Predicting Heart Disease from Eye Images: A Machine Learning Pipeline from Preprocessing to Model Evaluation

Katarina Barbosa

Cardiovascular disease remains one of the leading causes of death worldwide, and early, non-invasive screening methods could transform how clinicians identify individuals at risk. Recent work has explored whether subtle signals captured in facial, tongue, eye, or ear images may reflect underlying cardiovascular health. In this project, I focused specifically on eye images to investigate whether machine learning models could learn discriminative patterns associated with heart disease and predict cardiovascular status from unseen images.

The goal was not only to build prediction models, but also to document a clear, reproducible pipeline, from preprocessing and feature extraction to model training, evaluation, and interpretation. This blog-style write-up walks through each step of the process and summarizes the final findings.

Preparing the Eye Image Dataset

The dataset consisted of labeled eye photographs from participants with and without heart disease. The images were not standardized: lighting varied across photos, and some subjects were wearing glasses or had their eyes closed. These inconsistent or unusable images were manually removed to ensure dataset quality. Before any modeling could begin, the images were divided into three randomized subsets to ensure unbiased evaluation: 358 images for training, 120 for validation, and 120 for testing. The label distribution remained close to balanced across all splits, with the training set containing 175 controls and 183 heart-disease cases. The validation and test sets followed a similar pattern, with 59 controls and 61 cases in each.

This balanced split ensured that no model would be able to rely on simple class frequency patterns. Each image had to contribute meaningful visual information for the classifier to learn disease-related features.

Extracting Eye-Based Cardiovascular Features

To convert raw images into structured, learnable information, the first major step was feature extraction. A variety of computer vision techniques were used to transform each eye photo into a high-dimensional feature representation. These included texture descriptors such as Local Binary Patterns (LBP), structural shape descriptors like Histogram of Oriented Gradients (HOG), and vessel-related metrics that quantified redness, contrast, saturation, and vascular anisotropy. Measures of hyperemia (bloodshot intensity), red concentration, and color uniformity were also incorporated, as several ophthalmic studies have linked these indicators to systemic cardiovascular conditions.

With hundreds of extracted features, dimensionality reduction was essential. A recursive feature elimination (RFE) procedure was applied to identify the fifty most predictive features. This step helped reduce noise, improve model generalization, and highlight the features most strongly associated with cardiovascular health.

Training Machine Learning Models

Once features were selected, a suite of classical machine learning classifiers was trained. Each model underwent hyperparameter optimization using grid search on the validation set, ensuring a fair comparison across approaches. The models included random forests, gradient boosting, support vector machines (SVM), logistic regression, and XGBoost, all chosen for their proven performance on structured tabular features.

During validation, logistic regression produced one of the strongest AUC scores at 0.635, followed by the SVM at 0.618. Although these values reflect a modest signal, they indicated that the eye images did contain some learnable cardiovascular information. Random forest and XGBoost models achieved slightly lower validation AUCs, though they remained competitive.

An ensemble model combining the predictions of all five classifiers was also evaluated. While the ensemble performed similarly to the individual models, its validation AUC did not exceed the best single classifier.

Testing Model Performance on Unseen Eye Images

The true assessment of the pipeline came from evaluating each model on the independent test set. Here, several algorithms outperformed their validation results, suggesting strong generalization. The random forest emerged as the best-performing individual classifier with a test AUC of 0.732. It achieved an accuracy of 0.675, with a precision of 0.649 and a recall of 0.787 for identifying heart disease. This high recall indicated that the model successfully detected a large proportion of cardiovascular cases.

The SVM and logistic regression models also performed well. Logistic regression achieved an AUC of 0.680 and an F1-score of 0.713, while the SVM produced an AUC of 0.673 with particularly strong recall (0.885), reflecting its sensitivity to small visual differences in the image features. XGBoost also proved competitive with a test AUC of 0.722 and balanced precision-recall characteristics.

While the ensemble model reached an AUC of 0.720, it did not surpass the best individual model, confirming that the random forest alone captured the most robust structure in the data.

Interpreting the Most Informative Eye Features

Understanding *why* a model performs well is as important as the performance itself. To identify the visual traits most strongly associated with predicted heart disease, feature importance scores were generated. The top contributors included several HOG components, indicators of redness asymmetry between the left and right eye, measures of saturation variation, and markers of vascular hyperemia.

Features such as left-right red differences, red uniformity, center red dominance, and vessel anisotropy appeared prominently. These findings are consistent with clinical observations that vascular dilation, inflammation, and tissue oxygenation levels can manifest subtly in the eye. Texture-based features such as LBP and GLCM contrast and homogeneity also ranked highly, reinforcing the idea that micro-texture patterns in the sclera and conjunctiva may encode physiological information.

This interpretability step not only enhances trust in the models but also offers interesting avenues for future biomedical research.

Conclusion: Can Eye Images Predict Heart Disease?

The results of this project demonstrate that machine learning models can learn weak but meaningful patterns from eye images that correlate with cardiovascular health. While no single model reached diagnostic-level performance, the best classifier achieved a test AUC of 0.732, well above chance, and showed strong sensitivity to heart-disease cases. These findings suggest that ocular features may contain subtle biomarkers of cardiovascular physiology, and with larger datasets, improved feature-engineering methods, or deep learning approaches, predictive accuracy could increase further.

This work establishes a complete pipeline, from preprocessing and feature selection through model training, evaluation, and interpretation, that can serve as a foundation for future research in image-based cardiovascular assessment. As non-invasive imaging continues to expand in medical diagnostics, the potential for eye photographs to assist in early screening for heart disease remains a promising and exciting direction.