Early Genetic Disorder Prediction from Diagnosis and Inheritance Patterns

Emma Oo1, Sindhu Bhattarai2, and Dave Friesen3

University of San Diego

Master of Science, Applied Data Science

ADS-503 Summer 2022

June 27, 2022

Author Note

We have no conflicts of interest to disclose.

Correspondence concerning this paper should be addressed to Emma Oo, Sindhu Bhattarai , or Dave Friesen. Email: eoo@sandiego.edu, sbhattarai@sandiego.edu, or dfriesen@sandiego.edu

Abstract

[. . .]

*Keywords:* [. . .]

# Table of Contents

List of Figures n

List of Tables n

List of Equations n

Early Genetic Disorder Prediction from Diagnosis and Inheritance Patterns n

Objective n

Method n

Data Collection and Pre-Processing n

Sample Characteristics n

Multivariate Analysis n

Results n

[topic-specific section] n

[topic-specific section] n

[topic-specific section] n

Discussion n

Hypothesis Review n

Strengths, Weaknesses, and Opportunities n

References n

Appendix A – Project Code n

Appendix B – Presentation n

# List of Figures

**Figure 1** [figure name] n

# List of Tables

**Table 1** [table name] n

# List of Equations

**Equation 1** [equation name] n

# Early Genetic Disorder Prediction from Diagnosis and Inheritance Patterns

[introduction/overview; “Include a clearly defined problem statement.”; “Problem statement and justification for the proposed approach”]

Genetic diseases are the leading causes of childhood mortality (Clark, M. M, et al., 2018). In the United States, 15% of the mortality rate of the infants admitted to neonatal, pediatric, and cardiovascular intensive care units (ICUs) are due to the genetic diseases. Zarocostas, J. (2006) also reported that an estimated 3.3 million children under the age of 5 die each year from serious birth defects. Thus, rapid diagnosis and prognosis of genetic disorder of the children are very critical since it can alleviate the consequences of the getting late treatment. Wojcik, M. H., et al., 2018 mentioned that aapproximately one-third of diagnosed infants died before the diagnostic results highlighting the need to improve genetic diagnostic evaluation in the NICU, particularly to support end-of-life decision making.

Genetic disorders of the children can range from mitochondrial genetic inheritance disorders such as Alzheimer’s disease, diabetics, cancer, multifactorial genetic inheritance disorders such as cleft palate, Arthritis, high blood pressure and high cholesterol, single gene recessive disorders such as Cystic fibrosis, Sickle-cell anemia, Tay-Sachs, Hemochromatosis, Huntington disease, and developmental problems such as Autism, learning disability, mental retardation, speech problems, etc. There are many ways to perform the genetic testing as with many genetic disorders mentioned. These testing includes predictive genetic testing, Presymptomatic genetic testing, prenatal diagnosis, career testing, and newborn screening, etc. Among them, predictive genetic testing is beneficial since it can narrow down the disorders and tell the patients their chances of having it. The patients can then follow up with more thorough testing and alleviate the pain of late treatment. Thus, this study will implement the predictive genetic testing using predictive modeling techniques to classify the genetic disorders for the children.

Objective

[objective and hypothesis]

The goal of this study is to classify the accurate disorder from a group which includes Alzheimer’s, cancer, Cystic fibrosis, diabetics, Hemochromatosis, Leber’s hereditary optic neuropathy, Leigh syndrome, Mitochondrial myopathy, and Tay-Sachs.

# Method

[introduction; “Describe your analysis of the data.”; “EDA (graphical and non-graphical representations of relationships between response variable and predictor variables)”; “Data wrangling and pre-processing (handling of missing values, outliers, correlated features, etc.)”; “Data splitting (training, validation, and test sets)”]

## Data Collection and Pre-Processing

[. . .]

## Sample Characteristics

[. . .]

## Multivariate Analysis

[. . .]

# Results

[introduction; “Include a detailed solution.”; “Model strategies (describing main research questions and appropriate analytics methods)”; “Validation and testing (model tuning and evaluation)”; “Results and final model selection (performance measures, etc.)”]

## [topic-specific section]

[. . .; “You can use the case studies from Chapters 10 and 17 on how to build your predictive modeling analysis.”]

(1)

## [topic-specific section]

[. . .]

## [topic-specific section]

[. . .]

# Discussion

[“Discussion and conclusions (address the problem statement and suggestions/solutions could go beyond the scope of the course)”]

## Hypothesis Review

[. . .]

## Strengths, Weaknesses, and Opportunities

[. . .]

# References

Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., & Kingsmore, S. F. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. NPJ genomic medicine, 3(1), 1-10.

Wojcik, M. H., Schwartz, T. S., Yamin, I., Edward, H. L., Genetti, C. A., Towne, M. C., & Agrawal, P. B. (2018). Genetic disorders and mortality in infancy and early childhood: delayed diagnoses and missed opportunities. *Genetics in Medicine*, *20*(11), 1396-1404.

Zarocostas, J. (2006). Serious birth defects kill at least three million children a year. BMJ, 332(7536), 256.

# Appendix A

Project Code

# Appendix B

Presentation