Retinal dystrophies are a group of conditions that affect 1/2000 individuals. They are diagnosed based on the reported symptoms, the results of clinical testing, the course of disease, and the predicted mode of inheritance. There are currently only a handful of specialists in the United States who can effectively diagnose and manage patients with retinal dystrophies. These conditions are clinically and genetically heterogeneous. (Figure 1) Over 170 genes and thousands of different mutations have been implicated as disease-causing in individuals with various retinal dystrophies. Identifying the genetic diagnosis for a given patient is important for a number of reasons. It can confirm the diagnosis, direct clinical management, provide a more accurate prognosis, inform genetic counseling, and identify patients for whom molecular-based therapy might be available.

Genetic testing, however, is labor-intensive and expensive and is therefore only performed on a subset of patients! The cost of testing depends which and how many genes are analyzed and can be anywhere from several hundred to several thousand dollars. Thus, there is a need for improved diagnostic tools to help determine a patient’s primary clinical diagnosis and the appropriate targets for genetic testing. In addition, even when genetic testing is affordable, results may be inconsistent, and new methods are necessary to help interpret these test reports and identify which variants are most likely to be disease-causing.

Our purpose is to develop a machine learning program that predicts the most likely genetic cause of a patient’s retinal dystrophy using input information about patient demographics, electroretinogram (ERG) response, visual field, pattern of inheritance, and fundus autofluorescence (FAF) features. This program may help to inform appropriate genetic test panels to order and can be used as a tool with which to interpret genetic testing results, along with existing resources such as PolyPhen, SIFT, and reported causative mutations in the literature.

METHODS

Data on patient age, sex, ERG response, visual field, family history, FAF imaging, and genetic diagnosis was collected on 152 patients seen at the Kellogg Eye Center. After filtering out mutated genes that affected fewer than 5 patients, 252 patients were usable for machine training purposes. Machine learning algorithms were developed to predict the genes most likely to be mutated and causing the observed clinical features for a given patient. Multiple algorithms were applied, and the support vector machine (SVM) with linear and radial basis function (RBF) kernel was shown to perform best. Machine learning used 80/20 training/testing splits of the data. Imputation techniques were applied to compare the classification performance with that of a baseline classifier.

RESULTS

A prototype has been created that has greater than 60% accuracy for predicting the causative mutated gene in a given patient’s sample. The generation of this machine learning program provides physicians with a prediction of a causative mutated gene when evaluating a new patient in their clinic. This may help ophthalmologists refine their clinical diagnosis and identify the most relevant genetic testing to order. Additionally, genetic test results can often be difficult to interpret. Programs such as PolyPhen and SIFT help to determine whether or not a variant detected in a gene is truly pathogenic. The RetDegenDx program is one further tool that can be used to help predict the gene in which mutations are most likely to cause the clinical features of the patient. When used alongside existing programs, it can help to improve the interpretation of genetic test results.

CONCLUSIONS

The generation of the machine learning program provides physicians with a prediction of a causative mutated gene when evaluating a new patient in their clinic. This may help ophthalmologists refine their clinical diagnosis and identify the most relevant genetic testing to order. Additionally, genetic test results can often be difficult to interpret. Programs such as PolyPhen and SIFT help to determine whether or not a variant detected in a gene is truly pathogenic. The RetDegenDx program is one further tool that can be used to help predict the gene in which mutations are most likely to cause the clinical features of the patient. When used alongside existing programs, it can help to improve the interpretation of genetic test results.

REFERENCES

8. Sensitivity = 60% Specificity = 65% Positive Predictive Value = 64% Negative Predictive Value = 72%.