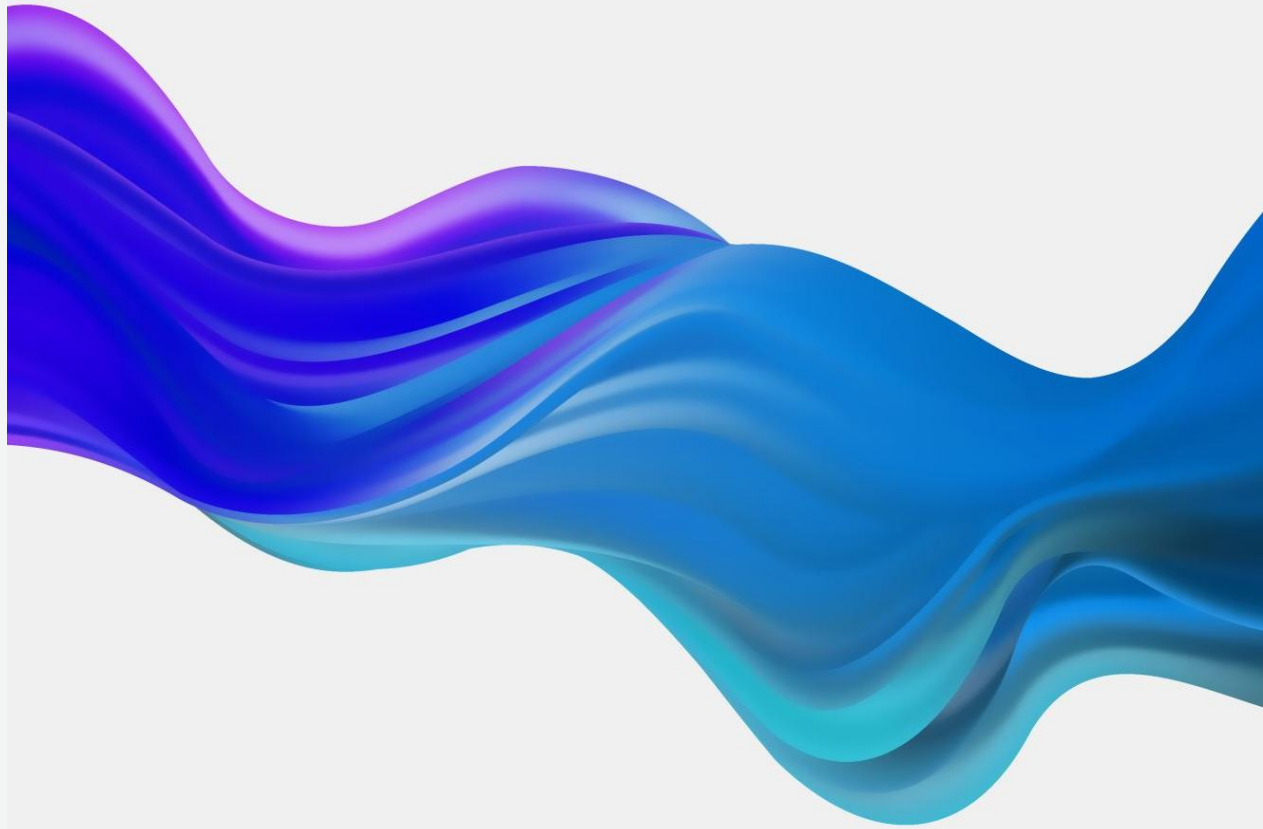
Exploratory data analysis

DISLEVE KANKU

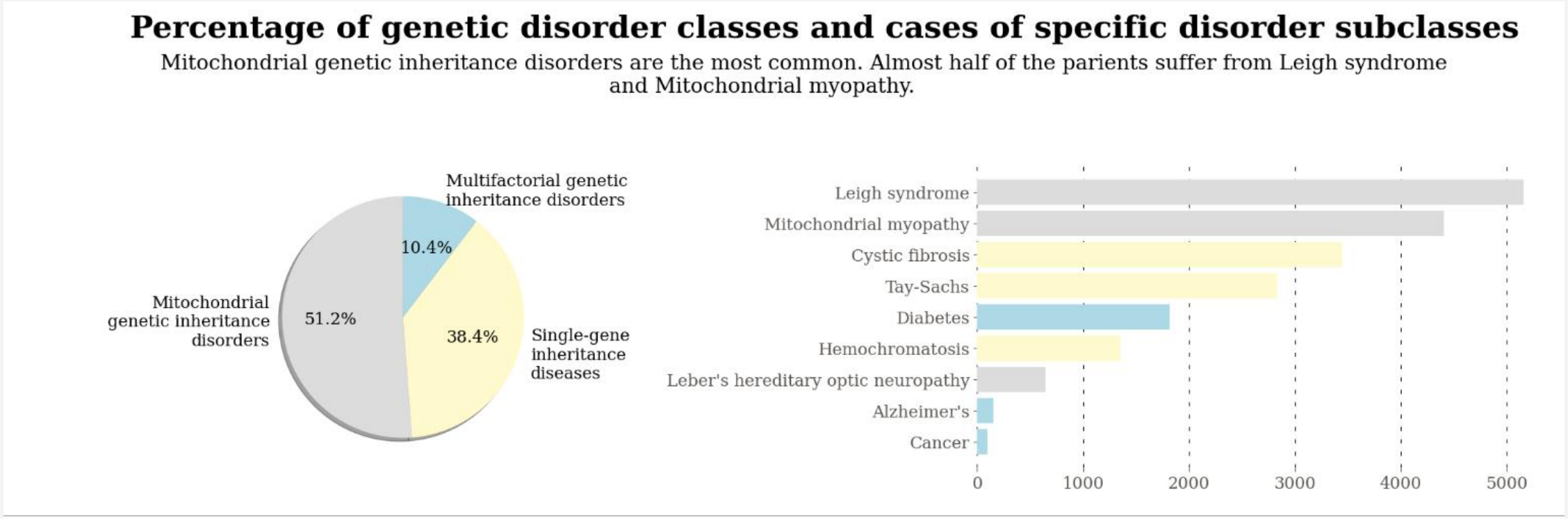


Introduction

- In this presentation, we will delve into the extensive Exploratory Data Analysis conducted as part of our Genetic Disorder Prediction project. Our aim is to unveil patterns and insights that contribute to the accurate prediction of genetic disorders

Dataset Overview

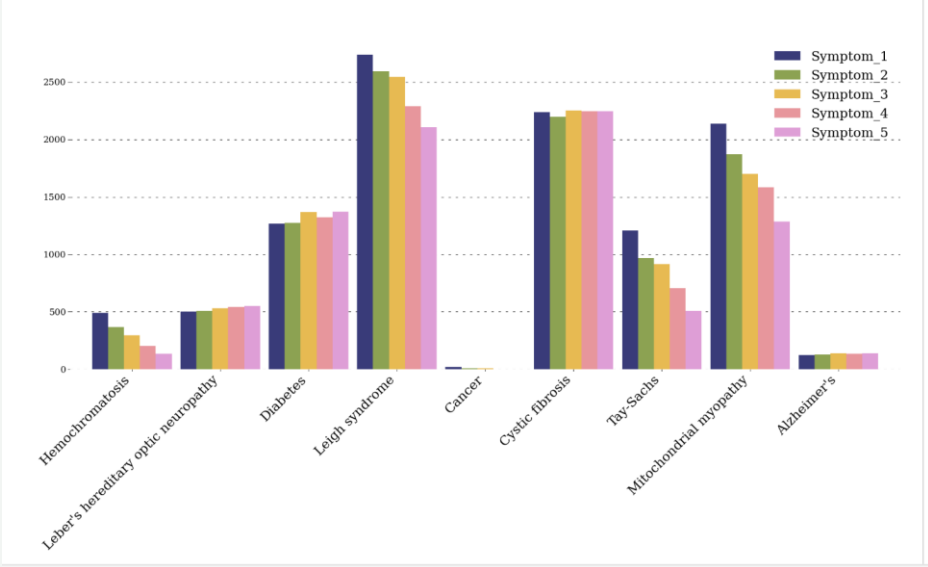
- The dataset comprises patient records, totaling 22,083 instances, sourced from diverse healthcare facilities.
- It contains a rich array of features, including patient age, genetic factors, medical test results, and, crucially, the presence or absence of a genetic disorder.



Key EDA Findings

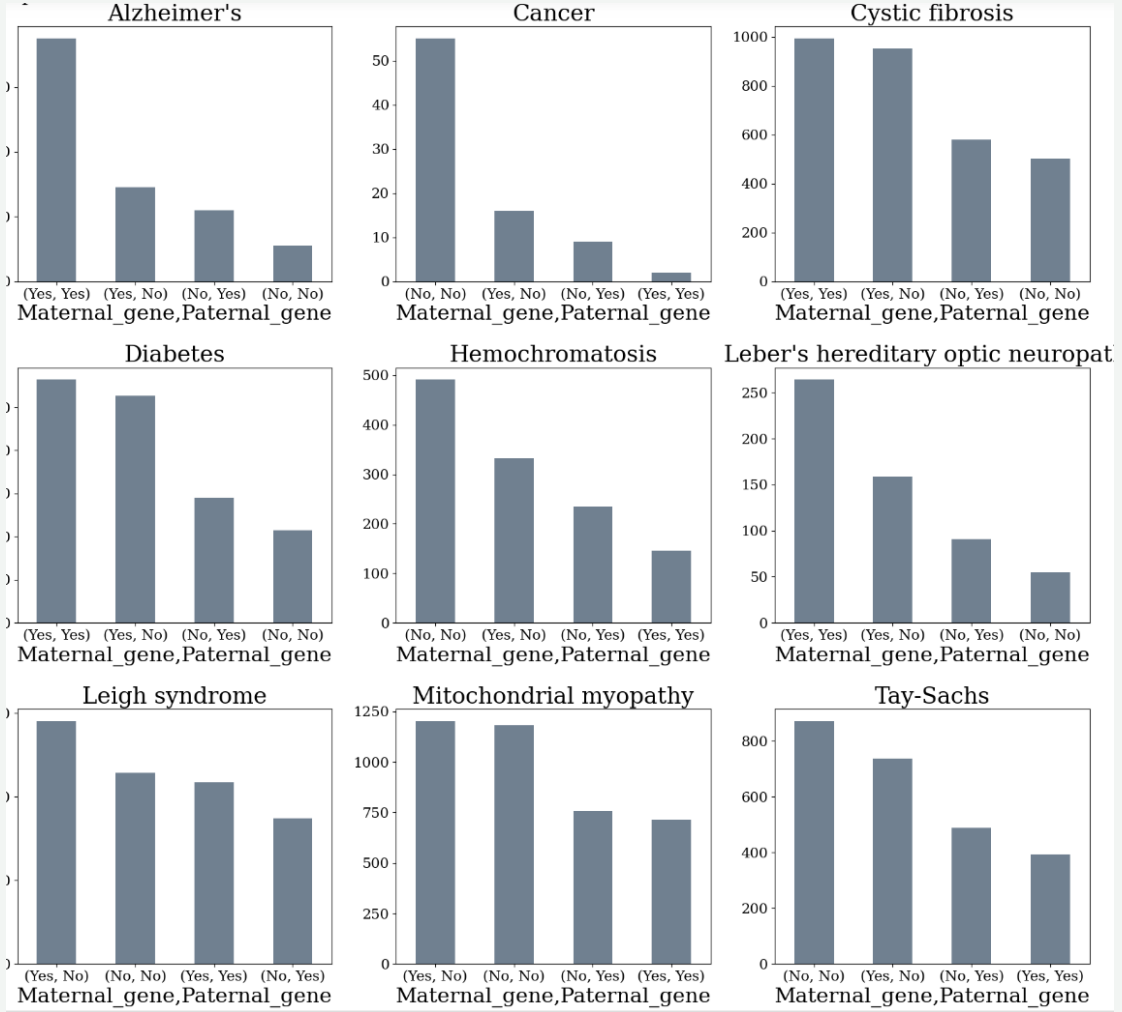
Key EDA findings

- All the symptoms appear in all of the disorder subclasses. Generally, symptoms 1, 2, and 3 are more common than symptom 4 and 5



Key EDA Findings

- Majority of diseases occur when patient carries maternal or parental defective gene, but there are some medical conditions that came as the effect of genetic mutations - most of the patients with cancer and hemochromatosis did not inherit defective genes from parents.



Conclusion

- Our comprehensive EDA has unearthed invaluable insights into patient demographics and genetic factors.
- These insights will serve as the foundation for building a robust Genetic Disorder Prediction model that holds immense potential for improved healthcare decision-making.