Semester Project Proposal

Machine Learning-Based Autism Prediction System

January 27, 2025

Submitted By:

Badar Rasheed Butt	211370113
Moaz Aslam	211370125
Ayesha Kanwal	211370178
Shamama Tarif	211370169

Submitted To:

Sir Usman Ali

Department of Computer Science

Gift University

Contents

1 Problem Statement 3

2 Introduction 3

3 Dataset Description 3

3.1 Files 3

3.2 Columns 4

4 References 4

1. Problem Statement

Autism Spectrum Disorder (ASD) is a developmental condition characterized by challenges in social skills, repetitive behaviors, and communication. The current diagnostic process is often time-consuming, requiring years for a confirmed diagnosis, which delays early intervention. Machine learning offers a powerful means to predict the likelihood of ASD and assist healthcare professionals in prioritizing resources for early diagnosis and treatment. This project aims to develop an AI-based prediction system for Autism using the given dataset to improve the efficiency and accuracy of autism screening.

2. Introduction

Autism Spectrum Disorder affects individuals in diverse ways, making early diagnosis crucial for effective intervention. High-quality early intervention can significantly improve learning, communication, and brain development. However, the diagnostic process—often influenced by a mix of genetic and environmental factors—remains complex. Machine learning can leverage patterns in behavioral data to predict autism likelihood efficiently. This project utilizes survey data collected from more than 700 individuals to build a predictive model for autism diagnosis, thereby aiding healthcare providers in early detection and resource optimization.

3. Dataset Description

The dataset for this project is sourced from REVA University's Autism Prediction Challenge. It includes survey results from over 700 individuals, labeled to indicate whether a person received an autism diagnosis. The key features of the dataset are:

3.1. Files

- **train.csv**: Training dataset
- **test.csv**: Testing dataset
- **sample_submission.csv**: Sample submission file in the correct format

3.2. Columns

- **ID**: Unique identifier for each patient
- **A1_Score to A10_Score**: Scores based on the Autism Spectrum Quotient (AQ) 10-item screening tool
- **age**: Patient's age in years
- **gender**: Patient's gender
- **ethnicity**: Patient's ethnicity
- **jaundice**: Whether the patient had jaundice at birth (Yes/No)
- **autism**: Whether an immediate family member has been diagnosed with autism (Yes/No)
- **country_of_res**: Patient's country of residence
- **used_app_before**: Whether the patient has undergone a screening test before
- **result**: Total score for AQ1-10 screening tool
- **age_desc**: Age category of the patient
- **relation**: Relationship of the person completing the test
- **Class/ASD**: Target variable indicating autism diagnosis (1 = Yes, 0 = No)

4. References

- Tabtah, F. (2017). Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment. *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*, pp.1-6. Taichung City, Taiwan, ACM.
- Thabtah, F. (2017). Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review. *Informatics for Health and Social Care Journal*. December 2017.

- Tensor Girl. Autism Prediction. https://kaggle.com/competitions/autismdiagnosis, 2022. Kaggle.