Cloud-Based Ocular Disease Diagnosis

Yaman Abdulhak, Ajaay Chandrasekaran, Edmond Cunningham, Lisa Jin, Levin Kim, Wenlu Yan, Xinghai Zhang
Thiran Jayasundera, Dana Schlegel, Andrew DeOrio
University of Michigan College of Engineering and Kellogg Eye Center, Ann Arbor, MI

University of Michigan Kellogg Eye Center

Background

Problem

- Around 400 patients are diagnosed with a genetic retinal dystrophy annually.
  - The process can become expensive and time consuming due to blood tests. It is very difficult to precisely diagnose rare cases of genetic retinal dystrophies due to hundreds of possible causes.
  - Specialized eye hospitals throughout the world have isolated patient data.
- How can we consolidate genetic retinal dystrophy data in one place and utilize this data to help clinicians worldwide narrow down the cause of a given patient's dystrophy?

Solution

- Create a webapp as a tool to assist with the diagnosis of genetic retinal dystrophies.

Stakeholder Objectives

KEC: Have an accurate cloud-based web app to diagnose retinal diseases. Build a large patient database and maintain patient privacy. Improve machine learning models.

Partner Hospitals: Have access to patient database and maintain patient confidentiality. Conduct further research on retinal diseases.

Dept. of Health & Human Services: Ensure HIPAA regulations. Maintain patient confidentiality.

Engineering Requirements

Machine Learning: Construct predictive model with >70% accuracy for predicting the gene causing a patient's dystrophy.

Inheritance Pattern: Construct predictive model with >75% accuracy for predicting the inheritance pattern causing a patient's dystrophy.

Frontend: Build a user-friendly web app that clinicians trust and value.

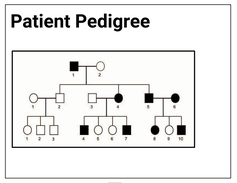
Backend: Build a web app backend that services user requests for webpages and data in <1 second.

HIPAA: Create a web app that securely manages sensitive patient data in accordance with the Health Insurance Portability and Accountability Act.

Data Analysis: Explore statistical correlations between a patient's dystrophy and other characteristic features, such as hearing loss.

Machine Learning Prediction Model

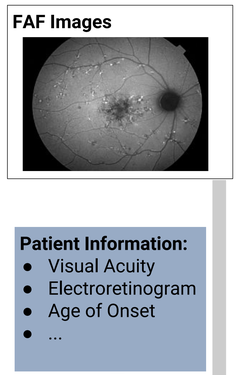
Inheritance Pattern Prediction



We asked a genetic specialist to come up with questions that would help us identify the inheritance pattern causing a patient's dystrophy. Then we answered those questions based on the patient's pedigree and used those answers to build our model. Then we used the output as features to feed into the gene prediction model.

Additionally, we built tools that allow users to input pedigrees electronically and answer the specialist's questions automatically.

Mutated Gene Prediction



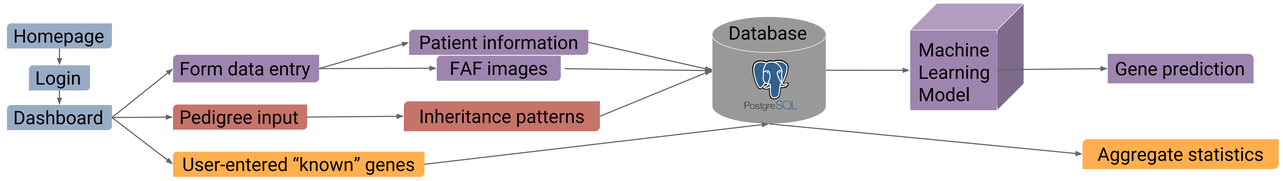
The model is built on top of last year's model, which was trained on 152 patient data points.

Modifications:
- Imputation method
- Feature encoding
- Feature standardization

Additions:
- Use inheritance pattern as features (see left)
- Create API for access from our web application

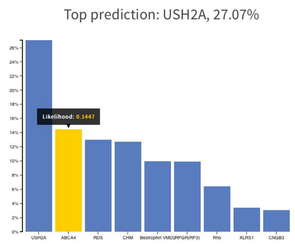
Web Application Prototype

Flow of Information: There are two primary functions at play in the web application: gene prediction and aggregate statistics. Given patient data, the former gives an informed prediction of what gene a patient's dystrophy may originate from. The latter relies on previously entered patient data, as well as a hindsight "known" gene, to produce statistical graphs that may expose novel relationships between the two.



Form Data Entry to Gene Prediction:

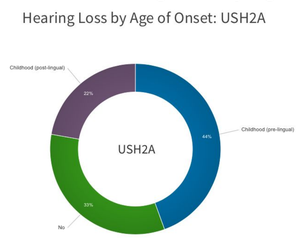
1. In creating a patient, the user will input into a form
   a. Demographics
   b. Ocular and syndromic features
   c. Pedigree: A GUI allows users to draw the pedigree.
   d. FAF images: Anomalous features in image are reported.
2. Patient information is stored in a PostgreSQL database.
3. The machine learning model receives data from the web application's database to predict the most likely genes.

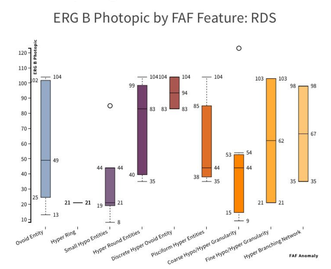


Along with the top prediction, the likelihoods of each of 9 genes (USH2A, ABCA4, RDS, CHM, Bestrophin VMD2, RPGR, Rho, XLRS1, and CNGB3) are reported to the user.

"Known" Genes to Aggregate Statistics:

- Once clinicians receive patients' verified diagnoses, this data may be used to give statistical insight to other users.
  - Filtered by gene: ERG B photopic/scotopic by FAF anomaly, FAF anomaly totals, multiple diagnoses, syndromic features, visual acuity by age of onset/progression
  - All genes: FAF anomaly totals, syndromic features, hearing loss by age of onset/progression





Validation Methodologies

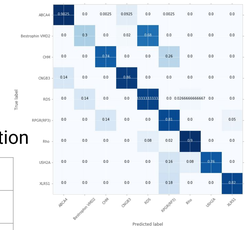
Predictive Model

5 - fold cross validation on 50 iterations:

Accuracy for Mutated Gene Prediction

Model	Mean (%)	Std (%)
Naive Guess	31.13	0
Previous Team	~67	?
SVM	80.15	2.08

Confusion Matrix for Gene Prediction

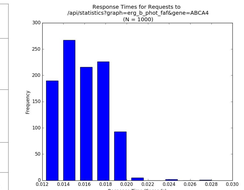


Accuracy for Inheritance Pattern Prediction

Model	Mean (%)	Std (%)
Naive Guess	40.09	0
Clinician Answers	77.10	1.61
Machine Answers	73.02	3.72

Web Prototype

Web App	
Database Latency	Database access time < 4s
Browser Page Load Timer	Page load times <= 1s
UI/UX	
Trust of website	Single Ease Question Average Score(7-point rating scale questions) = 6 > 5
Intuitive User Interface • Usability testing	Task Completion Rate = 85% > 78% Number of Errors/Task =0.5 < 0.7



Response times for requests to the statistics route were significantly under the 1s goal.

Pedigree App

Approximately 2.5 times faster to draw on paper from user testing. Aim to focus on user interface.

Conclusions

Future Work

- Collect more clean data on patients and their family history to better train the machine learning model.
- Deploy a robust web application that is:
  - Secure: The server is safeguarded against common web attacks.
  - Reliable: The patient information form saves entered data at various points of completion.
  - HIPAA-compliant: Stored patient data does not contain PHI, and the database aligns with HIPAA guidelines.

Acknowledgements

- We would like to thank our sponsors at Kellogg Eye Center, the U-M CSE Department, and the U-M Multidisciplinary Design Program for their generous support.