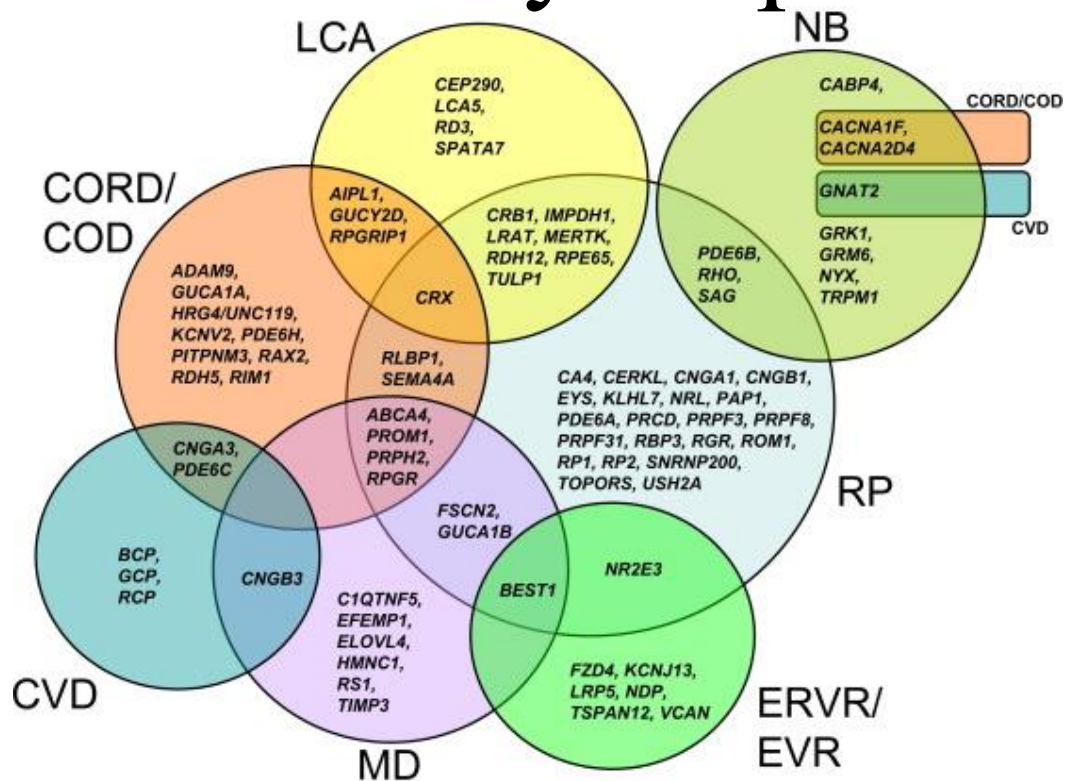# Inheritance Pattern Prediction of Retinal Dystrophies: A Machine-Learning Model

Dana Schlegel, MS, MPH, CGC; Edmond Cunningham; Xinghai Zhang; Yaman Abdulhak; Andrew DeOrio, PhD; K. Thiran Jayasundera, MD
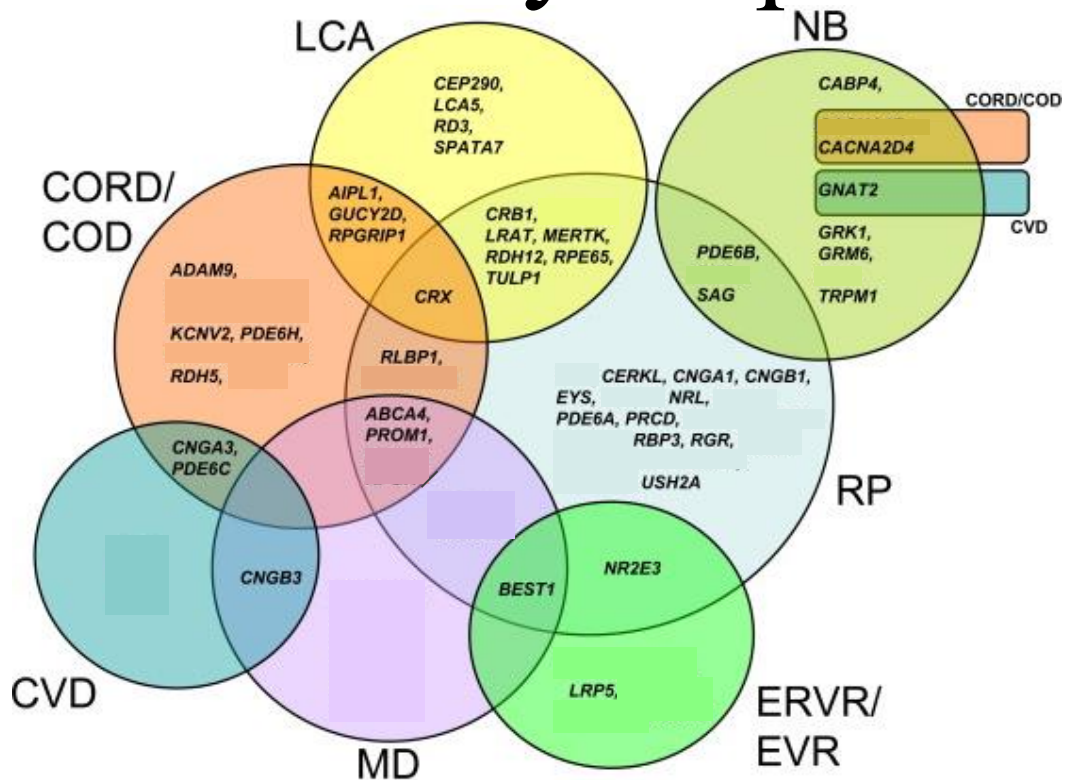
KELLOGG EYE CENTER
MICHIGAN MEDICINE

Retinal Dystrophies

Berger W et al, 2010.

Retinal Dystrophies

Autosomal Recessive (AR)
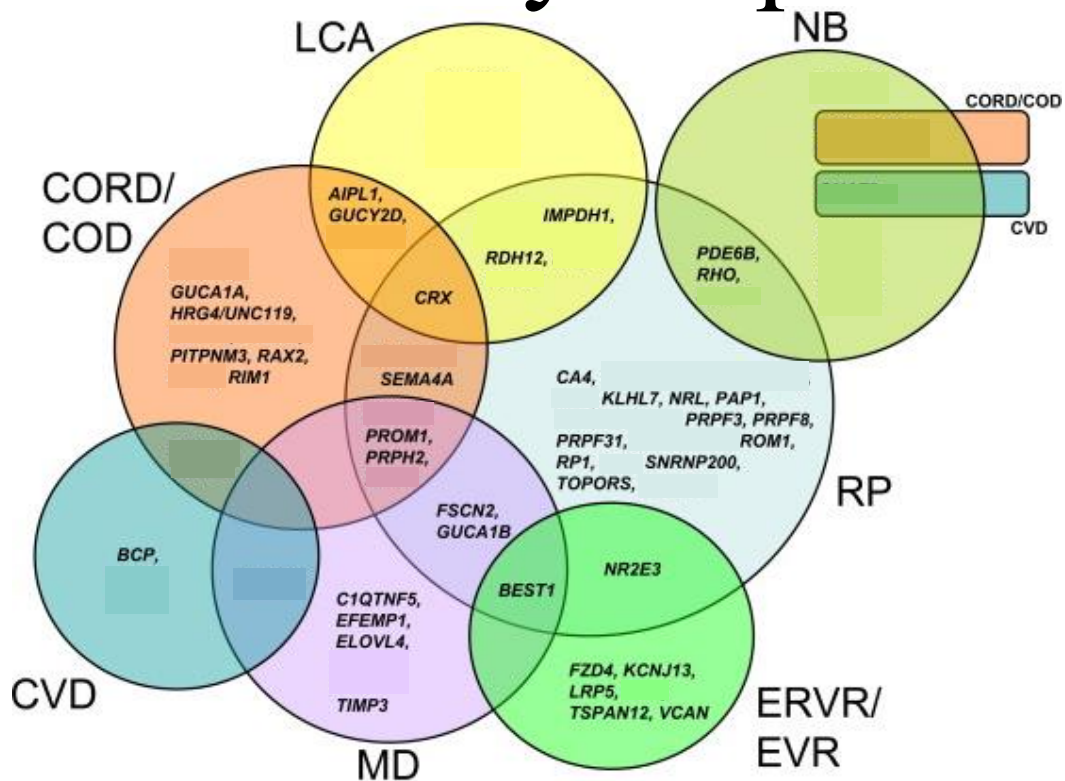
KELLOGG EYE CENTER
MICHIGAN MEDICINE

Retinal Dystrophies

Autosomal Dominant (AD)

Adapted from Berger W et al, 2010.
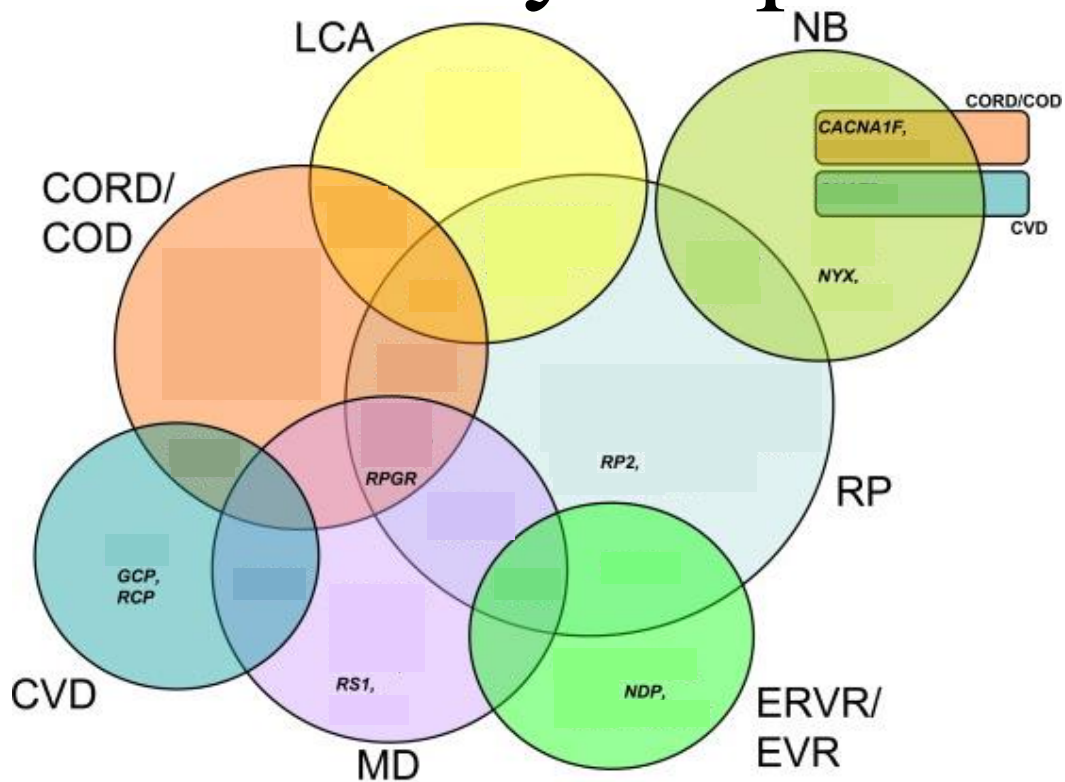
Retinal Dystrophies

X-linked (XL)

Adapted from Berger W et al, 2010.

Inheritance Pattern Prediction

- Can guide appropriate genetic testing
- May inform likely diagnosis
- Allows calculation of likely risks to relatives
- Required component of data collection for some retinal dystrophy studies

As far as we are aware, there is no current algorithm to predict pattern of inheritance for a given patient, and not all clinics have training in genetics or access to genetic specialists/genetic counselors

Aim

- Create a machine learning algorithm whose input is patient family history information and whose output is likely pattern of inheritance

- Used retrospective chart review on patients with genetically-proven retinal dystrophies
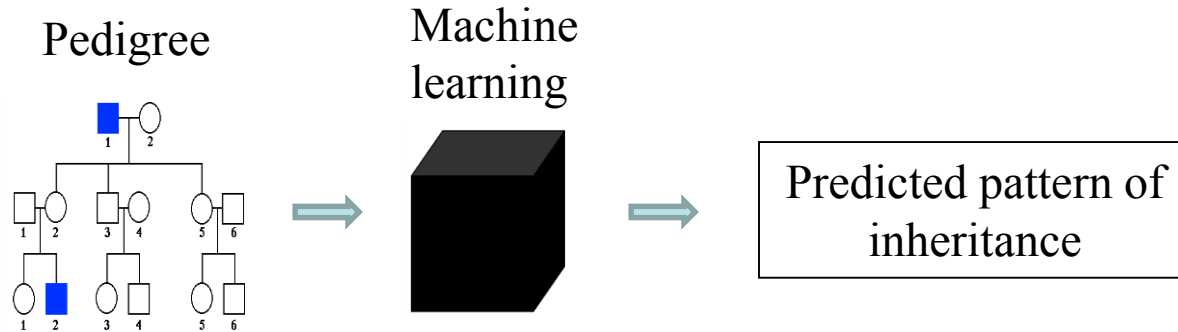
Data collection

- Kellogg Eye Center retinal dystrophy patients

- Family history obtained by genetic counselors (and, in rare cases, retinal dystrophy specialists) as a part of routine patient care

- Information collected by engineering and medical students trained by genetic counselors and retinal dystrophy specialists

- Pedigrees converted into digital computer-readable form

Data collection methodology

- Students trained in predicting pattern of inheritance based on interpretation of pedigree appearance evaluated likely pattern of inheritance for each patient (277 patients)

- Answers to 12 questions about family history were collected from each patient's pedigree and analyzed with machine learning (100 patients)

- Answers to the same 12 questions were collected through computer feature extraction of a digitized pedigree and analyzed with machine learning (90 patients)
  - Included tolerance for user input error

(Overlap of 70 patients between the three cohorts)

Family history features

|	Question	Possible Answers
1	Is more than one generation affected?	Yes/No
2	Do any affected males have affected sons?	Yes/No
3	Do any affected males have affected daughters?	Yes/No
4	Are there any unaffected individuals who are "skipped"? (Their parents or siblings or grandparents are affected and children or grandchildren are affected, but they themselves are unaffected.)	1. No 2. Yes - females only are skipped 3. Yes - at least some males are skipped
5	Are any siblings of the patient affected?	1. No 2. Yes, and no other relatives are affected 3. Yes, and other relatives are also affected
6	Are any cousins of the patient affected?	1. No 2. Yes - maternal cousins only 3. Yes - paternal cousins only 4. Yes - maternal and paternal cousins
7	Are both males and females affected?	1. Yes 2. No - only males 3. No - only females
8	Is onset of disease < or = 20yrs in males?	Yes/No
9	Do any females have asymmetric disease?	Yes/No
10	In general, do females have less severe or later onset of disease?	Yes/No
11	Is there more than one retinal diagnosis in the family? (ex. Stargardt and Pattern Dystrophy)	Yes/No
12	Is consanguinity present?	Yes/No

Family history features

	Question	Possible Answers
1	Is more than one generation affected?	Yes/No
2	Do any affected males have affected sons?	Yes/No
3	Do any affected males have affected daughters?	Yes/No
4	Are there any unaffected individuals who are "skipped"? (Their parents or siblings or grandparents are affected and children or grandchildren are affected, but they themselves are unaffected.)	1. No 2. Yes - females only are skipped 3. Yes - at least some males are skipped
5	Are any siblings of the patient affected?	1. No 2. Yes, and no other relatives are affected 3. Yes, and other relatives are also affected
6	Are any cousins of the patient affected?	1. No 2. Yes - maternal cousins only 3. Yes - paternal cousins only 4. Yes - maternal and paternal cousins
★**7**★	Are both males and females affected?	1. Yes 2. No - only males 3. No - only females
★**8**★	Is onset of disease < or = 20yrs in males?	Yes/No
★**9**★	Do any females have asymmetric disease?	Yes/No
10	In general, do females have less severe or later onset of disease?	Yes/No
11	Is there more than one retinal diagnosis in the family? (ex. Stargardt and Pattern Dystrophy)	Yes/No
12	Is consanguinity present?	Yes/No

KELLOGG EYE CENTER
MICHIGAN MEDICINE

Machine learning methodology

Gradient-Boosted Tree

Machine learns appropriate weight for each branch
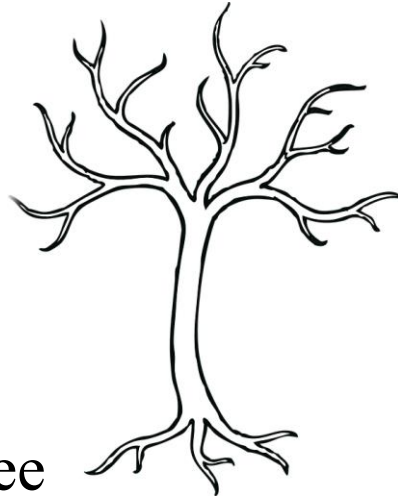
Decision tree

Machine learning methodology

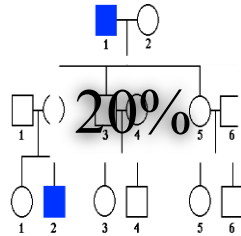## 80/20 training/testing split

Training

80%
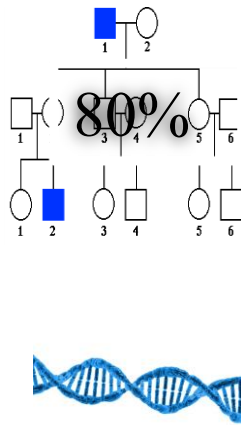
Machine learning

Classifier

Testing

20%

Predicted pattern of inheritance

KELLOGG EYE CENTER
MICHIGAN MEDICINE

Results

Method	Accuracy
Human-predicted	84%
Machine learning with human-entered answers	74%
Machine learning with computer-extracted answers	72%

Challenges

- Small dataset
  - Limited to patients with definitive genetic diagnosis
- Missing data values
  - Some questions were discarded for due to limited information
- Machine learning, but human-written questions
  - Our assumptions about the most important questions to ask may not always be correct
  - Is it better to ask more questions or fewer?
- Machines can make mistakes, too
  - Imputation bias
  - Attributing importance to unimportant features (worse with small dataset)
- Perfect prediction is impossible, even for experts
  - Ex. Isolated cases

Future Directions

- Collect more data from other institutions
  - Machine learning relies on large datasets for sufficient training
- As data collection increases, adjust questions that are informative/non-informative
  - Our expectations about what questions would be most useful might not have been correct
- Use machine learning directly on pedigree, without answering questions
  - Use statistical analysis to supplement or substitute for machine learning methodology

Acknowledgements

- Kellogg Eye Center
  - Thiran Jayasundera, MD
  - Kari Branham, MS, CGC
  - Naheed Khan, PhD
  - Abigail Fahim, MD, PhD
  - John Heckenlively, MD
- Eman Al-Sharif
- eyeGENE research project

- University of Michigan Multidisciplinary Program (MDP)
  - Andrew DeOrio, PhD
  - Edmond Cunningham
  - Xinghai Zhang
  - Yaman Abdulhak
  - Benjamin Leonard Cohen
  - Binghao Deng
  - Jason Dou
  - Jiayue Lu
  - Simeng Liu
  - Ajaay Chandrasekaran
  - Lisa Jin
  - Levin Kim
  - Wenlu Yan
  - Richmond Starbuck
  - Jacob Durrah
  - Benjamin Katz
  - Wei Xu
  - Vittorio Bichucher

Funding support from MDP

KELLOGG EYE CENTER
MICHIGAN MEDICINE