PATCH-BASED CLASSIFICATION FOR A SPARSE DATASET OF CORNEAL CONFOCAL MICROSCOPY IMAGES

Zane Z. Zemborain, MSc[1]

[1]Department of Biomedical Engineering, Duke University, Durham, NC

ABSTRACT

Dry Eye Disease (DED) and Neuropathic Corneal Pain (NCP) present as similar conditions but have different etiologies. Despite similar signs and symptoms, NCP patients do not respond well to DED therapies and vice versa. As a result, there is a critical need to find an objective measure to accurately differentiate between these conditions. Corneal Confocal Microscopy (CCM) images of the subbasal nerve plexus may facilitate this goal as both conditions involve disruptions to the corneal nerves. Further, images can be operated upon by an automated computer vision technique, such as Convolutional Neural Networks (CNN), which may find more success than clinicians in performing image classification. In this paper, I utilize an autoencoder to identify and cluster discriminative features within patches of CCM images and then I identify and quantify those features via an AlexNet classification architecture. Finally, I perform various classification operations on the resulting feature histograms in order to obtain a disease classification of DED or NCP. Unfortunately, the final model was unable to perform much better than random chance (40-60% accuracy) in terms of disease state classification.

***Index Terms*—**DED, NCP, CCM, CNN

1. INTRODUCTION/RELATED WORK

Dry Eye Disease (DED) is a common condition caused by a chronic lack of sufficient lubrication and moisture on the surface of the cornea which results in irritation [1]. Neuropathic Corneal Pain (NCP), by contrast, is an ill-defined disease in which the corneal nerve become irritated via an unknown cause [2]. Both conditions are difficult to diagnose and require detailed work ups via equipment that is not widely available [3]. Unfortunately, these conditions are often misdiagnosed for one another and have different associated therapies [3].

Corneal confocal microscopy, however, may allow for an objective method of distinguishing between these two conditions. It has proved moderately successful in tracking the progression of systemic neuropathies, such a diabetic retinopathy, which also affects the corneal nerves [4].

Given a sparse dataset of DED and NCP patients, a patch-based classification of the CCM images may be appropriate. Features, such as micro-neuromas, have been documented to be more frequent in NCP eyes [5], and features are more abundant than eyes. In one study, a group used a patch-based CNN to classify whole slide tissue images because (a) the images were far too large to efficiently pass through a CNN and (b) cellular-level visual features were discriminative [6]. I will follow a similar workflow (using different methods), as performed in the aforementioned paper, in performing this experiment.

2. METHODS

The Institutional Review Board of Duke University prospectively approved this retrospective, observational, cross-sectional study. The approval included the collection, deidentification, exporting and analysis of CCM scans and other ophthalmic-related records. It followed the tenets of the Declaration of Helsinki and the Health Insurance Portability and Accountability Act. Written informed consent was obtained from all patients.

2.1. Participants

The study included 69 eyes (45 DED; 24 NCP) from 43 consecutive patients (28 DED; 15 NCP), referred by a corneal specialist (V.P.) for CCM scans. We excluded eyes which lacked scans of the corneal vortex and/or had scans of exceptionally poor quality.

2.2. Corneal Confocal Microscopy (CCM)

All eyes were scanned on the HRT3 Rostock Cornea Module (Heidelberg Engineering GmbH, Heidelberg, Germany). As part of our imaging protocol, multiple volume and sequence scans of the subbasal nerve plexus were obtained at and around the corneal vortex.

2.3. Reference standard

A corneal specialist (Dr. Victor Perez), involved in the patient's care, judged whether each eye had dry eye disease (DED) or neuropathic corneal pain (NCP) using all of the available information (e.g. family history, patient history, repeat visits, Schirmer's test, TearLab osmolarity, tear breakup time, epithelial staining).

2.4. Physical Layers

The images were preprocessed with a homomorphic filter in order to reduce the illumination component and emphasize the reflection component of the images. The features of interest such as nerves and micro-neuromas have a higher reflectivity than the subbasal nerve plexus, but the illumination of the images was sometimes poor. This should help the subsequent networks train on these features.

Three image datasets were evaluated corresponding to 3 different size of field-of-view (FOV): (1) one CCM image of the corneal vortex for each eye and (2) one CCM image of the corneal vortex for each eye plus one unique, randomly-selected CCM images of the corneal nerves for each eye. (3) one CCM image of the corneal vortex for each eye plus two unique, randomly-selected CCM images of the corneal nerves for each eye.

2.5. Image Dataset Cycle

The dataset was very sparse. As such, only two patients (1 DED; 1 NCP) were randomly selected at the beginning of the workflow and their eyes were excluded until the final classification stage. This was repeated 5x for each dataset.

2.6. Patch extraction

Patches were iteratively extracted (row-wise and column-wise) from each image via overlapping windows of varying size (one-sixteenth, one-thirty-sixth, and one-sixty-fourth of the CCM image area). Consecutive patches (row-wise or column-wise) had 50% overlap (Fig. 1). Regardless of the initial size (at extraction), all patches were resized via bilinear interpolation to a standardized size in order to facilitate subsequent processing of the image patches.

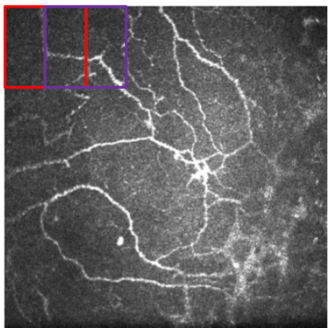

Figure 1: *A Corneal Confocal Microscopy (CCM) image of an eye with neuropathic corneal pain (NCP). The red and purple boxes correspond to consecutive patch windows at one-sixteenth scaling.*

2.7. Autoencoder

An autoencoder was utilized in order to perform a data-specific compression of the image patches to a 20-dimensional embedding space (Fig. 2). The compression and decompression functions were implemented via a convolutional neural network and trained on an image patch training set over 10 epochs.

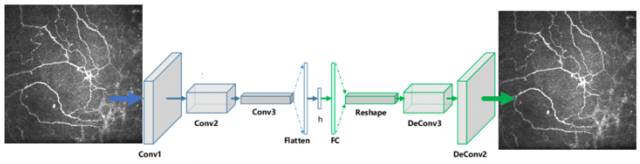

Figure 2: *A simplified diagram of the autoencoder architecture used to learn the embedding space of the corneal confocal microscopy (CCM) images.*

2.8. Discriminative patch selection

The elbow method was implemented in order to find the optimal k (and avoid overfitting) for k-means clustering of the 20-dimensional embeddings space (Fig. 3). Using the optimal k, the patches were partitioned into k_1 clusters. Each cluster contained patches with minimal within-cluster variances and effectively corresponded to common features within the CCM images. Once partitioned, the clusters with the greatest (weighted) disparity in occurrence between the DED and NCP training datasets were selected as the discriminative group of features. Subsequently, the elbow method was reimplemented on these "discriminative" clusters and their corresponding patches were re-partitioned into k_2 clusters. It is important to separate and identify these "discriminative clusters", as opposed to all of the clusters, in order to reduce the total number of classes for the AlexNet classification architecture, as well as to optimize it for these specific clusters.

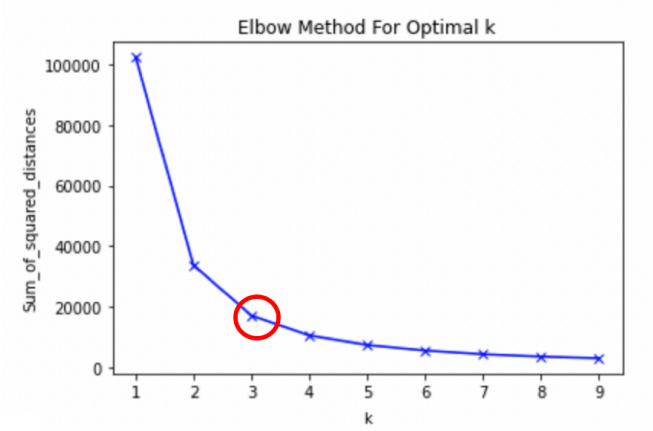

Figure 3: *The Elbow Method for finding the optimal K clusters in the CCM image patch dataset. The red circle corresponds to the elbow of the curve (or optimal k = 3) in the heuristic.*

2.9. AlexNet patch classification

An AlexNet [7] convolutional neural network architecture was designed to classify the patches into n number of classes, where n is equal to k_2 discriminative clusters plus 1 (all other patch types). Batch normalization and dropout were utilized in order to minimize overfitting on the relatively sparse dataset of patches. Furthermore, class imbalances were corrected by re-weighting via inverse class frequency. Class accuracy and loss were calculated throughout training.

2.10. Feature histogram

The trained AlexNet model was evaluated on the training/validation patches as well as on the untouched test patches (Note: split was performed patient-wise). The frequency of each class (regardless of accuracy) was quantified and compiled into a histogram for each eye.

2.11. Classification

Two different methods for classification were used upon the n-dimensional feature histograms (vectors), where n corresponds to the number of feature classes. (1) A k-means clustering method with 2 clusters i.e. DED and NCP (2) A simple non-linear neural network trained on the training/validation set histograms.

3. RESULTS

3.1. K-means clustering

For both image dataset (1) and (2), as well as for each of the 5 training/testing cycles, the optimal k_1 was 4 clusters and the optimal k_2 was 3 clusters. As such, the AlexNet model was always trained to detect n = 4 patch classes. Note: class labels from one cycle may not have directly corresponded to class labels of another cycle.

Interestingly, the common features in each "discriminative" cluster or patch class were not visually discernable. In figure 4, three patches with seemingly different features were grouped together. It is unclear as to how these patches were "nearby" in the embedding space.

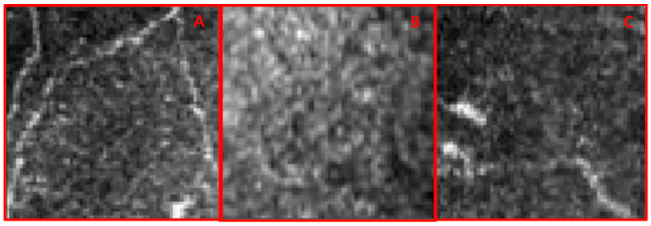

Figure 4: *Three Patches originating from a "discriminative" cluster*

3.2. AlexNet Training

The AlexNet training on the patch classes resulted in consistent training accuracy gains, as well as a steady decrease in training loss for each patch class (Fig. 5). Similarly, the validation accuracy increased as the epochs progressed albeit with some ups and downs (Fig. 6). The network managed to avoid overfitting while getting both the training and validation accuracies above ~70% for each class. This held true across datasets and cycles.

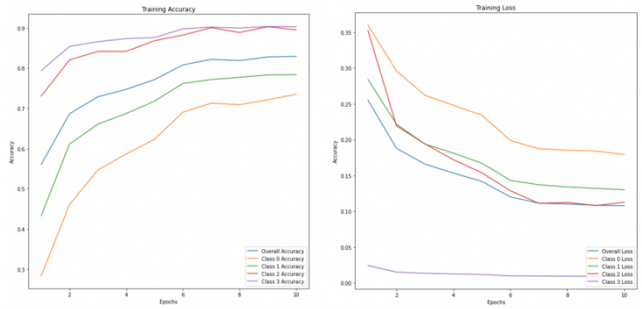

Figure 5: *Training Accuracy (left) increased and Training Loss decreased as the number of epochs increased*

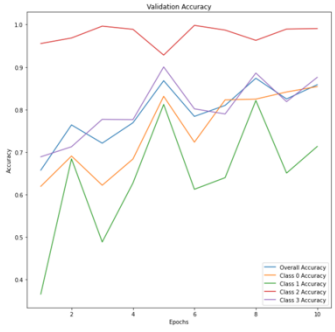

Figure 6: *Validation Accuracy increased (on average) as the number of epochs increased*

3.3. Classification via feature histogram

Tables 1-6 demonstrate the average disease-state classification accuracy of the cluster method and the non-linear neural network method when given the feature histograms. All of the accuracies hover around 40-60%, which is roughly equivalent to random chance.

DED Train	***NCP Train***	***Avg Train***	***DED Test***	***NCP Test***	***Avg Test***
50.4%	*57.3%*	*53.9%*	*29.0%*	*41.0%*	*35.0%*

Table 1: *Cluster Classification: Dataset 1 (averaged over 5 cycles)*

DED Train	***NCP Train***	***Avg Train***	***DED Test***	***NCP Test***	***Avg Test***
51.3%	*53.3%*	*52.3%*	*47.0%*	*53.0%*	*50.0%*

Table 2: *Cluster Classification: Dataset 2 (averaged over 5 cycles)*

DED Train	***NCP Train***	***Avg Train***	***DED Test***	***NCP Test***	***Avg Test***
43.6%	*49.6%*	*46.6%*	*50.0%*	*48.6%*	*49.3%*

Table 3: *Cluster Classification: Dataset 3 (averaged over 5 cycles)*

DED Train	***NCP Train***	***Avg Train***	***DED Test***	***NCP Test***	***Avg Test***
43.1%	*60.9%*	*52.0%*	*40.0%*	*58.5%*	*49.3%*

Table 4: *Non-Linear Classification: Dataset 1 (averaged over 5 cycles)*

DED Train	***NCP Train***	***Avg Train***	***DED Test***	***NCP Test***	***Avg Test***
49.2%	*53.3%*	*51.3%*	*47.5%*	*52.5%*	*50.0%*

Table 5: *Non-Linear Classification: Dataset 2 (averaged over 5 cycles)*

DED Train	***NCP Train***	***Avg Train***	***DED Test***	***NCP Test***	***Avg Test***
52.3%	*53.1%*	*52.7%*	*47.0%*	*51.1%*	*49.1%*

Table 6: *Non-Linear Classification: Dataset 3 (averaged over 5 cycles)*

4. DISCUSSION

Although not included in this paper, initially, I attempted to perform disease state classification of the CCM images directly (both at and away from the corneal vortex). The neural network struggled substantially with overfitting. In response, numerous steps were taken to prevent this issue. Data augmentation was performed; flips, rotations, crops, translations, and noise were added to the dataset. Regularization was implemented into the model architecture. i.e. batch normalization and drop out. The loss function was switched between cross entropy loss and cosine similarity loss. The class imbalances were worked on via inverse class frequency weighting and down sampling. The networks were also pretrained on natural images. Finally, different architectures such as AlexNet, VGG 16/19, and a Densely connected network were attempted. Under each and every modification, the validation set was unable to surpass 50% accuracy.

In order to address the sparse dataset issue, I moved to a feature detection approach. Image patches were more abundant than the CCM images. In terms of feature classification, the network was able to correctly classify >70% of the patches without overfitting (Fig. 6). Additionally, in theory, a histogram of features avoided the issue of a single image determining the disease state.

Unfortunately, I believe that the unsupervised learning of patch clusters was unsuccessful. It was difficult to determine the criteria by which the patches were grouped together as it did not seem to correspond to anatomical landmarks (e.g. nerves, micro-neuromas, bare space). Furthermore, I have low confidence in the criteria by which the most "discriminative" features were selected.

In the end, the patch-based approach performed roughly as well as the baseline approach (~50% accuracy) when using both the non-linear neural network and the cluster method on the histogram. This likely means one or more of three things: (1) the dataset is too small (2) the clustering method and discriminative feature determination is flawed (3) CCM cannot be used to distinguish between dry eye disease and neuropathic corneal pain.

Moving forward, I will likely need to increase the number of images per patient, increase the dataset of patients, and transition to a different discriminative feature approach (perhaps to a manual annotation of features).

5. REFERENCES

[1] Suvarna Phadatare, Munira Momin, Premanand Nighojkar, Sonali Askarkar, Kamalinder Singh, "A Comprehensive Review on Dry Eye Disease: Diagnosis, Medical Management, Recent Developments, and Future Challenges", Advances in Pharmaceutics, pp. 1-12, (2015)

[2] RD Treede, TS Jensen, JN Campbell, "Neuropathic pain: redefinition and a grading system for clinical and research purposes", Neurology, 70, pp. 1630–5, (2008)

[3] A Galor, HR Moein, C Lee, "Neuropathic pain and dry eye" Oculular Surface, 16, pp. 31-44, (2018)

[4] MS Jiang, Y Yuan, ZX Gu, SL Zhuang, "Corneal confocal microscopy for assessment of diabetic peripheral neuropathy: a meta-analysis", British Journal of Ophthalmology, 100, pp. 9-14, (2016)

[5] Hamid-Reza Moein, Anam Akhlaq, Gabriela Dieckmann, Alessandro Abbouda, Nicholas Pondelis, Zeina Salem, Rodrigo T. Müller, Andrea Cruzat, Bernardo M. Cavalcanti, Arsia Jamali, Pedram Hamrah, "Visualization of microneuromas by using in vivo confocal microscopy: An objective biomarker for the diagnosis of neuropathic corneal pain?", The Ocular Surface, Volume 18, Issue 4, pp. 651-656, (2016)

[6] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, J. H. Saltz, "Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, pp. 2424-2433, (2016)

[7] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks", Communications of the ACM, 60, 6 pp. 84–90, (June 2017),