Statistical Model To Identify Factors Leading To Autism In Children With Application On Pediatric Hospital Ain Shams University

A thesis submitted in partial fulfillment of the requirements for the Master Degree in Applied Statistics

By

Kholoud Ahmed Maher Hamed Ahmed

Demonstrator - Statistics, Mathematics & Insurance Department
Faculty of Commerce – Ain Shams University

Under supervision of

Prof. Dr. Medhat Mohamed Ahmed Abdel Aal
Professor – Statistics, Mathematics & Insurance Department
Faculty of Commerce – Ain Shams University

Prof. Dr. Menan Abd El Maqsoud Rabie
Professor – Psychiatry Department
Faculty of Medicine – Ain Shams University

Dr. Essam Fawzy Aziz
Associate Professor – Statistics, Mathematics & Insurance Department
Faculty of Commerce – Ain Shams University

2017

Approval sheet

Title of thesis: Statistical Model To Identify Factors Leading To Autism In Children With Application On Pediatric Hospital Ain Shams University

Academic Degree: M.Sc. in applied Statistics

Name of Student: Kholoud Ahmed Maher Hamed Ahmed

The thesis submitted in partial fulfillment of the requirements for the Master Degree in Applied Statistics has been approved by:

Examination Committee

1- **Professor Dr. Ibrahim Hassan Ibrahim**
Professor – Statistics, Mathematics & Insurance Department
Faculty of Commerce – Helwan University

2- **Professor Dr. Amr Ibrahim Abd El-Rahman El-Atraby**
Dean of Faculty of Commerce – Ain Shams University
Professor – Statistics, Mathematics & Insurance Department
Faculty of Commerce – Ain Shams University

3- **Professor Dr. Medhat Mohamed Ahmed Abdel Aal**
Professor – Statistics, Mathematics & Insurance Department
Faculty of Commerce – Ain Shams University

4- **Professor Dr. Menan Abd El Maqsoud Rabie**
Professor – Psychiatry Department
Faculty of Medicine – Ain Shams University

5- **Dr. Essam Fawzy Aziz**
Associate Professor – Statistics, Mathematics & Insurance Dept.
Faculty of Commerce – Ain Shams University

Date of thesis Defense 11 / 1 / 2017 **Approval date / / 2017**

To the Soul of my father.....

Kholoud Ahmed Maher

Acknowledgement

First and foremost, I would like to thank ALLAH for giving me the opportunity and strength to accomplish this research work.

I strongly acknowledge the contribution of **Prof. DR. Ibrahim Hassan Ibrahim** for his valuable time, patience, great support and professionalism which added a great value to the study.

I would like to express my sincere thanks and extreme gratitude to **Prof. DR. Amr Ibrahim Abdelrahman Elatraby** for all what I learnt from him throughout my academic years and I would like to thank him for agreeing to be a member of this committee and for his valuable time and advices which added a great value to the study.

No words can be sufficient to express my deep gratitude and appreciation to **Prof. DR. Medhat Mohamed Ahmed Abdel Aal** for granting me the privilege of working under his supervision and for his great support, valuable advices and continuous encouragement during all my academic years and throughout this work. Without his guidance and assistance I could never have been able to complete my thesis successfully.

I am so grateful and greatly thankful to my respected supervisor **Prof DR. Menan Abdel Maqsoud Rabie** for her generous support in the medical aspects of this work also for her great effort, careful supervision, continuous advice and guidance.

I sincerely would like to express my deepest gratitude and thanks to my respected supervisor **Prof. DR. Essam Fawzy Aziz** for

granting me the privilege of working under his supervision also for his careful supervision, continuous encouragement, and great support. His sincere effort and help will never be forgotten.

Many special thanks go to my colleagues for their help and goodwill.

Last but not least, I would like to express my warm gratitude to my mother and sister for their trust, kindness and unfailing support also I dedicate this thesis to the soul of my father may ALLAH grant him highest paradise (Ameen). Very special thanks to my beloved husband Ahmed and my children Mohammed and Omar who have always loved me, supported me and who have made untold number of sacrifices to help me accomplish this work.

Kholoud Ahmed Maher

Abstract

Kholoud Ahmed Maher Hamed Ahmed

Statistical Model To Identify Factors Leading To Autism In Children With Application On Pediatric Hospital Ain Shams University

Master Degree in Applied Statistics
Ain shams University, Faculty of Commerce
Statistics, Mathematics & Insurance Department
2017.

Autism spectrum disorders (ASDs) represent a group of neurodevelopmental disorders characterized by impairments in verbal and non-verbal communication, social withdrawal and stereotypical behaviors, which may or may not be associated with cognitive deficits, self-injurious behaviors and other neurological comorbidities. The increase of ASDs prevalence cannot be fully related to advances in diagnostics or sudden genetic shifts whereas there is a growing agreement among clinicians and scientists that ASDs resulted from an interaction between genetic and environmental factors.

One environmental factor that has received great attention is the body burden of heavy toxic metals as mercury, lead and aluminum. Heavy metals exposure is an increasing global problem and many previous studies demonstrated that heavy metals induce deleterious effects in humans.

This study drew a comparison between the results obtained on a given set of data gathered on a sample of Egyptian autistic children

against age and sex matched healthy controls using different statistical and data mining techniques being represented in parametric and nonparametric methods which are Logistic regression, Discriminant analysis, Classification and regression tree (CART), Artificial neural network and Random forests to determine the possible risk factors that may lead to autism and to reach the ideal model to be used as a tool to predict autism occurrence.

The variables studied in this thesis are House age, Age at conception, Consanguinity, Aluminum pans, Hair Lead level, Hair Mercury level, Dental amalgam, Passive smoking, Fish consumption and Gender. A split sample cross validation method is used to assess the validity of the classification models and it splits the sample into training and testing samples representing 75.7% and 24.3% of the total sample.

The traditional statistical techniques applied in this study represented in Forward Stepwise logistic regression and Stepwise discriminant analysis revealed that they have the same final model variables which are: House age, Age at conception, Aluminum pans, Consanguinity, and Hair lead level. Also, The data mining non parametric statistical techniques represented in Artificial neural network, Classification and regression tree (CART) and Random forests revealed better classification accuracy where the results sobtained according to the training data set showed that the artificial neural network has the best performance in the establishment of the prediction model and its classification accuracy is 82.08%, ***it is the recommended classifier for the diagnosis of autism.*** Moreover, the Classification and Regression trees (CART) as well as Random Forest models have the same classification accuracy which is equal to

78.30%. While the traditional models represented in Forward Stepwise logistic regression and Stepwise discriminant analysis models have accuracy rate of 77.40% and 76.40%% respectively showing less classification accuracy than those of the nonparametric data mining techniques.

Furthermore, the study showed that the classification accuracy of the cross validated group cases (Testing data set) for Logistic regression and Discriminant analysis is 76.50% and 79.40% while for Artificial Neural Networks, Classification and Regression Trees (CART) and Random forest is 79.41%, 76.47%, 79.41% respectively.

Key words: Autism, House age, Age at conception, Consanguinity, Hair Lead level, Hair Mercury level, Dental amalgam, Logistic regression, Discriminant analysis, Artificial Neural Networks, Classification and Regression Trees (CART) and Random forest.

Table of Contents

Chapter 3: Application of Statistical Techniques

List of Tables

List of Figures