to classify tissue states into homeostatic and non-homeostatic categories. To accomplish this, we utilized K-Means clustering, a popular algorithm for partitioning data into distinct clusters based on similarity.

Firstly, features related to tissue structure and morphology were extracted from the segmented images. These features serve as the foundation for unsupervised classification, capturing essential characteristics such as layer widths and the absence of breaks in layers. The features extracted from the trained U-Net models encapsulate relevant information about cell distribution and tissue organization, providing valuable insights into tissue states. Subsequently, the extracted features were used as input for K-Means clustering. K-Means clustering iteratively partitions the feature space into a predefined number of clusters, with each cluster representing a distinct tissue state. By optimizing cluster centroids to minimize intra-cluster variance, K-Means clustering identifies inherent patterns in the dataset, enabling automated classification of corneal epithelium tissue states based on cell count and structural integrity.

The choice of the number of clusters is critical in K-Means clustering and requires careful consideration. In our approach, we explored various classification schemes and analyzed the distribution of features to determine an optimal number of clusters. This process involved assessing the clustering performance across different numbers of clusters and selecting the configuration that best captured the underlying structure of the data. Through unsupervised clustering, corneal epithelium tissue images were categorized into distinct clusters representing homeostatic and non-homeostatic states. Overall, unsupervised clustering complements the segmentation capabilities of the U-Net models, facilitating a comprehensive understanding of corneal epithelium dynamics.

4. Results

Our methodology encompassed two distinct approaches for automating the classification of corneal epithelium tissue states. In the first approach, we considered a binary classification scheme, where tissue states were categorized into normal and abnormal classes. Following the segmentation of tissue images using the U-Net models, we achieved a remarkable test accuracy of 97.99% and a mean intersection over union (IoU) score of 94.89%, indicating precise delineation of cell boundaries. An example test image and the prediction can observed in Figure 4. Subsequently, for the

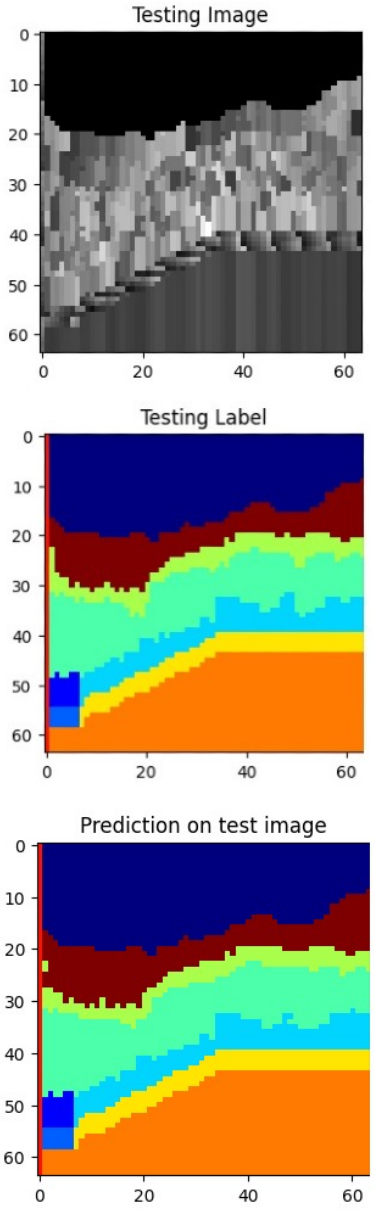

Figure 4. Figure showing the input test image and testing label and the corresponding prediction from the U-Net model

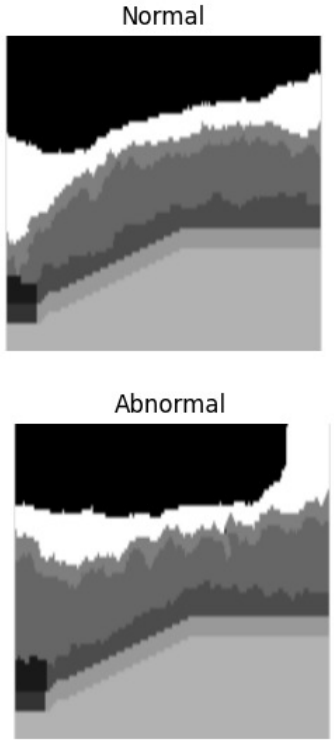

Figure 5. Figure representing sample images belonging to Normal and Abnormal states

binary classification, we utilized K-Means clustering to partition segmented images into two clusters, representing normal and abnormal tissue states which represented Homeostatic and Non-Homeostatic states which can be observed in Figure 5.

We also explored a second approach, wherein we employed the elbow method to determine the optimal number of clusters for classifying tissue states. This method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and selecting the point where the rate of decrease in WCSS slows down. From the elbow method graph in Figure 6, we observed that the rate of decrease in WCSS stabilized at around 8 clusters, suggesting that 8 clusters could be an appropriate choice. However, upon closer examination of the clustering results, we found that using 7 clusters led to better differentiation between tissue states. While the elbow method indicated 8 clusters as the optimal choice, further analysis revealed that 7 clusters provided more distinct separation between homeostatic and non-homeostatic tissue states. This discrepancy underscores the importance of evaluating clustering results beyond solely relying on the elbow method and considering the interpretability of the clusters in the context of the data.

Overall, our results demonstrate the efficacy of our methodology in automating the classification of corneal epithelium tissue states. Our methodology, integrating the precision of U-Net model segmentation with refined clustering techniques, streamlines the analysis of simulation results. This approach minimizes manual effort and bias, offering significant insights into tissue homeostasis and ocular health.

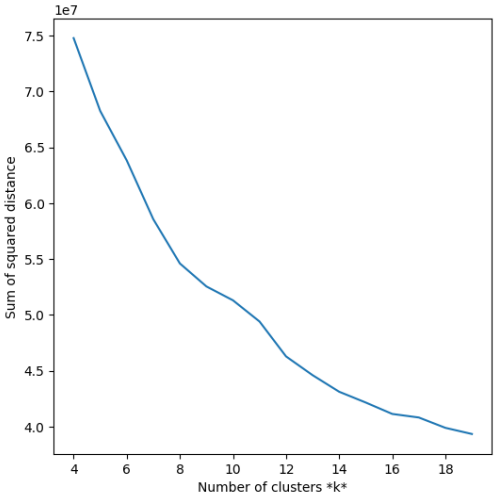

Figure 6. Graph showing the WCSS plot against the number of clusters

5. Discussion

Our methodology effectively addresses the research question by automating the classification of corneal epithelium tissue states. Through the integration of deep learning techniques with unsupervised clustering, our approach provides a streamlined framework for analyzing simulation outcomes and extracting insights into tissue dynamics. While our methodology demonstrates promising results, it is important to acknowledge certain limitations. Firstly, the reliance on simulated images may restrict the generalizability of our findings to real-world clinical scenarios. Additionally, the initial binary classification scheme may oversimplify the complexity of tissue states, warranting further exploration of multi-class classification approaches. Moreover, the interpretation of clustering results requires careful consideration, as determining the optimal number of clusters can be subjective and may vary depending on the dataset characteristics.

6. Conclusion

Our project significantly contributes to the advancement of automated classification techniques in biomedical image analysis, particularly in the context of corneal epithelium tissue assessment. By leveraging deep learning and unsupervised clustering, we have developed a robust methodology that enhances efficiency and accuracy in tissue state classification, offering valuable insights into ocular health. The potential impact of our work extends beyond academic research, with implications for clinical practice and patient care. Moving forward, future efforts could focus on refining the methodology, exploring additional features, and integrating real-world clinical data to further improve diagnostic and treatment strategies in the field of ophthalmology.

References

[1] Y. Gu, A. C. Zhang, Y. Han, J. Li, C. Chen, and Y. H. Lo. Machine learning based real-time image-guided cell sorting and classification. *Cytometry Part A*, 95(5):499–509, 2019. 2

[2] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen. Deep learning for cellular image analysis. *Nature Methods*, 16(12):1233–1246, 2019. 2

[3] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 9351:234–241, 2015. 2

[4] S. K. Sadanandan, P. Ranefall, S. Le Guyader, and C. Wählby. Automated training of deep convolutional neural networks for cell segmentation. *Scientific Reports*, 7(1):1–7, 2017. 2

[5] R. Vasan, M. P. Rowan, C. T. Lee, G. R. Johnson, P. Rangamani, and M. Holst. Applications and challenges of machine learning to enable realistic cellular simulations. *Frontiers in Physics*, 7:247, 2020. 2

[6] S. Wang, J. He, X. He, Y. Liu, X. Lin, C. Xu, L. Zhu, J. Kang, Y. Wang, Y. Li, S. Guo, Y. Zhang, Z. Luo, and Z. Liu. Aescsfs: an automatic evaluation system for corneal sodium fluorescein staining based on deep learning. *Therapeutic Advances in Chronic Disease*, 14:20406223221148266, 2023. 2

7. Github Link

https://github.iu.edu/cs-b657-sp2024/corneal_epithelium_project