Retinal Disease Detection Using Convolutional Neural Networks

Jusvin Charles, James Graham, Justin Jose, Nouhad Rizk

UNIVERSITY OF HOUSTON

College of Natural Sciences and Mathematics, University of Houston

Abstract

Retinal diseases, including conditions such as diabetic retinopathy, can lead to severe visual impairment if not detected and treated early. This research aims to deploy advanced neural network models to classify various types of retinal diseases using 1443 fundus images captured using three different fundus cameras with 4 conditions annotated through adjudicated consensus of two senior retinal experts. This study juxtaposes traditional machine learning techniques, such as Logistic Regression, Support Vector Machines (SVM) and Random Forest and applies deep learning techniques built with Convolutional Neural Networks (CNN) utilizing Transfer Learning. Additionally, the analysis and comparison of the effectiveness of the models' accuracies by focusing on the confusion matrices of each model and their retinal disease identification performances. The overarching goal is to develop a model that achieves high accuracy in retinal disease classification, potentially setting new benchmarks in diagnostic standards. This problem was split into two separate parts; first a binary identification of whether a given eye was diseased, followed by a multi-class identification for which disease an eye had.

Background

Fundus Images are photographs taken from interior surface of the eye and are crucial tools in the field of ophthalmology, used to diagnose several eye diseases. These common illnesses are either diseases or are a strong indicator of other serious conditions.

Diabetic Retinopathy (DR) - Microvascular complication of diabetes

Media Haze (MH) - Main indicator for cataracts

Drusens (DN) - Indicator of age-related macular degeneration

Optic Disc Cupping (ODC) - Main indicator for glaucoma

Research by Yonsei University College of Medicine used TL-SVM (Transfer Learning) to classify fundus images, their accuracy was reduced moving from 2 categories (87.4% Accuracy) to 10 categories (30.5%).

Methods

The decision to split the classification into two stages – one, identifying if an eye is healthy or unhealthy, and two, classifying the specific disease in unhealthy cases – was driven by clinical relevance and model efficiency. This method aids in handling the heavy data imbalance between healthy and unhealthy samples.

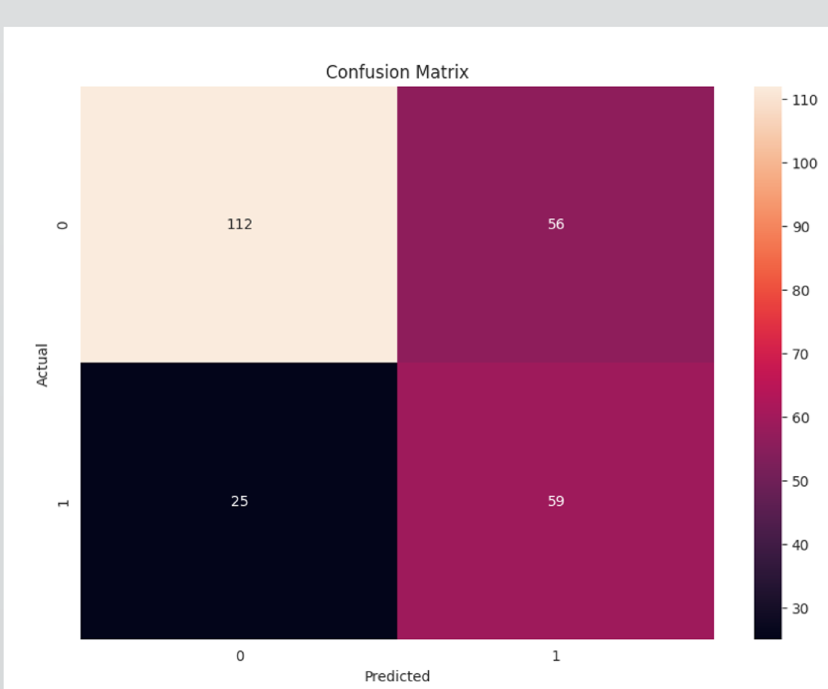
Logistic Regression: Grayscale images at 128x128 px as input, untuned and PCA-enhanced models.

Random Forest: Grayscale images at 128x128 px as input, untuned and PCA-enhanced models.

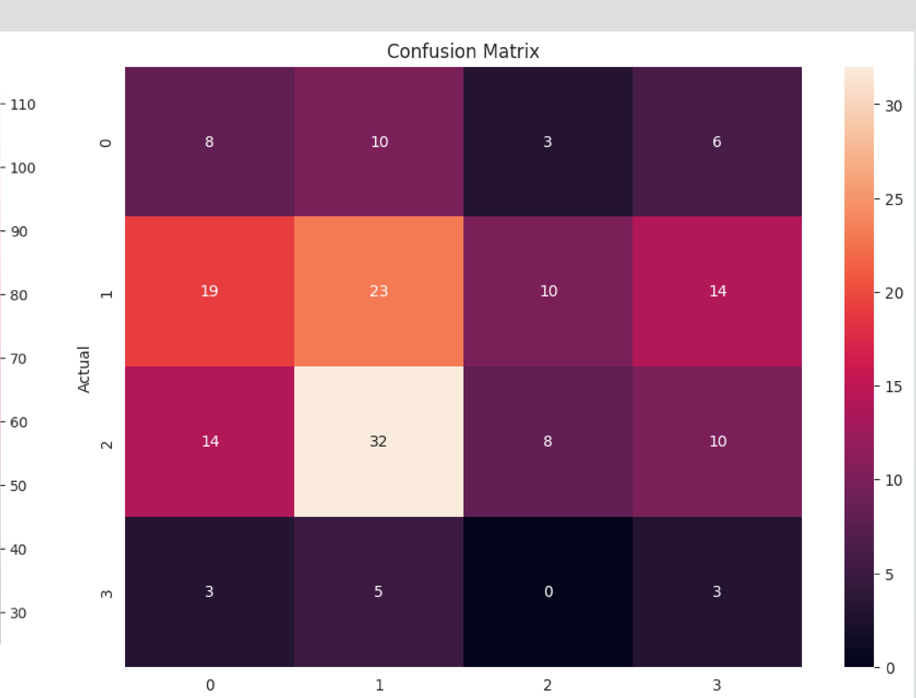
Support Vector Machine: Grayscale images at 128x128 px as input, using radial basis function as its kernel and optimized with class weighting.

InceptionV3: RGB images at 299x299 px as input, adapted with a learning rate scheduler, global pooling layer, dropout layer, and the last 5 layers unfrozen.

ResNet50: RGB images at 224x224 px as input, adapted with a learning rate scheduler, global pooling layer, dropout layer, and the last 5 layers unfrozen.



SVM Binary Classification
Unhealthy (0), Healthy (1)



SVM Multi-Class Classification
Drusens (0), Diabetic Retinopathy (1), Media Haze (2), Optic Disc Cupping (3)

Results

Model	Binary Accuracy	Multi-Class Accuracy	Brief Description
Logistic Regression	55%	27%	Performs at a basic level for basic tasks, struggles significantly with multi-label classification.
Logistic Regression + PCA	56%	21%	Slight improvement in binary tasks with PCA, but multi-label performance drops.
Random Forest	63%	40%	Shows improved binary classification; moderate capabilities in multi-label classification.
Random Forest + PCA	65%	40.3%	Best binary performance among non-deep learning models, with a marginal gain in multi-label accuracy.
SVM	68%	25%	Strong binary classification, but limited effectiveness for multi-class.
InceptionV3	56%	31%	Comparable to Logistic Regression in binary accuracy, with better-multi-label performance.
ResNet-50	82%	56%	Superior binary and multi-label accuracy for the dataset.

Conclusion

- The CNN models were prone to overfitting, even with dropout layers, due to the small dataset.
- ResNet-50 has a decisive edge compared to all the other models in this dataset.
- InceptionV3 fell behind more traditional models like Random Forest and SVM.
- PCA yielded minor improvement overall across the models.

Future Direction

In future phases of this research, there is a need for larger and higher-quality datasets, which will allow for more efficient model training and validation. Moreover, an avenue for the exploring how data augmentation, L2 Regularization, and hyper-parameter tuners could enhance the model's ability to generalize from the updated dataset. Additionally, access to larger computational resources would be optimal for performance and achieve higher accuracy results in retinal disease classification. Lastly, for faster results, an implementation that combines the binary classification with the multi-classification is necessary.

Acknowledgments

We would like to thank our TA, Kurmanbek Bazarov, for his assistance in preparing this project. We'd also like to thank Hewlett Packard Enterprise Data Science Institute for hosting this event, allowing us to experience the creation of a graduate-level research project. Finally, we'd like to thank the numerous medical professionals who assembled this dataset.