RPE Age Classification Using Scikit and CNNs

Adam Baig, Chun Lin, Herby Tzun

Georgia State University

MATH 3350: Scientific Computing

Dr. Yi Jiang

April 29th, 2025

Abstract

The Retinal Pigment Epithelium (RPE) is an essential component of retinal health as well as maintaining vision. Disease from RPE disfunction like Age-related Macular Degeneration is an increasing problem in developed countries as the average lifespan continues to grow. One way researchers are attempting to lessen the burden these diseases have by improving diagnostics and treatment is with AI algorithms. This project aims to compare several of the most common AI methods to see how they differ in the ability to classify samples of RPE crops into certain age groups. The three main algorithms used were Support Vector Machines (SVM), K Nearest Neighbors (KNN), and Convolutional Neural Networks (CNN). The results of these investigations provide a clear indication that CNNs can handle the specific type of data it was given better than SVMs and KNNs. It is hoped that these results can help to inform how further research can consider the methods that would better fit the data given.

I. BACKGROUND

The Retinal Pigment Epithelium (RPE) is a layer of cells in the outermost layer of the retina and connects to the Bruch's membrane. The RPE serves many vital functions in preserving retinal health such as: the decomposition and disposal of waste out of the retina, the facilitation nutrients into the retina, and protection of the retina from external sources. Since the RPE cells do not regenerate after birth, disfunction and the death of these cells have detrimental effects to vision (Yang S et al., 2021). One such disease related to the decline of RPE quality is Age-related Macular Degeneration (AMD) characterized by increased death of the RPE cells and damage to the macula. Currently AMD is the leading cause of vision loss in developed countries and there is no cure for it. Treatments and protective measures are available but mainly consist of

monitoring and the slowing down or stopping the progression of the disease (Age-Related Macular Degeneration, 2021). As such being able to diagnose early and understand the effects of age on the RPE are of vital importance to improving treatment.

One way that researchers have been attempting to study and improve the way that the RPE progresses with age as well as the development of any disease impacting the RPE is by using various AI models. In a paper by Frank-Publig et al (2024) there is a discussion into how AI models can improve the detection and monitoring of AMD progression. One notable way that these models are useful is through the fact that they can analyze huge datasets far faster than any human could ever hope to as well as being able to adapt to different imaging modalities with very little effort if any. Some ways that they were used were by separating high and low risk images as well as providing more accurate diagnoses of AMD and AMD related diseases. However, a limiting factor for all of these models is that, due to the reliance on good data, it is even more important for samples which are well representative of the population, and, with the nature of unsupervised learning, we are not always able to understand why a certain image is more indicative of disease as opposed to another (Frank-Publig et al., 2024).

As AI becomes more and more powerful and popular for use in the science communities, more methods continue to be developed and studied especially for use in certain fields. One example of this is the “detecting apoptosing retinal cells” or DARC detailed in Corazza et al. (2021). This model combines traditional CNNs with fluorescent biomarkers which highlight areas of apoptosis and use these images to assess risk and presence of AMD. This model detects what are called DARC spots which are notable areas of apoptotic cells and was found that images where more than five DARC spots were found had a strong correlation to wet AMD risk.

However, once again this method also faces the same issue of a lack of understanding behind why this model makes the selections it does (Corazza et al., 2021).

While the previous examples highlight the use of AI in prediction and diagnoses of disease in the RPE, specifically AMD, AI can also assist in analyzing the features of the RPE as it ages. One example is a morphometric analysis done on Mice RPE where machine learning methods were used to see if it was reliable to predict the age of an RPE based on the morphometric features of the image, which were just the cell borders isolated from other features in the crops. Unrelated to machine learning was the use of comparing the sizes and shapes of RPE cells between regions and age groups and some interesting results were found. For one, it was shown that the peripheral areas of the RPE show more signs of aging as the count and shape differ with age as the cells grow more elongated (Kim et al., 2021).

Previous studies have encapsulated how powerful and helpful various AI and machine learning models can be in assisting researchers. It is important to know and differentiate different AI models to know which model has a better performance in particular situations so that researchers can maximize their use. On the topic of using AI for prediction in images, these algorithms and models are the most commonly used: Convolutional Neural Networks (CNN), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN).

SVM is a supervised machine learning technique that aims to classify data by finding the optimal boundary that separates their classes with the maximum margin between data of different classes to classify new data. CNN on the other hand, is a type of neural network which falls under the deep learning category which is a subset of machine learning. It captures spatial hierarchies from images through its use of convolutional layer to predict image data. While KNN is also a supervised machine learning algorithm that uses proximity of data points to compare a

data point to the set that it's on which the algorithm memorizes and learns to produce an appropriate output when given new data.

In our project we look at flat mount images of healthy mice RPEs from various age groups. The data we used was obtained from the Emory Eye Center and were dissected and stained in a way that allows for a flat two-dimensional image. From these flat mounts crops were taken from the crops only those in good condition to limit defects or other artifacts. We set out to use some of these more common AI models using Python and its various libraries to see if we can reliably classify these images and how each model compares to one another.

II. METHODS

In this project, the dataset consisted of RPE images from wild-type(healthy) mice and our goals was to categorize age-related changes in normal retinal development using KNN, SVM and CNN scripts to study RPE cell progression. The images are to be pre-processed before they can be used as inputs in our AI models. Each pixel of the image is to be converted to numerical values corresponding to the RGB channel which are then stored within 3D tensors. To better load the data, we wrote a custom class within a separate python file to be used as a module which parses through the file path of the images to extract them as well as converting them into array of numbers. The original dataset compared six distinct age groups: 30, 45, 60, 180, 330, 720 days. In hopes of achieving higher accuracy, we group them into the following: Young (30,40,60), Middle (180), Old (330 and 720). The Young and Old groups initially had fewer data points compared to the Middle group. To address this imbalance, we applied some data augmentation techniques via cropping and rotations and flipping to generate additional samples to account for the skew in data, as shown in Chart 1. Notably, the dimensions of some of the sample images

varied, so all samples were standardized to be cropped into 149 by 149 pixels to maintain a more uniform dataset.

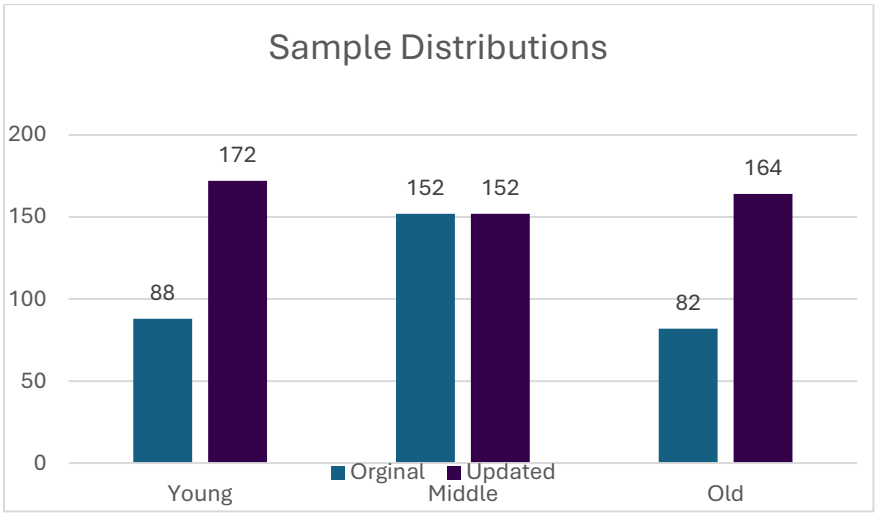

Chart 1:This shows the original amount of data points and the updated data through cropping and rotation and flipping

After data augmentation, we then split the whole dataset into training data and testing data with a ratio of 8:2 to prevent overfitting of our AI models which cause inaccurate evaluations.

When training CNN, we use PyTorch to separate the training data into batches of size 32 to be iterated through epochs which we chose to be 20. Each batch of data is then passed to the model to be trained through its layers: convolutional layer, max pooling layer, and fully connected layer. Each image of the data is convolved using kernels of subsequent size (3,32,3), (32,64,3), and (64,128,3) with padding. After each image is converted into a feature map from the kernel, it is then activated using the ReLU function to prevent linearity which is then passed to the max pooling layer where it further compactifies the feature map by choosing the maximum values. To further prevent overfitting, 25% of the elements from the output are dropped. The results are then flattened to be of same structure as neurons in a dense layer so that they could be inputted in the fully connected layer. Linear transformation with weights and biases are then performed on the neurons which are then put through the ReLU activation function again. 50% of the outputs are then dropped and the remaining go through linear combination which gives out

a final output array. We then use cross entropy loss to calculate the loss and the Adam optimizer to optimize the model's parameters so that it could minimize the loss and give better accuracy. After the data was trained through 20 epochs, we evaluate the model by passing though the testing data and display the confusion matrix for visual representation of model performance.

When training SVM, we set the type of kernel to be linear to be used for linearly separable data. Using Python's scikit learn library, it is easily implemented as a line of code to create the SVM model with set parameter. Since scikit learn doesn't have functionalities to split labels from the actual samples, we manually split them for both the training data and testing data. We then fit the model using the samples and labels from the training data to train it. We used the trained model to make predictions using the samples from the testing data. To measure the accuracy of the model, we compared the predicted labels with the actual labels from the testing data and put those metrics into a confusion matrix. The process to train KNN is similar to the process to train SVM since both of the algorithms are from the scikit learn library. The data is split in the same way as before and the model is also implemented as one line of code, but we used a technique called grid search for hyperparameter tuning of the KNN model each time it is trained. We then used the optimized model to make predictions which the model is then evaluated in the same manner as before. Of course, we also used the confusion matrix for this model.

III. RESULTS

Through this process, we achieved improved classification as seen in Figure 1, where our SVM and CNN models went past the 80% accuracy mark when classifying our images into their appropriate groups. The two models had a well-defined diagonal in the confusion matrix, indicating that the models were able to properly classify the images into the three groups.

However, it must be noted that our KNN model performed significantly worse, maintaining an accuracy of only around 60%.

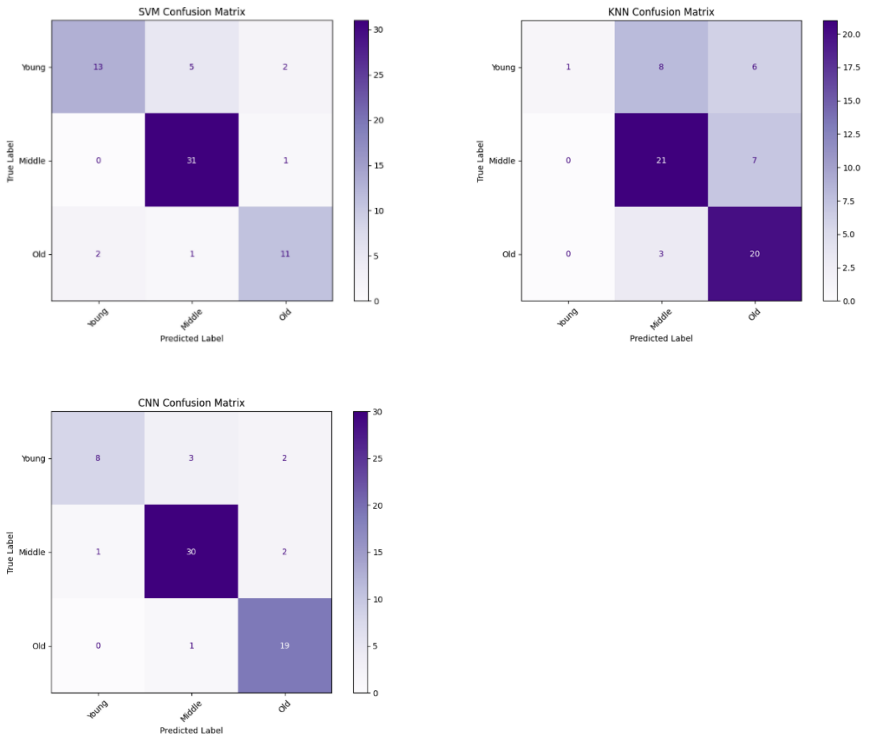

Figure 1:The SVM model achieved an accuracy range of 81%–85%, while the KNN model ranged from 60%–63%. The CNN model achieved the highest performance, with an accuracy range of 85%–88%. All models were trained with an 80-20 train-test split.

The SVM model had most of its misclassifications between the Young and Middle groups, likely due to overlapping visual features. In contrast, the CNN model showed a more balanced pattern across all groups, indicating a better overall generalization of features. It is worth mentioning that the increase in epoch cycles vastly improved in accuracy, where 10 cycles produced an average of 60% accuracy, and our final result of 20 epochs reached the range of 85%-88%. It was even tested for 50 epochs, where it would reach the 90% accuracy range, however, due to limitations on our hardware, run times took significantly longer, thus it was not appropriate to post those findings without multiple test runs. Unlike CNNs, KNN lacks the

ability to learn hierarchical features, and its performance degrades when visual differences between classes are subtle, as is the case with early- and mid-stage retinal aging. Additionally, KNN does not generalize well with imbalanced or noisy data, making it less suitable for this application.

IV. DISCUSSIONS

From the observations of our results, we can confidently state that CNN was the clear-cut winner when it came to improvement in accuracy and should be the clear choice when trying to identify normal retinal progress through visual images. However, our study also has limitations. The dataset size was relatively small, and although data augmentation was used, the model may not generalize well to broader, more diverse populations or to human retinal images without further validation. Additionally, the CNN's performance could be affected by subtle artifacts or variations in image acquisition methods.

Despite these challenges, our findings support the growing use of AI in biomedical imaging. By improving the ability to classify age-related changes in RPE cells, CNN-based models can contribute to early detection and tracking of AMD progression. As datasets grow and models become more refined, this approach could offer a valuable tool in both research and clinical settings for monitoring retinal health over time.

.

Acknowledgements

We would like to thank Emory Eye Center for providing us access to the flat mounts and crops of the data we used in this project. We would also like to thank our professor Dr Yi Jiang for allowing us to work on this as our class project.

Bibliography

Corazza, P., Maddison, J., Bonetti, P., Guo, L., Luong, V., Garfinkel, A., Younis, S., & Cordeiro, M. F. (2020). Predicting wet age-related macular degeneration (AMD) using DARC (detecting apoptosing retinal cells) ai (Artificial Intelligence) technology. *Expert Review of Molecular Diagnostics*, *21*(1), 109–118. https://doi.org/10.1080/14737159.2020.1865806

Frank-Publig, S., Birner, K., Riedl, S., Reiter, G. S., & Schmidt-Erfurth, U. (2024). Artificial Intelligence in assessing progression of age-related macular degeneration. *Eye*, *39*(2), 262–273. https://doi.org/10.1038/s41433-024-03460-z

Kim, Y.-K., Yu, H., Summers, V. R., Donaldson, K. J., Ferdous, S., Shelton, D., Zhang, N., Chrenek, M. A., Jiang, Y., Grossniklaus, H. E., Boatright, J. H., Kong, J., & Nickerson, J. M. (2021). Morphometric analysis of retinal pigment epithelial cells from C57BL/6J mice during aging. *Investigative Opthalmology & Visual Science*, *62*(2), 32. https://doi.org/10.1167/iovs.62.2.32

U.S. Department of Health and Human Services. (2021, June 22). *Age-related macular degeneration (AMD)*. National Eye Institute. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/age-related-macular-degeneration

Yang, S., Zhou, J., & Li, D. (2021). Functions and diseases of the retinal pigment epithelium. *Frontiers in Pharmacology*, *12*. https://doi.org/10.3389/fphar.2021.727870