Using Machine Learning To Mitigate/Predict Gender Biases in ADHD and ASD Diagnosis

Harleen Kaur Sethi Iyer

Submitted in partial fulfilment of the requirements of the degree

of

MSc Digital Business and Data Analytics

Henley Business School

September, 2024

Table of Contents

Acknowledgements:

I want to thank my Supervisor, Dr. Nico Biagi for supporting me through this immensely tough topic.

Abstract:

This study aims to explore how machine learning models can help address gender biases in the diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD). Current diagnostic criteria often result in fewer females being diagnosed, especially in disorders like ADHD and ASD where symptoms may vary between genders. The study assesses the performance of 6-7 machine learning models, such as Random Forest, Logistic Regression, and XGBoost, to see how effective they are at improving diagnostic accuracy and reducing these biases.

The results indicate that machine learning, specifically the Random Forest model, can significantly improve the diagnostic process for ADHD and ASD, leading to fairer outcomes for both genders. This research demonstrates the potential for machine learning to help reduce gender biases in medical diagnostics, offering a way to achieve more precise and timely diagnoses.

Chapter 1: Introduction

1.1 Background

ADHD is a developmental disorder that's characterised by a persistent pattern of inattention, hyperactivity, and impulsivity. These symptoms can begin in childhood and interfere with daily life, including social relationships and school or work performance. Some people with ADHD have mostly inattention symptoms, while others have mostly hyperactivity and impulsivity symptoms, and some have both (Data and statistics on ADHD 2024).

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) defines ADHD as a persistent pattern of inattention and/or hyperactivity-impulsivity that meets the following criteria:

Inattention: The person often has trouble paying attention to details, following instructions, or finishing tasks. They may also have difficulty organising tasks and activities and may avoid tasks that require sustained mental effort.

Hyperactivity-impulsivity: The person's symptoms must present in two or more settings, such as at home, school, or work, and negatively impact their social, academic, or occupational functioning.

ASD is a neurological and developmental disorder that affects how people learn, communicate, behave, and interact with others. People with ASD may have trouble with social interactions and interpreting and using verbal and nonverbal communication as per the About autism spectrum disorder (2024) Centers for Disease Control and Prevention.

According to DSM-5, to be diagnosed with ASD, a child must have persistent deficits in the following three areas of social communication and interaction: (i) social-emotional reciprocity; (ii) developing, understanding, and maintaining relationships; and (iii) nonverbal communication. In addition, at least two of the following four behaviours should be present: (i) inflexible to changes in routine; (ii) restrictive or fixated interests that may be abnormal in focus or intensity; (iii) hypo- or hyperactivity in response to sensory input or abnormal fixation with sensory aspects of the environment; and (iv) repetitive movements, speech, or use of items.

ADHD and ASD are neurodevelopmental disorders that significantly impact a person if they aren't diagnosed with them at the earliest possible. Misdiagnoses or underdiagnoses are common in this disorder as the comorbidities overlap with each other. Accurate and timely diagnosis is crucial for accessing appropriate interventions and support services. However, there is substantial evidence of gender disparities in the diagnosis of these disorders, with females being significantly less likely to be identified compared to males exhibiting similar symptoms and behaviours.

The progress in machine learning shows potential for tackling gender biases in diagnosis. ML can be used to create models that take into consideration gender differences in symptom presentation, leading to a fairer diagnostic process (Bahathiq et al., 2022). This study explores how ML can help reduce gender biases in diagnosing ADHD and ASD.

1.2 Research Problem

According to a study by Bruchmüller K et al., (2012), there is a significant issue of gender bias in the diagnosis of ADHD and ASD, with current diagnostic criteria being more focused on males. Research indicates that ADHD is commonly overdiagnosed in males, and gender plays a crucial role in diagnosis. The study suggests that traditional diagnostic criteria may not adequately identify symptoms in females. Limited inclusion of females in clinical data used for developing diagnostic tools results in inaccurate diagnoses. This issue is well recognized, as diagnostic assessments and tools have primarily been evaluated on male populations in the past.

1.3 Aims & Objectives:

The main objective of this study is to investigate the potential of machine learning in addressing gender biases in the detection of ADHD and ASD. This research will focus on:

- Investigating the impact of age and gender on the diagnostic procedures for ADHD and ASD is essential for enhancing the precision and impartiality of diagnostic frameworks.
- Formulate and assess machine learning algorithms that can anticipate ADHD and ASD diagnoses - The study will emphasize integrating gender and age as significant factors to guarantee more just outcomes in diagnoses.
- Assess the efficacy of machine learning algorithms in mitigating diagnostic prejudices - Through the comparison of algorithm outputs with actual diagnostic methodologies, the study will determine if machine learning can provide tangible enhancements in diagnostic equity.

1.4 Dissertation Structure

Following Chapter 1: Introduction,

Chapter 2: Literature Review - This chapter looks at what other studies have found about how gender biases affect the diagnosis of ADHD and ASD, and how machine learning technology can help improve healthcare diagnostics.

Chapter 3: Research Design and Methodology - This chapter explains the design of the study, how data was collected, and the machine learning methods used to meet the research goals.

Chapter 4: Data Analysis and Results - This part of the report shares the findings from looking at the data and assesses how well the machine learning models performed in this study.

Chapter 5: Discussion - This chapter talks about what the findings mean concerning the research questions and previous studies, and how they can impact healthcare practices.

Chapter 6: Conclusion and Recommendations - The last chapter gives a summary of the main discoveries, discusses any limitations in the study, and suggests ideas for future research and practical use of machine learning in medical diagnostics.

Chapter 2: Literature Review

2.1 Introduction

Attention-deficit/hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are neurodevelopmental disorders that can have a significant influence on a person's life (Makris G., et al, 2023). Precise and prompt diagnosis is essential for obtaining suitable therapies and support services.

This literature review intends to examine the present state of research on gender biases in the diagnosis of ADHD and ASD, the possible sources and effects of these biases, and the growing use of machine learning (ML) approaches to reduce these biases.

2.2 Independent Variables (Age & Gender)

2.2.1 Age

Age plays a crucial role in the determination and presentation of ADHD and ASD. In the initial stages of childhood, ADHD often presents with hyperactivity and impulsivity (Vos M, et al. 2022), while in older children and adults, inattention and issues with executive function become increasingly noticeable (Siddiqui U. *et al.*2024). Early ASD diagnosis focuses on developmental milestones, but later individuals may have social interaction and repetitive behaviour issues (Malwane M. 2022).

Kentrou et al. (2019) discovered that the gender factor influenced the age at which individuals were diagnosed with ASD, with females typically being diagnosed at a later age than males. The disparities in age of diagnosis between genders were further pronounced in the presence of co-occurring conditions, resulting in males being diagnosed 1.5 years later and females 2.6 years later compared to those without additional conditions in the ASD-only group.

2.2.2 Gender

There are examples indicating that the diagnosis of ADHD and ASD differs between girls and boys. The male-to-female ratio in ASD ranges from 1.33:1 to 16:1, with a recent estimate of 4:1 (Mahendiran, 2019). In girls and women with ADHD, the less conspicuous nature of inattentive symptoms compared to hyperactive ones leads to an underdiagnosis of the disorder (Mahendiran, 2019). Similarly, females within the autism spectrum often display subtler forms of social communication challenges than their male counterparts, who tend to exhibit more overt, repetitive behaviours (Fombonne, 2021).

Achieving more equitable diagnostic practices may involve leveraging machine learning models that factor in gender as an independent variable, as suggested by Rizvi & Mrini (2022) and Bahathiq et al. (2022).

To understand if there are any gender differences in misdiagnosis or delay in diagnosis in adults with ASD, research was conducted by Gesi, C. (2021) in which they made a comparison between the demographic and clinical characteristics of males and females with ASD. Categorical variables were analysed using the Chi-square test, while continuous variables were examined using the Student T-test. An examination of the data was conducted to explore the connections between symptom dimensions and delayed diagnosis of ASD in both women and men. An analysis was performed to examine the impact of gender, misdiagnosis, and their potential interaction on AdAS (Adult Autism Subthreshold Spectrum) Spectrum domain scores. This study utilised seven two-way ANOVA analyses to investigate these effects. The results of this study were females had a longer delay in mental health referral and a higher age at ASD diagnosis than males. Next, they investigated whether males and females had distinct diagnostic delays due to different causes. They conducted correlation studies between AQ (Autism-Spectrum Quotient) and AdAS Spectrum scores, age of ASD diagnosis, and time between first encounter and ASD diagnosis for women and men. A diagnostic delay negatively linked with AdAS Spectrum total, Verbal communication, empathy, inflexibility and routine domain scores in males, but not with ASD diagnosis age. Diagnostic delay was significantly connected with Attention to detail scores in women, whereas ASD age positively correlated with AdAS Spectrum Verbal communication. Looking at ASD diagnostic history, three circumstances were identified: 1. ASD was appropriately identified at first examination; 2. ASD was overlooked; 3. ASD was misdiagnosed with other mental problems. Women were more likely to be misdiagnosed than men at first examination. In this study, it was shown that age correlated with misdiagnosis or delay in diagnosis in women.

In a study conducted by Biederman (2002), it was viewed that gender had a modest impact on ADHD as a risk factor for dysfunction related to ADHD in children and adolescents who are referred. According to this study, gender does influence the disorder's clinical presentation. The fact that girls with ADHD had a greater prevalence of inattention symptoms and were less likely than boys to have comorbid disruptive behaviour issues, learning difficulties, and social dysfunction was a major contributing factor. To adequately assess this issue, further investigation in both referred and non-referred samples of children with ADHD is needed, as these factors may lead to a gender-based referral bias unfavourable to girls.

2.3 Dependent Variables: Diagnostic Time

Demographic data can be a great way to understand how, why and when a person is diagnosed with ADHD and ASD, what time was their 1st appointment, how long it took them to get diagnosed after the first appointment and what urged them to get themselves diagnosed. Ethnicity, age, income, education and socio-economic background play a role in the diagnoses. According to the Centers for Disease Control and Prevention (2024), the diagnosis of ADHD was more common in Black and White children (12% for both) compared to Asian children (4%). American Indian/Alaska Native children were also more likely to be diagnosed with ADHD (10%) than Asian children. Around 6% of Native Hawaiian/Pacific Islander children received a diagnosis of ADHD.

In general, non-Hispanic children were diagnosed with ADHD more frequently (12%) than Hispanic children (10%). The research found that being female, younger, having a higher socioeconomic status, and being part of certain racial groups were less likely to have this comorbidity (Casseus, 2022; Casseus et al., 2023).

2.3.1 Time of Diagnoses

An important dependent variable is the time between the onset of symptoms and the official diagnosis. Predicting the time to diagnosis using age and gender-based machine learning models can assist in identifying areas where diagnostic efficiency might be enhanced. Only 15.8% of children who have both autism and ADHD had been identified with both conditions before, suggesting possible delays in recognising comorbidity. Identifying comorbidity early and accurately is vital in meeting the clinical and socio-educational requirements of children with ADHD and ASD (Canals, J. et al. 2024).

2.3.2 Co-existing Conditions

The existence of other medical issues, such as anxiety, depression, or learning difficulties, is an important factor to consider. A national survey conducted by the CDC in 2022 found that almost 78% of children with ADHD also had another accompanying condition. Nearly half of the children diagnosed with ADHD experienced behavioural or conduct issues. Around 40% of children with ADHD also deal with anxiety. In addition to ADHD, other disorders commonly found in affected children are depression, autism spectrum disorder, and Tourette syndrome. People with both ADHD and ASD often have co-occurring disorders, which makes diagnosis and treatment more challenging. Developing more thorough treatment plans that cover several facets of a patient's health might be facilitated by an understanding of concurrent condition patterns. A study conducted by McGough (2005), shows that parents with ADHD had a higher prevalence of lifelong psychopathology, with 87% having at least one, and 56% having at least two additional mental illnesses. In comparison, the non-ADHD participants had rates of 64% and 27%, respectively. The ADHD participants had early onset of conduct disorder, dysthymia, oppositional defiant disorder, and major depressive disorder. The average ages at which oppositional defiant disorder (ODD) and conduct disorder manifest align with accounts of disruptive behaviour disorders appearing earlier in kids afflicted by ADHD. The average ages at which major depressive disorder and dysthymia begin do not include childhood, indicating that these diseases do not stem from childhood psychopathology. Sex differences were seen, with females exhibiting a larger susceptibility to mood and anxiety disorders, while males showed a greater vulnerability to substance use disorders, regardless of their ADHD status. However, it seems that individuals with ADHD are more susceptible to experiencing adult depressive disorders at an earlier stage of development. They couldn't find evidence supporting a younger age at which drug use problems begin. However, they did see a greater incidence of substance use disorders compared to unaffected individuals.

Casseus (2022) conducted a study using a U.S. sample to determine that 1.2% of children between 3 and 17 years old had both autism and ADHD, with sociodemographic factors playing a role in this combination. Patients who were diagnosed with ADHD as adults have reported a more comprehensive depiction. Shekim and colleagues (1990) evaluated 51 adults with ADHD who were referred for clinical assessment. They discovered that a significant number of these individuals had experienced other mental health conditions throughout their lives. Specifically, they found that 51% had also been diagnosed with generalised anxiety disorder, 34% had a history of alcohol abuse or dependence, 34% had a history of other drug abuse or dependence, 25% had experienced dysthymia, 18% had a separation anxiety disorder, 13% had obsessive-compulsive disorder (OCD), and 10% had a major depressive disorder. Downey and colleagues (1997) conducted a study where they analysed the psychological profiles of 78 persons with ADHD who were referred for professional treatment. They found that 37% of these individuals were now experiencing depression, 33% had a history of alcohol addiction or dependence, 22% had a history of other substance abuse or dependence, and 13% exhibited symptoms of antisocial personality disorder. These trials were conducted without any control group based on cohorts of patients requesting therapy.

2.3.3 Diagnostic Score

This score can help in understanding where a patient is currently in their journey. Using test results from instruments like the Autism Diagnostic Observation Schedule and the ADHD Rating Scale are examples of quantitative dependent variables (Bloomdes, 2019). These ratings are crucial for assessing the precision of diagnostic models, as they include comprehensive details on the severity of symptoms. The National Institute for Health and Care Excellence (NICE, 2012) recommends the AQ-10 for use with adults with possible autism who do not have a moderate or severe learning disability. However, it is emphasized that this is to help identify individuals for referral, not for diagnosis.

2.4 Relation between IV and DV

In ADHD and ASD, gender and age seem to correlate (Mahendiran, 2019). When symptoms appear, how long it takes to diagnose them, and whether or not an intervention is suitable are all affected by age. For instance, in a study by Mahendiran (2019), older females have a hard time being socially adaptive compared to younger females with ASD.

Because of biases in referrals and variations in symptom presentation, these diseases are more commonly diagnosed in men than in women. While males with ADHD are more likely to display obvious hyperactive behaviours, females with the same disorder are more likely to have inattentive symptoms, which make it more difficult to identify (ADHD in women: Symptoms, treatment, and support, 2024).

In a research conducted by Arnett, A.B. (2014) to verify the higher occurrence of ADHD in males within their sample, researchers randomly chose one participant from each household ("subsample"; n=1,074) to conduct group comparison studies. The subsample exhibited similar characteristics to the complete sample across all demographic factors. As anticipated, males showed more severe scores than females on the measures of inattention (male mean(SD) = 8.44(6.40); female mean(SD) = 5.15(5.31); $t(1072)=9.21$, $p<.001$), hyperactivity/ impulsivity (male mean(SD) = 5.24(5.07); female mean(SD) = 3.19(3.92); $t(1072)=7.44$, $p<.001$), and total ADHD (male mean(SD) = 13.72(10.40); female mean(SD) = 8.34(8.49); $t(1072)=9.32$, $p<.001$) scores.

By examining larger datasets to discover patterns and enhance diagnostic accuracy, ML presents a potential strategy for tackling these biases. Incorporating age and gender as independent variables into ML models allows for adjusting developmental and gender-specific differences, resulting in more accurate and fair diagnoses.

2.4.1 Possible gaps

The underrepresentation of females in ADHD and ASD studies is a consequence of the lower sample sizes employed for research. Including women in studies and research would be a good starting point for reducing gender bias in diagnosis.

Since there is a lack of substantial data that covers the symptoms that females experience in ADHD and ASD, ML might also be biased in this regard.

Due to multiple co-existing diseases, it could get tough to diagnose ADHD and ASD as they may hinder their diagnosing process. For instance, in a study that looked at the link between early-life allergies and later-life ASD and ADHD diagnoses in children, researchers discovered that atopic disorders in infancy were associated with a higher risk of ASD and ADHD later in life (Lamanna, 2017).

2.5 Conclusion

The literature review stresses the significance of age and gender in the diagnosis of ADHD and ASD, emphasising how these characteristics impact the manifestation of symptoms and the timing of diagnosis. Machine learning (ML) can enhance diagnostic accuracy and promote fairness by incorporating these factors. The necessity for demographic variety, the absence of longitudinal data, and the underrepresentation of women in research are all major gaps that have not been adequately addressed. To advance diagnostic techniques and improve patient outcomes, it is crucial to address these gaps by conducting inclusive research and designing ML models that can be easily understood.

Chapter 3: Research Design and Methodology

3.1 Introduction

As per the American Psychiatric Association (2013), Attention Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) are neurodevelopmental conditions that present significant challenges for affected individuals. ADHD is characterised by symptoms of inattention, hyperactivity, and impulsivity, whilst ASD impacts communication, behaviour, and social interaction. These disorders are typically diagnosed during childhood and profoundly affect social, academic, and occupational functioning throughout one's life. Early diagnosis and intervention are critical for managing these conditions and enhancing quality of life (Avlund, et al., 2021). However, considerable gender disparities exist in the diagnosis of ADHD and ASD, with females being more likely to be underdiagnosed or misdiagnosed compared to males. (Gesi, et al., 2021) (Moura, et al., 2023).

The criteria used for diagnosing ADHD and ASD were mainly established from research conducted on males, leading to a tendency to overlook these conditions in females (Dan, 2021). Research has shown that the conventional diagnostic models may not effectively identify symptoms in women (Dan, 2021). The consistent underdiagnosis of ADHD and ASD in females can be attributed to their symptoms being less noticeable compared to males, such as subtle social difficulties in ASD and less obvious signs of inattentiveness in ADHD, resulting in delayed or inaccurate diagnoses (Dan, 2021).

Autistic women appear to exhibit a "camouflage" effect, displaying a high level of functioning, fewer abnormal play patterns or restricted interests, improved socio-emotional reciprocity, and coping mechanisms. As a result, women with ASD frequently receive multiple diagnoses before receiving the necessary support and care, causing delays (Moura, et al., 2023). Females often display more subtle symptoms, such as inattentiveness in ADHD and more nuanced social impairments in ASD, which can lead to delayed or inaccurate diagnoses (Gesi, et al., 2021).

Recent advancements in Machine Learning (ML) offer an opportunity to mitigate such biases by developing predictive models that account for gender differences in symptom presentation. This research design focuses on leveraging ML techniques, such as decision trees and neural networks, alongside bias-mitigation algorithms, to develop fair models for diagnosing ADHD and ASD across genders.

3.2 Project Background

A 2021 study by Gesi explored gender differences in the misdiagnosis and delayed diagnosis of ASD in adults, comparing males and females based on demographic and clinical characteristics. The study used the Chi-square test for categorical data, the Student T-test for continuous data, and seven two-way ANOVA analyses to examine the impact of gender, misdiagnosis, and their interaction on AdAS Spectrum scores.

Key findings indicated that females experienced longer delays in mental health referrals and were diagnosed with ASD at a later age compared to males. The study further investigated the reasons behind these delays, discovering that in males, diagnostic delays were negatively associated with AdAS Spectrum scores in areas such as verbal communication, empathy, inflexibility, and routine, but not with the age of diagnosis. In females, delays were linked to higher attention to detail scores, and older age at diagnosis was positively correlated with verbal communication scores.

The study identified three diagnostic scenarios: correct initial diagnosis, overlooked ASD, and misdiagnosis with other mental conditions. It was found that women were more likely to be misdiagnosed at their first examination. Additionally, age was found to correlate with the likelihood of misdiagnosis or delayed diagnosis in women.

Similarly for ADHD, a study was conducted by Biederman (2002) found that while gender modestly impacts ADHD as a risk factor for dysfunction, it significantly influences its clinical presentation. Girls with ADHD were more likely to exhibit inattention symptoms and less likely than boys to have comorbid disruptive behaviour issues, learning difficulties, and social dysfunction. The study suggests a need for further research in both referred and non-referred children with ADHD, as these gender differences may result in a referral bias that disadvantages girls.

Current ML approaches in diagnosing ADHD and ASD focus on behavioural data, cognitive tests, neuroimaging and sMRI to predict symptoms. Techniques such as Support Vector Machines (SVMs), decision trees, and neural networks have been employed with moderate success (Ghasemi et al., 2022; Rizvi & Mrini, 2022). However, few studies address gender biases directly. Research suggests that the underrepresentation of females in training datasets leads to biased outcomes (Mikolas et al., 2022).

Finding datasets that accurately distribute the population between both genders including diagnosed and control groups. Typically finding such datasets is tough as there aren't many females who are diagnosed with either of the disorders easily.

3.3 Project Aims & Objective:

Aims: The primary aim of this research is to develop and evaluate ML models that mitigate gender biases in diagnosing ADHD and ASD, providing more equitable diagnostic tools for healthcare practitioners.

As I cannot get datasets from NHS, open-source datasets are my only option. I can merge datasets to get an accurate number of participants.

Objectives:

a) Diverse datasets: Search for a dataset that represents either ADHD/ASD symptoms or their diagnosis score to understand the level of severity. This dataset should include diverse demographic information like ethnicity, country of residence, income, comorbidities, and family history of the disease.
b) Explore Gender-related patterns: Use Exploratory Data Analysis to understand the relationship between gender and diagnosis. Identifying specific patterns between age, gender and disorder.
c) Exploring Other Patterns: If age and gender do not influence the diagnosis, demographic variables may play a role. Identifying patterns in family history and diagnosis. If ethnicity plays a role in the condition or if a country of residency influences the diagnosis.
d) Develop Interpretable ML Models: Create interpretable models (e.g., decision trees, random forests, and neural networks) to uncover the particular elements that cause biased predictions across genders.

3.4 Research Approach and Methodology:

The research methodology will follow a structured approach combining data collection, exploratory data analysis (EDA), ML model development, and validation.

3.4.1 Data Collection and Preparation:

The initial phase of the process entails gathering extensive and varied datasets that accurately reflect the differences in ADHD and ASD between genders. These datasets may be sourced from a range of places such as open-access databases, clinical records, and publicly accessible neuroimaging datasets. Once collected, thorough cleaning and preprocessing procedures will be applied to the datasets. This will involve eliminating any biases, addressing missing data points, and ensuring an equal distribution of gender, age, and demographic variables within the datasets.

3.4.2 Exploratory Data Analysis (EDA)

After gathering the data, EDA will be carried out to examine how gender is related to ADHD/ASD symptom presentation. Statistical techniques will be used to find any significant trends such as whether there are milder signs of ASD among girls and lack of focus among boys which fail to detect the disorders early enough. The EDA stage also requires feature selection and ranking methods that can help identify the variables (e.g., age, symptom intensity, family history) most predictive of diagnostic outcomes between both genders. Correlation matrices will be used along with other tools like heat maps and regression analysis to visualize and quantify these age-related and gender-related differences.

3.4.3 Model Development

Developing the ML models with transparent techniques will allow clinicians to easily grasp the decision-making process. Train models like decision trees, XGBoost, and neural networks will be made on the gathered data to forecast ADHD and ASD diagnoses. In healthcare, interpretable models shine, fostering transparency and trust in AI-driven decisions. For ASD, 6 ML models will be made to understand how gender and age impact the diagnosis, those models will be tuned to check if there is any overfitting in training or testing split. For

ADHD 7 ML models will be made including the Gaussian Naïve Bayes model which is primarily used for the classification of data points into different categories based on probability estimates, highly recommended for medical diagnosis. Followed by hyperparameter tuned models to check for an even split between training and testing.

3.4.4 Model Validation and Evaluation

The main goal of this study is to develop machine learning algorithms that can reduce gender biases when diagnosing ADHD and ASD. In addition to looking at accuracy, the research will also focus on fairness, precision, recall, and disparate impact analysis. Fairness is important to ensure that diagnostic results are equally beneficial for individuals of all genders. Precision and recall are essential for finding the right cases and avoiding false positives. Disparate impact analysis will examine how different demographic groups are affected by the predictions made by the models, to achieve unbiased outcomes. These metrics play a critical role in creating diagnostic tools that are inclusive, transparent, and ethical. The ultimate goal is to make sure that these models serve everyone equally regardless of gender and to address any underlying biases in healthcare.

3.5 Expected Research Outcomes:

The results of this study are expected to have implications in both academic and real-world settings.

Enhanced Precision in Diagnosis for Females: The study aims to reduce any prejudices present in machine learning algorithms, which in turn will enhance the accuracy and promptness of diagnoses for females with ADHD and ASD.

Currently, these individuals are underserved by the existing diagnostic methods, highlighting the necessity for improvement in diagnostic processes.

Development of Fair Diagnostic Instruments: The new models being created will equip healthcare professionals with advanced tools that can effectively diagnose ADHD and ASD in individuals of all genders. This advancement is crucial in ensuring that healthcare outcomes are fair and unbiased across the board.

3.6 Deliverables:

Building on the example of NHS costs: If the average price for a consultation is £35, and patients typically need 3-5 consultations for a correct diagnosis, the current cost per patient varies from £105 to £175. As per the NHS, the referral waiting times in June 2024, were 187,567 patients with an open referral for suspected autism. Of these, 163,666 (87.3%) had a referral that had been open for at least 13 weeks. Assuming that 100,000 patients seek diagnosis yearly, the overall expense to the healthcare system could reach £10.5 million to £17.5 million. If the new AI-powered diagnostic tool could decrease the number of consultations needed by just one, it could potentially save the healthcare system £3.5 million on an annual basis.

Furthermore, quicker and more precise diagnoses could result in timelier interventions, potentially lowering the long-term healthcare expenses linked to untreated or incorrectly diagnosed conditions.

Enhanced Diagnostic Precision Guidelines: A collection of performance standards that healthcare providers can utilise to evaluate and enhance their diagnostic precision for both genders. These standards could be assimilated into the current healthcare quality enhancement strategies.

Decreased healthcare expenses due to more effective diagnosis procedures and timelier interventions. Possible boost in productivity as individuals receive suitable assistance and treatment sooner.

Social Consequences: Improved standard of living for individuals with ADHD and ASD as a result of prompt and accurate diagnoses. Reduction in gender inequalities in mental health diagnosis and therapy.

Chapter 4: Data Analysis and Results:

4.1 Autism Spectrum Disorder:

4.1.1. Overview

This section outlines the data analysis and findings about ASD. The main objective of this analysis was to uncover any gender biases in the diagnosis of ASD and to create machine learning algorithms that are better equipped at identifying ASD in females, who frequently receive inadequate diagnoses or incorrect diagnoses as a result of more subtle symptom manifestations. In addition to investigating gender imbalances, I am seeking to explore potential correlations between age and ASD diagnoses, as well as other contributing factors. The dataset utilised in this study was sourced from Kaggle. (https://kaggle.com/competitions/autismdiagnosis)

4.1.2. Data Preparation

The Autism Spectrum Quotient has 10 questions with 4 choices for each statement (definitely agree, slightly agree, slightly disagree, definitely disagree). The questions range from sensory issues to social issues. This test is recommended in ‘Autism: recognition, referral, diagnosis and management of adults on the autism spectrum’ (NICE clinical guideline CG142). www.nice.org.uk/CG142

This dataset comprises survey responses from over 700 individuals who completed an application form. The dataset includes indicators indicating whether the individual has been diagnosed with autism, enabling machine learning algorithms to forecast the probability of autism diagnosis (Tensor Girl, 2022). The dataset has 800 entries with 22 columns. The average age of participants is 28.45 with a standard deviation of 16.31. There are 530 male participants and 270 female participants. The dataset includes 639 participants with autism and 161 participants without autism.

Columns like ‘ID’ and ‘age_desc’ had no use for the dataset and were hence removed to lessen the number of columns.

A Chi-square test for independence was conducted to examine the relationship between gender and ASD diagnosis. The test yielded a non-significant result, $\chi^2(1, N = [800]) = 0.00$, $p = .976$, indicating that there is no statistically significant association between gender and ASD diagnosis in this dataset.

```
# get data info
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 800 entries, 0 to 799
Data columns (total 22 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   ID               800 non-null    int64  
 1   A1_Score         800 non-null    int64  
 2   A2_Score         800 non-null    int64  
 3   A3_Score         800 non-null    int64  
 4   A4_Score         800 non-null    int64  
 5   A5_Score         800 non-null    int64  
 6   A6_Score         800 non-null    int64  
 7   A7_Score         800 non-null    int64  
 8   A8_Score         800 non-null    int64  
 9   A9_Score         800 non-null    int64  
 10  A10_Score        800 non-null    int64  
 11  age              800 non-null    float64
 12  gender           800 non-null    object 
 13  ethnicity        800 non-null    object 
 14  jaundice         800 non-null    object 
 15  family_autism    800 non-null    object 
 16  contry_of_res    800 non-null    object 
 17  used_app_before  800 non-null    object 
 18  result           800 non-null    float64
 19  age_desc         800 non-null    object 
 20  relation         800 non-null    object 
 21  ASD              800 non-null    int64  
dtypes: float64(2), int64(12), object(8)
memory usage: 137.6+ KB
```

Figure 1. Data information

4.1.2.1 Data Cleaning

In the 'Ethnicity' section, some values were marked with a question mark and then replaced with 'Missing'. Dummy variables were made for this category.

Variables like gender, ethnicity, jaundice, and family_autism were likely converted into numerical forms, possibly using methods such as Label Encoding, where categorical values are turned into numerical labels (e.g. male = 1, female = 0). One-Hot Encoding has been used to create binary columns for each unique category, especially beneficial for non-ordinal variables.

4.1.3. Exploratory Data Analysis

4.1.3.1 Visualisations

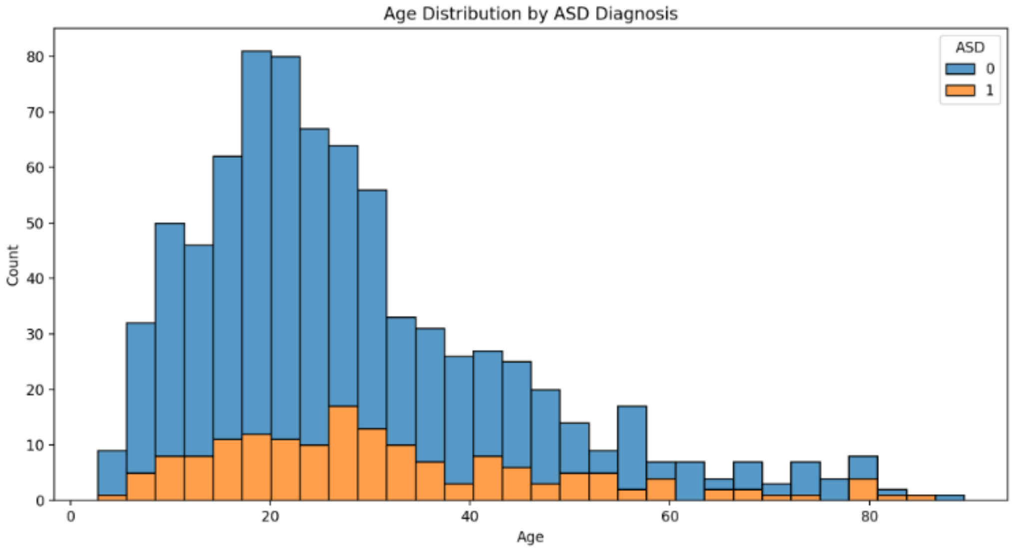


Figure 2. Age distribution by ASD diagnosis.

Figure 2 illustrates the age breakdown of people who have been diagnosed with ASD (coloured in orange) and those who have not (coloured in blue). The dataset includes a wide range of ages, but the majority of individuals are aged between 20 and 40 years old. There appears to be a higher prevalence of ASD diagnoses in younger age groups, especially in children and young adults. The distribution of non-ASD cases (coloured in blue) seems to be more evenly distributed among various age groups.

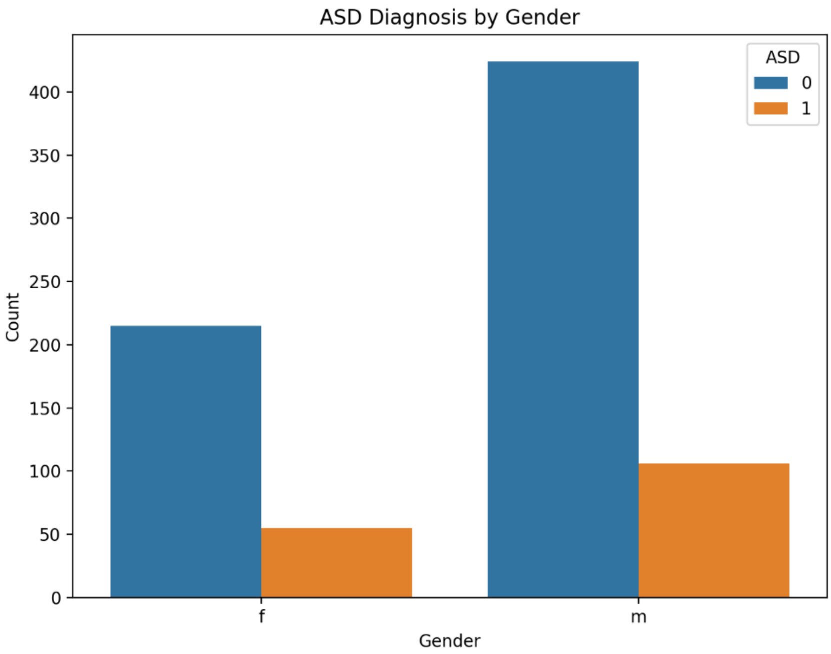


Figure 3. ASD diagnosis by Gender

Figure 3 displays the number of ASD diagnoses based on gender. In the dataset, there are 530 males and 270 females, with more males than females. ASD diagnoses are more frequent among males than females. Yet, the proportion of ASD diagnoses is similar in both gender groups (Women account for 20.37% of individuals diagnosed with ASD, while men make up 20.00% of those with an ASD diagnosis.).

Furthermore, in this dataset, 20.13% (639 individuals) are not diagnosed with ASD, while the remaining 79.87% (131 individuals) are diagnosed with ASD.

Male individuals make up 530 of the individuals in the dataset, while female individuals account for 270. The dataset shows a significant overrepresentation of males, with a ratio approaching 2:1 between the genders.

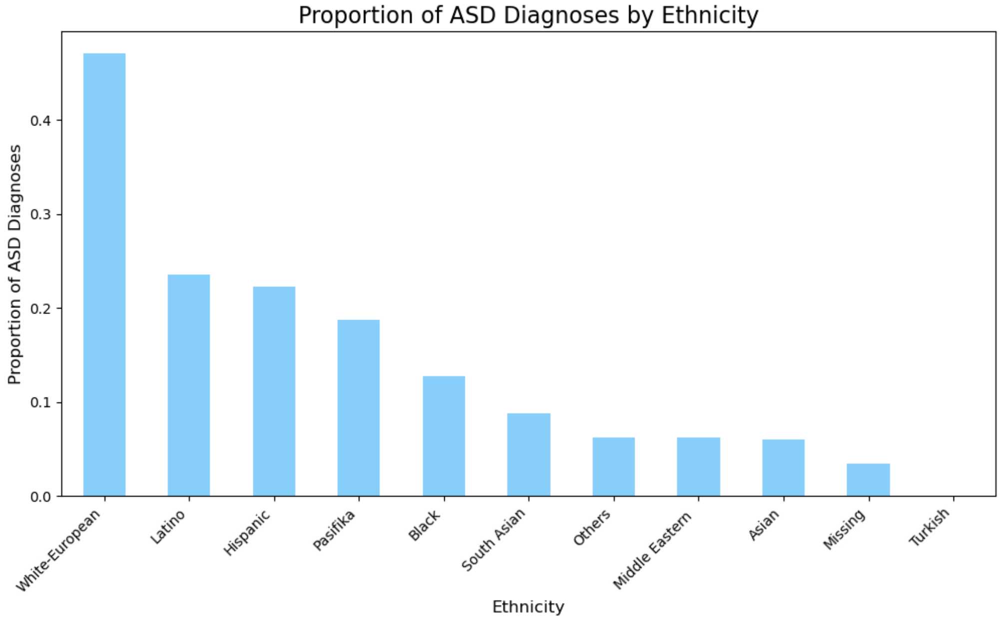


Figure 4. ASD diagnosis by Ethnicity

Figure 4 shows the ethnicity distribution within the dataset. White Europeans make up the largest proportion at 47.08%. Latino and Hispanic groups account for approximately 23-24%. Asian and Middle Eastern groups have smaller proportions, around 6%. Certain groups, such as Turkish individuals, do not have any ASD diagnoses in this particular dataset. This shows that the dataset has potential sampling biases as there are no participants included in Turkish ethnicity.

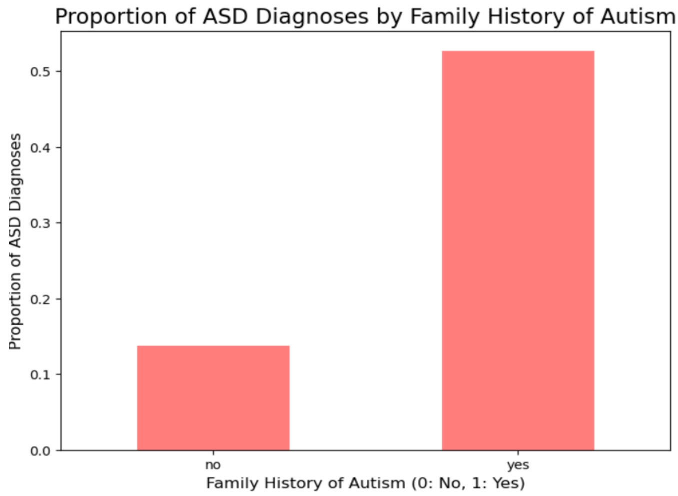


Figure 5. Family history of autism

Figure 5 shows the family history of autism. A family history of autism is strongly associated with a higher proportion of ASD diagnoses. In fact 52.67% of people with relatives who have autism also have an ASD diagnosis. In contrast, only 13.75% of individuals without a family history of autism are diagnosed with ASD. This suggests that genetics play a significant role in ASD development. Possibly, participants with family members having autism are more open to getting diagnosed compared to the other.

4.1.3.2 Correlation Analysis

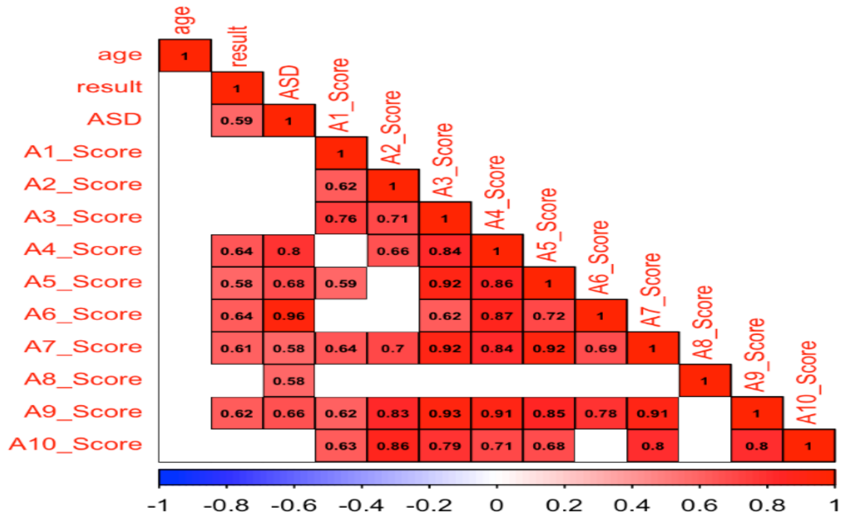


Figure 7. Correlation Heatmap

There is a non-significant correlation between age and ASD diagnosis ($r=0.11$, $p>0.05$), indicating that age does not have a significant direct association with ASD diagnosis within this set of data. Overall, age does not appear to be a significant factor in predicting ASD diagnosis or influencing the other variables in the dataset.

The A-Scores, specifically A6, A4, and A9, appear to be the most significant indicators of an ASD diagnosis. It is crucial to understand that correlation does not necessarily indicate causation, and these associations should be viewed in light of existing clinical understanding of ASD.

4.1.4 Machine Learning Models

I tested the dataset with 6 machine-learning algorithms, using the entries mentioned above. I initialised the machine learning pipeline by splitting the dataset into train and test data with a ratio of 8:2. The accuracy of the 6 models in the training and testing set is reported in Table 2

Table 2. Machine Learning Models

No	Model	Training Accuracy %	Testing Accuracy %
0	Logistic Regression	89.21	86.25
1	K-nearest neighbors	87.03	83.75
2	Support Vector Machine	91.71	87.50
3	Decision Tree Classifier	100.00	83.12
4	Random Forest Classifier	100.00	87.50
5	XGBoost Classifier	100.00	81.87

The Logistic Regression model demonstrates high accuracy on both the training and testing datasets, indicating good generalization without overfitting.

The performance of the K-nearest neighbors model is decent on both datasets, although slightly lower than Logistic Regression. There is a small difference between the training and testing accuracy of the K-Nearest Neighbors model, suggesting some generalization ability but could benefit from parameter tuning.

The SVM model performs very well on both datasets, with minimal discrepancy between training and testing accuracy, indicating strong generalization and suitability of the model.

The Decision Tree model achieves perfect accuracy on the training data, but experiences a significant decline in testing accuracy, a common sign of overfitting where the model memorizes the training data instead of learning general patterns.

Similar to the Decision Tree, the Random Forest model also achieves perfect accuracy on the training data yet performs better on the testing data. Despite some degree of overfitting, as indicated by the perfect training score, the Random Forest model is more resilient and generalizes better than the decision tree.

The XGBoost model exhibits perfect training accuracy but a noticeable decrease in testing accuracy, pointing towards overfitting, whereby the model is too specifically tailored to the training data and lacks generalization to new data.

4.1.4.1 Overall Analysis and Choosing the Best Model

Decision Tree, Random Forest, and XGBoost exhibit signs of overfitting, evident from achieving perfect accuracy on training data but a decline in accuracy on testing data.

The SVM model stands out with the highest testing accuracy (87.50%) compared to other models. It also shows a minimal difference in performance between training and testing data, indicating its ability to adapt well to new information. Based on the results presented in Table 2, SVM appears to be the most promising model, offering a good balance between training and testing performance.

4.1.4.2 Hyperparameter Tuned Models

Table 3. Hyperparameter tuned models training and testing performance:

No.	Model	Training Accuracy %	Testing Accuracy %
0	Tuned Logistic Regression	87.65	85.62
1	Tuned K-nearest neighbours	88.43	86.87
2	Tuned Support Vector Machine	95.15	85.62
3	Tuned Decision Tree Classifier	87.50	86.87

4	Tuned Random Forest Classifier	92.65	86.87
5	Tuned XGBoost Classifier	92.18	85.00

The accuracy of the 6 models in the training and testing set after hyperparameter tuning are reported in Table 3.

The Logistic Regression model demonstrates a good balance between accuracy in both training and testing, showing its ability to apply well to new data. There is a small difference between the accuracy in training and testing, indicating minimal overfitting.

The tuned KNN model performs slightly better than Logistic Regression, with a small margin between training and testing accuracy, suggesting effective fine-tuning and generalization to new data.

The SVM model has strong training accuracy, signifying a good fit with the training data. However, a larger gap between training and testing accuracy implies a possible risk of overfitting, even after adjustments.

The Decision Tree model displays similar performance on both the training and testing datasets, showing that tuning has helped prevent overfitting and allowing for effective generalization to new data.

The Random Forest model exhibits relatively high accuracy on both datasets, with a moderate gap suggesting a slight tendency to overfit. Nonetheless, it still generalizes well after tuning.

The XGBoost model shows high training accuracy but a notable decrease in testing accuracy, indicating potential overfitting even after tuning.

4.1.4.3 Choosing the Best Model after Hyperparameter Tuning

The SVM that has been adjusted, the Random Forest Classifier that has been adjusted, and the XGBoost Classifier that has been adjusted exhibit indications of overfitting (with high accuracy in training but lower accuracy in testing). Nevertheless, the Random Forest's performance remains fairly stable.

4.1.4.4 Overall Model Performance

Table 4. Overall model performance

No	Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
0	Logistic Regression	0.80	0.80	0.80	0.73	0.91
1	Random Forest	0.87	0.87	0.87	0.87	0.92
2	SVM	0.77	0.60	0.77	0.67	0.90
3	Decision Tree	0.83	0.82	0.83	0.83	0.79
4	KNN	0.81	0.82	0.81	0.81	0.88
5	XGBoost	0.81	0.85	0.81	0.76	0.90

The overall model performance is presented in Table 4.

The Logistic Regression model demonstrates satisfactory overall accuracy and a high ROC AUC value (0.91), which indicates strong discriminatory ability. However, the F1 Score is somewhat lower, indicating a potential trade-off between Precision and Recall.

The Random Forest model boasts the highest accuracy (87%), a good balance between Precision and Recall, and the highest ROC AUC (0.92), making it the top performer overall in terms of accuracy and discriminatory power.

On the other hand, the SVM model exhibits lower accuracy (77%) and a significant difference between Precision (60%) and Recall (77%). This suggests the model may produce many false positives and might not be the ideal choice when seeking a balance between Precision and Recall.

The Decision Tree model delivers solid overall performance with balanced Precision and Recall. However, its ROC AUC (0.79) is lower than others, signifying a potential weakness in distinguishing between classes.

The KNN model showcases balanced Precision and Recall with a moderate level of accuracy at 81%. The ROC AUC (0.88) is reasonable, indicating commendable overall performance.

Lastly, the XGBoost model displays high Precision (85%) and a good ROC AUC (0.90), denoting strong discriminatory ability. Nevertheless, the F1 Score is lower, hinting at a potential trade-off between Precision and Recall.

Random Forest emerges as the top performer with the highest accuracy rate of 87.50%, balanced Precision and Recall, a high F1 Score of 87%, and the highest ROC AUC of 0.92.

Both Logistic Regression and XGBoost also exhibit high ROC AUC values, implying good discriminatory capability, but Random Forest demonstrates superior overall balance.

4.1.5 Feature Importance:

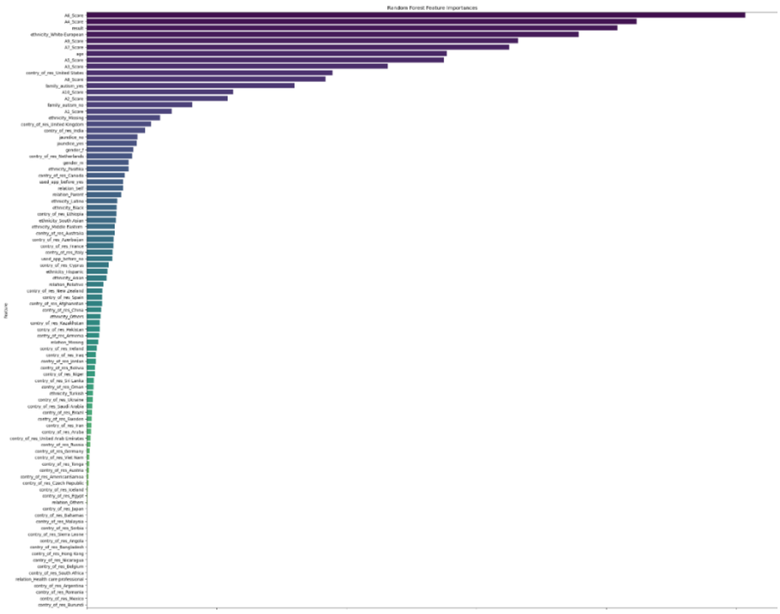


Figure 9 Feature importance of Random Forest

Figure 9 presented the feature importance of the Ranod Forest model.

The most important variables are 'A6_score' and 'A4_score', which have significantly higher importance than the rest. 'ethnicity_White_European' is the third most important feature. Other notable features include 'A9_Score', 'age', and 'A7_Score'. Country-specific variables (e.g., 'country_of_res_United States', 'country_of_res_United Kingdom') have lower importance but are still relevant.

4.1.6 Feature Selection

```
Selected Features with Tuned Random Forest: ['A1_Score', 'A2_Score', 'A3_Score', 'A4_Score', 'A5_Score', 'A6_Score', 'A7_Score', 'A8_Score', 'A9_Score',
'A10_Score', 'age', 'result', 'ethnicity_Missing', 'ethnicity_White-European', 'family_autism_no', 'family_autism_yes', 'contry_of_res_United States']
C:\Users\harle\anaconda3\Lib\site-packages\sklearn\base.py:432: UserWarning: X has feature names, but SelectFromModel was fitted without feature names
  warnings.warn(
C:\Users\harle\anaconda3\Lib\site-packages\sklearn\base.py:432: UserWarning: X has feature names, but SelectFromModel was fitted without feature names
  warnings.warn(
Tuned Random Forest Training Accuracy: 93.12%
Tuned Random Forest Testing Accuracy: 85.00%
Tuned Random Forest Precision: 0.70
Tuned Random Forest Recall: 0.58
Tuned Random Forest F1 Score: 0.64
Tuned Random Forest ROC AUC: 0.91
```

Fig. 11 Feature Selection of tuned Random Forest.

The code shows specific traits like scores and demographic information (age, ethnicity, family_autism, and country_of_res). These traits are essential for the model's predictions.

Training Accuracy: 93.12% - Shows how many predictions were correct in training data.

Accuracy in Testing is 85%, indicating the proportion of correct predictions in the test dataset.

The small drop in testing accuracy compared to training accuracy indicates that the model generalises effectively but may be slightly overfitting the training data. Precision (0.70): It shows the ratio of accurate positive predictions to all positive predictions. A precision of 0.70 means 70% of the model's positive predictions were correct. "Recall (0.58): This measures how many actual positive results were correctly identified by the model, with a recall of 0.58 indicating 58% accuracy in identifying positives." F1 Score (0.64): The F1 score shows a balance between Precision and Recall, but suggests room for improvement in addressing false positives and false negatives. ROC AUC of 0.91 shows the model's ability to differentiate between positive and negative classes is excellent. A score of 0.91 indicates the model effectively distinguishes between classes.

The optimized Random Forest model achieves high accuracy, especially on the training dataset, showing it effectively captures data patterns. However, the lower recall score of 0.58 indicates that the model may be missing some positive cases. The model has a high ROC AUC score of 0.91, indicating good performance. Improving the balance between Precision and Recall could make the model more reliable for all types of predictions.

4.1.7 Conclusion

In this dataset, gender shows less impact on diagnosis of ASD but age has a significant impact on it. The average age of individuals with Autism Spectrum Disorder (ASD) is 32.02 years. The average age of individuals without ASD is 27.55 years. There is a positive correlation, but it is relatively weak, suggesting a slight inclination for older individuals to be diagnosed with ASD. The examination shows that the age range of 31-50 has a greater proportion of ASD diagnoses compared to other age brackets. Furthermore, individuals diagnosed with ASD typically tend to be slightly older on average than those who do not have ASD. The statistical evaluation validates the significance of this discrepancy.

Gender does not appear to have a significant impact on ASD diagnosis in this particular dataset. The rates of ASD diagnosis are nearly identical for males and females, and the statistical test confirms that any small difference observed is likely due to random chance rather than a genuine effect of gender.

It's important to note that this finding is specific to this dataset and may not necessarily reflect broader population trends. In many other studies and clinical observations. The lack of gender difference in this dataset could be due to various factors, such as sampling methods, diagnostic criteria used, or the specific population studied.

4.2 Attention Deficit Hyperactivity Disorder

4.2.1 Overview

The dataset seems to contain different characteristics related to the diagnosis of ADHD and other cognitive tests. The main focus appears to be on a binary or multiclass label that shows the status of ADHD (referred to as DX and ASD in previous discussions). Important attributes in the dataset consist of demographic, clinical, and psychological measurements such as age, gender, IQ scores, and various cognitive indicators. The dataset contains 222 entries with 23 columns, including information about participants' demographics, ADHD measures, IQ scores, and quality control data for various scans.

Age: Indicates the age of the individuals.

Gender: Categorical variable indicating the gender of the individuals (Male/Female).

Handedness: Indicates whether the individual is left-handed or right-handed.

ADHD Index: A clinical tool used to evaluate the likelihood of ADHD.

Inattentive & Hyperactive/Impulsive Scores: Clinical ratings that assess specific behaviours associated with ADHD.

IQ Scores: Comprises Verbal IQ, Performance IQ, and Full IQ scores, which reflect the individual's cognitive abilities.

Diagnosis/Target Variable (DX/ADHD): The variable targeted for predicting ADHD or other diagnostic categories.

Table 5. ADHD dataset variables

This dataset contains various columns related to phenotypic data, including:

- `scandir_id`: Identifier for the scan directory
- `site`: Site number
- `gender`: Gender of the participant
- `age`: Age of the participant
- `handedness`: Handedness score
- `dx`: Diagnosis code
- `secondary_dx`: Secondary diagnosis
- `adhd_measure`: ADHD measure type
- `adhd_index`: ADHD index score
- `inattentive`: Inattentive score
- `hyper_impulsive`: Hyperactive/impulsive score
- `iq_measure`: IQ measure type
- `verbal_iq`: Verbal IQ score
- `performance_iq`: Performance IQ score
- `full2_iq`: Full IQ score (2 subtests)
- `full4_iq`: Full IQ score (4 subtests)
- `med_status`: Medication status
- `qc_rest_1` to `qc_rest_4`: Quality control scores for resting state scans
- `qc_anatomical_1` and `qc_anatomical_2`: Quality control scores for anatomical scans

The dataset utilised in this study was sourced from Nolan Nichols (2014). https://data.world/nicholsn/adhd-200

Total number of participants: 149

4.2.2 Data Preparation:

The ADHD diagnosis counts in the dataset are represented by the dx column, which contains numerical codes. Here's what these codes typically mean:

0: No ADHD diagnosis

1: ADHD diagnosis

2: Other diagnosis (not ADHD)

3: Possible ADHD or sub-threshold ADHD symptoms

In the dataset:

99 participants have a diagnosis code of 0, indicating no ADHD diagnosis.

77 participants have a diagnosis code of 1, indicating an ADHD diagnosis.

2 participants have a diagnosis code of 2, indicating another diagnosis.

44 participants have a diagnosis code of 3, indicating possible ADHD or sub-threshold symptoms.

There were a few missing values in 'Handedness', QC_Rest_1, QC_Rest_2, QC_Anatomical_1, and QC_Anatomical_2 were replaced.

Summary Statistics:

	ScanDir ID	Site	Gender	Age	Handedness	DX	\
count	1.490000e+02	149.0	149.000000	149.000000	149.000000	149.000000	
mean	1.682075e+06	5.0	0.590604	11.749128	0.641409	0.718121	
std	2.484528e+06	0.0	0.493381	3.045170	0.273090	1.090998	
min	1.000200e+04	5.0	0.000000	7.170000	-0.200000	0.000000	
25%	1.006500e+04	5.0	0.000000	9.160000	0.480000	0.000000	
50%	1.011600e+04	5.0	1.000000	11.410000	0.670000	0.000000	
75%	2.991307e+06	5.0	1.000000	14.200000	0.830000	1.000000	
max	9.907452e+06	5.0	1.000000	17.960000	1.000000	3.000000	

	ADHD Measure	ADHD Index	Inattentive	Hyper/Impulsive	IQ Measure	\
count	149.0	149.000000	149.000000	149.000000	149.0	
mean	2.0	55.127517	54.912752	54.818792	2.0	
std	0.0	14.197587	14.141148	13.628296	0.0	
min	2.0	40.000000	40.000000	41.000000	2.0	
25%	2.0	42.000000	43.000000	43.000000	2.0	
50%	2.0	49.000000	49.000000	48.000000	2.0	

Fig 12. Statistics of the dataset

The age distribution ranges from approximately 7 to 18 years, with a mean age of around 11.7 years. In gender, the mean of 0.590 suggests that approximately 59% of the participants are male.

A t-test was conducted to compare the age difference between ADHD and non-ADHD groups. The results indicated no statistically significant difference in age between the two groups, t(174) = -1.33, p = .185. This p-value (> 0.05) suggests that there is no statistically significant difference in age between ADHD and non-ADHD groups.

Further, categorical variables were One Hot Encoded converting them to 0 or 1.

4.2.3 Exploratory Data Analysis:

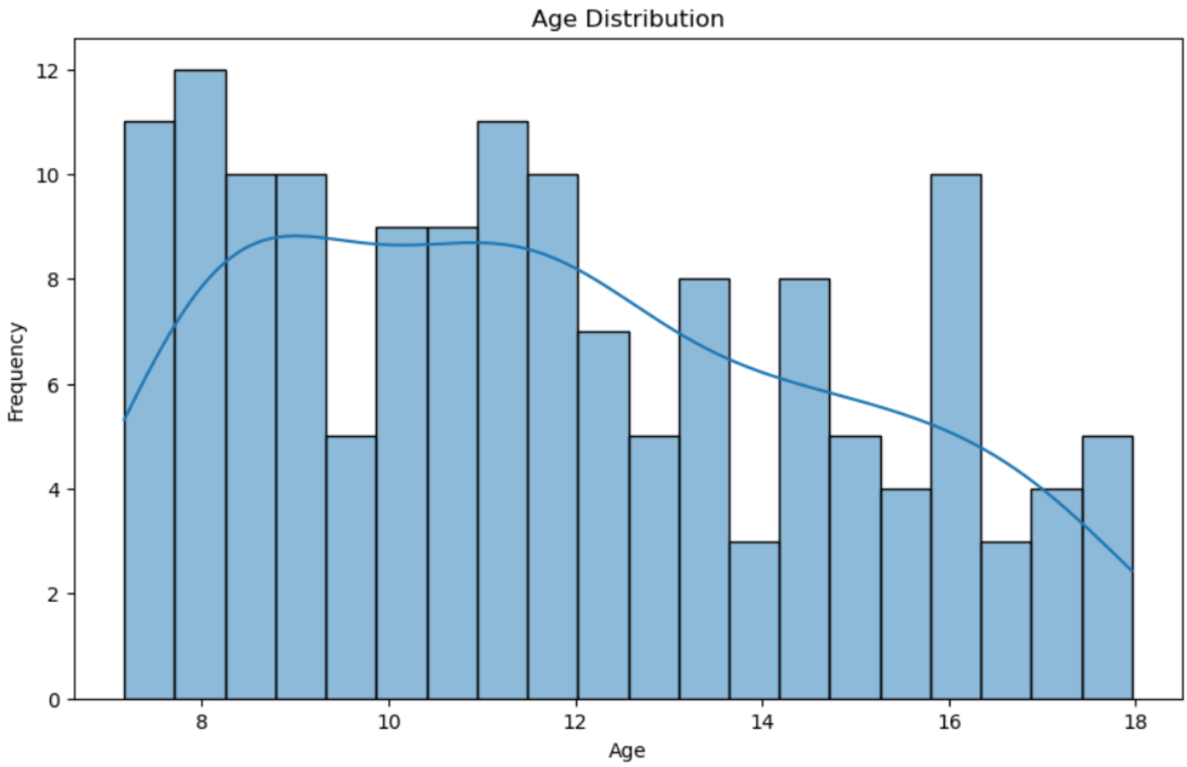


Fig 13. Age distribution histogram

The age range of participants is approximately 7 to 18 years old. There's a higher concentration of participants between 8 and 14 years old.

The distribution is relatively similar for both ADHD and non-ADHD groups, with ADHD having slightly more participants in most age ranges.

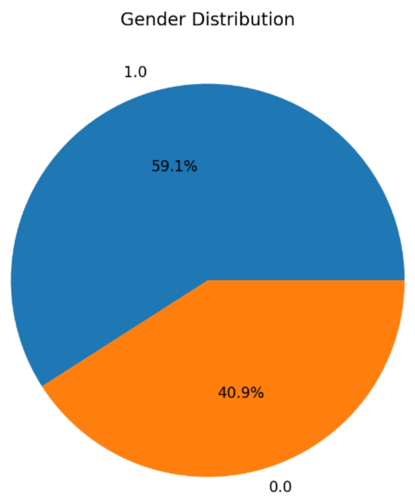


Fig 14. Gender distribution

There are more male participants (88, 59.06%) than female participants (61, 40.94%). For non-ADHD participants (0), the gender split is relatively balanced (52.75% female, 47.25% male). For ADHD participants (1), there's a strong male predominance (84.85% male, 15.15% female). Categories 2 and 3 have very few participants, which might represent different subtypes or severity levels of ADHD.

There is a significant overrepresentation of males in the ADHD group. This is consistent with the common pattern seen in ADHD research, where males are more commonly diagnosed with ADHD compared to females. The considerable gender disparity in ADHD diagnosis could be an indication of actual variations in ADHD incidence between males and females, or it could also suggest potential biases in ADHD detection and diagnosis. Certain studies propose that ADHD may be underdiagnosed in females due to differences in symptom manifestation. It should be highlighted that the sample size, particularly for females with ADHD, is relatively small. This restricts the applicability of these results and raises the likelihood of error in our statistical analyses.

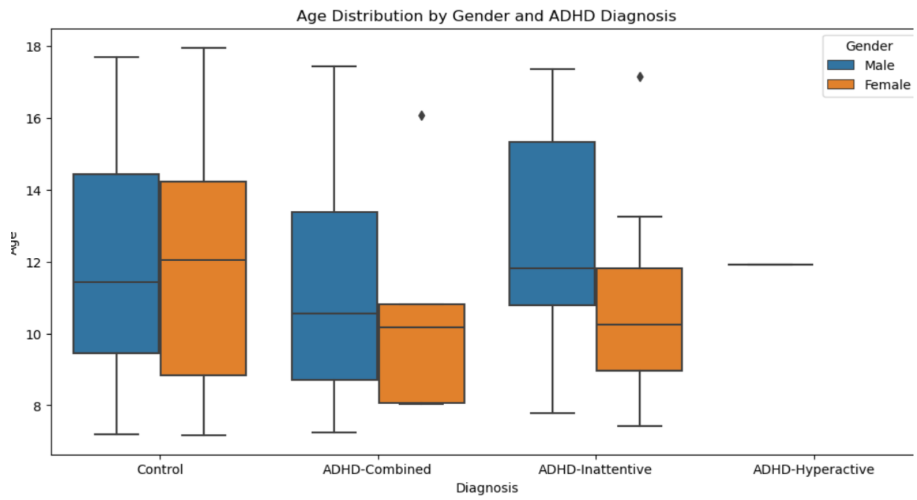


Fig 15. Age distribution by Gender and ADHD Diagnosis

Control Group: Both males and females have a similar median age, which is around 12-13 years. The age range (from the lower to the upper quartile) is also similar between genders, spanning from about 10 to 14 years.

ADHD-Combined: Males in this group have a wider age range compared to females, with the median age slightly lower than that of females. The median age for females is approximately 11 years, while for males it is around 10 years. The age range for females is narrower in this category.

ADHD-Inattentive: The median age for both males and females is around 12 years, but males have a slightly wider interquartile range than females. This indicates a broader age variation among males with this type of ADHD.

ADHD-Hyperactive: Both males and females in this group have a very narrow age range, with the median age being approximately 12 years. It is important to note that there are fewer data points in this category, shown by the shorter whiskers and lack of outliers, indicating that fewer participants are classified under this category.

Outliers: In the ADHD-Combined category for males, there are a small number of cases that fall outside the typical age range, showing that some males with ADHD are significantly younger or older than expected. The box plot provides insight into the variations in age distributions not just between different types of ADHD, but also between males and females within those categories. Despite some differences in age distributions and ranges between genders, the median ages are fairly consistent across the board. This implies that although ADHD symptoms may manifest differently in males and females, the age at which it is usually identified or observed in clinical settings is not vastly different.

4.2.4 Machine Learning Models:

I tested the dataset with 6 machine-learning algorithms, using the entries mentioned above. I initialised the machine learning pipeline by splitting the dataset into train and test data with a ratio of 8:2.

Table 6. Machine Learning Models

No.	Model	Training Accuracy %	Testing Accuracy %
0	Logistic Regression	89.915966	83.333333
1	K-nearest neighbors	84.873950	63.333333
2	Support Vector Machine	100.000000	60.000000
3	Decision Tree Classifier	100.00000	76.666667
4	Random Forest Classifier	100.00000	76.666667
5	Gaussian Naive Bayes	87.394958	70.000000

6	Gradient Boosting Classifier	100.000000	70.000000

Logistic Regression:

The accuracy of training is 89.92% while the accuracy of testing is 83.33%. Logistic Regression analysis indicates a satisfactory balance in accuracy between training and testing, implying that the model is generalising effectively. There are no clear indications of either overfitting or underfitting. This model is likely one of the top performers in this comparison.

K-nearest neighbours (KNN):

Training Accuracy: 84.87%

Testing Accuracy: 63.33%

Evaluation: There is a significant difference between the accuracy of training and testing. This suggests that the model might be too focused on the training data and not performing well on new data. KNN might have difficulty with intricate data, particularly when dealing with many features.

Support Vector Machine (SVM):

Training Accuracy: 100.00%

Testing Accuracy: 60.00%

Evaluation: The Support Vector Machine (SVM) model demonstrates flawless accuracy in training, however, it falters in testing accuracy, indicating a problem of overfitting. The model excels in memorizing the training set but struggles to apply its learnings to new, unseen data.

Decision Tree Classification:

Training Accuracy: 100.00%

Testing Accuracy: 76.67%

Evaluation: The Decision Tree model demonstrates flawless accuracy during training but a noticeably lower accuracy during testing. This points to overfitting, a common occurrence in decision trees as they often learn specific patterns in the training data that may not apply well to other data.

Random Forest Classifier:

Training Accuracy: 100.00%

Testing Accuracy: 76.67%

Evaluation: Much like the Decision Tree model, Random Forest achieves flawless training accuracy but indicates some overfitting with a decrease in testing accuracy. Nonetheless, Random Forests tend to excel in generalization compared to decision trees, hence this outcome is somewhat anticipated.

Gaussian Naive Bayes:

Training Accuracy: 87.39%

Testing Accuracy: 70.00%

Evaluation: Despite having a lower testing accuracy compared to some models, Naive Bayes demonstrates good performance in training and indicates that the model is not overfitting.   Naive Bayes is effective for solving straightforward problems, particularly when the features are mostly independent.

Gradient Boosting:

Training Accuracy: 100.00%

Testing Accuracy: 70.00%

Evaluation: The model achieved flawless accuracy during training but experienced a considerable drop in accuracy during testing, indicating a tendency towards overfitting.   Gradient Boosting is susceptible to overfitting, particularly in the absence of proper regularization or when trained too intensively.

Main Points:

Overfitting: Support Vector Machines, Decision Trees, Random Forests, and Gradient Boosting are all prone to overfitting, as evidenced by their flawless or nearly flawless training accuracy contrasted with significantly lower testing accuracy. Top Generalisation: Logistic Regression demonstrates the most effective generalisation between the performance in training and testing, with a small difference in accuracy between the two. This quality renders it as one of the most dependable models in this analysis. Under-performance: KNN and SVM exhibit the most significant performance disparities, suggesting inadequate generalization and under-performance on the test data.

4.2.4.1 Model Performance:

Table 7. ML models with Accuracy, F1 score, Recall and Precision

Model	**Train Accuracy (%)**	**Test Accuracy (%)**	**Train F1 Score**	**Test F1 Score**	**Train Recall**	**Test Recall**	**Train Precision**	**Test Precision**
Logistic Regression	89.92	83.33	0.8986	0.8292	0.8992	0.8333	0.8983	0.8676
K-nearest neighbors	84.87	63.33	0.8300	0.6100	0.8500	0.6300	0.8300	0.6200
Support Vector Machine	100.00	60.00	1.0000	0.4500	1.0000	0.6000	1.0000	0.3600
Decision Tree Classifier	100.00	76.67	1.0000	0.7750	1.0000	0.7667	1.0000	0.8000
Random Forest Classifier	100.00	76.67	1.0000	0.7629	1.0000	0.7667	1.0000	0.7907
Gradient Boosting Classifier	100.00	70.00	1.0000	0.6862	1.0000	0.7000	1.0000	0.6971
Gaussian Naive Bayes	87.39	70.00	0.8779	0.7121	0.8740	0.7000	0.8831	0.7600

Model Performance Analysis:

Logistic Regression:

Train Accuracy: 89.92%, Test Accuracy: 83.33%

Logistic Regression shows a balanced performance between training and testing data, with only a small drop in accuracy, indicating good generalization.

F1 Score (train: 0.8986, test: 0.8292) shows that precision and recall are well balanced on both datasets.

Conclusion: Logistic Regression performs well, with consistent scores across train and test data. It is a reliable model with good generalization capabilities.

K-nearest neighbours (KNN):

Train Accuracy: 84.87%, Test Accuracy: 63.33%

There is a notable drop in accuracy when moving from training to testing data, indicating potential overfitting or underperformance in generalizing to unseen data.

F1 Score (train: 0.8300, test: 0.6100) also shows a significant drop, suggesting the model struggles with unseen data.

Conclusion: KNN may be overfitting the training data and failing to generalize well, as indicated by the large drop in performance on the test set.

Support Vector Machine (SVM):

Train Accuracy: 100.00%, Test Accuracy: 60.00%

The perfect accuracy on the training set coupled with the much lower accuracy on the test set shows clear signs of overfitting.

F1 Score (train: 1.0000, test: 0.4500) also indicates that while the model memorizes the training data, it fails to generalize to new data.

Conclusion: The SVM is likely overfitting the training data and struggling with test data. It requires tuning or regularization to improve generalization.

Decision Tree Classifier:

Train Accuracy: 100.00%, Test Accuracy: 76.67%

Similar to SVM, the Decision Tree also achieves perfect accuracy on the training set but drops in accuracy on the test set, indicating overfitting.

F1 Score (train: 1.0000, test: 0.7750) shows that, while it performs perfectly on the training set, it doesn't generalize as well to the test data.

Conclusion: The Decision Tree overfits the training data and may need pruning or regularization to improve its generalization to test data.

Random Forest Classifier:

Train Accuracy: 100.00%, Test Accuracy: 76.67%

Like the Decision Tree, Random Forest also overfits the training data but performs slightly better on the test data than the Decision Tree, indicated by the high test precision (0.7907).

F1 Score (train: 1.0000, test: 0.7629) shows overfitting on the training set, though it maintains a reasonable balance on the test set.

Conclusion: Random Forest overfits the training data but is more robust on the test set compared to the Decision Tree. It might benefit from tuning the hyperparameters.

Gradient Boosting Classifier:

Train Accuracy: 100.00%, Test Accuracy: 70.00%

The perfect training accuracy indicates overfitting, but the drop in test accuracy suggests it struggles with unseen data.

F1 Score (train: 1.0000, test: 0.6862) indicates that the model overfits and could benefit from parameter tuning, regularization, or early stopping.

Conclusion: Gradient Boosting is overfitting the training data but maintains reasonable performance on the test set. It could benefit from further tuning.

Gaussian Naive Bayes:

Train Accuracy: 87.39%, Test Accuracy: 70.00%

Naive Bayes is not overfitting as much as other models, as indicated by the smaller gap between training and testing accuracy. However, the test accuracy is lower than Logistic Regression or Random Forest.

F1 Score (train: 0.8779, test: 0.7121) shows that Naive Bayes has good precision and recall balance but is less powerful on the test data than some other models.

Conclusion: Naive Bayes offers a balance between performance on training and testing data and may be a good option if simplicity is preferred.

Hyperparameter Tuning:

Table 8. Tuned ML Models

	Model	Training Accuracy %	Testing Accuracy %
0	Tuned Logistic Regression	92.44	80.00
1	Tuned K-nearest neighbors	82.35	73.33
2	Tuned Support Vector Machine	93.28	70.00
3	Tuned Decision Tree Classifier	89.92	70.00
4	Tuned Random Forest Classifier	89.08	73.33

Explanation and Performance Comparison:

1. **Tuned Logistic Regression**:
    - **Training Accuracy**: 92.44%
    - **Testing Accuracy**: 80.00%
    - **Analysis**: Logistic Regression shows strong generalization, with the training accuracy relatively close to the testing accuracy. This suggests that the model is not overfitting and performs well on unseen data. The small gap (around 12%) between training and testing accuracy indicates good model performance.
2. **Tuned K-nearest neighbours (KNN)**:
    - **Training Accuracy**: 82.35%
    - **Testing Accuracy**: 73.33%
    - **Analysis**: KNN has moderate accuracy on both training and testing sets. The small difference between training and testing accuracy suggests that the model generalizes reasonably well but may not be

capturing complex patterns in the data compared to models like Logistic Regression.

3. **Tuned Support Vector Machine (SVM)**:
   - **Training Accuracy**: 93.28%
   - **Testing Accuracy**: 70.00%
   - **Analysis**: SVM shows a gap between training and testing accuracy, which could indicate slight **overfitting**. The model performs well on the training data but struggles more with unseen data. This suggests that SVM might have memorized parts of the training data and is not generalizing as well.
4. **Tuned Decision Tree Classifier**:
   - **Training Accuracy**: 89.92%
   - **Testing Accuracy**: 70.00%
   - **Analysis**: The Decision Tree shows similar performance to SVM, with a reasonable training accuracy but lower testing accuracy. This suggests potential **overfitting**, as Decision Trees can often memorize the training data without generalizing well to new data.
5. **Tuned Random Forest Classifier**:
   - **Training Accuracy**: 89.08%
   - **Testing Accuracy**: 73.33%
   - **Analysis**: Random Forest shows a relatively balanced performance, with close training and testing accuracies. Random Forest, known for its ability to generalize well, performs similarly to KNN but slightly better due to its ability to reduce overfitting through multiple decision trees (ensembling).
6. **Tuned Gaussian Naive Bayes**:
   - **Training Accuracy**: 87.39%

- **Testing Accuracy**: 70.00%
- **Analysis**: Naive Bayes shows a good balance between training and testing accuracy, but the overall performance is slightly lower than models like Logistic Regression and Random Forest. It suggests that the model generalizes reasonably well but doesn't capture complex patterns as effectively as the other models.

General Observations:

- **Logistic Regression** has the best balance between training and testing accuracy, making it the **best generalizing model** in this comparison.
- **K-nearest neighbors** and **Random Forest** also perform well with close training and testing accuracy, indicating that they generalize well but may not capture all the complexity of the data.
- **Support Vector Machine (SVM)** and **Decision Tree** show signs of **overfitting**, as indicated by the higher training accuracy but lower testing accuracy.
- **Gaussian Naive Bayes** performs moderately, with a balanced performance but slightly lower accuracy compared to Logistic Regression and Random Forest.

4.2.5 Feature Importance:

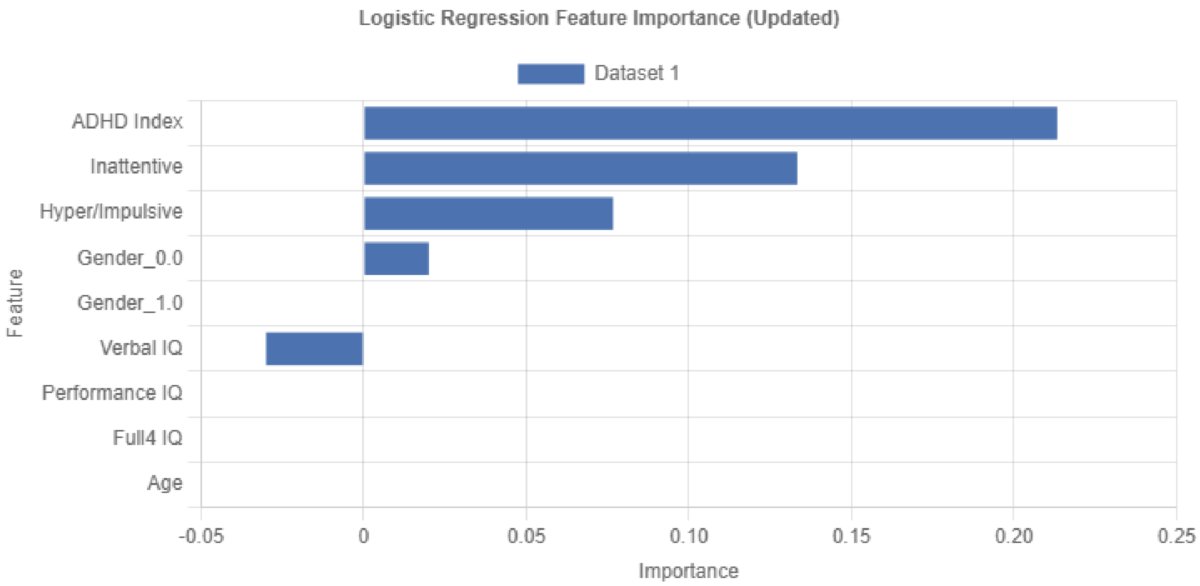


Fig 16. Logistic Regression Feature Importance

1. **ADHD Index** has the highest importance, which suggests that this variable is the most influential in the model's predictions.
2. **Inattentive** and **Hyper/Impulsive** also have relatively high importance, indicating that they contribute significantly to the model's decision-making.
3. **Gender**: The gender categories (female) and (Male) have moderate importance, meaning that gender plays some role in the predictions, but not as much as the ADHD-related features.
4. **IQ measures**: Variables like **Verbal IQ**, **Performance IQ**, and **Full IQ** have less influence compared to ADHD symptoms and gender.
5. **Age** appears to have the least importance in the model, indicating that it has little influence on the model's predictions.

Chapter 5: Discussion/ Evaluation:

5.1: Introduction:

This study aimed to address gender biases in the diagnosis of ADHD and ASD through the use of machine learning (ML) techniques. The primary objective was to develop and evaluate machine learning models that could mitigate the underdiagnosis of females, who often present subtler symptoms compared to males, such as inattentiveness in ADHD or more nuanced social impairments in ASD. Additionally, the study sought to examine how age and gender influence diagnostic procedures and whether ML models can predict more equitable diagnostic outcomes by considering these variables.

The key findings of this research demonstrated that certain machine learning models, such as Support Vector Machines (SVM) and Random Forest Classifiers, showed promise in detecting ADHD and ASD symptoms across genders. However, some models, particularly decision trees and XGBoost classifiers, displayed signs of overfitting during training, leading to less accurate results when tested on new data.

5.2 Interpreting Finding:

5.2.1 Gender Bias

The research conducted in this study revealed that females, who tend to display more subtle symptoms, are often overlooked in diagnosis due to the current male-centric diagnostic criteria. For instance, in cases of ADHD, females tend to exhibit inattentiveness rather than hyperactivity, a symptom more commonly associated with males. The study on ADHD had a higher number of male participants than females, indicating a potential bias in diagnosis. In contrast, the control group for ADHD had an almost equal distribution of male and female participants, highlighting the gender imbalance in diagnosed cases.

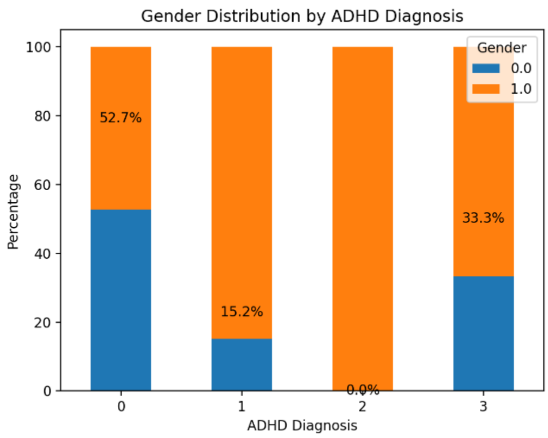


Fig 19. ADHD diagnosis by Gender

When it comes to Autism Spectrum Disorder (ASD), gender was found to have minimal to no impact on the accuracy of machine learning models. Gender does not seem to significantly influence the diagnosis of ASD in the specific dataset analysed. The prevalence of ASD diagnosis appears to be similar between males and females, with statistical analysis suggesting that any minor differences may be due to random chance rather than a direct correlation with gender.

5.2.2 Age Bias

Logistic regression has been identified as a dependable model for the dataset ADHD due to the close matching of its training and testing scores. Within this model, age appears to have a limited impact, suggesting that it only has a minor influence on the predictions made by the model.

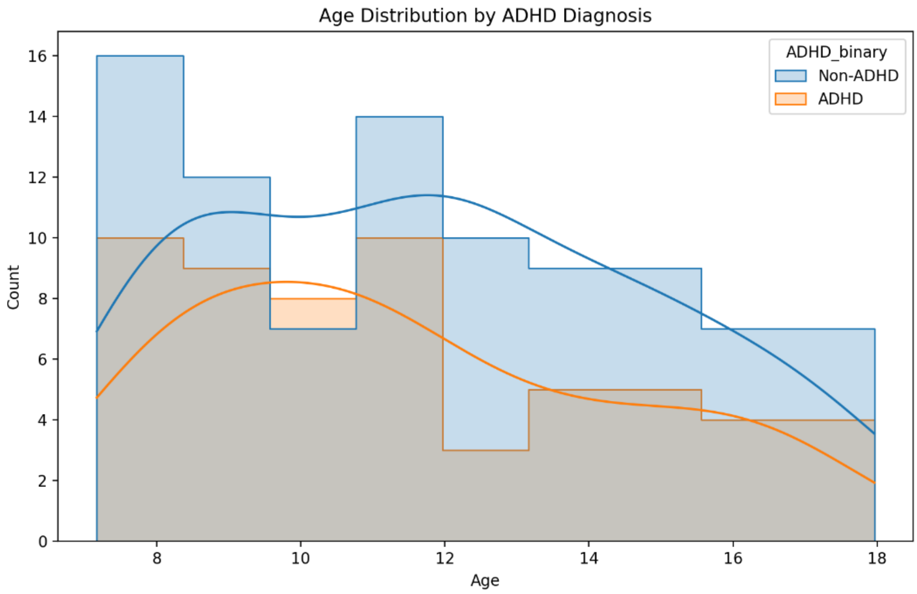


Fig 20. ADHD Age distribution by diagnosis

The histogram shows substantial overlap in age distribution between ADHD and non-ADHD groups. Both distributions cover similar age ranges and have similar shapes, making it difficult to distinguish between the groups based on age alone.

In the Autism Spectrum Disorder (ASD) dataset, the average age of individuals with ASD is 32.02 years, compared to around 27.55 years for those without ASD. The correlation between age and ASD diagnosis is weak, suggesting older individuals are slightly more likely to be diagnosed with ASD. Analysis shows that the 31 to 50-year-old age group has a higher percentage of ASD diagnoses. On average, individuals with ASD are slightly older than those without. Statistical analysis confirms the significance of this age difference.

5.2.3. Other Affecting Variables:

Demographic variables like ethnicity, country of residence and whether the family has a history of autism were shown as important features in ASD. Overall, the Autism Quotient Score had the most impact on the diagnoses. The AQ Score, also known as the Autism Quotient Score, is a key indicator in all models. Elements like age and ethnicity, especially White European background, play important roles. While the significance of variables specific to certain countries may vary across models, they tend to be ranked lower than the primary features.

In ADHD, the ADHD index has the highest importance which comprises ADHD symptoms and behaviours. Symptoms like inattentiveness and hyperactivity were common in both genders but females had more inattentive traits rather than hyperactive.

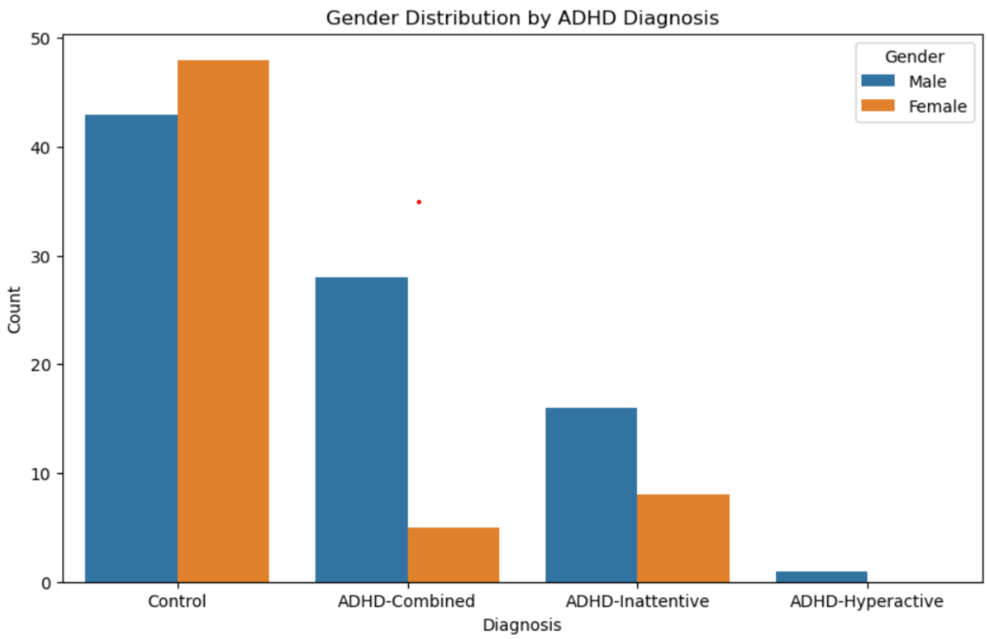


Fig 21. Gender distribution by ADHD diagnosis

There are more females than males diagnosed with ADHD-Inattentive type. This is in line with research suggesting that females are more likely to show inattentiveness, leading to underdiagnosis compared to males. Few or no females are diagnosed with ADHD-Hyperactive type, showing that hyperactive symptoms are less common or less recognised in females. Very few females, around 5 or fewer, are diagnosed with ADHD-Combined type, which includes both inattentiveness and hyperactivity/impulsivity symptoms. This indicates a lower number of females being diagnosed with this subtype.

5.2.4 Model Performance:

In ASD, the performance of the machine learning models varied, with some models demonstrating better generalizability and lower risks of overfitting. The Random Forest model, in particular, stood out with a testing accuracy of 87.5%, indicating its strong potential for generalizing across different datasets. This model was able to strike a balance between sensitivity and specificity, making it well-suited for clinical applications where both false positives and false negatives carry significant risks. Additionally, with a high ROC AUC score of 0.921, the model is highly effective at distinguishing between classes, making it the best-performing model in terms of both accuracy and its ability to differentiate between outcomes.

Both the Logistic Regression and XGBoost models show strong discriminatory ability with high ROC AUC values (0.911 and 0.908, respectively), meaning they are good at distinguishing between classes. However, XGBoost has a higher Precision (85.31%) but a lower F1 Score, indicating a potential trade-off between Precision and Recall. While both models perform well in terms of discrimination, Random Forest outperforms them with better overall balance, as it maintains high accuracy, a strong Precision-Recall balance, and the highest ROC AUC, making it the top-performing model.

In ADHD, the Logistic Regression model demonstrates strong generalisation, as the training accuracy is quite similar to the testing accuracy. This implies that the model is not overfitting and performs effectively on new data. The narrow difference (approximately 12%) between training and testing accuracy suggests a good model performance. Followed by Random Forest and KNN models.

However, other models, such as the decision tree and Decision Tree classifiers, showed signs of overfitting, where they performed well on the training data but struggled when applied to new data. Overfitting remains a common issue in machine learning, especially in healthcare settings where datasets are often imbalanced or biased.

5.3 Limitations:

This study has several limitations that should be acknowledged. First, the availability of diverse datasets remains a challenge. Although this research made use of open-source datasets, the underrepresentation of females in many ADHD and ASD studies limits the ability of machine learning models to fully address gender biases. Additionally, the datasets used in this study did not account for other important variables, such as socioeconomic status, or comorbid conditions, all of which can influence the diagnostic process.

As the data is collected from an open-source platform, its reliability is questionable in the realm of whether the real data was captured properly or not. Biases in participation could also have been an issue or less number of diverse participants.

Another limitation is the issue of model overfitting, particularly with decision tree and XGBoost models, which performed well on training data but poorly on testing data. This suggests that the models may not generalize well to new, unseen data, underscoring the need for more robust validation techniques.

Chapter 6: Conclusion and Future Work:

6.1 Summary of Results/Findings:

ADHD Results and Findings:

Gender Discrepancy in ADHD Diagnosis: ADHD is often over diagnosed in males and underdiagnosed in females due to differences in symptoms. Males show hyperactive behaviours while females have subtler symptoms, leading to missed diagnoses (Lin, H Y. et al.2023).

Comorbidities: ADHD individuals often have anxiety, depression, and learning difficulties, complicating the diagnosis. Girls with ADHD are less likely to have disruptive behaviours, contributing to underdiagnosis (Biederman et al. 2002).

Machine Learning Impact: The research suggests using ML models to improve ADHD diagnosis accuracy and reduce gender biases. Logistic regression, Random Forest, and SVM models showed that ML can identify missed patterns.

ASD Results and Findings:

Gender Bias: ASD diagnosis is often focused on males, leading to underdiagnosis or misdiagnosis of females who may mask their symptoms.

Age and ASD: Older individuals are slightly more likely to be diagnosed with ASD, but gender does not affect diagnosis rates (Moura, et al., 2023).

ML in ASD Diagnosis: ML models like Random Forest and Logistic Regression can improve accuracy in diagnosing ASD, especially for subtle symptoms in females.

Family History and Genetics: There is a strong link between a family history of autism and being diagnosed with ASD, suggesting a genetic component that could help predict diagnosis.

6.2 Contribution to research:

6.2.1 Addressing Gender Bias in ADHD and ASD Diagnoses:

This research focuses on gender biases in the diagnostic processes for ADHD and ASD. Diagnostic criteria have been based on male symptoms, leading to underdiagnosis or misdiagnosis of females. The study highlights the importance of including gender-specific symptomatology in diagnostic frameworks. It explores how symptoms appear differently in males and females, specifically the subtler symptoms in females. This research offers insights into improving diagnostic criteria to be more inclusive.

6.2.2 Integration of Machine Learning for Diagnostic Precision

One key finding of this study is that machine learning models can boost the precision and impartiality of ADHD and ASD diagnoses. The research shows that algorithms like Logistic Regression, Random Forest, and KNN can greatly improve diagnostic accuracy by identifying patterns in symptoms and demographic data. This is particularly useful for females and other underrepresented groups. The study highlights the potential of machine learning to forecast diagnostic results based on gender and age, making diagnoses less biased and more tailored.

6.2.3 Economic Implications for Healthcare Systems

This study shows how using machine learning in diagnostics can save money by reducing consultations and shortening the diagnostic timeline. Significant cost savings are estimated, particularly in large healthcare systems like the NHS.

6.3 Conclusions and Suggestions

The research highlights gender biases in ADHD and ASD diagnosis, leading to underdiagnosis and misdiagnosis of females. Machine learning can help reduce biases and improve accuracy in diagnosis by considering gender and other factors. The study suggests the importance of diagnostic frameworks that consider gender differences in symptoms. Using machine learning can help save money by reducing the need for multiple consultations. Challenges still exist, including avoiding overfitting in models and ensuring diversity in datasets.

Suggestions

Healthcare providers and guidelines should update criteria for ADHD and ASD to identify gender-specific symptoms, reducing underdiagnosis in females and ensuring fairer treatment.

Medical facilities should use clear Machine Learning models for diagnosing ADHD and ASD, taking into consideration gender differences to ensure a fair and effective process. Enhance data collection by using more diverse datasets for training machine learning models. Include balanced numbers of males and females and individuals from various ethnic and socio-economic backgrounds to improve model accuracy and fairness.

Healthcare providers need to regularly evaluate AI diagnostic tools to make sure they are accurate and unbiased. It is important to continuously train the models with new data to keep them effective.

Healthcare providers need more training on using ML tools and understanding gender-specific signs of ADHD and ASD. This will help them use technology-based diagnostic methods in clinics.

Public Health Policy Adjustments: Policymakers should back AI and ML tools in healthcare to reduce diagnostic delays and improve outcomes for all patients, irrespective of gender, by providing infrastructure and funding.

6.4 Future Work

Future research needs to concentrate on overcoming the restrictions found in this study. One key aspect for future investigations is creating more extensive datasets that encompass a wider range of people in terms of gender, ethnicity, and comorbidities. This would enhance the applicability of machine learning models and decrease biases in diagnosis. Moreover, more sophisticated methods like deep learning or bias-mitigation algorithms could be investigated to enhance the precision and fairness of diagnostic tools. One potential research direction is incorporating machine learning models into clinical practice. Working with healthcare providers can improve the models and create easy-to-use interfaces for clinicians. Regular updates with new data can help keep the models up-to-date.

References:

About autism spectrum disorder (2024) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/autism/about/index.html (Accessed: 13 September 2024).

Arnett, A.B. *et al.* (2014) *Sex differences in ADHD symptom severity*, *Journal of Child Psychology and Psychiatry, and allied disciplines*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4385512/ (Accessed: 03 June 2024).

Autism spectrum quotient (AQ-10) test: Autism spectrum disorder in adults: Diagnosis and management: Guidance (2012) NICE. Available at: https://www.nice.org.uk/guidance/cg142/resources/autism-research-centre-autism-spectrum-quotient-aq10-test-186582493 (Accessed: 13 September 2024).

ADHD in women: Symptoms, treatment, and support (2024) HelpGuide.org. Available at: https://www.helpguide.org/mental-health/adhd/adhd-in-women (Accessed: 10 September 2024).

Avlund, S.H. et al. (2021) Factors associated with a delayed autism spectrum disorder diagnosis in children previously assessed on suspicion of autism - journal of autism and developmental disorders, SpringerLink. Available at: https://link.springer.com/article/10.1007/s10803-020-04849-x#citeas (Accessed: 27 August 2024).

Bahathiq, R.A. et al. (2022) 'Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging', Frontiers in Neuroinformatics. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9554556/ (Accessed: 15 April 2024).

Belz, F.M. and Peattie, ken (2009) *(PDF) Sustainability Marketing — an innovative conception of marketing*, *Research Gate*. Available at: https://www.researchgate.net/publication/225723866_Sustainability_marketing_-_An_innovative_conception_of_marketing (Accessed: 03 June 2024).

Biederman, J. et al. (2002) Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic, American Journal of Psychiatry. Available at: https://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.159.1.36 (Accessed: 08 June 2024).

Bloomdes (2019) A brief overview of the ADOS-2: An assessment for autism spectrum disorder, Children’s Resource Group - A Multi-Specialty Behavioral Health Practice. Available at: https://www.childrensresourcegroup.com/a-brief-overview-of-the-ados-2-an-assessment-for-autism-spectrum-disorder/ (Accessed: 05 September 2024).

Bruchmüller, K., Margraf, J. and Schneider, S. (2012) *Apa PsycNet*, *American Psychological Association*. Available at: https://psycnet.apa.org/doiLanding?doi=10.1037%2Fa0026582 (Accessed: 27 August 2024).

Cao, O. (2023) 'Machine Learning for Brain Disorders', SpringerLink. Available at: https://link.springer.com/book/10.1007/978-1-0716-3195-9#about-this-book (Accessed: 07 April 2024).

Casseus, M. (2022) Prevalence of co-occurring autism spectrum disorder and attention-deficit/hyperactivity disorder among children in the United States, Autism : the International Journal of research and Practice. Available at: https://pubmed.ncbi.nlm.nih.gov/35362330/ (Accessed: 13 September 2024).

Casseus, M., Kim, W. and Horton, D.B. (2023) Wiley Online Library | Scientific Research Articles, journals, ..., Wiley Online Library. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2894 (Accessed: 13 September 2024).

Canals, J. et al. (2024) Prevalence of comorbidity of autism and ADHD and associated characteristics in school population: Epined Study, Autism research : official journal of the International Society for Autism Research. Available at: https://pubmed.ncbi.nlm.nih.gov/38695661/ (Accessed: 13 September 2024).

Colliot, O. (2023) 'Feature selection bias in machine learning models for ADHD and ASD diagnosis', Cognitive Neurodynamics. Available at: https://link.springer.com/article/10.1007/s11571-023-09765-8 (Accessed: 10 April 2024).

Dan, B. (2021) *Sex differences in neurodevelopmental disorders*, *Developmental medicine and child neurology*. Available at: https://pubmed.ncbi.nlm.nih.gov/33792030/ (Accessed: 27 August 2024).

Data and statistics on ADHD (2024) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/adhd/data/index.html (Accessed: 03 September 2024).

DOWNEY, K.K. et al. (1997) Adult attention deficit hyperactivity disorder:... : The Journal of nervous and mental disease, The Journal of Nervous & Mental Disease. Available at: https://journals.lww.com/jonmd/abstract/1997/01000/adult_attention_deficit_hyperactivity_disorder_.6.aspx (Accessed: 16 June 2024).

Fombonne, E. (2021) Epidemiological surveys of autism and other pervasive developmental disorders: An update - journal of autism and developmental disorders, SpringerLink. Available at: https://link.springer.com/article/10.1023/A:1025054610557#citeas (Accessed: 30 May 2024).

Gesi, C. et al. (2021a) Gender differences in misdiagnosis and delayed diagnosis among adults with autism spectrum disorder with no language or intellectual disability, Brain sciences. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8306851/ (Accessed: 15 August 2024).

Gesi, C. et al. (2021b) Gender differences in misdiagnosis and delayed diagnosis among adults with autism spectrum disorder with no language or intellectual disability, MDPI. Available at: https://www.mdpi.com/2076-3425/11/7/912 (Accessed: 06 June 2024).

Ghasemi, E., Ebrahimi, M. and Ebrahimie, E. (2022) 'Machine learning models effectively distinguish attention-deficit/hyperactivity disorder using event-related potentials', Cognitive Neurodynamics. Available at:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9666608/ (Accessed: 10 April 2024).

Kentrou, V., de Veld, D. M., Mataw, K. J., & Begeer, S. (2019). Delayed autism spectrum disorder recognition in children and adolescents previously diagnosed with attention-deficit/hyperactivity disorder. *Autism : the international journal of research and practice*, *23*(4), 1065–1072. https://doi.org/10.1177/1362361318785171

Lamanna, A.L. et al. (2017) Risk factors for the existence of attention deficit hyperactivity disorder symptoms in children with autism spectrum disorders, Neuropsychiatric disease and treatment. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5478272/ (Accessed: 29 May 2024).

Lin, H.-Y. *et al.* (2023) *Gender differences in auditory and visual attentional performance in children with and without ADHD*, *OUP Academic*. Available at: https://academic.oup.com/acn/article-abstract/38/6/891/7042575?redirectedFrom=fulltext&login=false (Accessed: 01 September 2024).

Mahendiran, T. et al. (2019) Sex differences in social adaptive function in autism spectrum disorder and attention-deficit hyperactivity disorder, Frontiers in psychiatry. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6751776/ (Accessed: 01 June 2024).

Makris, G., Eleftheriades, A. and Pervanidou, P. (2023) *Early life stress, hormones, and neurodevelopmental disorders*, *Karger Publishers*. Available at: https://karger.com/hrp/article/96/1/17/841578/Early-Life-Stress-Hormones-and-Neurodevelopmental (Accessed: 02 September 2024).

Malwane, M. I., Nguyen, E. B., Trejo, S., Jr, Kim, E. Y., & Cucalón-Calderón, J. R. (2022). A Delayed Diagnosis of Autism Spectrum Disorder in the Setting of Complex Attention Deficit Hyperactivity Disorder. *Cureus*, *14*(6), e25825. https://doi.org/10.7759/cureus.25825

McGough, J.J. et al. (2005) Psychiatric comorbidity in adult attention deficit hyperactivity disorder: Findings from Multiplex Families, American Journal of Psychiatry. Available at: https://ajp.psychiatryonline.org/doi/full/10.1176/appi.ajp.162.9.1621 (Accessed: 04 June 2024).

Mikolas, P. et al. (2022) 'Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records', Nature Scientific Reports. Available at: https://www.nature.com/articles/s41598-022-17126-x (Accessed: 04 April 2024).

Moura, J.D.C. *et al.* (2023) *Autism spectrum disorders - gender differences and the diagnosis dilemma*, *European Psychiatry*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10661480/ (Accessed: 10 August 2024).

Rizvi, N. and Mrini, K. (2022) 'Using HCI to tackle race and gender bias in ADHD diagnosis', arXiv.org. Available at: https://arxiv.org/abs/2204.07900 (Accessed: 08 April 2024).

Shekim, W. O. et al. (1990). A clinical and demographic profile of a sample of adults with attention deficit hyperactivity disorder, residual state. Comprehensive psychiatry, 31(5), 416–425. https://doi.org/10.1016/0010-440x(90)90026-o

Siddiqui , U. et al. (2024) Sex differences in diagnosis and treatment timing of comorbid depression/anxiety and disease subtypes in patients with ADHD: A database study, Journal of Attention Disorders. Available at: https://pubmed.ncbi.nlm.nih.gov/38756010/ (Accessed: 13 September 2024).

Sumner, M. (2018) *Written evidence submitted by Dr. Mark Sumner*, *SFI0026 - evidence on sustainability of the Fashion Industry*. Available at: https://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/environmental-audit-committee/sustainability-of-the-fashion-industry/written/88396.html (Accessed: 03 June 2024).

Ter-Minassian, L. et al. (2022) 'Assessing machine learning for fair prediction of ADHD in school pupils using a retrospective cohort study of linked education and Healthcare Data', BMJ Open. Available at: https://bmjopen.bmj.com/content/12/12/e058058 (Accessed: 05 April 2024).

Thomas, A.W. et al. (2019) 'Analyzing neuroimaging data through recurrent deep learning models', Frontiers in Neuroscience. Available at: https://www.frontiersin.org/articles/10.3389/fnins.2019.01321/full (Accessed: 11 April 2024).

Vos, M., Rommelse, N. N. J., Franke, B., Oosterlaan, J., Heslenfeld, D. J., Hoekstra, P. J., Klein, M., Faraone, S. V., Buitelaar, J. K., & Hartman, C. A. (2022). Characterizing the heterogeneous course of inattention and hyperactivity-impulsivity from childhood to young adulthood. European child & adolescent psychiatry, 31(8), 1–11. https://doi.org/10.1007/s00787-021-01764-z

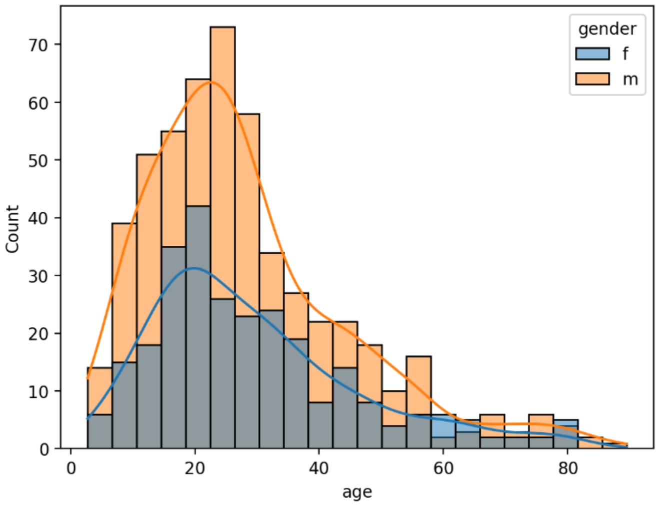
Appendix:

ADHD200.ipynb

Autism2.ipynb

	0	1	2	3
Female	52.75	15.15	0	33.33
Male	47.25	84.85	100	66.67

ADHD Gender statistics



ASD gender and age distribution