DEVELOPMENT OF MACHINE LEARNING MODELS FOR THE BEHAVIOURAL ASSESSMENT OF AUTISM SPECTRUM DISORDER

A THESIS SUBMITTED TO THE INSTITUTE OF GRADUATE STUDIES
OF
NEAR EAST UNIVERSITY

By
ABDULMALIK AHMAD LAWAN

In Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy
in
Computer Information Systems

NICOSIA, 2021

ABDULMALIK AHMAD LAWAN
DEVELOPMENT OF MACHINE LEARNING MODELS FOR THE BEHAVIOURAL ASSESSMENT OF AUTISM SPECTRUM DISORDER
NEU 2021

DEVELOPMENT OF MACHINE LEARNING MODELS FOR THE BEHAVIOURAL ASSESSMENT OF AUTISM SPECTRUM DISORDER

A THESIS SUBMITTED TO THE INSTITUTE OF GRADUATE STUDIES
OF
NEAR EAST UNIVERSITY

By
ABDULMALIK AHMAD LAWAN

In Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy
in
Computer Information Systems

NICOSIA, 2021

ABDULMALIK AHMAD LAWAN: DEVELOPMENT OF MACHINE LEARNING MODELS FOR THE BEHAVIOURAL ASSESSMENT OF AUTISM SPECTRUM DISORDER

Approval of Director of the Institute of Graduate Studies

PROF.DR. HÜSNÜ CAN BAŞER

We certify this thesis is satisfactory for the award of the Degree of Doctor of Philosophy in Computer Information Systems

Examining Committee in Charge:

ASSOC. PROF. DR. MÜESSER NAT	Committee Chairperson, Department of Management Information Systems, NEU
PROF. DR. NADİRE ÇAVUŞ	Supervisor, Department of Computer Information Systems, NEU
PROF. DR. GÖNÜL AKÇAMETE	Department of Special Education, NEU
PROF.DR. HÜSEYİN UZUNBOYLU	Higher Education Planning, Evaluation, Accreditation and Coordination Council, YODAK
PROF. DR. FEZİLE ÖZDAMLI	Department of Computer Information Systems, NEU

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: ABDULMALIK AHMAD LAWAN

Signature:

Date: 07/06/2021

ACKNOWLEDGEMENT

This thesis would not have been possible without the help, support and patience of my main supervisor, my deepest appreciation goes to Prof. Dr. Nadire Cavus, for her relentless encouragement and guidance. She has walked me through all my postgraduate studies. Without her consistent and illuminating instruction, this thesis could not have reached its present form.

My profound gratitude goes to my place of work; Kano University of Science and Technology and the Kano state government through scholarship board for their financial support throughout my PhD program.

I wish to acknowledge all the NEU teachers that have taught me during my coursework.

Above all, my unlimited thanks and heartfelt love would be dedicated to my dearest family for their loyalty and their great confidence in me. I am greatly indebted to my late father Malam Ahmad Lawan for the excellent upbringing and educational inspiration. I would like to thank my daring mother Malama Aisha Musa for giving me all the support, encouragement and constant love since from the onset of my life. I would like to thank my lovely wife Surayya Muhammad Hassan for her personal support and great patience at all times.

Outstanding gratitude to Dr. Muideen Bakare and the Childhood Neuropsychiatric Disorders Initiative (CNDI) for providing me with the data collection instrument and relevant clinical supports. Special gratitude to my friends that helped me in data collection. To all esteemed acquaintances that supported in one way or the other, I say thank you all.

Special thanks to the thesis defense jury members for their understanding and insightful comments that improved the quality of the thesis document and shade more light on the research endeavor.

To my parents…

ABSTRACT

Autism Spectrum Disorder (ASD) is a neuropsychiatric disorder associated with significant social, communication, and behavioral challenges. Neither the cause nor the cure of ASD is clearly understood. The insufficient number of trained clinicians with the lack of accurate and accessible diagnostic tools resulted in overlooking symptoms of ASD in children around the world. Several studies suggested machine learning (ML) systems for quick and accurate assessment of ASD. However, despite the promising metrics achieved by the ML algorithms, numerous challenges limit the real-life implementation of the ML-based systems. The challenges are related to misalignments of the data pre-processing techniques and the ML algorithms with the concepts upon which professionals assess ASD. Specifically, compromising the validity of the assessment tools by the common dimensionality reduction techniques is among the key challenges identified. The aim of this study is to propose a clinically valid ML-based ASD screening approach. The present study conducted a data collection using a novel tool titled Child Development for Household Survey to Estimate Burden of ASD (CDHSEBA). The collected data contains 171 ASD and 209 control cases gathered based on purposive sampling approach. The collected data was utilized in developing multiple machine learning models using various combination of the CDHSEBA questionnaire items. A comparative analysis of the performances of the machine learning models and an empirical CDHSEBA-based scoring algorithm was conducted. The best performing ML model based on Naïve Bayes classification algorithms achieved the highest classification accuracy of 88% while the empirical scoring algorithm achieved the classification accuracy of 56%. Overall, the study findings revealed that the empirical scoring approach preserves the clinical validity of the screening instruments based on the high true positives rate of 97% obtained. On the other hand, most of the machine-learning models outperformed the empirical scoring method in the correct classification of non-ASD cases because of its low true negatives rate of 23%. The study provides a roadmap for developing effective ML-based ASD screening and diagnostic systems that comprises few behavioral features and preserve clinical relevance. The study will also guide researchers, neuropsychiatrists, and relevant stakeholders on the advances in ASD assessment with ML.

Keywords: Autism spectrum disorder; screening; diagnosis; artificial intelligence; machine learning

ÖZET

Otizm Spektrum Bozukluğu (OSB), önemli sosyal, iletişim ve davranışsal zorluklarla ilişkili bir beyin bozukluğudur. Otizmin nedeni ve tedavisi net olarak anlaşılmamıştır. Yeterli sayıda eğitimli klinisyen, doğru ve erişilebilir teşhis araçlarının eksikliği ile birleştiğinde, dünya çapında çocuklarda otizmin erken semptomlarının gözden kaçmasına neden oldu. Birçok çalışma, OSB'nin hızlı ve doğru bir şekilde değerlendirilmesi için makine öğrenimi (MÖ) sistemlerini önermektedir. Bununla birlikte, OSB'nin davranışsal değerlendirmesinde makine öğrenimi algoritmaları tarafından elde edilen umut verici ölçütlere rağmen, çok sayıda zorluk makine öğrenimi tabanlı sistemlerin gerçek hayattaki uygulamasını sınırlamaktadır. Zorluklar, temel olarak veri ön işleme tekniklerinin ve MÖ algoritmalarının, profesyonellerin OSB değerlendirme araçlarını oluşturduğu ve kullandığı kavramsal temele sahip yanlış hizalamalarıyla ilgilidir. Özellikle, ortak boyut indirgeme teknikleri ile değerlendirme araçlarının geçerliliğinden ödün vermek, belirlenen temel zorluklar arasındadır. Bu çalışmada, OSB Yükünü Tahmin Etmek için Hanehalkı için Çocuk Gelişimi Anketi (CDHSEBA) adlı yeni bir değerlendirme aracı kullanarak veri toplama gerçekleştirilmiştir. Amaçlı örnekleme yaklaşımına dayalı olarak 171 OSB ve 209 kontrol vakasını içeren veri toplanmıştır. Toplanan veriler, CDHSEBA anket maddelerinin çeşitli kombinasyonlarını kullanarak birden çok makine öğrenimi modelinin geliştirilmesinde kullanıldı. Çalışma, makine öğrenimi modellerinin performanslarının karşılaştırmalı analizi ve deneysel bir CDHSEBA tabanlı puanlama algoritması ile yürütüldü. Naïve Bayes sınıflandırma algoritmalarına dayanan en iyi performans gösteren MÖ modeli, % 88 ile en yüksek sınıflandırma doğruluğunu elde ederken, ampirik puanlama algoritması % 56 sınıflandırma doğruluğuna ulaştı. Sonuç olarak, çalışma bulguları ampirik puanlama yaklaşımının, elde edilen % 97'lik yüksek gerçek pozitif oranına dayalı olarak tarama araçlarının klinik geçerliliğini koruduğunu ortaya koymuştur. Ayrıca, makine öğrenme modellerinin çoğu, % 23'lük düşük gerçek negatif oranı nedeniyle OSB olmayan vakaların doğru sınıflandırılmasında ampirik puanlama yönteminden daha iyi performans göstermiştir. Çalışma, birkaç davranışsal özelliği içeren ve klinik alaka düzeyini koruyan etkili MÖ tabanlı OSB tarama ve teşhis sistemleri geliştirmek için bir yol haritası sağlayacaktır. Ayrıca çalışmanın araştırmacılara, nöropsikiyatristlere ve ilgili paydaşlara makine öğrenimi ile OSB değerlendirmesindeki ilerlemeler konusunda rehberlik edeceği de ümit edilmektedir.

Anahtar Kelimeler: Otizm spektrum bozukluğu; tarama; teşhis; yapay zeka; makine öğrenme

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC: Boston Autism Consortium

ADI-R: Autism Diagnostic Interview-Revised

ADOS: Autism Diagnostic Observation Schedule

ADTree: Alternative Decision Tree

AGRE: Autism Genetic Resource Exchange

ANN: Artificial Neural Network

APADA: Association of Parents and Friends for the Support and Defense of the rights of people with Autism

AQ: Autism Quotient

ASD: Autism Spectrum Disorder

AUC: Area Under the Curve

BACO: Binary Ant Colony Optimization

CATC: Clustering-based Autistic Trait Classification

CDHSEBA: Child Development for Household Survey to Estimate Burden of ASD

CNN: Convolutional Neural Network

FS: Feature Selection

FT: Feature Transformation

ICTs: Information Communication Technologies

KNN: K-Nearest Neighbor

LDA: Linear Discriminant Analysis

LR: Logistic Regression

MFCM: Multilayer Fuzzy Cognitive Maps

MGOA: Modified Grasshopper Optimization Algorithm

ML: Machine Learning

NB: Naïve Bayes

NDAR: National Database for Autism Research

OBS: Observation-based Classification

PASS: Pictorial Autism Assessment Schedule

Q-CHAT: Quantitative Checklist for Autism Toddlers

RF: Random Forest

RIPPER: Repeated Incremental Pruning to Produce Error Reduction

SRS: Social Responsiveness Scale

SSC: Simons Simplex Collection

SVIP: Simons Variation in Individuals Project

SVM: Support Vector Machine

CHAPTER 1

INTRODUCTION

This chapter provides theoretical background on the research topic including statements on the limitation of the previous relevant studies, the key motivation for the present study as well as the research aim and objectives. Specifically, this introductory chapter contains seven subsections including background, problem statement, aim and objectives, significance of the study, limitations of the study, and overview of the thesis.

1.1 Background

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental disorder characterized by communication impairment, restrictive and compulsive behavior. Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) termed ASD as a constellation of four disorders that shared similar deficits in communication, social interaction, and behavior. These disorders were autistic disorder, Asperger's syndrome, childhood disintegrative disorder, and Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) (American Psychiatric Association, 2013). The primary symptoms for diagnosing ASD are deficits in social communication and the presence of restricted, repetitive patterns of behavior, interests, or activities. These symptoms must be present in early childhood and impair the child's everyday functioning (American Psychiatric Association, 2013; Bakare & Munir, 2011). The rising prevalence of ASD necessitates the need for early and cost-effective ASD diagnosis to set the path for appropriate, and efficient treatment (Baio et al., 2018; Chauhan et al., 2019). Early diagnosis of ASD also leads to improved outcomes in communication and social interaction and guides parents to the right interventions in school, home, and clinic (Case-Smith, Weaver, & Fristad, 2015; Durkin et al., 2015; Matson & Konst, 2014). Thus, the need for cost-effective assessments coupled with the global rise in ASD cases necessitates extensive research. This is because the current clinical assessment of ASD is not cost-effective; studies have shown that the cost of identifying one child with ASD in universal screening settings is about 700,000 USD (Yuen, Carter, Szatmari, & Ungar, 2018). Additionally, the assessment instruments perform poorly (Guthrie et al., 2019; Øien et al., 2018; Surén et al., 2019); there is a high number of false negatives and false positives. The tradeoff was to create a tool that reduced false positives (such as the M-CHAT R) and implement robust scoring methods. However, the tradeoff leads to more false-negative cases. Addressing these challenges lead to several suggestions including the so-called quick and accurate Machine Learning

(ML)-enabled ASD assessment systems (Campbell et al., 2017; Shahamiri & Thabtah, 2020; Thabtah, 2019; Wingfield et al., 2020). The promising results realized with ML algorithms across various research fields motivated these suggestions and made it a vital step toward cost-effective assessment of ASD.

In other words, the research aim of this study will explore the advances in the application of machine learning in the behavioral assessment of ASD and propose a novel ML-based approach that preserves the clinical validity of the screening and diagnostic instrument by adhering to the conceptual basis upon which professionals diagnose ASD. In line with the research aim, the present study systematically reviewed recent articles on the application of machine learning algorithms toward quick and accurate assessment of ASD. Thus, challenges were identified and future directions toward real-life ML-based assessment systems were proposed. Nonetheless, the existing literature have shown that the fundamental aim of the previous studies was on increasing diagnostic speed by reducing the input parameters and improving classification performance by achieving high evaluation metrics of accuracy, sensitivity and specificity among others. However, their little scientific evidence on the validity of the findings against the conceptual basis upon which professionals built and utilized ASD assessment tools.

1.2 Problem Statement

The key challenges in the research area are the inherent discrepancies within the research data used in the previous studies as well as the lack of conceptual understanding on the relevance of the data-centric approaches utilized with the basis upon which professionals diagnose the disorder. Filling the research gap entails providing a definitive explanation on the relevance of the data-centric approach used especially in the data pre-processing stages as well as the machine learning modeling with the conceptual basis used by professionals in building and utilizing the standard ASD diagnostic instruments. Accordingly, a definitive explanation of the clinical validity and sufficiency of the reported findings would lead to a viable pathway toward real-life implementation of the ML-based assessment systems.

Recently, there is an increasing application of ML in assessing ASD based on either genetic (Ghafouri-Fard et al., 2019; Sekaran & Sudha, 2021), brain imaging (Fu & Costafreda, 2013; Jack, 2018; Moon, Hwang, Kana, Torous, & Kim, 2019), physical biomarkers (Raya et al., 2020; Dahiya et al., 2020; Hashemi et al., 2018; Liu et al., 2016; Sarabadani et al., 2020), or behavioral data. Worthy of note, despite the excellent evaluation metrics reported in the ML-based behavioral studies, it is evident that the research methods and the resulting ML models are liable to fail professionals' assessments. For instance, apart from improving the diagnostic accuracy, most of the studies focused on reducing or transforming the items of the assessment instruments using various data-centric approaches. However, most of the studies fail to probe the relevance of the data-centric approaches, the sufficiency of the input parameters as well as the resulting ML models against the basic assumptions for the clinical assessment of ASD symptoms.

In other words, the data-centric approaches utilized in the studies, and the resulting ML models, missed the human knowledge upon which ASD is assessed; a combination of the optimized parameters, most often, violates the fundamental assumptions upon which clinicians diagnose ASD (Thabtah, 2018). Secondly, studies have revealed that discrepancies within the data repositories are among the key challenges that limit the reliability of the results reported in the machine learning studies (Abdelhamid et al., 2020; Alahmari, 2020). For instance, Torres et al. (2020) studied the statistical properties of ADOS scores from 1324 records and identified various factors that could undermine the scientific viability of the scores. Particularly, the empirical distributions in the generated scores violate the theoretical requirements of normality and homogeneous variance, which are essential for independence between bias and sensitivity. Thus, they suggested readjusting the scientific use of ADOS due to the variation in the distribution and dispersion of the scores, the lack of proper metrics to define similarity measures to characterize change, and the impact that these elements have on sensitivity-bias codependencies and longitudinal tracking of ASD. Thirdly, misalignment of the ML algorithms, and the resulting models with the scientific basis of the assessment tools is another factor that limits the reliability of the reported evaluation metrics (Thabtah, 2018). Specifically, most of the studies employed various data optimization techniques for parameter reduction. However, the sufficiency of the reduced parameters in identifying ASD symptoms is questionable and liable to fail professionals' assessment. Additionally, most of the studies overlooked the inherent limitations associated with the assessment instruments especially the reported high number of false negatives and false positives (Guthrie et al., 2019; Øien et al., 2018; Surén et al., 2019). Thus, the need for ascertaining and improving the reliability of the assessment instruments remains.

1.3 Aim and Objectives

The main aim of the present study is to propose a quick, accurate and clinically valid ASD screening method based on machine-learning models. Accordingly, three sub-aims are derived to address the key aim of the study. The sub-aims are itemized as follows:

1. What is the classification performance of the empirical scoring algorithm?
2. What is the classification performance of the various
    - machine-learning models for the experimental Scenario 1
    - machine-learning models for the experimental Scenario 2
    - machine-learning models for the experimental Scenario 3
    - machine-learning models for the experimental Scenario 4

   based on the complete as well as the reduce set of input parameters?
3. What is the comparative performance of the machine-learning models and the empirical scoring algorithm based on the evaluation metrics of
    - True positives?
    - False positives?
    - True negatives?
    - False negatives?
    - Sensitivity?
    - Specificity?
    - Classification accuracy?

The objectives of the study are identified as follows:

- To analyse the performance of the empirical scoring approach.
- To experiment with multiple machine-learning models based on various input scenarios that align/misalign with the empirical/clinical approach.
- To make comparative analyses on the performance of the experimental approaches.

1.4 Significance of the Study

Despite several studies on the application of machine learning algorithms in quick and accurate screening and diagnosis of autism spectrum disorder, there is no evidence on the real-life implementation of the machine-learning models for clinical use. This might be related to the misalignments between the conceptual understanding upon which professionals diagnose ASD and the data-centric approaches employed in developing the machine-learning models. The present study will purposely address the misalignments by demonstrating how machine-learning models for quick and accurate behavioral screening and diagnosis of ASD could be implemented by upholding the clinical procedures used by professionals in administering diagnostic instruments. Accordingly, the novel approach will utilize the advantages in machine-learning techniques while preserving the clinical validity of the assessment instrument.

1.5 Limitations of the Study

Common to other scientific studies, the present study is limited in its various stages especially during the literature review, and data collection. Firstly, the findings of the literature were constrained by the systematic approach followed, which considered only journal articles from the popular scientific databases and published in the English language. A detailed explanation on the limitations within the literature review process is provided in the third chapter. Secondly, the data collection process was initially planned to include both quantitative and qualitative data collection. Specifically, the qualitative data collection was planned to involve expert interviews with professionals in the clinical assessment of ASD from the eight Federal Neuropsychiatric Hospitals in Nigeria. Unfortunately, the sudden advent of the Corona Virus (COVID-19) and the resulting containment measures made it impossible to conduct qualitative data collection. Thirdly, the planned data analysis has not captured the history of the previous diagnostic tools used in the clinical diagnosis of the ASD patients as well as the class of participants with other non-ASD neuropsychiatric disorders. Capturing the stated diagnostic history could help in understanding the relevance of the previous diagnostic instruments with the novel instrument utilized in the present study.

1.6 Overview of the Thesis

The present thesis writeoff contains six distinct chapters described as follows:
The first chapter is the introductory part of the thesis report that explains the main study aim and statement of the problems that motivated embarking on the study, it also highlighted the significance of carrying out the study, as well as some of the limitations faced.

The second chapter entitled “related research” is the backbone of the research that updates readers on what was done on the research topic, what is the research gap, and what we could foresee from the future. Accordingly, the second chapter portrays the systematic approach followed in searching, analyzing, and discussing related works thematically, methodologically, and chronologically.

The third chapter “theoretical framework” provides the theoretical basis and understanding of the key concepts based on which the study was carried out including the conceptual understanding of the ASD screening and diagnostic instruments, the machine learning algorithm as well as the evaluation metrics.

Nonetheless, the fourth chapter entitled "methodology" describes the proposed research procedure utilized during the research data collection, data analysis, as well as report writing.

Consequent to the methodology, the fifth chapter provided a detailed explanation with the help of tables and figures on the results obtained and further discussed the findings.

The sixth, which is the final chapter, concludes the research work with summarized findings, and recommendations for future studies.

CHAPTER 2

RELATED RESEARCH

This chapter as the backbone of the research updates readers on what was done on the research topic, what is the research gap, and what we could foresee from the future. Accordingly, the third chapter portrays the systematic approach followed in searching, analyzing, and discussing related works thematically, methodologically, and chronologically. Specifically, this chapter explains the systematic review process adopted in the present study based on the PRISMA framework and the chapter contains numerous subheadings that explained trend of studies that utilizes ML in ASD screening and diagnosis, the commonly employed data processing and modeling techniques, as well as details of studies that utilized the common dimensionality reduction techniques. The chapter ended up with brief explanation on the novelty of the present research.

2.1 Background of the Literature

This section aims to explore the advances in the application of machine learning in the behavioral assessment of ASD. In this section, a systematic review was conducted on recent publications on the application of machine learning algorithms toward quick and accurate assessment of ASD. Thus, challenges were identified and future directions toward real-life ML-based assessment systems were explored. Certain literature reviews highlighted the accuracy and efficiency of ML algorithms in ASD assessment based on the reported evaluation metrics (Song, Kim, Bong, Kim, & Yoo, 2019; Thabtah, 2018). However, none of the existing literature reviews systematically analyzed the subject area and validates its findings against the scientific basis upon which professionals built and utilized ASD assessment tools. For instance, Thabtah (2018) identified some limitations within the research methodologies and proposed intuitive stages toward appending the ML models into ASD screening tools. Similarly, Song et al. (2019) reviewed 13 relevant studies on varying data types and discussed the possibility of achieving effective classification of ASD based on the study findings.

2.2 Literature Review Process

For this systematic search, conducted in October 2020, careful planning and scheduling of tasks were conducted at each stage to identify the most relevant studies. The search strategy was tailored to four electronic databases: IEEEXplore, PubMed, Scopus, Web of Science, and the search terms used are “Autism Spectrum Disorder” OR “Autistic Disorder” OR “Autism” AND “Screening” OR “Assessment” OR “Identification” OR “Test” OR “Detection” AND “Machine Learning” OR

“Artificial Intelligence”. All the searches spanned a decade (i.e. from 2011 to date) and included papers published with English titles. Beyond the database search, relevant publications on the advances in ASD assessment were accessed from other online sources.

PRISMA statement (Moher et al., 2009) was followed in the selection process. Relevant studies utilized PRISMA in critical appraisal and summary of literature to inform researchers on the advances in the assessment of autism and other neuropsychiatric disorders (Dahiya et al., 2020; Low, Bentley, & Ghosh, 2020; Moon et al., 2019; Song et al., 2019). The phenomena of interest in the criteria of inclusion included any published paper on the use of ML in ASD assessment. At the initial screening stage, apart from duplicates removal, the records were probed against the inclusion criteria to decide on whether or not to include the study in the literature review. The decisions for inclusion vs. exclusion on the records were coded under a designated column in the excel sheet imported from the databases.

Records that met the initial inclusion criteria were retrieved for the next screening stage. Thus, for records whose titles and corresponding abstracts aligned with the preset inclusion criteria, full-text articles of the studies were retrieved for the subsequent screening stage. In the next PRISMA screening stage, the researcher with the help of colleagues reviewed the downloaded papers, independently, to ascertain their relevance with the search query used as well as the set research question. Specifically, three hundred and sixty-seven records were carefully assessed for eligibility. One hundred and eighty studies out of the 367 records were discarded due to the following reasons: book chapters (n = 17), conference papers (n=138), editorial materials (n = 11), literature reviews (n = 15), not written in English (n = 9). The remaining one hundred and seventy-seven studies were further assessed; one hundred and forty-four records were eliminated because they are either based on brain imaging data (n = 57), genetic data (n = 35), or physical/metabolic biomarkers (n = 32) while others are intervention studies (n = 20). Consequently, thirty-three full-text articles were retrieved, read, and qualitatively assessed. Nonetheless, additional articles were excluded because ML is not the main method employed (n = 7) and ASD is not the main neuropsychiatric disorder assessed (n = 4). Finally, 22 studies met the inclusion criteria. The PRISMA flow diagram (Figure 2.1) summarized the abovementioned systematic literature review process and

Table 2.1 itemized the key items of the inclusion and exclusion criteria of the study.

Table 2.1: Inclusion and exclusion criteria of the study

Inclusion criteria
• Journal articles published in the English language • Documents published within the last ten years from 2011 to date • Full-text papers that are accessible and downloadable • Studies that utilized behavioral data • Studies that employed machine learning as the main technique • Studies that considered autism as the main disorder assessed
Exclusion criteria
• Papers that are written in other languages • Duplicated papers • Full-text of the document is not accessible on the internet • The study aim is not clearly defined • Studies that are not relevant to the stated research question • Relevant studies but machine learning are not the main methods • Relevant studies but autism is not the main disorder assessed • Conferences papers, editorial materials, and literature reviews • Studies that utilized data from either brain imaging, genetic, or physical/metabolic biomarkers. • Intervention studies

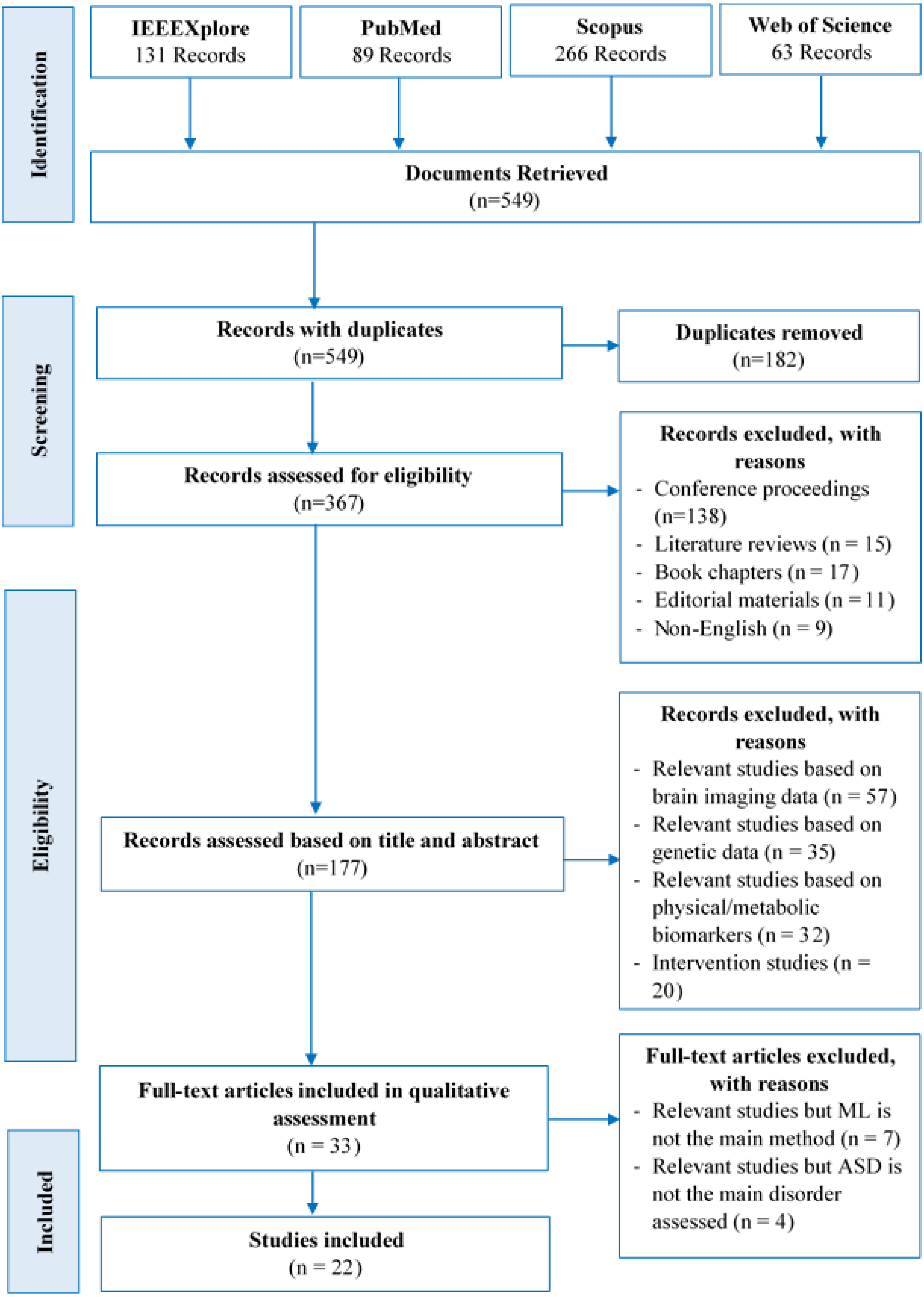


Figure 2.1: PRISMA flow diagram of the search results

In the present study, to ensure the quality of the systematic approach, the researcher carefully adhered to the planned systematic literature review process to maintain the quality of the study. Particularly, at every phase of the systematic literature review, the researcher ensured careful planning and allocation of tasks. The researcher created an online Mendeley repository for referencing and the researcher monitored the progress of the review based on preset milestones to ensure that all tasks complied with the scheduled deadlines. The Mendeley repository was also used in keeping track of the data extraction stages, noting essential observations and sharing vital contents related to the study. The researcher further upheld peer-reviewing with friends and rereading the text at each phase of the study to enhance the systematic literature review. Nevertheless, unbiased and constructive assessments on the systematic approach used in this study were sought from my supervisor and academic colleagues.

As the final stage of the study's PRISMA, the data extraction stage, 22 articles were appraised critically and the following information was extracted from the studies:

- Author(s) (year)
- Number of citations
- Source(s) of the research data
- Data collection/assessment instrument
- ML model(s)developed
- Best performing model(s)
- The key finding(s)

2.3 Descriptive Analysis on Trends and Status of the Study on ML in ASD Assessment

Based on the exported data, the trend of studies on the use of ML in the behavioral assessment of ASD showed the most cited references, the most cited journals, as well as citation and publication frequencies across the years.

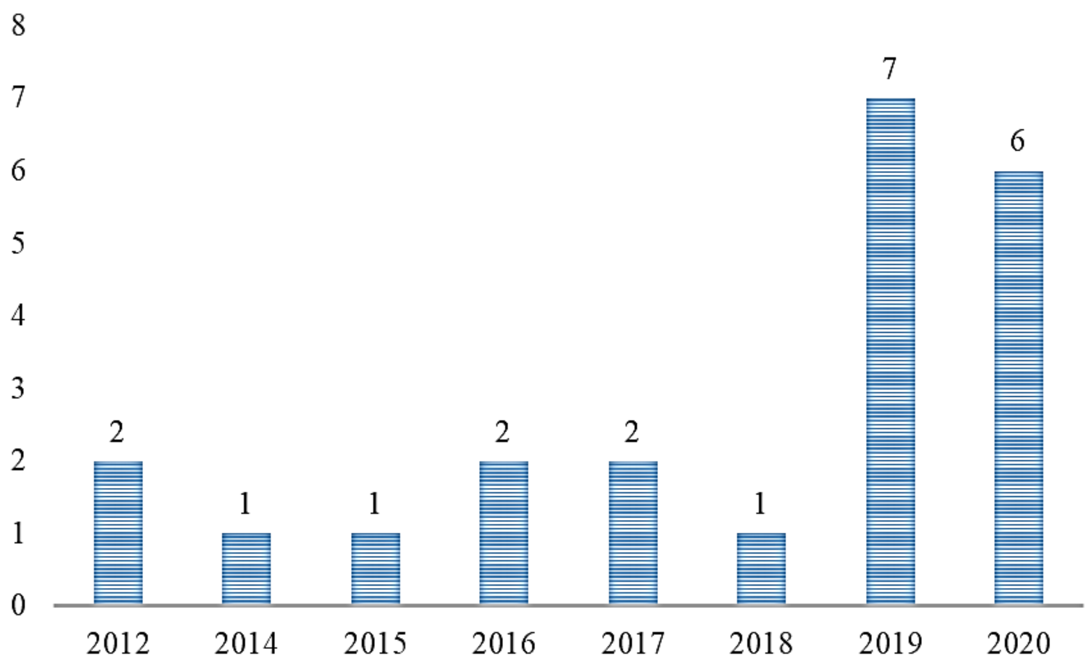


Figure 2.2: Articles distribution over the years

With the increasing application of ML in healthcare studies, as shown in

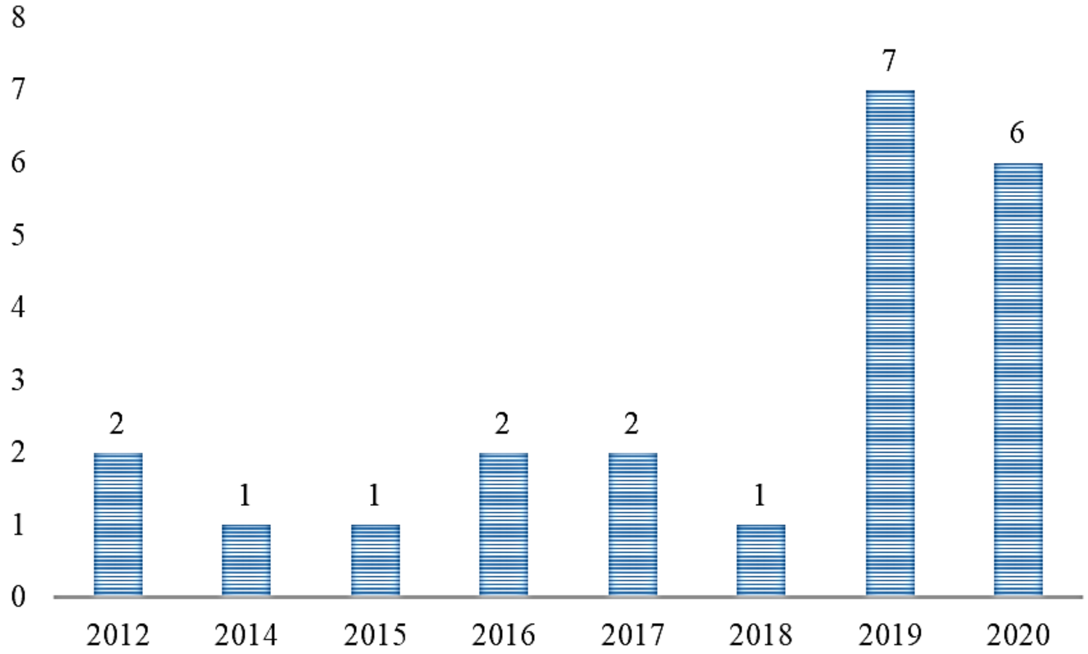


Figure **2.2**, there are more publications on ML and ASD assessment. From 2012 to 2018, not so many studies cared about the application of ML in ASD assessment. However, with the recently increased patronage of ML techniques across various fields, there is an increasing demand for intelligent tools for accurate assessment of ASD. From Figure 2.3, most of the articles contributing to the area were published in Translational Psychiatry (n = 5), followed by the Health Informatics Journal (n = 3). The remaining fifteen journals depicted published one article, each.

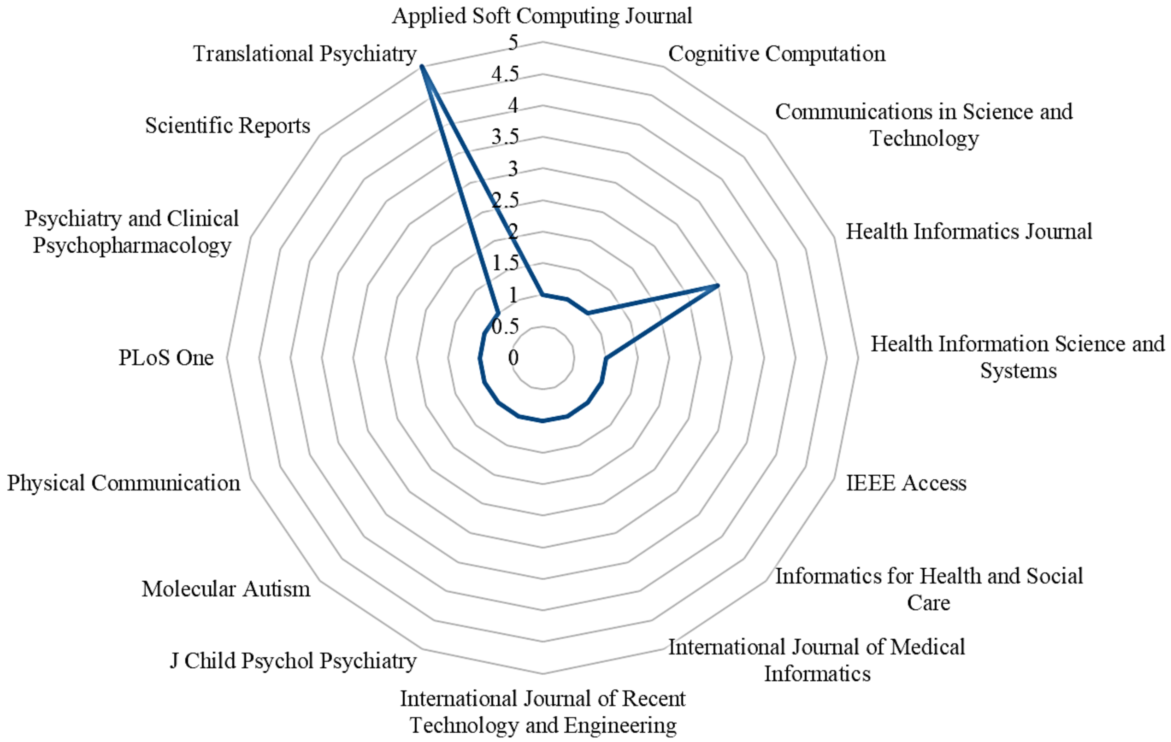


Figure 2.3: Number of articles published by journals

Based on the citation data exported, as shown in Table 2.2, we can see that the most cited references are Wall et al. (2012) (n = 160), Wall et al. (2012) (n = 106), Duda et al. (2016) (n = 89), Kosmicki et al. (2015) (n = 84), and Bone et al. (2016) (n = 77). Most of the significant references; with the highest number of citations, were published in Translational Psychiatry (Duda et al., 2016; Kosmicki et al., 2015; Wall, Kosmicki, et al., 2012) (Figure **2.4**2.4, n = 408) in the years 2012 (Figure 2.52.5, n = 266), 2015 (Figure **2.5**2.5, n = 84), and 2016 (Figure **2.5**2.5, n = 166). Figure **2.4**2.4 highlighted the citation data of the eight most cited journals involved in the study; Translational Psychiatry (n = 408), PLoS One (n = 106), Journal of Children Psychological Psychiatry (n = 77), and so on.

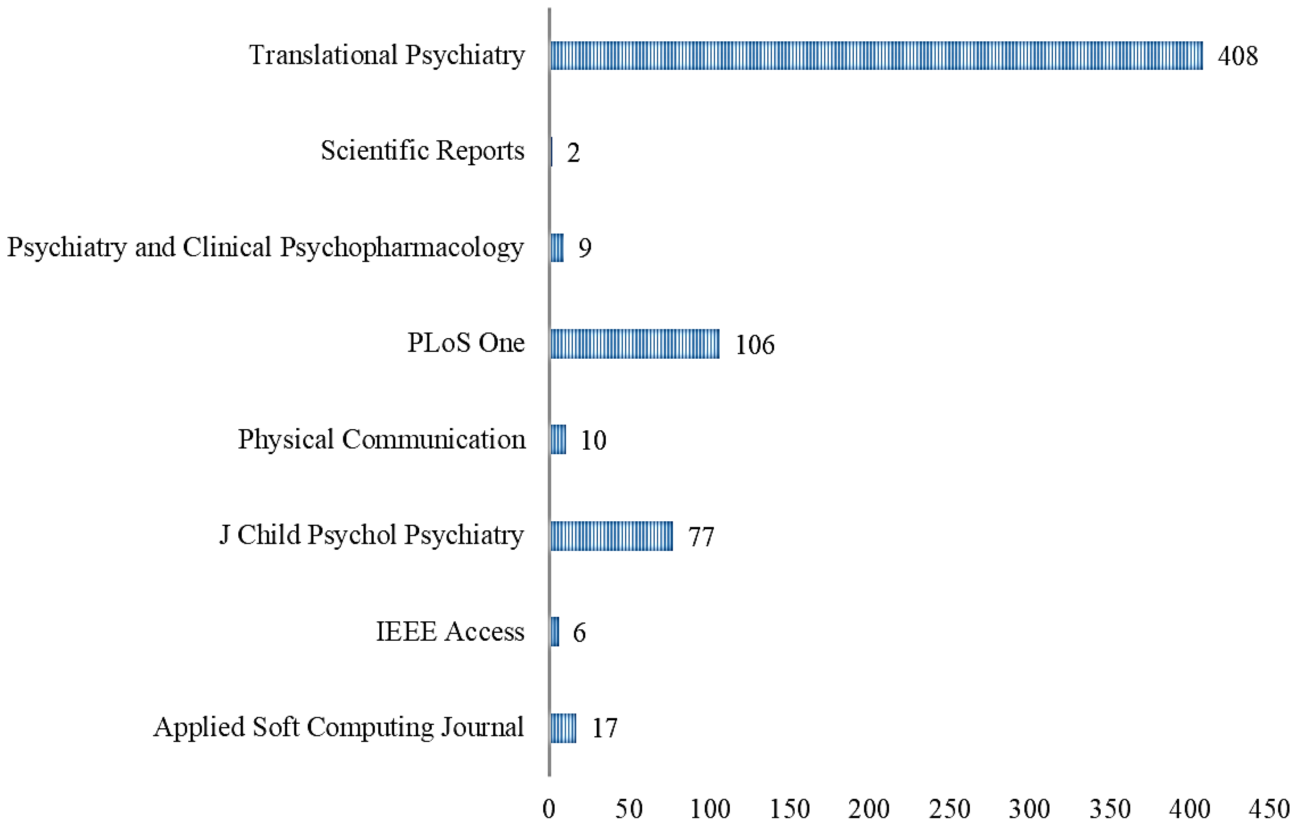


Figure 2.4: Sum of citations per journal

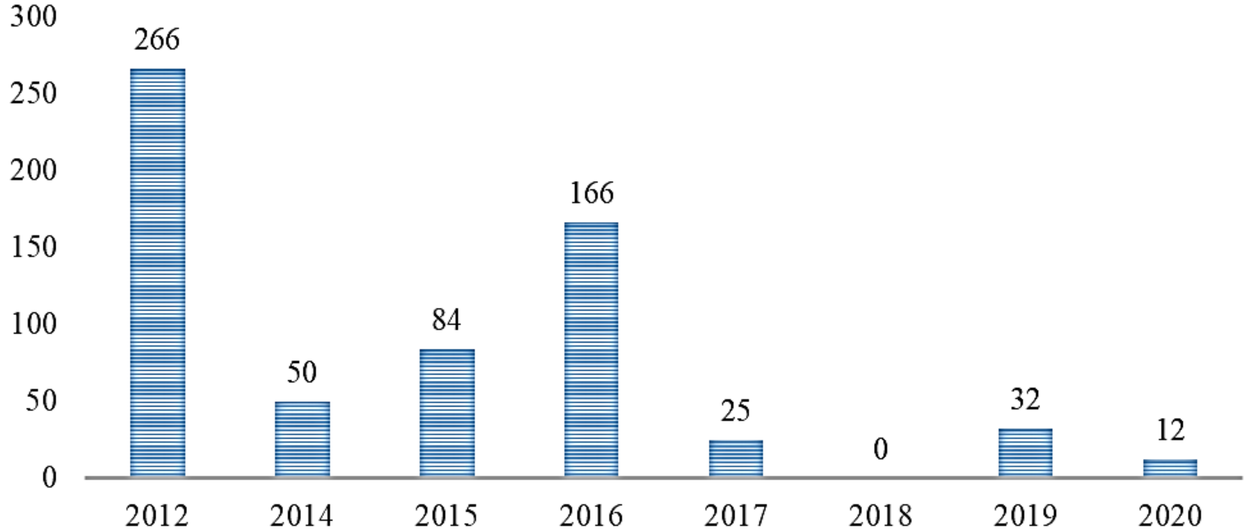


Figure 2.5: Number of citations across years

2.4 Trends in the Data Pre-Processing and ML Modeling Techniques

Dimensionality reduction is among the common data pre-processing technique employed in the previous studies. Most of the studies primarily aimed at streamlining the data collection instruments followed by evaluating the performance of various ML algorithms on the streamlined datasets (Bellesheim et al., 2018; Duda, Kosmicki & Wall, 2015; Kosmicki et al., 2015; Küpper et al., 2020; Wall, Dally, et al., 2012). While various feature selection methods were applied in streamlining the most influential features of the data collection instruments from the datasets, other studies utilized various feature transformation techniques in reducing the input parameters. For instance, in the work of Puerto et al. (2019), the inputs were fuzzified into membership values before applying the classification algorithms. Similarly, before implementing the classification models, Baadel et al. (2020) and Akter et al. (2019) transformed the inputs using clustering and feature transformation functions, respectively. Nonetheless, other studies employed a trial-error approach in selecting the most influential features. The trial-error approach involves repetitive evaluation of the ML models using a varying combination of the features; the most influential combination achieves superior results with fewer input parameters. Specifically, the studies utilized various feature selection techniques including trial-error (Duda et al., 2015; Thabtah, 2019; Usta et al., 2019; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012), Variable Analysis (Va) (Pratama et al., 2019; Thabtah et al., 2018), information gain (IG) and chi-square testing (CHI) (Thabtah et al., 2019), sequential feature selection (SFS) (Suresh & Renugadevi, 2019), correlation-based feature selection (CFS) and minimum redundancy maximum relevance (mRMR) (Wingfield et al., 2020). Additionally, ML-based feature selection techniques employed include recursive feature selection (Küpper et al., 2020), sparsity/parsimony enforcing regularization techniques (Levy et al., 2017), stepwise backward feature selection (Kosmicki et al., 2015), and forward feature selection (Duda et al., 2016).

The data-centric approaches reviewed in this section have employed at least one machine-learning algorithm in the model implementation and evaluation. As shown in Table **2.2**2.2, the commonly implemented ML algorithms are Random Forest (RF) (Baadel et al., 2020; Goel et al., 2020; Pratama et al., 2019; Wingfield et al., 2020), Support Vector Machines (SVM) (Bone et al., 2016; Kosmicki et al., 2015; Küpper et al., 2020; Levy et al., 2017; Suresh & Renugadevi, 2019), Alternative Decision Tree (ADTree) (Duda et al., 2015; Usta et al., 2019; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012), and Logistic Regression (LR) (Kosmicki et al., 2015; Thabtah, 2019; Thabtah et al., 2019). To achieve comparative results, most of the studies employed several algorithms such as Adaboost, Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Naïve Bayes, and K-Nearest Neighbor (KNN).

Data collection or assessment instruments are the bedrock in the behavioral studies in ASD assessment. The most utilized data collection instruments are AQ-10 (Akter et al., 2019; Baadel et al., 2020; Goel et al., 2020; Pratama et al., 2019; Shahamiri & Thabtah, 2020; Suresh & Renugadevi, 2019; Thabtah, 2019; Thabtah et al., 2019, 2018; Thabtah & Peebles, 2020), Q-CHAT-10 (Akter et al., 2019; Shahamiri & Thabtah, 2020; Thabtah et al., 2018; Thabtah & Peebles, 2020), ADOS (Duda et al., 2015; Kosmicki et al., 2015; Küpper et al., 2020; Levy et al., 2017; Puerto et al., 2019; Wall, Kosmicki, et al., 2012), ADI-R (Bone et al., 2016; Puerto et al., 2019; Wall, Dally, et al., 2012) and Social Responsiveness Scale (SRS) (Bone et al., 2016; Duda et al., 2017; Duda et al., 2016). Others include Autism Behavior Checklist, Aberrant Behavior Checklist, Clinical Global Impression (Usta et al., 2019), and MCHAT-based Pictorial Autism Assessment Schedule (PASS) (Wingfield et al., 2020). Thus, the need for improving the reliability of these assessment instruments and ascertaining their relevance in ML modeling remains.

Most of the studies utilized retrospective data that were publicly accessible. The most prominent sources of data utilized in the studies include Boston Autism Consortium (AC), Autism Genetic Resource Exchange (AGRE), Simons Simplex Collection (SSC) (Duda et al., 2017, 2015, 2016; Kosmicki et al., 2015; Levy et al., 2017; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012), National Database for Autism Research (NDAR) (Duda et al., 2015; Kosmicki et al., 2015) and Simons Variation In Individuals Project (SVIP) (Duda et al., 2015; Kosmicki et al., 2015; Levy et al., 2017). Other studies utilized data sets from ASDTest: Kaggle and UCI ML repository (Akter et al., 2019; Baadel et al., 2020; Goel et al., 2020; Pratama et al., 2019; Shahamiri & Thabtah, 2020; Suresh & Renugadevi, 2019; Thabtah, 2019; Thabtah et al., 2019, 2018; Thabtah & Peebles, 2020), Association of Parents and Friends for the Support and Defense of the rights of people with Autism (APADA) (Puerto et al., 2019), PASS app (Wingfield et al., 2020), Ondokuz Mayis University Samsun (Usta et al., 2019) and ASD outpatient clinics in Germany (Küpper et al., 2020). To achieve standardized comparative results, there is a need for standardized ASD data repositories for machine learning studies (Thabtah, 2018).

2.5 The Commonly Employed Experimental Procedures

Apart from the common aim of streamlining the various data collection instruments followed by model evaluation, other studies focused on either optimizing the machine-learning algorithms (Goel et al., 2020; Suresh Kumar & Renugadevi, 2019), proposing input optimization techniques (Akter et al., 2019; Baadel et al., 2020; Pratama et al., 2019; Thabtah et al., 2018), or implementing ML-based screening apps (Shahamiri & Thabtah, 2020; Wingfield et al., 2020). For instance, Goel et al. (2020) proposed Modified Grasshopper Optimization Algorithm (MGOA) for improved performance over common ML algorithms. The proposed MGOA (GOA with Random Forest classifier) outperformed other basic models and predicted ASD with approximate accuracy, specificity, and sensitivity of 100%. Similarly, Suresh Kumar and Renugadevi (2019) proposed Differential Evaluation (DE) Algorithm to find the optimal solution of SVM parameters. The proposed DE tuned SVM achieved better performance over SVM, ANN and DE optimized ANN in classifying ASD. As stated earlier, apart from trial-error, studies employed either feature selection or transformation techniques for dimensionality reduction. For instance, Thabtah et al. (2018) demonstrated the superiority of variable analysis (Va) over information gain (IG), Correlation, and chi-square (CHI) in reducing AQ-10 items. Variable analysis (Va) derived fewer features while maintaining competitive predictive accuracy, sensitivity, and specificity rates. A replicated study by Pratama et al. (2019) produced a higher sensitivity of 87.89% in Adults AQ with RF and an increased specificity level of 86.33% in Adolescents AQ with SVM. Despite the good performance of the above-mentioned techniques in automating feature selection processes across various applications (Alhaj et al., 2016; Roobaert et al., 2006), none of the previous studies justified the conformity of the feature selection methods with the conceptual basis upon which professionals built and utilize ASD diagnostic instruments.

Furthermore, unlike other medical diagnoses, the absence of definitive measures and medical tests for diagnosing ASD makes it difficult to numerical quantify the disorder based on few parameters. Notably, accurate assessment of ASD relied on precise application of the commonly used behavioral scales built based on knowledge and expertise of the professionals. Thus, application of the human knowledge is imperative to reliable ASD diagnosis. Based on that, there is need for quantifying the trade-offs of dimensionality reduction (ensuring fewer items for quick assessment) and validity (preservation of the human knowledge for correct diagnosis). Specifically, a machine-learning model built based on fewer behavioral features that does not sufficiently capture the human knowledge of the assessment instrument, will not be valid for clinical use. Thus, there is need for applying dimensionality reduction techniques that professionals could track its ability in preserving the validity of the assessment instruments. Figure **2.6**2.6 described the experimental stages toward clinical implementation of the ML models for ASD assessment.

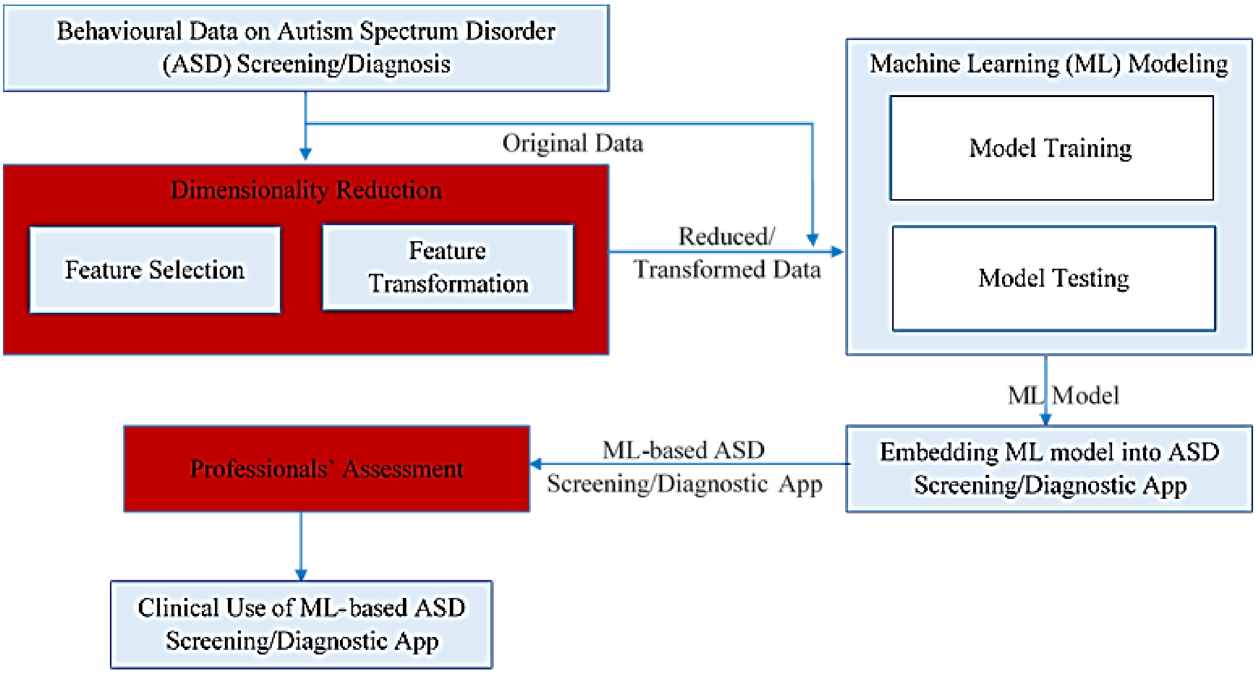


Figure 2.6: Commonly employed experimental stages

Nonetheless, various feature transformation techniques were equally utilized in the dimensionality reduction processes. For instance, Akter et al. (2019) utilized three feature transformation techniques; Log, Z-score, and Sine functions, and evaluated the performance of nine different ML models on the transformed datasets. Log, Z-score, and Sine functions normalize data by converting excessively skewed entities into a normal distribution, converting features into -1 to 1 value range, and transforming instances to the sine 0-2π value intervals, respectively. In addition, they recorded varying superior performances of the ML models and the feature transformation approaches across the datasets. The feature transformations resulting in the best classifications were Z-score and Sine function on children, adolescents, and toddlers' datasets, respectively.

However, despite the reported improved performances of the ML models on the transformed datasets and the theoretical understanding of the capabilities of the transformation functions, studies have demonstrated how these transformations compromise the relevance of the original data to the transformed data (Curtis et al., 2016; Feng et al., 2014; Lapteacru, 2016; Wiesen, 2006). Researchers ought to be mindful of the limitations in using these transformations in terms of the relevance of the original to the transformed data during results interpretation. For instance, Feng et al. (2014) demonstrated such irrelevancies between the statistical findings of standard tests performed on original and log-transformed data. Similarly, several studies have highlighted some of the pitfalls and inconsistencies in the application of Z-scores and its concepts that overlooked the meaning of the original data, its standard deviations, and confusing applications (Curtis et al., 2016; Lapteacru, 2016; Wiesen, 2006).

Recent studies further demonstrated how ML-enable ASD screening and diagnostic models could be developed, evaluated, and implemented. Recently, Baadel et al. (2020) proposed Clustering-based Autistic Trait Classification (CATC) which identifies ASD-based traits' similarity, unlike the commonly used scoring functions. CATC showed significant improvement in the ASD classification based on clustered inputs. Comparative evaluation of various classification algorithms showed better improvement with the Random Forest classifier. On the implementation of mobile apps for ASD screening, Wingfield et al. (2020) and Shahamiri & Thabtah (2020) embedded RF and CNN-based scoring models, respectively, while Thabtah (2019) employed ML to validate Autism Spectrum Disorder Test (ASDTest); a mobile screening app embedded with non-ML functions. In all the foregoing studies, the commonly used evaluation metrics are classification accuracy, sensitivity, and specificity. Specificity is the ratio of non-ASD cases that are correctly classified (i.e. true negatives rate) and sensitivity is the ratio of true ASD cases that are correctly classified (i.e. true positives rate) while classification accuracy is derived from sensitivity and specificity; as the measure of precisely classified cases from the total number of the cases.

2.6 ML-based Studies and the Dimensionality Reduction Approaches

2.6.1 Trial-error approaches

Goel et al. (2020) appraised how the time-consuming clinical diagnostic process is aggravating the severity of ASD among patients and highlighted how early and precise identification could remediate the disorder. Accordingly, the authors proposed a nature-inspired optimization algorithm entitled Modified Grasshopper Optimization Algorithm (MGOA); which is a combination of GOA and random forest classifier to detect the symptoms of ASD across various age groups. The algorithm was employed in modeling three distinct ASD screening datasets sampled using Autism Quotient (AQ); AQ-children, AQ-adolescents, and AQ-adults. The comparative study results indicated that the proposed MGOA outperformed other common machine learning algorithms by achieving approximate specificity, accuracy, and sensitivity of 100% in detecting ASD across all age groups.

Shahamiri and Thabtah (2020) cited the significant time-ineffectiveness, subjectivity, and fiscal costs of the conventional processes used in the early identification of ASD. Based on the identified challenges, the authors suggested an intelligent ASD screening method for accurate pre-diagnostic classifications, accessibility, and improved efficiency. Specifically, the study proposed the implementation of an online application named Autism AI, which prompts a reduced version of the

Quantitative Checklist for Autism Toddlers (Q-CHAT) and AQ-10 questionnaires aimed at ASD diagnosticians and busy medical clinics. The machine-learning model embedded in the application was trained with Convolutional Neural Network (CNN) on historical ASD screening datasets available at the UCI ML repository. In the experimental runs, the comparative evaluation of CNN against C4.5, Bayes Net, and RIDOR models revealed the superior performance of the CNN-based classification of ASD.

Thabtah and Peebles (2020) identified the resulting consequences in the current lengthy and cost-ineffective ASD diagnostic approaches that solely relied on the subjective judgment of the limited number of licensed specialists. The authors proposed novel machine learning techniques entitled Rules-Machine Learning (RML) in an attempt to address the identified challenges bedeviling the conventional diagnostic methods and the misalignment that exists between the conceptual basis of ASD diagnosis and the data-driven techniques employed in the existing machine learning studies. Experimental evaluation of multiple machine learning models such as decision trees, Bagging, and Boosting against the RML based on the ASDTest datasets of children, adolescents, and adults revealed the superior performance of RML. Particularly, RML achieved higher predictive specificity, harmonic mean, sensitivity, and accuracy and it provides knowledge bases (rules) that can be utilized by experts in tracking the reason behind case classification.

Wall, Dally, et al. (2012) studied the responses on the 93 Autism Diagnostic Interview–Revised (ADI-R) parameters archived in the Autism Genetic Resource Exchange (AGRE) retrospective dataset to streamline the responses and implement machine learning models from the reduced data. From the 891 ASD and 75 non-ASD responses contained in the dataset, the study analysis extracted 7 out of the 93 items of ADI-R to be sufficient in the ML classification of ASD with substantial predictive accuracy. The study further tested the performance of the developed 7-items classification models on two independent datasets containing 1654 and 322 ASD cases sourced from the Simons Foundation and Boston Autism Consortium, respectively. In both cases, Alternative Decision Tree (ADTree) achieved higher predictive performance over other variants of decision tree classifiers.

Duda et al. (2015) evaluated the correlation between the ASD scores of the ADOS algorithm and machine learning model termed as an observation-based classifier (OBC). The best performing classifier with more than 97% statistical accuracy was derived using fewer than 30% of the standard ADOS-G items. The accuracy of the OBC was tested on an independent sample of 2333 children with ASD and 283 non-ASD cases gathered from five data repositories; namely AC, AGRE, SSC, NDAR, and SVIP. The comparative analysis of the results provided by the OBC and the original ADOS algorithms revealed the existence of statistically significant indirect correlation with above 75% magnitude. The predictive accuracy of the classifier and its significant correlation with the outcome of the gold-standard ADOS-G scores demonstrated the capability of OBC in measuring the severity of ASD phenotype.

Wall, Kosmicki, et al. (2012) identified the dichotomy in the four ADOS modules, each of which takes at least 30 minutes to administer, that are tailored across a various group of participants based on their language and developmental levels. The authors experimented on multiple machine-learning algorithms on the reduced set of ADOS Module 1 responses retrieved from AGRE and AC repositories; which contained 612 ASD cases and 15 controls. The study results analysis indicated that 8 of the 29 behavioral features contained in Module 1 of the ADOS were sufficient for the best performing ADTree classifier in detecting ASD cases with a predictive accuracy of 100%. Furthermore, the eight-item classifier was validated against the complete set features in 110 and 336 ASD cases retrieved independently from AC and Simons Foundation, respectively. The validation result yielded a sensitivity of 100% and specificity of 94%.

Thabtah (2019) implemented an ASD screening and data collection app named ASDTest with the key aim of enhancing accessibility and alleviate the healthcare costs and delays in the current diagnostic practices. The app was used in collecting the retrospective AQ-10 and Q-CHAT-10 data utilized in numerous ML studies. The author also utilized 1400 instances from the data in training and evaluating the predictive performance of two different machine-learning models of Logistic Regression and Naïve Bayes algorithms. Using the trail-error technique the author was able to implement the ML models with reduced AQ-10 and Q-CHAT-10 parameters. The predictive performance of the classifiers in terms of the classification accuracy, true negative, and true positive rates was found to approach 100%.

Duda et al. (2017) conducted a validation study on the ML classification of the ASD comorbid disorders especially ADHD by incorporating the SRS responses used in Duda et al. (2016) with the crowdsourced dataset to improve the model’s capability on ‘real-world’ data. The crowdsourced data involved responses to 15 most influential SRS items on 248 children with ASD and 174 ADHD. The combined dataset with 3417 cases was subjected to subsampling and repeated cross-validations in the modeling of SVM, LR, and LDA algorithms. LDA as the best performing algorithm achieved an AUC of 89% with the 15 SRS items.

2.6.2 Feature selection approaches

Küpper et al. (2020) highlighted the various complications concerning the current ASD diagnosis and particularly in older individuals. The authors suggested a machine learning approach for the detection of ASD among adolescents and adults using fewer items from the ADOS Module 4. The data modeling approach employed SVM in examining possible improvement in ASD diagnostic accuracy and improving diagnostic speed by identifying the most relevant behavioral features from the diagnostic instrument using recursive feature selection. The study utilized a retrospective dataset sourced from ASD outpatient clinics in Germany comprising 673 adolescents with high-functioning ASD, 385 adults with ASD, and 288 controls. The study findings revealed that for the adolescents and adult groups as well as the combined groups, 5 behavioral features were identified as sufficient in classifying ASD with good predictive accuracy with no significant difference from the performance of the complete ADOS items and its conventional screening algorithm.

Levy et al. (2017) highlighted how shorter, mobile-based ASD diagnostic instruments could alleviate the bottlenecks of delay and inaccessibility to accurate diagnostic practices. The authors claimed the clinical sufficiency of the subsets of ADOS behavioral features in the previous studies based on the reported classification accuracies. The study expanded the claim toward achieving generalizable ML models for the clinical population. Accordingly, the study involved ADOS module 2 and 3 responses sampled from 1319 ASD cases, 70 controls and 2870 ASD cases, 273 controls, respectively, gathered from AC, AGRE, SSC, and SVIP repositories. Furthermore, stable subsets of the ADOS predictive features were extracted by utilizing sparsity/parsimony enforcing regularization techniques. The study augmented demographic features of age and gender unto the ADOS shorter items. The study findings of the 17 unique classification models explored yielded 5 to 10 relevant features. The best performing models achieved AUC of 0.95 and 0.93 with the reduced features from ADOS's Module 3 and Module 2, respectively. The authors claimed the clinical stability of the reduced items as well as the potential generalizability of the models based on their sparsity, and interpretability with the augmented parameters.

Kosmicki et al. (2015) emphasized the value of implementing quick and accurate ASD diagnostic measures by looking at the persistent prevalent raise. The authors ran eight different machine-learning algorithms in the data modeling of ADOS modules 2 and 3 behavioral data containing retrospective data of 4540 cases. During the experimental runs, stepwise backward feature selection captured 9 and 12 relevant items out of the 28 behavioral features in module 2, and module 3, respectively. The reduced features were utilized in evaluating the predictive performance of seven distinct machine-learning models, namely; ADTree, SVM, Logistic Model Tree, LR, NB, NBTree, RF. The best performing models, SVM and LR, achieved a classification accuracy of almost 100%. In essence, at least a 55% reduction in the behavioral features was recorded while maintaining the predicted performance of the models. The authors recommended utilizing similar models in the real-life implementation of objective, self, and/or parent-administered mobile apps for preliminary evaluation of ASD risk factors.

Thabtah et al. (2018) highlighted the demand for identifying fewer parameters that could sufficiently classify ASD in the same way as the complete set of features in the commonly used screening instruments. Thus, the authors proposed the application of intelligent feature selection techniques called variable analysis (Va) on the AQ (AQ-10 Adult, AQ-10 Adolescent, and AQ-10 Child) and Q-CHAT-10 datasets. The authors demonstrated the superiority of Va over other commonly used feature selection methods such as information gain, and correlation based on the performance of the resulting ML models on the streamlined datasets. The resulting classifiers based on Repeated Incremental Pruning to Produce Error Reduction (RIPPER), and C4.5 verified the efficacy of Va in terms of the achieved performances concerning classification accuracy, positive predictive values, sensitivity, and specificity.

Thabtah et al. (2019) presented a novel framework for time-efficient ASD screening for adults and adolescents based on logistic regression analysis. Vital and most influential features from the screening instrument were extracted using various feature selection approaches. Comparative analysis on the performance of the feature selection methods on the two datasets of adults and adolescents showed the efficacy of dimensionality reduction approaches in improving the classification efficiency of the logistic regression model. The comparative analysis between the dimensionality reduction techniques of information gain (IG) and Chi-square testing (CHI) yielded fewer influential features that are capable of assessing the symptoms of ASD sufficiently and maintaining the classification performance of the classifiers in terms of accuracy, specificity, and sensitivity among others.

Suresh and Renugadevi (2019) proposed an algorithm optimization technique called Differential Evolutionary (DE) to improve the predictive performance of the commonly used machine learning algorithms. The authors evaluated the performance of four distinct machine-learning models, namely SVM, ANN, DE optimized SVM, and DE optimized ANN on the AQ-child, AQ-adolescent, and AQ-adult responses retrieved from the UCL ML repository. The responses to the screening instruments were reduced using Sequential Feature Selection and the resulting DE optimized SVM models outperformed ANN and DE optimized ANN in classifying the ASD symptoms quantified using the data collection instruments.

Pratama et al. (2019) utilized the AQ-10 (children, adolescents, and adults) datasets in ML modeling with support vector machine (SVM), artificial neural network (ANN), and random forest (RF) to reduce the misclassifications of the commonly used AQ scoring algorithms. The study employed Va in streamlining the AQ-10 items and 10-fold cross-validation was equally applied in the experimental runs. The study findings revealed the performance of RF and SVM with higher adult AQ sensitivity of 87.89% and adolescents AQ specificity level of 86.33%, respectively.

Usta et al. (2019) evaluated the performance of multiple machine learning algorithms in the modeling of behavioral data of 433 children with ASD containing numerous demographic predictors as an addendum to the items of Autism Behavior Checklist, Aberrant Behavior Checklist, and Clinical Global Impression from the better predictive outcome. The historical ASD diagnostic data was sourced from Ondokuz Mayis University Samsun. The comparative assessment on the performance of the modeling algorithms, namely Naive Bayes, Generalized Linear Model, Logistic Regression, and Decision Tree showed a significant effect of the demographic factors on the models' performance. Demographic factors of paternal and maternal age, birth weight, and severity level significantly influence the better prediction of the disorder. Thus, the classification approach revealed several other influential factors that affect the predictive performance of the models.

Wingfield et al. (2020) proposed a culturally sensitive picture-based screening app called Pictorial Autism Assessment Schedule (vPASS) to alleviate cultural variations in the interpretation of the behavioral symptoms of ASD as noted in the sensitivity of MCHAT among Sri Lankan children. The authors demonstrated the possibility of removing feature redundancy and overcoming the cultural variation in the interpretation of ASD symptoms. vPASS was based on the items of the PASS checklist reduced using correlation-based feature selection (CFS) and minimum redundancy maximum relevance (mRMR). PAAS checklist was derived by modifying the cultural considerations in DSM-V and M-CHAT. The vPASS app embeds a machine-learning model based on the Random Forest classifier, which achieved superior performance in the experimental run over NB, Adaboost, Multilayer Perceptron, J48, PART, and SMO with a receiver operating characteristic of 98%.

Duda et al. (2016) proposed an ML-based ASD screening for quick and accurate assessment of ASD and ADHD symptoms to mitigate the negative consequences of the subjective, cumbersome, and time-intensive conventional approaches. The study employed forward feature selection and undersampling on the retrospective datasets on Social Responsiveness Scale (SRS) responses, archived in AC, AGRE, and SSC repositories. The study analyzed 2775 ASD and 150 ADHD cases that were modeled using six different machine-learning algorithms. 10-fold cross-validation was employed in each experimental run on five of the 65 SRS items, which were found to be sufficient in the classification task. The experimental run was tailored on ADTree, RF, SVM, LR, Categorical lasso, and LDA, and all the models were able to classify ASD from ADHD by utilizing 5 of the 65 items of SRS with high average accuracy and AUC of 96.5%.

2.6.3 Feature transformation approaches

Bone et al. (2016) targeted best classification of ASD cases and non-ASD cases by combining the items of ADI-R and SRS scores from Balanced Independent Dataset. Accordingly, a robust SVM classifier was modeled using 1264 ASD cases and 462 non-ASD cases with parameter tuning across multiple levels of cross-validation. The experimental findings revealed how the SVM model utilized 5 of the fused ADI-R and SRS items in sufficiently classifying ASD with approximate sensitivity of 90%, and specificity above 50%. The authors highlighted that, despite the good performance of the machine-learning model, the study findings pinpointed the limitations in the current parent-report instruments.

Puerto et al. (2019) proposed a novel machine learning approach to the behavioral assessment of ASD entitled Multilayer Fuzzy Cognitive Map (MFCM-ASD). The authors utilized ADOS and ADI-R responses from the retrospective dataset of the Association of Parents and Friends for the Support and Defense of the rights of people with Autism. MFCM-ASD model was based on fuzzified inputs and it achieved better predictive performance over other comparative models of SVM, Random forest, and NB. The authors stated that the superior performance of MFCM-ASD characterized by its robustness makes it an effective ASD diagnostic technique and medical decision support system for children with different cognitive characteristics.

Akter et al. (2019) gathered early-detected ASD risk factors relating to different age groups quantified using AQ-10 Adult, AQ-10 Adolescent, and AQ-10 Child as well as Q-CHAT-10 screening questions. The authors transformed the datasets using Z-score, logarithmic, and sine functions. The transformed features were passed as inputs to nine distinct classification algorithms including Adaboost, FDA, C5.0, LDA, MDA, PDA, SVM, and CART. The comparative results showed varying superior performance of the ML models and FT approaches across the datasets. The ML models resulting in the best classifications were SVM (toddlers dataset), Adaboost (children dataset), Glmboost (adolescents dataset), Adaboost (adults dataset). The FTs resulting in the best classifications were sine function (toddlers dataset), and Z-score (children and adolescent datasets).

Baadel et al. (2020) conducted an experimental study to improve diagnostic speed through dimensionality reduction and redundancy elimination using a novel semi-supervised machine-learning framework entitled Clustering-based Autistic Trait Classification (CATC). The proposed clustering approach was validated using classification techniques in which the influential input parameters were identified as clusters as opposed to the commonly used scoring functions. The AQ-10 Adult, AQ-10 Adolescent, and AQ-10 Child were equally utilized in the comparative evaluation of the resulting models. The study findings revealed the efficacy of the clustering technique in producing classifiers with higher predictive accuracy, true negative, and true positive rates over other machine learning algorithms including Random Trees, Rule Induction, Random Forest, and Artificial Neural Network (ANN). The detailed literature is summarized with the help of Table 2.2.

2.7 Advances toward Quick and Accurate Assessment of ASD Symptoms

The search for cost-effective ASD assessment coupled with the global rise in ASD cases attracted the implementation of quick and accurate assessment measures based on data intelligence techniques including machine-learning algorithms. Despite the various attempts in ML-based ASD assessment using functional magnetic resonance imaging (MRI), eye tracking, and genetic data among others, the promising results based on behavioral data call for further research. For instance, Plitt, Barnes, and Martin (2015) found that ASD classification via behavioral measures consistently surpassed rs-fMRI classifiers. Accordingly, in line with the common research aim of the behavioral studies, various dimensionality reduction techniques were employed to improve the diagnostic speed of the resulting ML models. However, unlike the reduced dimensions, there is enough evidence on the good reliability, high internal consistency, and convergent validity between the common assessment instruments within large samples (Becker et al., 2012; Chan et al., 2017; Chojnicka & Pisula, 2017; Falkmer et al., 2013; Medda et al., 2019).

Furthermore, studies have ascertained the robustness of the common assessment instruments in the quantitative measurement of the various dimensions of communication, interpersonal behavior, and stereotypic/repetitive behavior associated with ASD. Therefore, it will be difficult to sufficiently measure the key dimensions of the instruments using the fewer items generated by the common dimensionality reduction techniques. For instance, while professionals interpret SRS scores based on the sum of its 65 items, Bone et al. (2016), Duda et al. (2016), and Duda et al. (2017) implemented SRS-enabled machine-learning models with at most 5, 5, and 15 items, respectively. Specifically, Duda et al. (2016) and Duda et al. (2017) focused on classifying ASD from ADHD using the SRS data from AC, AGRE, SSC. Duda et al. (2016) implemented ADTree, RF, SVM, LR, Categorical lasso, and LDA models and achieved the highest area under the curve (AUC) of 0.965 in classifying ASD from ADHD by utilizing five of the 65 items of SRS identified using forward feature selection. Duda et al. (2017) validated the findings of Duda et al. (2016) with crowdsourced data to improve the model's capability on 'real-world' data and the findings revealed that LDA outperformed LR and SVM by achieving an AUC of 0.89 with 15 items. Despite the high metrics reported by the studies, based on the standard clinical procedures for ASD diagnosis, the ML models are neither clinically sufficient nor readily implementable for real-life use.

Similarly, Wall, Dally, et al. (2012) compared the performance of 15 different ML algorithms on AGRE, SSC, and AC datasets and found ADTree to outperformed other models by utilizing 7 of the 93 items contained in the ADI-R in classifying ASD with 99.9% accuracy. In a similar study by Wall, Kosmicki, et al. (2012), ADTree outperformed 17 comparative models by achieving 100% accuracy with 8 of the 29 items in Module 1 of ADOS. Moreover, Duda et al. (2015) demonstrated the superior performance of ADTree in achieving 97% classification accuracy with a 72% reduction in ADOS-G items. Nonetheless, Levy et al. (2017) and Kosmicki et al. (2015) reduced the items of ADOS using sparsity/parsimony enforcing regularization and stepwise backward feature selection techniques, respectively and reported the superior performance of LR and SVM over other ML algorithms. Specifically, in the study by Levy et al. (2017), with at most 10 features from ADOS's Module 3 and Module 2, AUC of 0.95 and 0.93 was achieved, respectively. While Kosmicki et al. (2015) recorded an accuracy of 98.27% and 97.66% with 9 of the 28 items from module 2, and 12 of the 28 items from module 3, respectively. Recently, Küpper et al. (2020) utilized ADOS data from a clinical sample of adolescents and adults with ASD and reported good performance of SVM on fewer items reduced using the recursive feature selection technique. The foregoing studies have demonstrated how ML-enabled ASD screening and diagnostic models could be developed and evaluated. However, numerous challenges associated with the behavioral assessment instruments, data repositories, and applied data intelligence algorithms need to be understood and addressed.

Although ML-based approaches are data-centric and are expected to improve objectivity and automation (Achenie et al., 2019), with the global rise in ASD cases, the capacity to quickly and accurately assess ASD requires a careful understanding of the conceptual basis of the assessment instruments, as well as their relevance to the logical concepts of the ML algorithms. Nonetheless, discrepancies within the data repositories such as data imbalance limit the clinical relevance of the high evaluation metrics reported in the studies (Abdelhamid et al., 2020; Alahmari, 2020). For instance, Torres et al. (2020) studied the statistical properties of ADOS scores from 1324 records and identified various factors that could undermine the scientific viability of the scores. Particularly, the empirical distributions in the generated scores break the theoretical conditions of normality and homogeneous variance, which are critical for independence between bias and sensitivity. Thus, Torres et al. (2020) suggested readjusting the scientific use of ADOS due to the variation in the distribution of the scores, lack of appropriate metrics for characterizing changes, and the impact of both on sensitivity-bias codependencies and longitudinal tracking of ASD.

In essence, the applied data intelligence algorithms, and the resulting models, missed the human knowledge upon which the assessment instruments were built and applied by the professionals (Thabtah, 2018). Additionally, most of the studies overlooked the inherent limitations associated with the dimensionality reduction techniques, and the assessment instruments (Guthrie et al., 2019; Øien et al., 2018; Surén et al., 2019). Thus, the need for ascertaining the clinical relevance of the data-centric approaches and readjusting the scientific use of the assessment instruments remains. Obviously, in the future, it can be said that the trend in the application of ML in the behavioral assessment of ASD will go on. On the other hand, the pressing demands for cost-effective assessment of ASD remain. Thus, future studies need to revisit the relevance of the data collection instruments to ML algorithms.

Conclusively, machine learning has been broadly applied in the behavioral assessment of ASD based on a variety of data types as input to data-intelligence algorithms. Commonly utilized inputs include the items of screening tools such as ADI-R and ADOS-G. Popular ML algorithms used are SVMs, variants of the decision trees, random forests, and neural networks. However, the multitudes of challenges in accurate ASD assessments are yet to be addressed by the suggested machine learning approaches. Specifically, the high metrics achieved with the data-intelligence techniques have not guaranteed the clinical relevance of the ML models. Additionally, the commonly used evaluation measures of classification accuracy, specificity, and sensitivity among others cannot sufficiently reflect the human knowledge applied by professionals in assessing behavioral symptoms of ASD. Consequently, understanding the clinical basis of the assessment tools and the logical concepts of the data-intelligence techniques will lead to promising studies on the real-life implementation of cost-effective ASD assessment systems.

The novelty in the present study is that while previous literature reviews focused on the performance of various data intelligent techniques on different data sets, this study systematically reviewed the literature and provide a definitive explanation on the relevance of the reported findings toward the real-life implementation of the ML-based assessment systems. The findings of this systematic literature review serve as my guide in identifying the research gap and novelty of my novel approach and will hopefully guide researchers, caregivers, and relevant stakeholders on the advances in ASD assessment with ML. Nonetheless, a few of the limitations associated with the present systematic literature review include overlooking other non-English documents. Thus, possible excellent studies reported in other languages might have been missed. Secondly, the search filters spanned ten years and were limited to the four scientific databases mentioned. Furthermore, the records retrieved relied on the few search terms utilized in the search query. Therefore, relaxing the search filters across additional databases could yield additional relevant studies. Lastly, the present study considered only full-text online journal articles. Consequently, the findings are limited to the studies included. Future research agenda could relax the search criteria to incorporate other scholastic databases for further comparative results. Besides, future studies could relax the search filters to include books, conference papers, and so on.

2.8 Research Novelty

None of the previous studies aimed at preserving the clinical validity of the assessment instruments while benefitting from the precision of the machine learning algorithms. Specifically, unlike the common feature selection approaches, in the present study, none of the items is discarded, instead, each item serves as an integral part in generating the input dimensions. Noteworthy, to build on or replicate the reviewed studies, the present study explored novel data-intelligence techniques that will achieve not only excellent evaluation metrics but also adhere to the conceptual basis upon which professionals diagnose ASD.

Table 2.2: Information Extracted from the Articles

Article/ Citations	Aim	Tool	Data Source	FS/FT	FS/FT Method	Modeling Algorithms	Key Findings
Goel et al. (2020) C = 10	Proposed Optimization Algorithm for improved performance over common ML	AQ-10 (child, adolescent, adult)	ASDTest	-	-	GOA, BACO, LR, NB, KNN, RF-CART + ID3, *MGOA	The proposed MGOA (GOA with Random Forest classifier) predicted ASD cases with approximate accuracy, specificity, and sensitivity of 100%.
Shahamiri & Thabtah (2020) C = 0	Implementation and evaluation of CNN-based ASD scoring system	Q-CHAT-10, AQ-10	ASDTest	-	-	C4.5, Bayes Net, RIDOR, *CNN	The performance evaluation showed the superior performance of CNN over other algorithms; indicating the robustness of the implemented system.
Thabtah & Peebles (2020) C = 28	Demonstrate the superiority of Rules-based ML over other models	Q-CHAT-10, AQ-10 (child, Adolescent, adult)	ASDTest	-	-	RIPPER, RIDOR, Nnge, Bagging, CART, C4.5, and PRISM, *RML	Empirically evaluated rule induction, Bagging, Boosting, and decision trees algorithms on different ASD datasets. The superiority of the RML model was reported in not only classifying ASD but also offer rules that can be utilized in understanding the reasons behind the classification.

Wall, Dally, et al. (2012) C = 106	Streamlining ADR-I and evaluate ML performance	ADI-R	AGRE, SSC, AC	FS	Trial-error	*ADTree, BFTree, ConjunctiveRule, DecisionStump, FilteredClassifier, J48, J48graft, JRip, LADTree, NNge, OneR, OrdinalClassClassifier, PART, Ridor, and SimpleCart	The best model utilized 7 of the 93 items contained in the ADI-R in classifying ASD with 99.9% accuracy.
Duda et al. (2015) C = 50	Streamlining ADOS and demonstrate the superior performance of ADTree over common hand-crafted methods	ADOS	AC, AGRE, SSC, NDAR, SVIP	FS	Trial-error	ADTree	72% reduction in the items from ADOS-G with >97% accuracy.
Küpper et al. (2020) C = 2	Streamlining ADOS and demonstrate the performance of SVM	ADOS	ASD outpatient clinics in Germany	FS	Recursive Feature Selection	SVM	SVM achieved good sensitivity and specificity with fewer ADOS items pointing to 5 behavioral features.

Wall, Kosmicki, et al. (2012) C = 160	Streamlining ADOS and evaluate ML performance	ADOS	AC, AGRE, SSC	FS	Trial-error	*ADTree, BFTree, Decision Stump, Functional Tree, J48, J48graft, Jrip, LADTree, LMT, Nnge, OneR, PART, Random Tree, REPTree, Ridor, Simple Cart	The ADTree model utilized 8 of the 29 items in Module 1 of the ADOS and classified ASD with 100% accuracy.
Levy et al. (2017) C = 21	Streamlining ADOS and evaluate ML performance	ADOS	AC, AGRE, SSC, SVIP	FS	Sparsity/parsimony enforcing regularization techniques	LR, Lasso, Ridge, Elastic net, Relaxed Lasso, Nearest shrunken centroids, LDA, *LR, *SVM, ADTree, RF, Gradient boosting, AdaBoost	With at most 10 features from ADOS's Module 3 and Module 2 AUC of 0.95 and 0.93 was achieved, respectively.
Kosmicki et al. (2015) C = 84	Streamlining ADOS and evaluate ML performance	ADOS	AC, AGRE, SSC, NDAR, SVIP	FS	Stepwise Backward Feature Selection	ADTree, *SVM, Logistic Model Tree, *LR, NB, NBTree, RF	The best performing models have utilized 9 of the 28 items from module 2, and 12 of the 28 items from module 3 in classifying ASD with 98.27% and 97.66% accuracy, respectively.

Thabtah (2019) C = 31	Propose ASDTest; AQ-based mobile screening app, streamline AQ-10 items and evaluate the performance of 2 ML models	AQ-10 (child, adolescent, adult)	ASDTest	FS	Trial-error	NB, *LR	Feature and predictive analyses demonstrate small groups of autistic traits improving the efficiency and accuracy of screening processes.
Thabtah et al. (2018) C = 47	Demonstrate the superiority of Va over other FS methods based on the performance of ML models on the streamlined datasets	Q-CHAT-10, and AQ-10 (child, adolescent, adult)	ASDTest	FS	Va, IG, Correlation, CFS, and CHI	Repeated Incremental Pruning to Produce Error Reduction (RIPPER), C4.5 (Decision Tree)	Va derived fewer features from adults, adolescents, and children datasets with optimal model performance. Demonstrate the efficacy of Va over IG, Correlation, CFS, and CHI in reducing AQ-10 items
Thabtah et al. (2019) C = 13	Streamlining AQ-10 and demonstrate the superior performance of LR over common hand-crafted methods	AQ-10 (adolescent, adult)	ASDTest	FS	IG, CHI	LR	LR showed acceptable performance in terms of sensitivity, specificity, and accuracy among others.

Suresh Kumar & Renugadevi (2019) C = 0	Algorithm Optimization (improvement in accuracy compared to common ML)	AQ-10 (child, adolescent, adult)	ASDTest	FS	SFS	SVM, ANN, *DE SVM, DE ANN	DE optimized SVM outperformed ANN and DE optimized ANN in classifying ASD. DE is effective.
Pratama et al. (2019) C = 0	Input Optimization using Va	AQ-10 (child, adolescent, adult)	ASDTest	FS	Va	SVM, *RF, ANN	RF succeeded in producing higher adult AQ sensitivity (87.89%) and a rise in the specificity level of AQ-Adolescents was better produced using SVM (86.33%).
Usta et al. (2019) C = 9	ML Performance Evaluation	Autism Behavior Checklist, Aberrant Behavior Checklist, Clinical Global Impression	Ondokuz Mayis University Samsun	FS	Trial-error	NB, LR, *ADTree	The ML modeling revealed the significant influence of other demographic parameters in ASD classification.

Wingfield et al. (2020) C = 3	Propose PASS; a culturally sensitive app embedded with ML model	PASS	VPASS app	FS	CFS, mRMR	*RF, NB, Adaboost, Multilayer Perceptron, J48, PART, SMO	PASS app overcomes the cultural variation in interpreting ASD symptoms and the study demonstrated the possibility of removing feature redundancy.
Duda et al. (2016) C = 89	ML Performance Evaluation in classifying ASD from ADHD	SRS	AC, AGRE, SSC	FS	Forward Feature Selection	ADTree, RF, SVM, LR, Categorical lasso, LDA	All the models could classify ASD from ADHD by utilizing 5 of the 65 items of SRS with high average accuracy (AUC = 0.965).
Duda et al. (2017) C = 25	Improve models' reliability using expanded datasets for classifying ASD from ADHD	SRS	AC, AGRE, SSC, & crowdsourced data	FS	-	SVM, LR, *LDA	LDA model achieved an AUC of 0.89 with 15 items.
Bone et al. (2016) C = 77	Demonstrate the improved accuracy of SVM over common hand-crafted rules	ADI-R, SRS	Balanced Independent Dataset	FT	Tuned parameters across multiple levels of cross-validation	SVM	The SVM model utilized 5 of the fused ADI-R & SRS items and classified ASD sufficiently with below (above) 89.2% (86.7%) sensitivity and 59.0% (53.4%) specificity.

Puerto et al. (2019) C = 17	Propose MFCM-ASD and evaluate its performance against other ML models	ADOS, ADI-R	APADA	FT	Inputs fuzzification	*MFCM-ASD, SVM, Random forest, NB	The superior performance of MFCM characterized by its robustness makes it an effective ASD diagnostic technique.
Akter et al. (2019) C = 6	Compare FT methods and evaluate the performance of ML models on the transformed datasets	Q-CHAT-10, and AQ-10 (child, adolescent, adult)	ASDTest	FT	Log, Z-score, and Sine FT	Adaboost, FDA, C5.0, LDA, MDA, PDA, SVM, and CART	Varying superior performances of the ML models and FT approaches were achieved across the datasets.
Baadel et al. (2020) C = 2	Input Optimization using a clustering approach	AQ-10 (child, adolescent, adult)	ASDTest	FT	CATC	OMCOKE, RIPPER, PART, *RF, RT, ANN	CATC showed significant improvement in screening ASD based on traits' similarity as opposed to scoring functions. The improvement was more pronounced with RF classifier.

ASD: Autism Spectrum Disorder. FS: Feature Selection. FT: Feature Transformation. ML: Machine Learning. ANN: Artificial Neural Network. SVM: Support Vector Machine. CNN: Convolutional Neural Network. RF: Random Forest. LR: Logistic Regression. ADTree: Alternative Decision Tree. LDA: Linear Discriminant Analysis. MGOA: Modified Grasshopper Optimization Algorithm. BACO: Binary Ant Colony Optimization. NB: Naïve Bayes. KNN: K-Nearest Neighbor. RIPPER: Repeated Incremental Pruning to Produce Error Reduction. ADOS: Autism Diagnostic Observation Schedule. ADI-R: Autism Diagnostic Interview-Revised. Q-CHAT: Quantitative Checklist for Autism Toddlers. AQ: Autism Quotient. SRS: Social Responsiveness Scale. PASS: Pictorial Autism Assessment Schedule. AC: Boston Autism Consortium. AGRE: Autism Genetic Resource Exchange. SSC: Simons Simplex Collection. NDAR: National Database for Autism Research. SVIP: Simons Variation In Individuals Project. APADA: Association of Parents and Friends for the Support and Defense of the rights of people with Autism. MFCM: Multilayer Fuzzy Cognitive Maps. CATC: Clustering-based Autistic Trait Classification. * Best performing models.

CHAPTER 3

THEORETICAL FRAMEWORK

This chapter provides explanations on the theoretical basis of the key elements that aid the promising approach of diagnosing ASD using machine-learning models. This research endeavor comprises expertise in computer science, data science and neurodevelopmental psychiatry. Specifically, the present chapter will aid understanding of the key concepts upon which this study was carried out including the conceptual understanding of the ASD screening and diagnostic instruments, the machine-learning algorithms as well as the evaluation metrics.

3.1 ICT-based ASD Screening

Reliable diagnosis qualifies patients to access public services available for ASD. However, current diagnostic practices are prone to numerous challenges that warrant extensive research. Information and communication technologies (ICTs) play a significant role in ASD research and intervention including different means of curving the inherent drawbacks associated with ASD screening and diagnosis. Figure **3.1**3.1 highlighted the various roles of ICTs in ASD research and intervention. However, the proliferation of simple platforms for end-user app development coupled with the pressing need for quick ASD assessment tools makes it easy to find numerous apps for ASD tests, though not necessarily reliable. Thus, there is an urgent need for reliable ASD screening apps embedded with genuine instruments and effective scoring algorithms. This study reviewed the functional and non-functional features of the existing ASD screening apps and demonstrates a concise procedure for implementing reliable apps.

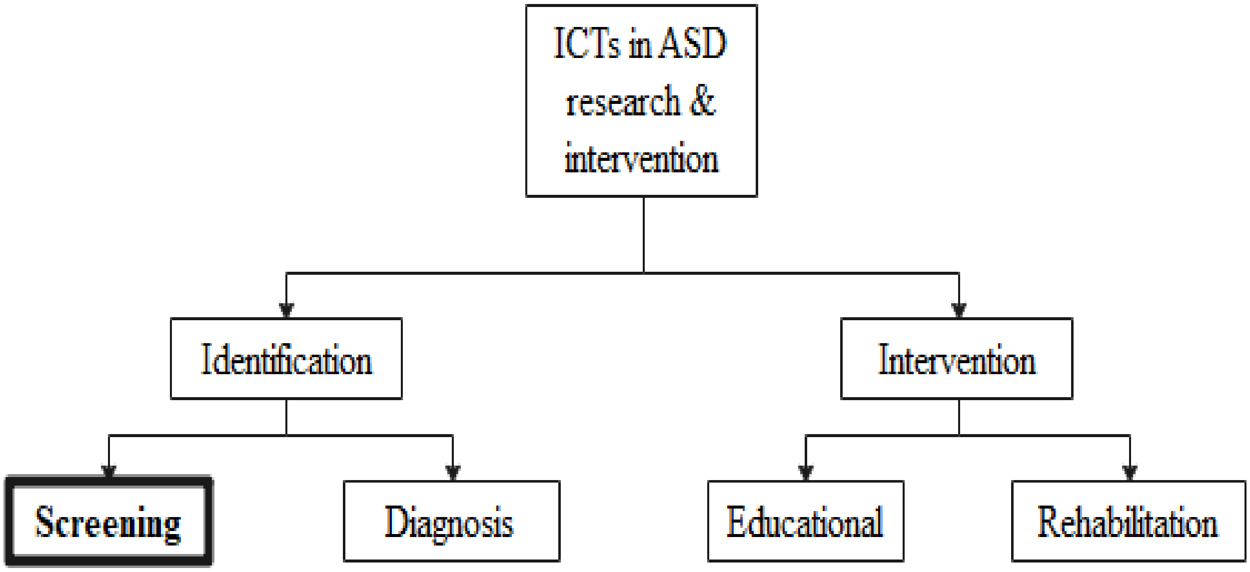


Figure 3.1: Role of ICTs in ASD research and intervention summarized base on the literature

Mobile platforms are the commonly used ICTs in the health sector, education, entertainment, and so on. Consequently, in this study, we reviewed the functional and non-functional features of the existing mobile apps for ASD screening. Table 3.1 highlighted the existing mobile apps meant for ASD screening and their features. However, a reliable mobile app for ASD screening must embed reliable instruments, implement an effective scoring algorithm, and maintain usability. Furthermore, even with the right instrument embedded, an effective and reliable ASD screening app must be robust in predicting concisely whether the basic conditions for “at-risk” of ASD are meet (i.e. a deficit in social communication and the presence of restrictive and repetitive patterns of behavior).

Table 3.1: Analysis of common mobile apps for ASD screening

Author & Year	App Name	Media	Research	Platform		Target Group		Items
				Android	iOS	Children	Adults	
Thabtah (2019)	ASDTests	Questionnaire	√	√	×	√	√	10
Cambodia (2019)	Autism Test	Questionnaire	×	√	×	√	√	13
Consurgo (2015)	Autism Test	Questionnaire	×	√	×	×	×	20
Patra and Arun (2011)	ISAA	Questionnaire	×	√	√	×	×	40
Tollet (2019)	AQ-10 Test	Questionnaire	×	√	×	×	×	10
Xie (2019)	CHAT-23 Scale	Questionnaire	×	×	√	√	×	23
La Trobe (2017)	ASDetect	Video	×	√		√	×	NA
Nazneen et al. (2015)	NODA	Video	√	√	√	√	×	NA
Egger et al. (2018)	AaB	Video	√	×	√	√	√	NA
This study		Questionnaire	√	-	-	√	-	<30

The apps depicted in

Table 3.13.1 utilized various handcrafted rules in quantifying the symptoms of ASD. However,

Author & Year	**App Name**	**Media**	**Research**	**Platform**		**Target Group**		**Items**
				Android	**iOS**	**Children**	**Adults**	
Thabtah (2019)	ASDTests	Questionnaire	√	√	×	√	√	10
Cambodia (2019)	Autism Test	Questionnaire	×	√	×	√	√	13
Consurgo (2015)	Autism Test	Questionnaire	×	√	×	×	×	20
Patra and Arun (2011)	ISAA	Questionnaire	×	√	√	×	×	40
Tollet (2019)	AQ-10 Test	Questionnaire	×	√	×	×	×	10
Xie (2019)	CHAT-23 Scale	Questionnaire	×	×	√	√	×	23
La Trobe (2017)	ASDetect	Video	×	√		√	×	NA
Nazneen et al. (2015)	NODA	Video	√	√	√	√	×	NA
Egger et al. (2018)	AaB	Video	√	×	√	√	√	NA
This study		Questionnaire	√	-	-	√	-	<30

designing the handcrafted rules to compute the scores of the screening instrument requires extensive human experience and knowledge. Replacing the handcrafted rules with a data-driven machine learning approach seems advantageous. Unlike the handcrafted rules, machine-learning approaches are objective and bound to improve classification efficiency, predictive accuracy, specificity, as well as sensitivity. Thus, embedding ML-based classifiers into the mobile screening apps can offer an automatic classification of ASD at homes, schools, and clinics where clinicians could verify the screening results against their experience and knowledge.

Although there is no reported evidence on the clinical application of the machine-learning models, recent studies have investigated the applicability of ML in improving the classification time of an ASD diagnosis as well as detecting the most influential items from the commonly used diagnostic instruments. Machine learning is a promising research endeavor that brought researchers from computer science, statistics, biology, and other arts and applied sciences intending to mimic the human brain in the discovery of hidden knowledge from unseen datasets. ML techniques, such as SVM, decision trees, neural network, and their ensembles, do not solely depend on users in the classification processes or model training. Detail on the application of machine learning in ASD studies will be revisited in the subsequent sections.

3.2 Current ASD Screening Practices

Instances of ASD are rapidly increasing. For instance, in the United State, the prevalence of the disorder reached 1 in every 68 children with developmental delays in terms of communication, socialization and repetitive behavioral patterns. While globally the prevalence of ASD is estimated at 1.5% of the entire world population. Furthermore, the current clinical diagnostic process is tedious, time-inefficient, for instance the average delay for clinical diagnosis is 3 years in UK and the commonly used diagnostic instruments such as ADI and ADOS contain numerous items. Consequent to that, many ASD cases remain undetected around the world. To alleviate these challenges, neuropsychiatrists, behavioral scientists and relevant researchers proposed several home-administered screening tools for preliminary assessment of people at risk of ASD.

Exemplary ASD screening instruments included the ADI-R, AQ and SRS among others as explained with the help of Table **3.2**3.2. In essence, ASD screening questionnaires provide a quick and vital threshold to diagnostic procedures. Unfortunately, universal guidelines have identified that no questionnaire-based tool can accurately diagnose ASD (Penner et al., 2018). Therefore, the current screening practices involve asking caregivers to complete a paper-based questionnaire about a child's behavior and if the caregiver indicates a certain number of symptoms, the trained professional refers the child for a clinical assessment. In other words, the responses determine if the child is "at-risk" of ASD and needs to be referred for clinical assessments or semi-standardized behavioral tasks that will diagnose ASD symptoms and their severity. However, current screening practices share similar drawbacks with diagnostic practices such as limited awareness and inaccessibility to reliable screening tools, inconsistent application as well as subjective interpretations across professionals (Bartolotta & Rizzolo, 2019; Murphy et al., 2016; Ruparelia et al., 2016).

Table 3.2: Common ASD screening instruments

Instrument	Description
Checklist for Autism in Toddlers (CHAT)	Established in the 1990s and proposed for screening toddlers at 18 months
Modified-CHAT	It is an enhanced CHAT to provide a comprehensive parent-report questionnaire
M-CHAT-Revised/Follow-up	It is the most updated version of CHAT established in 2014 to reduce the number of false positives
Social Communication Questionnaire (SCQ)	SCQ is a 40-item, true/false questionnaire. It parallels the ADI-R in content and is used for brief screening to determine the need to conduct a full ADI-R interview
Autism Quotient (AQ) – Children	It is a parent-report questionnaire for quantifying ASD symptoms in children of 4-11 years
Autism Quotient (AQ) – Adolescents	Adapted version to situate adolescents (age 9.8–15.4 years) on the ASD continuum from autism to normality
Autism Quotient (AQ) – Adults	Developed to quantify symptoms of ASD in Adults
Infant Toddler Checklist (ITC)	24-item parent questionnaire rated on a 3 point Likert scale aimed for toddlers within 6-24 months
Systematic Observation of Red Flags	It provides an observational screening measure for toddlers within 16-24-months
Test of Nonverbal Intelligence	Multiple versions of a multi-dimensional questionnaire for estimating the intelligence of neurologically impaired individuals. Studies adapted it in ASD screening.

3.3 Common Problems with Current ASD Screening and Diagnostic Practices

The absence of medical tests for ASD and limited professionals made questionnaires and structured interviews the common standards for ASD screening and diagnosis. Furthermore, existing screening and diagnostics procedures are not cost-effective; they are subjective and bound to inconsistent interpretations across professionals. In the rigorous diagnostics procedures, individuals need to be observed under a variety of activities. However, the limited time and constrained clinical environment will not favor an actual assessment of challenging ASD symptoms and lead to insufficient elicitation of information. Several studies suggested ICT-based ASD screenings to address the persisting problems associated with the long period of diagnosis and waiting, inconsistent measurement, as well as inadequate practitioners (Campbell et al., 2017; Duda et al., 2016; Levy et al., 2017; Rudra et al., 2014; Ruparelia et al., 2016; Thabtah & Peebles, 2019; Ward et al., 2018).

There exist another stem of challenges, despite the overreliance on the current assessment instruments for early diagnosis of ASD which leads to improved outcomes in communication and social interaction and guides parents to the right interventions in school, home, and clinic (Case-Smith et al., 2015; Durkin et al., 2015; Matson & Konst, 2014). Thus, the need for cost-effective assessments coupled with the global rise in ASD cases necessitates extensive research. This is because the current clinical assessment of ASD is not cost-effective; studies have shown that the cost of identifying one child with ASD in universal screening settings is about 700,000 USD (Yuen et al., 2018). Additionally, the assessment instruments perform poorly (Guthrie et al., 2019; Øien et al., 2018; Surén et al., 2019); there is a high number of false negatives and false positives. The tradeoff was to create a tool that reduced false positives (such as the M-CHAT R) and implement robust scoring methods. However, the tradeoff leads to more false-negative cases.

3.4 Commonly used Datasets for AI-Based Behavioral Assessment of ASD

Studies on the application of machine learning in the behavioral assessment of autism spectrum disorder have utilized numerous datasets in evaluating the models' performances as well as optimizing the items used in the data collection. Specifically, some of the popular datasets employed include Boston Autism Consortium (AC), Autism Genetic Resource Exchange (AGRE), Simons Simplex Collection (SSC) (Duda et al., 2017, 2015, 2016; Kosmicki et al., 2015; Levy et al., 2017; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012), National Database for Autism Research (NDAR) (Duda et al., 2015; Kosmicki et al., 2015) and Simons Variation In Individuals Project (SVIP) (Duda et al., 2015; Kosmicki et al., 2015; Levy et al., 2017). Other studies utilized data sets from ASDTest: Kaggle and UCI ML repository (Akter et al., 2019; Baadel et al., 2020; Goel et al., 2020; Pratama et al., 2019; Shahamiri & Thabtah, 2020; Suresh Kumar & Renugadevi, 2019; Thabtah, 2019; Thabtah et al., 2019, 2018; Thabtah & Peebles, 2020), Association of Parents and Friends for the Support and Defense of the rights of people with Autism (APADA) (Puerto et al., 2019), PASS app (Wingfield et al., 2020), Ondokuz Mayis University Samsun (Usta et al., 2019) and ASD outpatient clinics in Germany (Küpper et al., 2020). To achieve standardized comparative results, there is a need for standardized ASD data repositories for machine learning studies (Thabtah, 2018). However, discrepancies within the data repositories such as data imbalance limit the reliability of the high evaluation metrics reported in the studies (Abdelhamid et al., 2020; Alahmari, 2020). For instance, Torres et al. (2020) studied the statistical properties of ADOS scores from 1324 records and identified various factors that could undermine the scientific viability of the scores. Particularly, the empirical distributions in the generated scores violate the theoretical requirements of normality and homogeneous variance, which are essential for independence between bias and sensitivity. Thus, they suggested readjusting the scientific use of ADOS due to the variation in the distribution and dispersion of the scores, the lack of proper metrics to define similarity measures to characterize change, and the impact that these elements have on sensitivity-bias codependencies and longitudinal tracking of ASD.

3.5 Dimensionality Reduction

Dimensionality reduction techniques are processes applied in reducing the number of input features, columns, or variables from a given dataset. Most often, the complexity of the data modeling task depends on the number of input parameters in the various cases. Studies employ various dimensionality reduction techniques to mitigate the difficulty of data visualization and predictions for the training dataset. Specifically, previous ML studies on ASD classification problems have utilized various dimensionality reduction techniques for obtaining the most relevant features and better predictive models.

The key to reducing ASD diagnostic delays entails optimizing items of the screening instrument as well as improving their accessibility across clinicians and carers. Furthermore, accessible screening instruments will provide better understanding to parents on the status of their children and the needed services and appropriate early treatments that could mitigate the challenges associated with the disorder. However, as stated earlier, the bulk of items in most of the current screening instruments are tedious to administer especially for parents and other unprofessional administrators. Accordingly, researchers have been criticizing the existing screening instruments as time-inefficient. Consequently, recent studies employed various dimensionality reduction techniques to summarize the items of the screening instruments. However, studies have cautioned the applicability of the commonly used dimensionality reduction techniques without a careful understanding of the veracity of the selected or transformed features.

Researchers ought to understand the applicability, pros, and cons of the commonly used feature selection and transformation techniques. Feature selection techniques automate the manual process of extracting the most influential parameters in static datasets across various applications (Alhaj et al., 2016; Roobaert et al., 2006). However, there is little or no evidence that could justify the conformity of the feature selection methods with the conceptual basis upon which professionals built and utilize ASD diagnostic instruments.

Furthermore, unlike other medical diagnoses, the absence of definitive measures and medical tests for diagnosing ASD makes it difficult to numerically quantify the disorder based on few parameters. Notably, accurate assessment of ASD relied on the precise application of the commonly used behavioral scales built based on the knowledge and expertise of the professionals. Thus, application of the human knowledge is imperative to reliable ASD diagnosis. Based on that, there is a need for quantifying the trade-offs of dimensionality reduction (ensuring fewer items for quick assessment) and validity (preservation of the human knowledge for correct diagnosis). Specifically, a machine-learning model built based on fewer behavioral features that do not sufficiently capture the human knowledge of the assessment instrument, will not be valid for clinical use. Thus, there is a need for applying dimensionality reduction techniques that professionals could be tracked on its ability to preserve the validity of the assessment instruments.

Similarly, the commonly used feature transformation techniques are reported to support the learning process of the common machine learning algorithms utilized in previous ASD studies.

Feature transformation techniques such as Log, Zscore, and Sine functions normalize data by converting excessively skewed entities into a normal distribution, converting features into -1 to 1 value range, and transforming instances to the sine 0-2π value intervals, respectively. However, studies have demonstrated how these transformation techniques could significantly compromise the relevance of the original data to the transformed data (Curtis et al., 2016; Feng et al., 2014; Lapteacru, 2016; Wiesen, 2006). Thus, researchers must be mindful of how these transformations could undermine the clinical validity of the resulting ML-based ASD screening and diagnostic models. For instance, Feng et al. (2014) demonstrated such irrelevancies between the statistical findings of standard tests performed on original and log-transformed data. Similarly, several studies have highlighted some of the pitfalls and inconsistencies in the application of Z-scores and how its concepts overlooked the meaning of the original data, as well as its vital statistics (Curtis et al., 2016; Lapteacru, 2016; Wiesen, 2006). In essence, the applicability of dimensionality reduction techniques must align with the conceptual basis of ASD screening and diagnosis.

3.6 Classification Algorithms

The common practice of identifying people at-risk of ASD was based on observable behavioural traits quantified in terms of questionnaire items (Cavus et al., 2021; Thabtah, 2018). The quantification of the ASD traits rely on handcrafted arithmetic rules that lead to either of the two screening outcomes; ASD and non-ASD. Thus, the reliability of these screening outcomes depend on the items used, expertise of the administrator as well as the scoring rules. Both designing and administering tasks involves application of experience and extensive human-knowledge. In response to the dependencies on human-knowledge, studies demonstrated the value of deriving automatic models based on retrospective ASD data and controls (Achenie et al., 2019; Cavus et al., 2021; Moon et al., 2019). Such automated models are data-centric and are believed to improve the objectivity in scoring ASD symptoms. Accordingly, the commonly used data-modelling techniques include machine learning (ML). However, neither clinicians nor parents can easily verify the usefulness of the data-centric models. Recently, researchers on the application of ML in ASD assessment have demonstrated the improved accuracy of the data-centric models and their efficacy toward achieving time-efficient and cost-effective screening of ASD cases (Moon et al., 2019; Thabtah, 2019; Thabtah & Peebles, 2020; Zhang et al., 2020). The field of machine learning encompasses knowledge of computer science, statistics and data science among others. Furthermore, the human-independent machine-learning techniques includes SVM, decision trees, neural network, and their hybrids.

Behavioral assessment of ASD is primarily a classification problem. This is because; the learning objective of the machine learning models is to provide accurate classification between ASD and non-ASD cases. Worthy of noting, classification is the process of training and testing a machine-learning model to automate two-directional (binary classification problem) or more directional (multi-class classification problem) discrete decision processes based on certain input parameters. For instance, the binary classification model can differentiate between ASD and no-ASD cases, while a multi-class classification model can be trained to differentiate between the multiple ASD constellations based on certain behavioral input parameters. Commonly used classification algorithms employed in the behavioral assessment of ASD include Random Forest (RF) (Baadel et al., 2020; Goel et al., 2020; Pratama et al., 2019; Wingfield et al., 2020), Support Vector Machines (SVM) (Bone et al., 2016; Kosmicki et al., 2015; Küpper et al., 2020; Levy et al., 2017; Suresh Kumar & Renugadevi, 2019), Alternative Decision Tree (ADTree) (Duda et al., 2015; Usta et al., 2019; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012), and Logistic Regression (LR) (Kosmicki et al., 2015; Thabtah, 2019; Thabtah et al., 2019). To achieve comparative results, most of the studies employed several algorithms such as Adaboost, Multilayer Perceptron, Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA), Naïve Bayes, and K-Nearest Neighbor (KNN).

3.7 Model Selection

One of the most difficult decisions is on choosing the best models for a particular dataset. This is because with numerous supervised and unsupervised classification algorithms one must have enough understanding of the data at hand as well as the conceptual understanding of the classification algorithms to deliver a generalizable model (Jadhav & Channe, 2016; Navada et al., 2011; Wang et al., 2010). Furthermore, classification algorithms are data-centric and have different data processing style; as such, no single algorithm is best for every dataset (Brunner & Kim, 2016; Wang, Qin, Jin, & Zhang, 2010). However, in both classification and regression problems, by following simple data visualization techniques and some theoretical understandings of the data processing style of the algorithms, it will be possible to recognize the most suitable algorithm for a particular dataset (Brunner & Kim, 2016). Visualization can be achieved with a pair plot, which reveal the patterns that will lead to a reduced error rate and best case in the algorithm's complexity. Identifying the probable best-case complexity of an algorithm requires understanding the logical concept of how it works. In general, by visualizing the data and understanding the working style of the algorithms, it will be easy to select the most appropriate machine-learning algorithm for a specific problem. In essence, with clearly defined input and target parameters, the choice will be from the pool of supervised machine-learning algorithm. Otherwise, unsupervised learning algorithms are preferred when the target parameters are not defined and the aim is on grouping the data points. Classification or regression approach can be considered when the target parameter is discrete or continuous, respectively.

The problem at hand, ASD and no-ASD assessment is a binary classification problem. As stated earlier, for the data-centric approach, selecting the best machine-learning algorithm that will suit the dataset requires a conceptual understanding of how the algorithms work and analyzing the data using data-visualization techniques (Thabtah, 2019; Thabtah et al., 2018; Wingfield et al., 2020). This is because each of the algorithms has its advantages and disadvantages (Achenie et al., 2019; Baadel et al., 2020; Poyarkov et al., 2016; Thabtah et al., 2019). The present study included multiple algorithms to have more comparative results. Specifically, in each of the four data modeling scenarios adopted in this study at least 23 different classifiers were trained and evaluated. Exemplary classification algorithms included in this experimental study are multiple variants of decision trees, SVM, and K-Nearest Neighbor. This is more closely related to the model selection approach of some of the previous studies that applied trial-error, which involves repetitive data modeling with different machine-learning algorithms, and the model with the highest accuracy is considered the best. The following subsections provide a conceptual understanding of the learning styles of the different machine learning algorithms.

3.7.1 Decision trees

The decision tree is among the most popular supervised learning algorithms employed in solving data classification problems. It is a logical, flowchart-like structure, as shown in Figure **3.**3.2 and Figure , which represent test on attributes using internal nodes and test outcomes or class labels using leaf or terminal nodes, while the uppermost node is termed as the root node. Decision trees have been employed in the explicit visualization of decisions as well as decision-making processes. The phrase "decision tree" implies the tree-like model of decisions and its simplicity in understanding, visualizing and interpreting classification decisions as well as the implicit feature selection or variable screening that decision trees perform (Su & Zhang, 2006; Wang et al., 2010). Another important capability of decision trees is their ability handling both categorical and numerical inputs as well as multi-output problems.

However, decision trees relatively depend on some human effort during data preparation and adjusting possible nonlinear relationships between parameters (Liu et al., 2020; Su & Zhang, 2006). Additionally, decision tree algorithms create complex trees that do not provide sufficient generalization for the training data (Jadhav & Channe, 2016; Navada et al., 2011). This is also known as overfitting. Nonetheless, the high variance in decision trees is another key disadvantage that is usually alleviated or lowered by methods of bagging and boosting (Poyarkov et al., 2016). Variance lead to unstable decision trees in which small variations in the data might result in a completely different tree being generated. As part of the dependence on human effort required by decision trees in mitigating bias from dominant classes, prior to fitting with the decision tree, manual balancing of datasets is recommended (Jadhav & Channe, 2016).

Furthermore, fitting decision trees with greedy algorithms cannot guarantee optimal decision tree. Thus, to mitigated that multiple trees are trained with random selection and replacement of features and samples. Another scenario that lead to the creation of bias trees is when there is dominance between the classes and multiple algorithms were used in deciding how a node should be split into two or more sub-nodes. Thus, the creation of the sub-nodes increases the homogeneity of results. In contrast, in line with the data, cases with similar traits are grouped into regions, nodes on all available variables are split in the decision tree and the split that results in the most homogeneous sub-nodes is selected. In the present study, multiple variants of the decision tree learner were utilized in the data modeling as can be seen in the experimental settings.

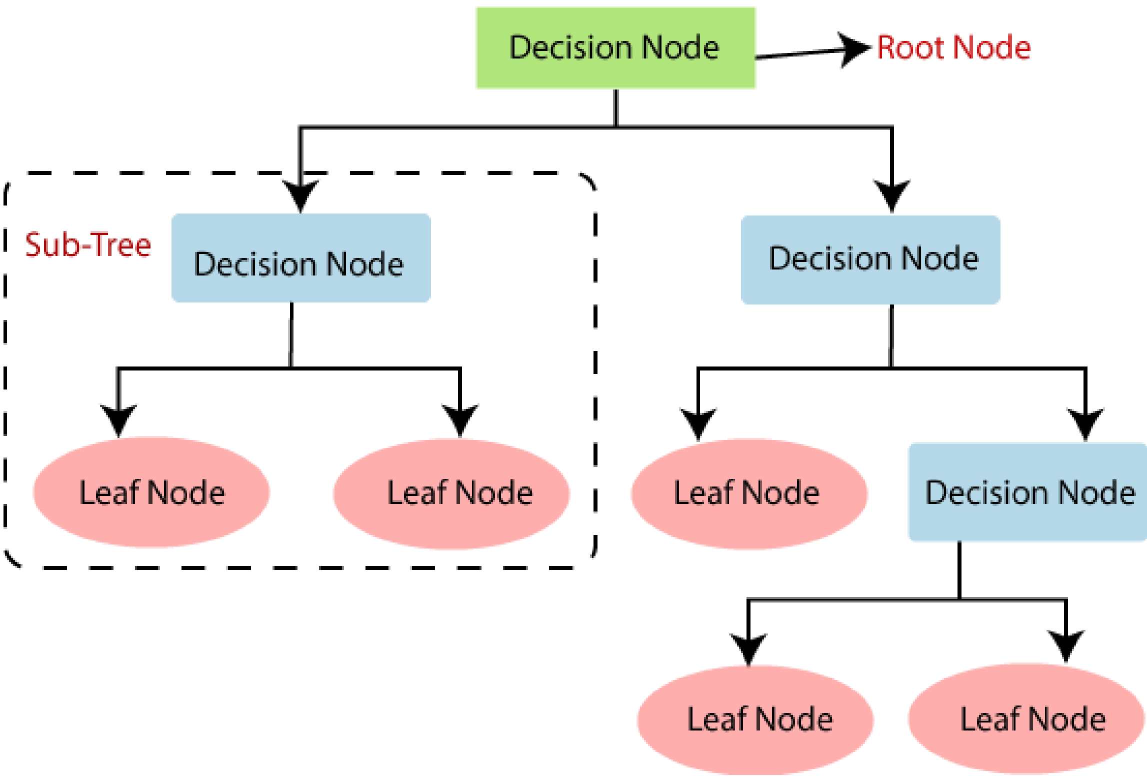


Figure 3.2: Decision tree terminologies (Poyarkov et al., 2016; Su & Zhang, 2006; Wang et al., 2010)

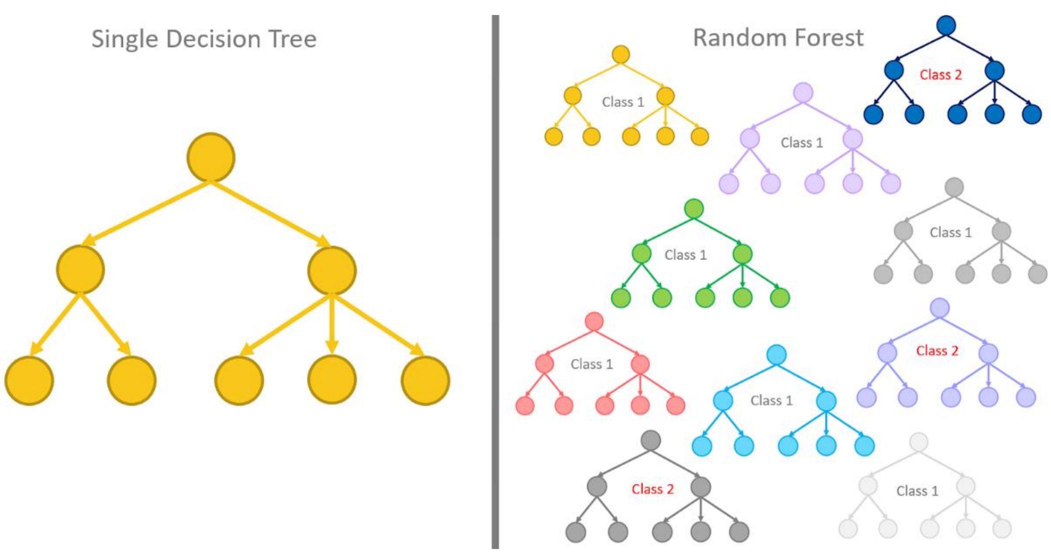


Figure 3.3: Decision tree vs random forest (Jadhav & Channe, 2016; Poyarkov et al., 2016)

3.7.2 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is also among the simplest supervised ML algorithms for solving both regression and classification problems. KNN categorizes unseen cases based on their similarity to the available categories (Peterson, 2009). Similar to NB, KNN is a lazy learner algorithm. However, KNN is a non-parametric algorithm; meaning that it does not assume the underlying data. KNN is a lazy learner because it does not immediately learn during training rather the training dataset is just stored and when unseen data arrived new instances are classified based on their similarities with the stored cases (Batista & Silva, 2009; Jadhav & Channe, 2016). For instance, to put a new case into the required category, we start by specifying the number of neighbors, followed by calculating the Euclidean distance between the neighboring cases. Based on the calculated Euclidean distance the nearest neighbors could be categorized into their respective similar instances (Jiang et al., 2007; Peterson, 2009). This is further explained with the help of Figure **3.4*1***3.4.

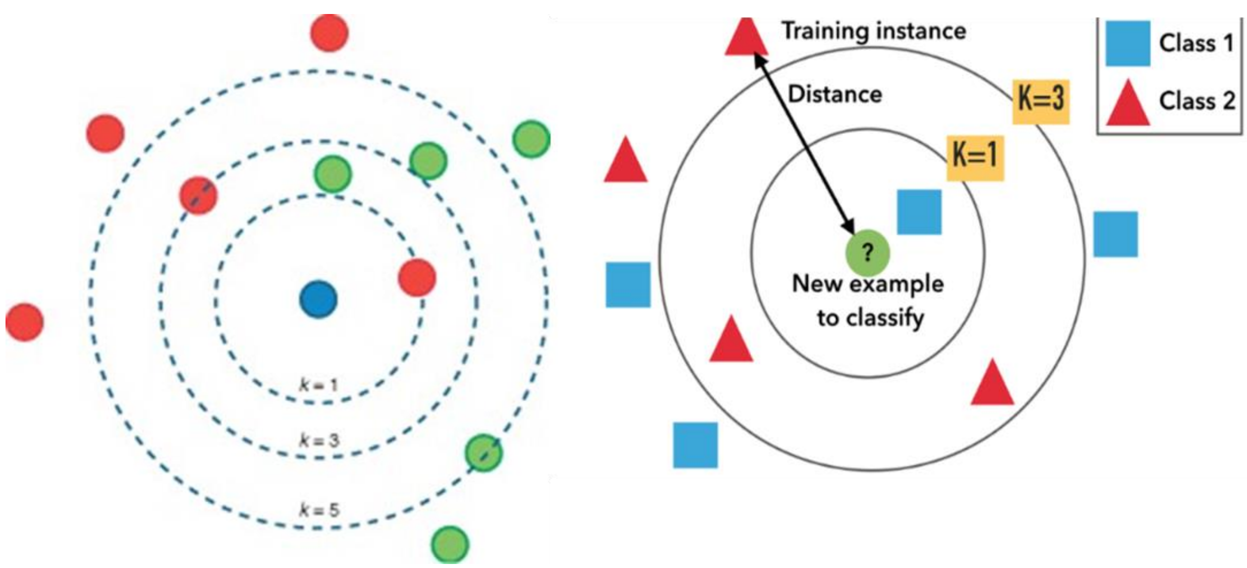


Figure 3.41: K-nearest neighbour (Batista & Silva, 2009; Jadhav & Channe, 2016; Jiang et al., 2007; Peterson, 2009)

KNN Algorithm is easy to implement, effective on the large training dataset, and robust to the noisy training data (Batista & Silva, 2009; Jadhav & Channe, 2016; Jiang et al., 2007; Peterson, 2009). The main limitation of KNN is the complexity in determining the best value of k as well as the computation cost in calculating Euclidean distance (Batista & Silva, 2009; Ertuğrul & Tağluk, 2017; Jiang et al., 2007; Lubis et al., 2020).

3.7.3 Naïve Bayes

Naïve Bayes is among the most effective as well as simplest supervised learning algorithms for both binary and multi-class classification problems (Rish, 2001; Webb, 2016). NB is probabilistic and is commonly used in text classification of multi-dimensional training datasets on Sentimental analysis, and spam filtration among others (Singh et al., 2019; Xu, 2018). As its name implies, Naïve Bayes assumes independence between the features occurrences and it is based on the Bayes Theorem. Thus, Naïve implies that the classification features contribute independently to the predictive processes. While Bayes Theorem implies the probabilistic future feature occurrence based on prior knowledge (Jadhav & Channe, 2016; Webb, 2016). In essence, NB classifiers are the family of simple “probabilistic classifiers” based on applying Bayes’ Theorem, as shown in the following equation, with strong (naive) independence assumptions between the futures.

$$\mathrm{P(ASD|nASD)} = \frac{\mathrm{P(nASD|ASD)\ P(ASD)}}{\mathrm{P(nASD)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.1)$$

Where,

P (ASD|nASD) is the probability of ASD occurring given evidence nASD has already occurred.

P (nASD|ASD) is the probability of nASD occurring given evidence ASD has already occurred.

P (ASD) is the probability of ASD occurring.

P (nASD) is the probability of nASD occurring.

In machine learning terminology, we can rewrite the equations as follows:

$$P(class|data) = \frac{P(data|class)\ P(class)}{P(data)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.2)$$

Or using Bayesian probability terminology as:

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3.3)$$

However, one of the key limitations of the NB classifiers is the assumption that all features are unrelated or independent (Jadhav & Channe, 2016; Rish, 2001). Thus, NB classifiers cannot learn the possible relationships that exist between features. The three types of NB classifiers are

Gaussian, Multinomial and Bernoulli (Singh et al., 2019). Gaussian models assume that the sampled cases contain features that are normally distributed. This assumption implies that the input parameters are non-discrete or continuous values; which might not always be the case. Nevertheless, both Multinomial and Bernoulli NB classifiers assumed that the sampled cases are based on a multinomial distribution and both are popularly applied in document classification problems. In contrast, while Multinomial NB classifiers work based on the frequency of input case against predictor variable, Bernoulli NB classifiers treat predictor variables as independent Booleans variables. Both could classify to which categories such as sports, politics, education, and so on a document belongs.

3.7.4 Logistic Regression

Logistic Regression is among the commonly used supervised learning models for predicting categorical or discrete target variable when given a set of input variables. Despites its similarity with Linear Regression, Logistic Regression is mainly used in solving classification problems by fitting an "S" shaped logistic function (Sigmoid Function) and providing probabilistic values within the range of 0 and 1 (Schein & Ungar, 2007; Thabtah et al., 2019). Logistic Regression could be either Binomial, Multinomial or Ordinal when the targets are only two, 3 or more unordered or 3 or more ordered possibilities, respectively.

3.7.5 Support Vector Machine

SVM is among the most widely used supervised learning algorithms for both classification and regression problems. The learning objective of classification SVM is to create the most optimum decision boundary or line called a hyperplane, which segregates n-dimensional dataset into classes for accurate categorization of unseen cases. SVM creates a hyperplane by choosing the extreme data points/vectors called support vectors (Pradhan, 2012; Suthaharan, 2016; Zhou, Zhang, & Wang, 2016). The two major types of SVM algorithms; linear and non-linear SVMs are popularly applied in facial recognition, text categorization, and image classification among others. Linear SVM applies to data cases that can be linearly separated using a single straight line. While non-Linear SVMs apply to cases that cannot be separated using a straight line. In non-linear SVMs, multiple hyperplanes are needed in separating n-dimensional space and the best decision boundary also known as optimal hyperplane provides the best classification of the data points (Pradhan, 2012; Wang et al., 2010). As a data-centric technique, the dimension of the dataset determines the dimension of the hyperplane. For instance, as shown in Figure **3.5*2***3.5 and Figure **3.6*3***3.6, if the dataset has 2 features, the best decision boundary will be a straight line, while a 2-dimensional hyperplane will be needed for a dataset with 3 features. SVM algorithms create hyperplanes that optimize the gap between the support vectors. In essence, support vectors are the cases that affect the position of the hyperplane and are closest to it from both dataset classes.

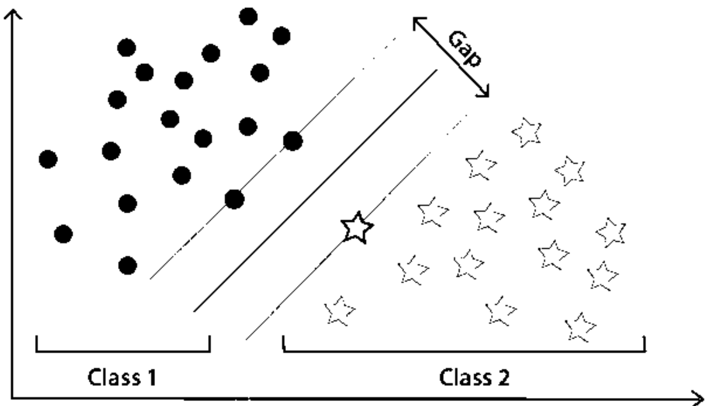


Figure 3.52: Support Vector Machine (SVM) (Suthaharan, 2016; Zhou et al., 2016)

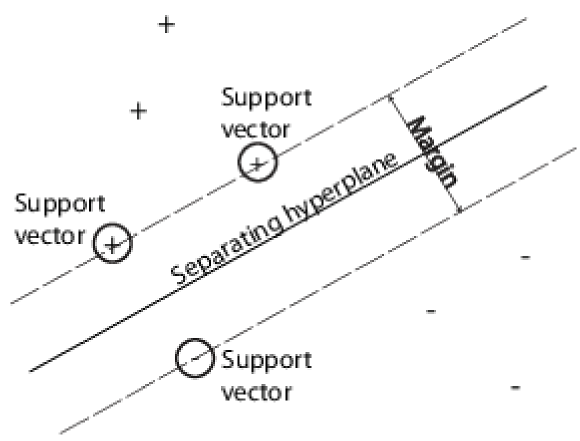


Figure 3.63: SVM terminologies (Pradhan, 2012; Suthaharan, 2016; Zhou et al., 2016)

The question of which algorithm is the best is usually approached with a trial-error approach by evaluating the performance based on different input combinations (Duda et al., 2015; Thabtah, 2019; Usta et al., 2019; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012). Trial-Error Approach is becoming very difficult because there are so many supervised and unsupervised algorithms in both classification and regression problems. Fortunately, with the help of simple data visualization techniques and some theoretical understanding of the algorithms, it will be easier to select the most appropriate machine-learning algorithm for a specific problem (Brunner & Kim, 2016). Noteworthy, by selecting a supervised machine-learning algorithm, it means the data set has clearly defined input and target parameters. Otherwise, unsupervised learning algorithms are preferred when the target parameters are not defined and the aim is on grouping the data points.

Furthermore, the classification or regression approach can be considered when the target parameter is discrete or continuous, respectively (Brunner & Kim, 2016; Jadhav & Channe, 2016). For instance, differentiating ASD from no-ASD cases is a binary classification problem. However, in selecting the best machine-learning algorithm that will suit a particular data set, conceptual understanding of how the algorithms work and application of data-visualization techniques are of paramount importance because each of the algorithms has its advantages and disadvantages. Although previous studies usually apply a trial-error approach (Duda et al., 2015; Usta et al., 2019; Wall, Dally, et al., 2012; Wall, Kosmicki, et al., 2012); which involves repetitive data modeling with different machine-learning algorithms and the model with the highest accuracy is considered the best, data visualization can be achieved with a pair plot. Pair plots will reveal the patterns that will lead to a reduced error rate and best case in the algorithm's complexity. Thus, identifying the probable best-case complexity of an algorithm requires understanding the pair plots concerning the logical concept of how the algorithm works.

3.8 Confusion Matrix

The confusion matrix is a cross-tabulation employed in depicting the basic parameters based on which predictive performance of the classification model can be obtained (Luque et al., 2019; Susmaga, 2004; Tharwat, 2020). The performance of every classification model could be inferred based on how much it categorize cases correctly or conduct misplacements. As shown in

Table **3.3**, case categorization could fall under any of the following basic classification performance metrics; True Negative, True Positive, False Negative, or False Positive.

Table 3.3: Confusion matrix

	Predicted class POSITIVE ("at-risk" of ASD)	**Predicted class NEGATIVE (control group)**
Actual class POSITIVE ("at-risk" of ASD)	TRUE POSITIVES (TP)	FALSE NEGATIVES (FN)
Actual class NEGATIVE (control group)	FALSE POSITIVES (FP)	TRUE NEGATIVES (TN)

Where,

TP is the number of patients already diagnosed with ASD and the screening instrument classified them as ASD positive.

FP is the number of patients that are truly non-autistic (i.e. belonging to the control group) but the screening instrument classified them as ASD positive. FP is also called a Type-I error.

TN is the number of patients that are truly non-autistic (i.e. belonging to the control group) and the screening instrument classified them as ASD negative.

FN is the number of patients already diagnosed with ASD but the screening instrument classified them as ASD negative. FN is also called a Type-II error.

With the help of the confusion matrix, we can calculate the different parameters for the model, such as accuracy, sensitivity, specificity, and precision among others.

3.8.1 Sensitivity and specificity

Sensitivity and specificity are statistical measures that indicate the predictive value of an instrument in classifying positive and negative cases in a test (Lalkhen & McCluskey, 2008). In this study, the data is presented using the confusion matrix depicted in the results section while Equations (3.4, 3.5 and 3.6) were followed in providing the scoring metrics of the predictive performance of the screening instrument based on the algorithm shown in **Error! Reference source not found.**4.2.

$$Sensitivity = \frac{TP}{TP+FN} \text{ ……………………………………………….. (3.4)}$$

$$Specificity = \frac{TN}{TN+FP} \text{ ………………………………………………… (3.5)}$$

Classification accuracy is among the most important parameters in determining the predictive performance of the model in a classification problem. The frequency of how often correct outputs are predicted by the model is expressed in terms of the classification accuracy. Computation of the classification accuracy is achieved in terms of the ratio of the number of correct classifications yielded by the model to the sum of cases correctly classified by the model as shown in Equation 3.6.

$$Overall\ accuracy = \frac{TP+TN}{TP+TN+FP+FN} \text{ ………………………………….....(1.6)}$$

CHAPTER 4

METHODOLOGY

This chapter described the proposed research methodology utilized during formulation of the screening instrument, data collection, data analysis as well as the machine-learning experimentation and data modeling. Categorically, this chapter provides explanation on the research participants, history and usability of the data collection instrument as well as the ML experimental settings. The chapter contains the following sections: participants, data collection instrument, data analysis, experiment setting as well as duration and resources. The chapter contains two subsections under the data analysis that described how the evaluation metrics of the manual scoring algorithm of the data collection instrument and the machine learning models could be calculated.

4.1 Proposed Research Methodology

The key aim of this thesis is to demonstrate the predictive performance of various ML models on a novel screening instrument and its promising approach that ensure quick and accurate screening of ASD cases as well as preserving the clinical validity of the screening instrument. Consequently, careful implementation of a scientific procedure was upheld toward achieving the research aim and objectives. The proposed research methodology was described with the help of **Error! Reference source not found.**.

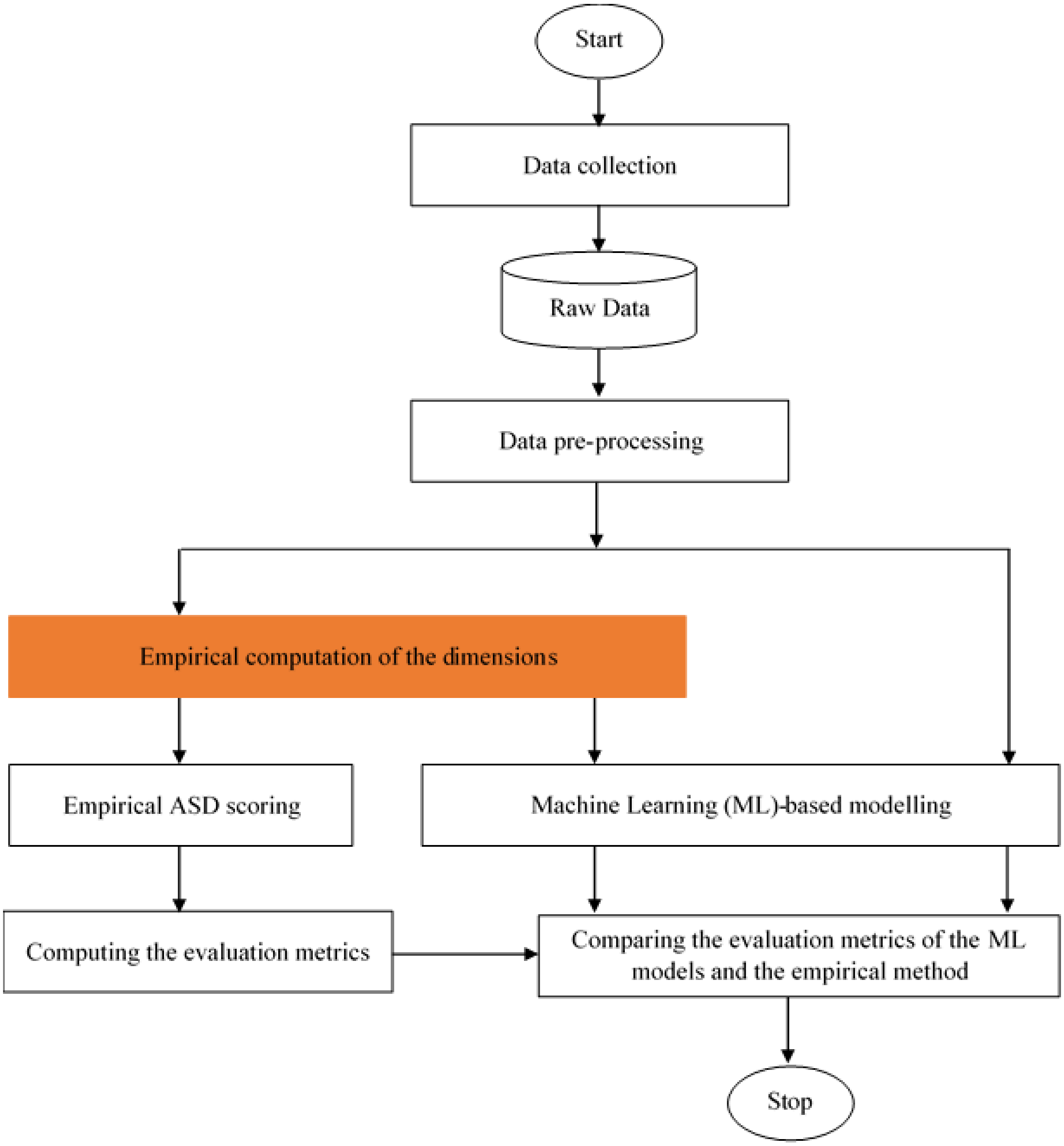


Figure 4.14: Flowchart of the proposed research methodology

4.1 Participants

The study data was collected from caregivers, parents, and other relatives of children with a diagnosis of neurodevelopmental disorders, including ASD based on purposive sampling approach. However, some of the control cases were drawn from participants with neither symptoms of ASD or comorbid neurodevelopmental disorders. Nonetheless, some of the responses on ASD cases were collected from teachers and clinicians. This is due in part to the lack of direct access to enough ASD cases through parents and caregivers. Initially, the data collection was planned to be conducted through psychiatrists and other clinicians from the eight federal neuropsychiatric hospitals in Nigeria. Nevertheless, due to the sudden advent of COVID-19 and the resulting restrictions on travels and hospital visits among other containment measures the researcher could not execute the initial data collection plan. This limited the number of responses realized. Thus, the data were partly collected through KoboToolbox (an internet-based data collection system) and paper-based questionnaires. Through both data collection means, 411 responses were gathered. Cases with missing values were eliminated and that reduced the responses to 380 valid cases containing 171 ASD cases and 209 controls.

4.2 Data Collection Instrument

The proposed data collection instrument named Child Development for Household Survey to Estimate Burden of ASD (CDHSEBA) is a questionnaire with its empirical scoring algorithm for assessing children "at-risk" of ASD. CDHSEBA is meant to be used by parents, caregivers, clinicians, and researchers in screening ASD symptoms in children at the developmental age of at least three years. The questionnaire was developed by researchers at the Childhood Neuropsychiatric Disorders Initiative (https://cndinitiatives.org/) under the leadership of Dr. Muideen Bakare (https://scholar.google.com/citations?user=TenqIhAAAAAJ&hl=en&oi=ao); who is a Chief Consultant Psychiatrist and Head, Training and Research at Child and Adolescent Unit, Federal Neuro-Psychiatric Hospital, Enugu State, Nigeria. CDHSEBA was developed based on the diagnostic criteria described in DSM-5 and it has been in use by clinicians at the Federal Neuro-Psychiatric Hospital, Enugu State, Nigeria (https://fnhe.gov.ng/). The scoring algorithm was also derived from DSM-5 to provide logical and numerical measures on the symptoms of ASD. Similar to other diagnostic instruments, manual procedures have been used in computing the ratings on “at-risk” of ASD and providing decisions based on the responses given on the items of the CDHSEBA questionnaire. Moreover, sensitivity, specificity and the commonly used evaluation metric of classification accuracy are the key statistics that have been used in ascertaining the scientific rigor of a diagnostic instrument in health-related researches (Trevethan, 2017). In the present study, the data collection instrument achieved a staggering sensitivity of 97% with classification accuracy and specificity of 56% and 23%, respectively. Figure 5 depicts the procedure based on which the ratings of ASD symptoms are computed. The rating scale is based on 0s and 1s; if the response is NO (i.e. behavior asked is not present), it is coded as 0 while YES (i.e. behavior asked is present), is coded as 1. The total score for the symptoms is then calculated and YES or NO decision is provided on each section of the questionnaire. Consequently, the overall decision is computed following the criteria given for the diagnosis of ASD in DSM-5 as depicted in Figure 5; which summarizes how the empirical scoring validates if the questionnaire responses meet the two conditions for “at-risk” of ASD.

Furthermore, the proposed questionnaire contains less than 30 items upon which symptoms of ASD are scored and the scoring algorithm follows section-by-section computations to meet the DSM-5 diagnostic criteria. Part 1 of the questionnaire captures demographic information (i.e. items 1, 2, and 3) while part 2 is categorized into sections A and B, whose description follows.

A) Deficits in social communication

This section of the questionnaire contains items 4, 5, 8, 9, 10, 11, 12, 13, 14, and 15, which covers deficits in social communication and can further be grouped into three major categories following the DSM-5 criteria:

A1: Deficits in socio-emotional reciprocity (items 4, 5, 8, 9)

A2: Deficits in non-verbal communication (items 10, 11, 12)

A3: Deficits in developing, maintaining, and understanding relationships (items 13, 14, 15)

Condition A: The patient can be said to be presenting with social communication deficits if they get a score of YES in 3/10 of these symptoms and the symptoms must be from at least two different categories i.e. must have a YES in at least A1 and A2, A1 and A3, or A2 and A3.

B) Restricted behavior

This section of the questionnaire captures information on the presence of restricted and repetitive patterns of behavior, activities, or interests. Items in the questionnaire capturing these are 6, 7, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 and can further be grouped into four major subcategories in accordance with DSM-5:

B1: Stereotyped movements, language, or use of speech (items 6, 7, 16, 17, and 18)

B2: Insistence on sameness and inflexibility thought (item 19)

B3: Highly restricted, fixated interests and abnormal intensity in focus (items 20, 21, 22, and 23)

B4: Sensitivity to sensory input (items 24, 25)

Condition B: The patient would be said to be presenting with repetitive and stereotyped behavior if they present with 3 of the listed symptoms with the symptoms being elicited from two different sub-categories; that's a combination of the positive screen in B1 and B2, B1 and B3, B1 and B4, B2 and B3, B2 and B4 or B3 and B4.

CDHSEBA was chosen in the present study due to so many reasons that align with the research aim and settings. Firstly, CDHSEBA has fewer items and that is in line with the need for a quick screening instrument that has fewer items than the common gold standards. Secondly, the clinical empirical scoring method of the data collection instrument involves some form of dimensionality reduction; specifically, hand-crafted rules for feature transformation, in which the complete set of items is transformed into fewer dimensions (i.e. A1, A2, A3, B1, B2, B3, and B4), which in turn lead to the main conditions upon which cases at-risk of ASD are identified. Last but not the least, the data collection instrument has been in use in an environment similar to the data collection units. Thus, there will be little or no environmental effect on the interpretability of the study findings.

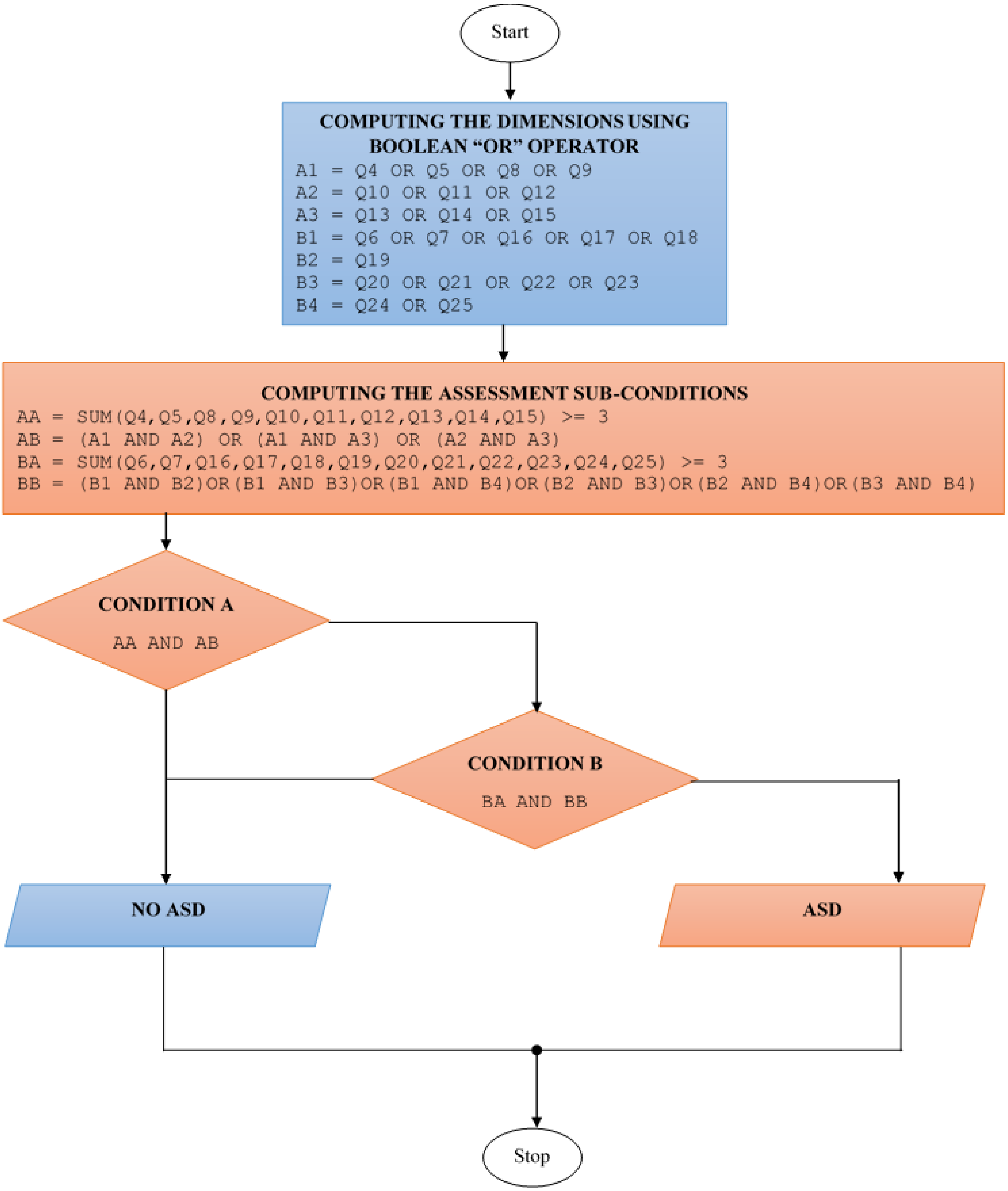


Figure 5: Flowchart of the empirical scoring algorithm of the study

4.3 Data Analysis

In the data analyses stage SPSS 25, Microsoft Excel 2016, and MATLAB R2019b were used. Before the machine learning modeling of the collected data, additional variables were computed using the data transformation feature of SPSS. The SPSS syntax employed in the computation is shown in Table 4.1.

Table 4.1: SPSS syntax for the computation of the variables

```
GET
  FILE='C:\Users\hp\Google Drive\Researches\PHD CIS\Thesis report\Data\Data.sav'.
DATASET NAME DataSet1 WINDOW=FRONT.
COMPUTE Q18=Q18A | Q18B | Q18C | Q18D | Q18E.
EXECUTE.
COMPUTE Q24=Q24A | Q24B.
EXECUTE.

COMPUTE A1=Q4 | Q5 | Q8 | Q9.
EXECUTE.
COMPUTE A2=Q10 | Q11 | Q12.
EXECUTE.
COMPUTE A3=Q13 | Q14 | Q15.
EXECUTE.
COMPUTE B1=Q6 | Q7 | Q16 | Q17 | Q18.
EXECUTE.
COMPUTE B2=Q19.
EXECUTE.
COMPUTE B3=Q20 | Q21 | Q22 | Q23.
EXECUTE.
COMPUTE B4=Q24 | Q25.
EXECUTE.
COMPUTE conditionAA=(Q4 + Q5 + Q8 + Q9 + Q10 +  Q11 + Q12 + Q13 + Q14 + Q15) >= 3.
EXECUTE.
```

```
COMPUTE conditionAB=(A1 & A2) | (A1 & A3) | (A2 & A3).
EXECUTE.
COMPUTE conditionBA=SUM(Q6,Q7,Q16,Q17,Q18,Q19,Q20,Q21,Q22,Q23,Q24,Q25) >= 3.
EXECUTE.
COMPUTE conditionBB=(B1 & B2) | (B1 & B3) | (B1 & B4) | (B2 & B3) | (B2 & B4) | (B3 &
B4).
EXECUTE.
COMPUTE conditionA=conditionAA & conditionAB.
EXECUTE.
COMPUTE conditionB=conditionBA & conditionBB.
EXECUTE.
COMPUTE computedASDstatus=conditionA & conditionB.
EXECUTE.


COMPUTE TP=(clinicalStatus = 1) & (computedASDstatus = 1).
EXECUTE.
COMPUTE TN=(clinicalStatus = 0) & (computedASDstatus = 0).
EXECUTE.
COMPUTE FP=(clinicalStatus = 0) & (computedASDstatus = 1).
EXECUTE.
COMPUTE FN=(clinicalStatus = 1) & (computedASDstatus = 0).
EXECUTE.
```

Where,

Q18 is the summarized value derived from the sub-items Q18A, Q18B, Q18C, Q18D, and Q18E using the OR (i.e. |) Boolean operator. Similarly, Q24 was also computed based on Q24A and Q24B. A shown in the code listing, the seven sub-dimensions of the data collection instrument (i.e. A1, A2, A3, B1, B2, B3, and B4) were equally derived based on their corresponding items. Nonetheless, the sub-conditions for assessing the disorder were equally computed as conditionAA, conditionAB, conditionBA, and conditionBB. Where, conditionAA tests whether at least three responses were YES on the items under section A of the data collection instrument and conditionAB ascertain if the YES responses were from either of the combinations of items under A1 and A2, A1 and A3, or A2 and A3 as clearly explained while describing the manual scoring algorithm. Equally, conditionBA tests if at least three responses from the items in section B of the data collection instrument are YES while conditionBB=ascertained if the YES items are from either of the combinations B1 and B2, B1 and B3, B1 and B4, B2 and B3, B2 and B4, or B3 and B4 as described in the manual scoring algorithm.

Furthermore, the code listing showed the computation of the main conditions for diagnosing the disorder (i.e. conditionA and conditionB), where conditionA is computed as TRUE if both conditionAA and conditionAB are TRUE and conditionB was equally computed in similar passion based on conditionBA and conditionBB. Finally, the code listing captured the key variable used in identifying the screening status of the participants (i.e. computedASDstatus). Accordingly, the computedASDstatus is TRUE if both conditions conditionA and conditionB are TRUE. Additionally, the basic evaluation metrics were equally shown in the code listing. Specifically, true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were identified based on the computed status (i.e. computedASDstatus) and the previous status (i.e. clinicalStatus) as indicated in the questionnaire response. Evaluation metrics of the manual scoring algorithm were computed based on the computed TP, TN, FP, and FN values.

4.3.1 Sensitivity and specificity analysis of the manual scoring algorithm

Sensitivity and specificity are statistical measures that indicate the predictive value of an instrument in classifying positive and negative cases in a test (Lalkhen & McCluskey, 2008). In this study, the data on the predictive performance of the manual scoring algorithm is presented using the confusion matrix depicted in

Table **3.3** while Equations (3.4, 3.5 and 3.6) were followed in providing the commonly used evaluation metrics. However, the predictive performance of the scoring algorithm depicted in Figure 5 is empirical; based on linear Equations. Machine learning models were usually built to validate the empirical findings provided by manual scoring algorithms. Notably, studies have shown the improved accuracy of machine learning algorithms over manual scoring algorithms (Baadel et al., 2020; Bone et al., 2016; Thabtah et al., 2019). However, some of the previous studies that employed varying data pre-processing techniques do not preserve the clinical validity of the data collection instrument. Accordingly, in the present study, an alternative approach to manual scoring was employed based on the machine learning algorithms. Specifically, to provide comparative findings, the present study employed both linear and non-linear machine learning classification algorithms to capture possible non-linear patterns in the data and to evaluate the performance of the models in classifying cases “at-risk” of ASD without compromising the conceptual validity of the data collection instrument. The technique proposed in the present study grouped items of the collection instrument into distinctive dimensions that align with the human knowledge use in the clinical assessment of ASD. Thus, the derived dimensions are utilized in training the machine-learning models. However, various data scenarios, with a reduced and extended list of items, were experimented on to provide more comparative results.

4.3.2 Sensitivity and specificity analysis of the machine learning models

Model development in machine learning is a data-centric process that involves training the model with one part of the data and testing with the other part. Figure 4.36 depicts the workflow diagram based on which the multiple machine learning models were built and evaluated. Similar metrics computed by Equations (3.4, 3.5 and 3.6) were employed in scoring the predictive performance of the classification models. The higher predictive performance of the ML classifiers validates the scientific value of the empirical scoring algorithm.

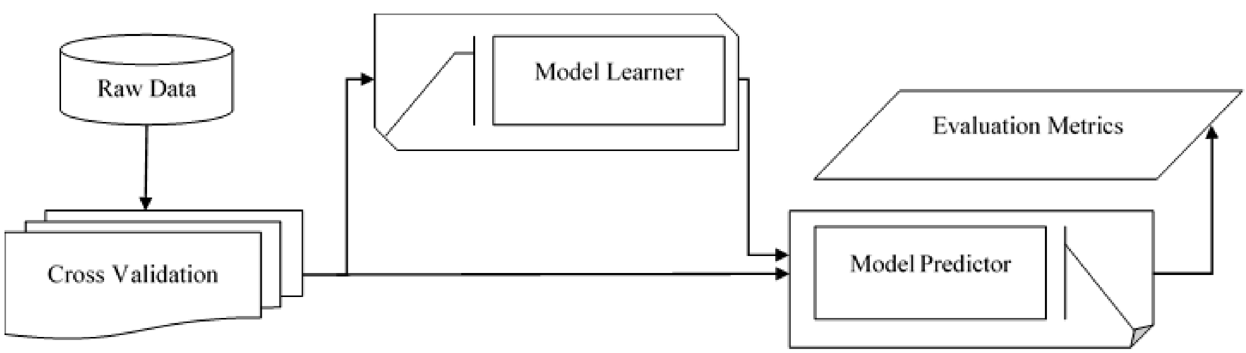


Figure 4.36: Machine Learning-based classification of cases “at-risk” of autism for the study

4.4 Experiment Setting

This section presents the experimental settings for the comparative analysis of the machine learning algorithms and the empirical scoring algorithm. While the empirical scoring algorithm utilized the items described under the data collection tool (i.e. Q4-Q25), multiple machine learning algorithms were implemented using a different combination of the CDHSEBA raw and processed parameters. Specifically, twenty-five different machine-learning algorithms were implemented based on four different data scenarios. The data Scenarios 1 and 3 involved the raw items of the CDHSEBA, while Scenarios 2 and 4 contain the transformed CDHSEBA dimensions, as described in Table . These scenarios were meant to provide comparative results on the impact of the clinical data transformation on the performance of the machine learning algorithms based on the commonly used evaluation metrics and weigh the results against the tradeoff of preserving the clinical validity of the data collection tool as well as the developed ML models. The study experimented on multiple machine learning algorithms because each of the algorithms has different learning styles in processing the dataset, as explained in the third chapter "theoretical framework". Specifically, simple variants of decision trees, variants of KNN and SVM machine learning algorithms were considered in the present study. Although, these variants are not the most sophisticated for real-world classification problems but have indicated their efficacy in terms of efficiency and predictive performance. Various distinct evaluation metrics were utilized in revealing the comparative performance of the empirical and ML algorithms in identifying autistic traits from the datasets of the present study. Specifically, classification accuracy, specificity, as well as sensitivity were used.

The proposed empirical scoring algorithm was implemented on SPSS version 25 as shown in the Table 4.1. The variable computation function of SPSS was utilized in implementing the empirical scoring algorithm. While the multiple machine learning classifiers were implemented on MATLAB version R2019b. The classification learning module of the MATLAB package was utilized in training the machine learning models. In testing the models generated by the multiple learning algorithms under consideration, in each of the four experimental data scenarios a 10-fold cross-validation was adopted. Therefore, in each of the 10 cross-validations, the training dataset is partitioned into 10 subsets. Then the remaining nine data subsets are randomly utilized by the classification algorithm in the process of testing the classifier. This validation processes is iterated ten times before averaging the classification error rates. Moreover, no hard coding was carried out as the algorithms module as well as the cross-validation procedures are embedded in the MATLAB R2019b platform and were selected from the graphical user interface before the learning phase. Finally, all the experimental runs were conducted on a HP-branded personal computer (Elite Book) that has x64-based Intel® processor Core™ i5-2410M dual CPU with 2.30GHz speed, and an installed RAM of 4.0GB capacity running on 64-bit Microsoft Operating System. The different parameter combination employed in each experimental scenario is described with the help of Table and the experimental stages employed in the machine learning modelling using MATLAB workspace is depicted with the help of Figure **4.4**4.4 – Figure 4.9.

Table 4.1: Highlights of the different data scenarios for the models' development

Data scenario	**Input variables**	**Target variable**	**Description**
Scenario 1	Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18, Q19, Q20, Q21, Q22, Q23, Q24, Q25	clinicalStatus	The inputs utilized were similarly used in the manual scoring method.
Scenario 2	A1, A2, A3, B1, B2, B3 and B4	clinicalStatus	The inputs used are the dimensions derived by categorizing the items in scenario 1 and averaging the responses.
Scenario 3	Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18A, Q18B, Q18C, Q18D, Q18E, Q19, Q20, Q21, Q22, Q23, Q24A, Q24B, Q25, and Q26 (severity level)	clinicalStatus	In this scenario, the complete set of items were passed to the machine learning algorithms without any processing or categorization. Unlike scenario 1, Q26 (severity level) was included and the sub-items in Q18 and Q24 items were not averaged.
Scenario 4	A1, A2, A3, B1, B2, B3, B4 and Q26 (i.e. severity level)	clinicalStatus	This is similar to scenario 2 with the addition of demographic variable of severity level (i.e. Q26)

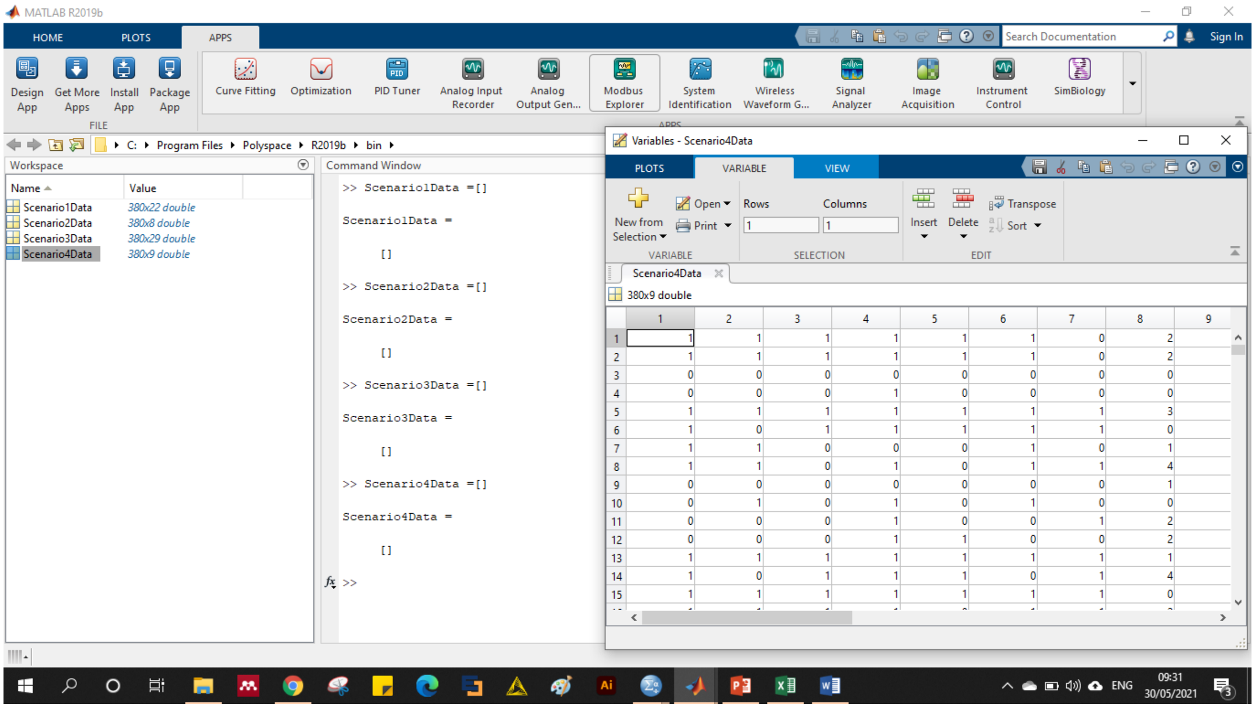


Figure 4.4: Loading the modelling data into the MATLAB workspace

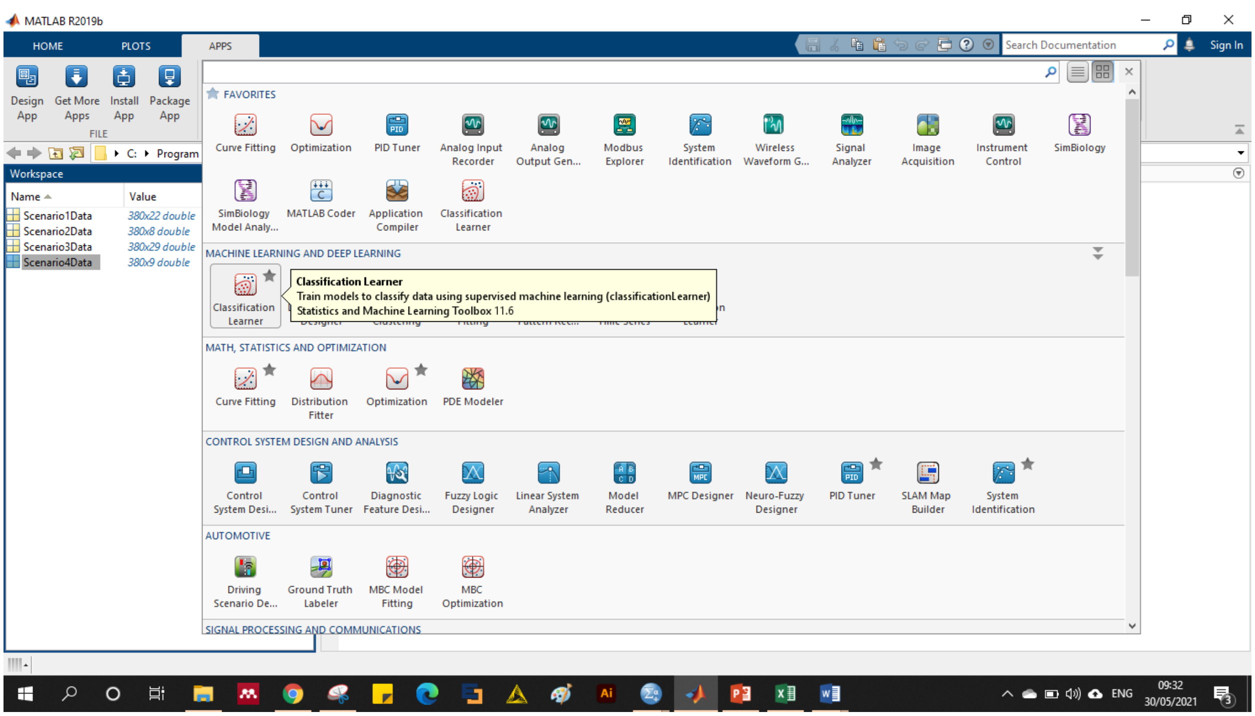


Figure 4.57: Selecting the classification module

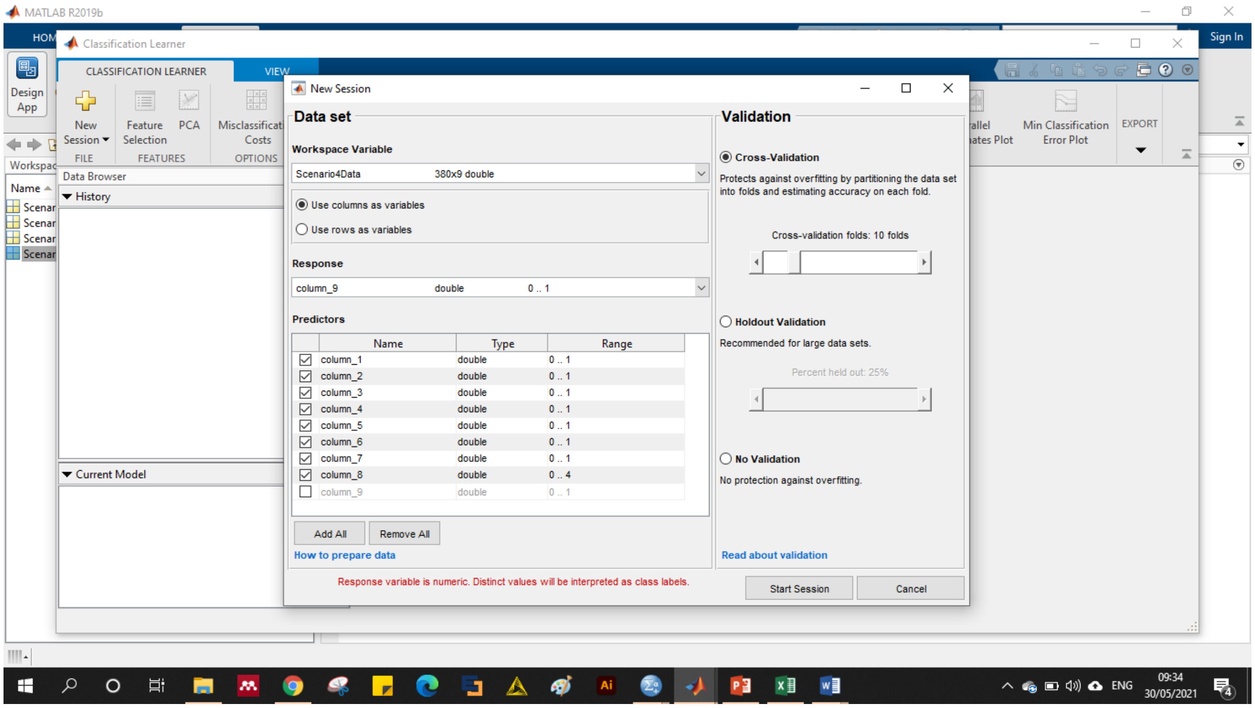


Figure 4.6: Setting the model training session

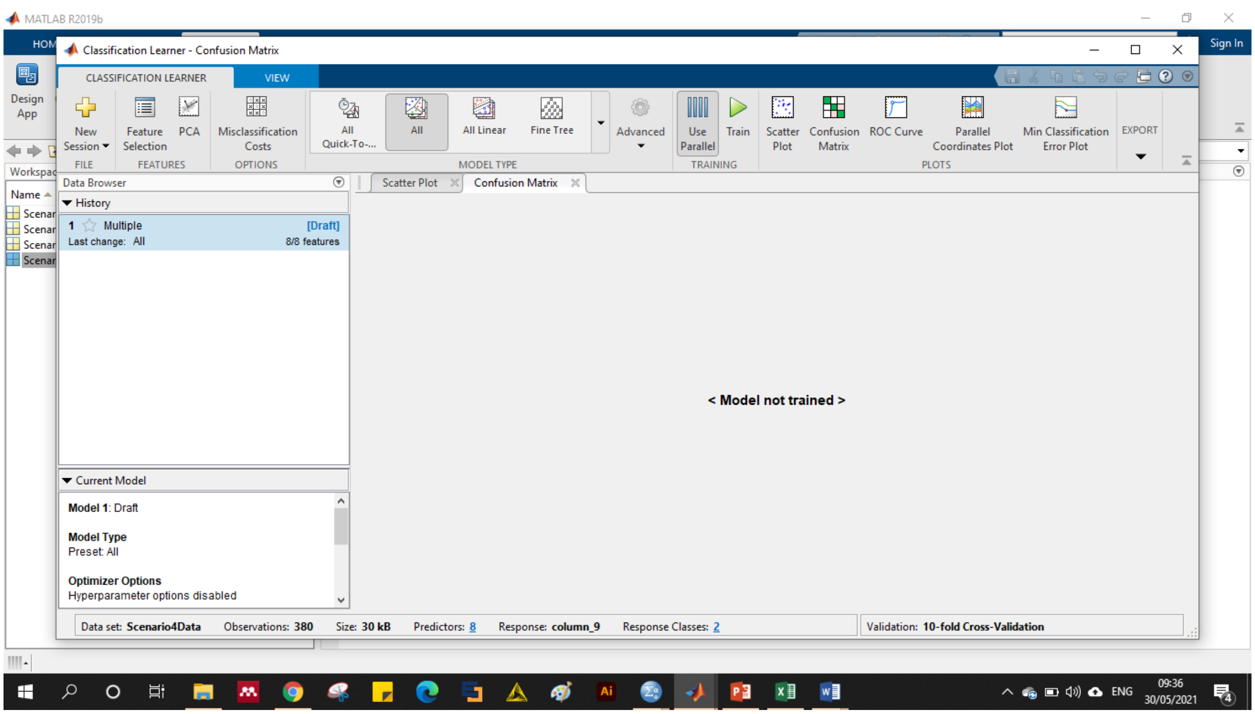


Figure 4.7: Selecting the ML classification algorithms

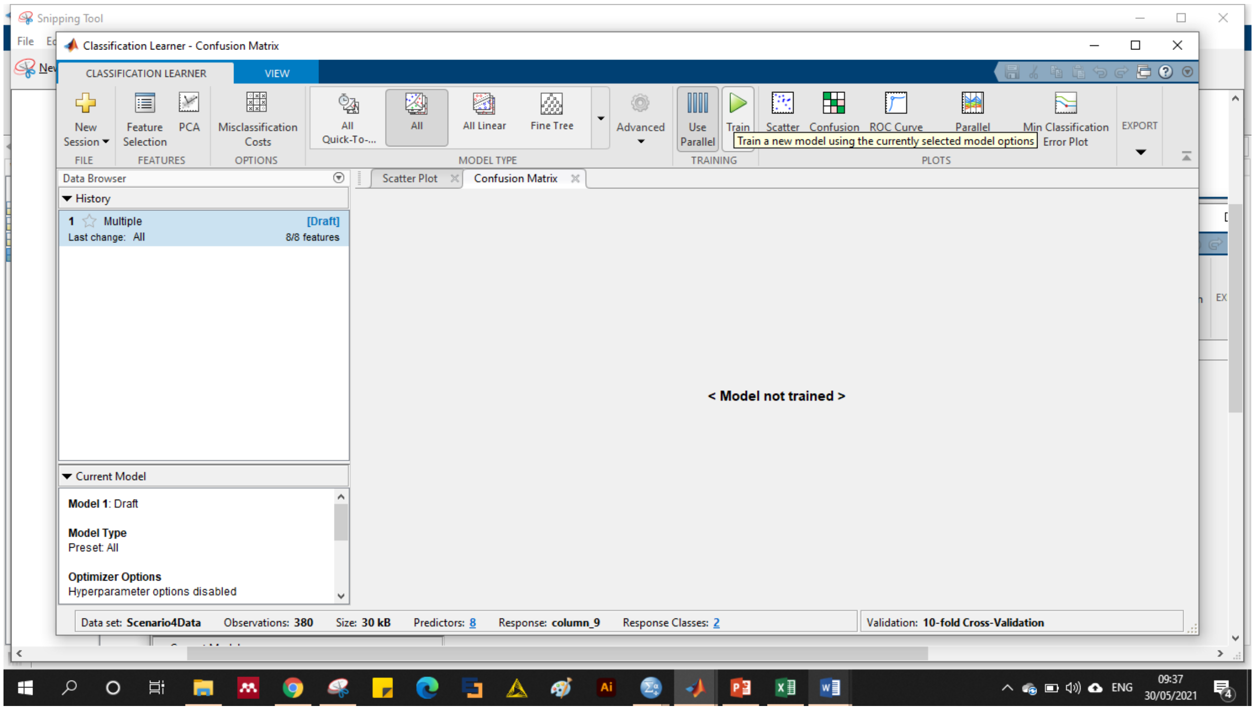


Figure 4.8: Start model training

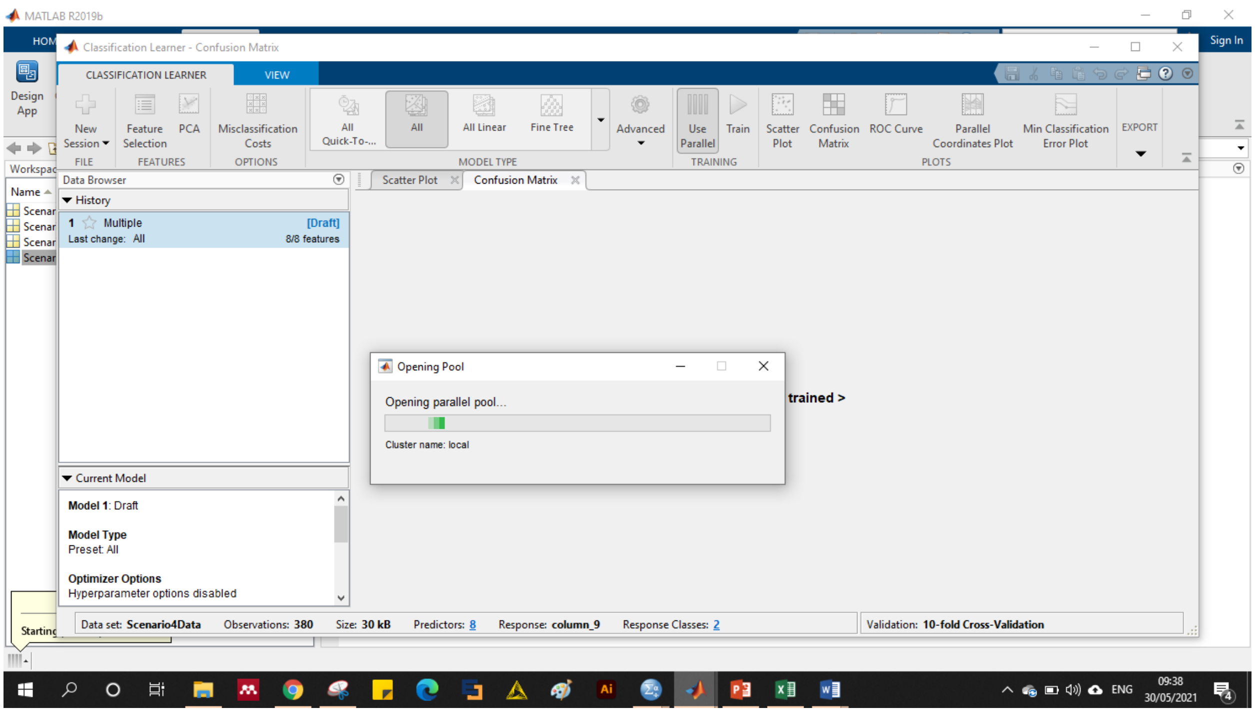


Figure 4.98: Running state

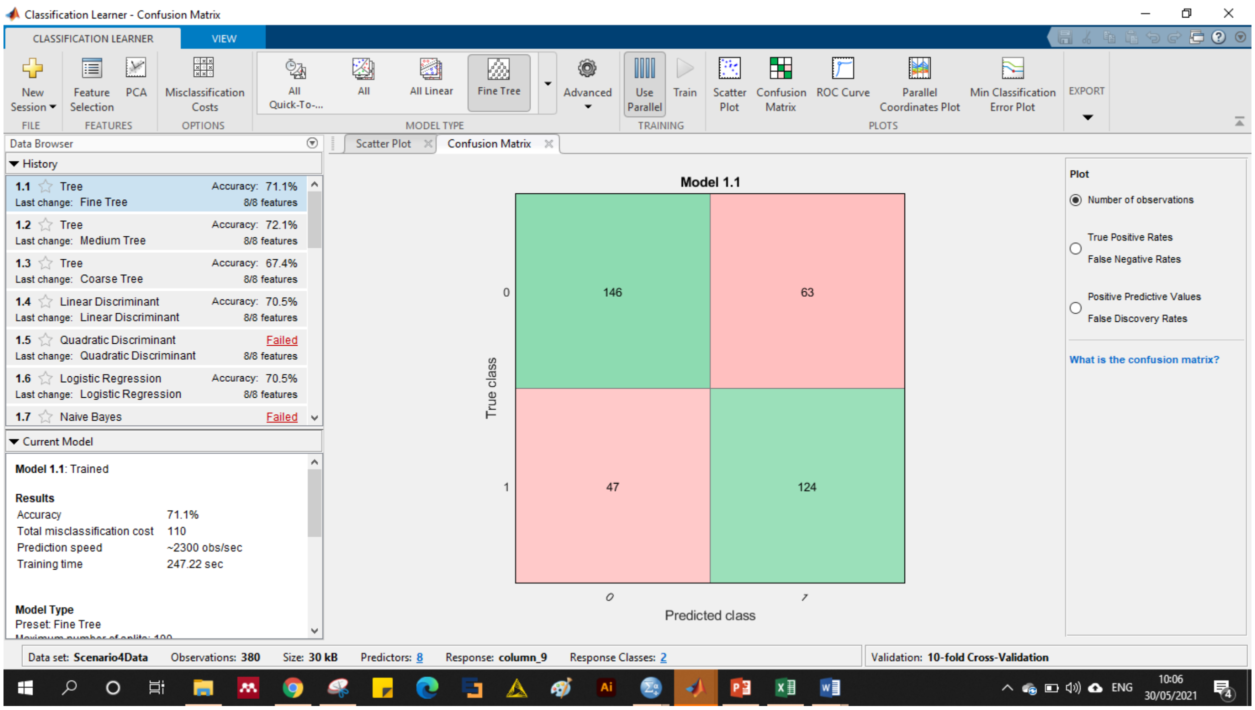


Figure 4.10: Results display

4.5 Duration and Resources

This study started with a literature search around mid 2018 after deciding on the research area and the first draft of the thesis document was completed in the month of March of 2021. All the research tasks were carried within this timeframe. Some of the key tasks and their corresponding weekly durations are depicted with the help of Table 4.. Supplementary tasks, readings as well as familiarization training that are paramount to the successful completion of the work were equally carried out within the same timeframe. All the financial expenses incurred during the research and related tasks were completely shouldered by the researcher.

Table 4.2: Schedule of the Thesis

Work done	**Duration**
• Literature Search and Readings	48 weeks
• Formulation of the Research Proposal	15 weeks
• Preparation of Data Collection Instrument	5 weeks
• Research Data Collection/Entry and Quality Inspections	6 weeks
• Data Analysis, Results Interpretation, and Discussion	16 weeks
• Thesis Report Writing and Compilation	20 weeks
• Revision of the thesis report based on the Supervisor's feedback	8 weeks

CHAPTER 5

RESULTS AND DISCUSSION

This chapter provided a detailed explanation with the help of tables and figures on the results obtained based on the research methodology adopted in the present study. Accordingly, the chapter was opened with the depiction of the confusion matrix for the empirical scoring algorithm followed by computation of the evaluation metrics achieved by the empirical approach. The results for the comparative evaluation of the empirical scoring algorithm and the multiple machine-learning models was depicted based on the four experimental data scenarios.

5.1 Confusion Matrix of the Empirical Scoring Algorithm

The ASD screening process is a binary classification problem since individuals are classified as either having ASD traits or No ASD traits using quantifiable behavioral variables. Therefore, performance evaluation methods that align with the binary classification problem in ML have been used. The basic parameters of true positives, false positives, true negatives, and false negatives, as explained in the confusion matrix (

Table **3.3**3.3), are used in deriving different evaluation metrics including classification accuracy, specificity, and sensitivity to evaluate the performance of both empirical and machine learning algorithms.

As stated earlier, using the confusion matrix, a test case will be assigned a predicted class in the classification step of the screening. Accordingly, Table 5.1 depicted the true positives, false positives, true negatives, and false negatives rates achieved by the empirical scoring algorithm. Consequently, the manual computations of the commonly used evaluation metrics were executed based on the Equations (3.4, 3.5 and 3.6).

Table 5.1: Confusion matrix of the empirical scoring algorithm

	Predicted class POSITIVE (“at-risk” of ASD)	**Predicted class NEGATIVE (control group)**
Actual class POSITIVE (“at-risk” of ASD)	166	5
Actual class NEGATIVE (control group)	162	47

Data: TP = 166, FN = 5, FP = 162, TN = 47

(TP: true positives, FN: false negatives, FP: false positives, and TN: true negatives)

Thus,

$$Sensitivity = \frac{TP}{TP+FN}$$

$$= \frac{166}{166+5}$$

$$= \frac{166}{171}$$

= 0.971

$$Specificity = \frac{TN}{TN+FP}$$

$$= \frac{47}{47+162}$$

$$= \frac{47}{209}$$

= 0.225

$$Overall\ accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{166+47}{166+47+162+5}$$

$$= \frac{213}{380}$$

= 0.561

$$Precision = \frac{TP}{TP+FP}$$

$$= \frac{166}{166+162}$$

$$= \frac{166}{328}$$

= 0.506

$$F - measure = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}$$

$$= 2 \times \frac{0.506 \times 0.971}{0.506 + 0.971}$$

$$= 2 \times \frac{0.491}{1.477}$$

$$= 2 \times 0.333$$

= 0.666

Therefore, the evaluation metrics of the empirical scoring algorithm can be expressed as percentages as follows:

- Sensitivity = 97.1%,
- Specificity = 22.5%,
- Accuracy = 56.1%,
- Precision = 50.6%,
- F-measure = 66.6%.

5.2 Evaluation Criteria of the Machine-Learning Models of the Experimental Scenarios

Similarly, Table5.2,

Table 5.3*1*, Table 5.42, and Table 5.53 summarized the true positives, false positives, true negatives, and false negatives rates achieved by the multiple machine learning models implemented under data Scenarios 1, 2, 3, and 4, respectively. Consequent to that, classification specificity, sensitivity, and accuracy metrics were calculated for each of the machine-learning algorithms based on Equations (3.4, 3.5 and 3.6) respectively. Recall that classification accuracy, as opposed to error rate, shows the percentage of the test cases that were correctly classified from the dataset considered in each of the scenarios. On the other hand, sensitivity indicated the number of test cases that are truly autistic based on the recorded clinical diagnoses (cases with ASD), while specificity shows the number of the test cases that are truly non-autistics based on the recorded clinical diagnoses (cases with no ASD). To be exact, various parameter combinations were made and the paramount evaluation metrics such as TP, FP, TN, FN, sensitivity, specificity, and accuracy were recorded based on the performance of the individual models. Table-Table 5.5 depicted the recorded metrics with an addendum of the comparative results computed based on the empirical scoring algorithm. This is further explained in the following subsections and the different data scenarios as summarized with the help of Table 4.2.

5.2.1 Results of machine-learning models for Scenario 1

The parameters utilized in this scenario were equally used in the empirical scoring algorithm. The input parameters passed to the multiple machine learning algorithms are Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18, Q19, Q20, Q21, Q22, Q23, Q24, Q25 while the output parameter considered is clinicalStatus. Table depicted the metrics achieved by each of the machine learning algorithms in this scenario. It worth noting that the training sessions of Quadratic Discriminant and Gaussian Naïve Bayes algorithms failed in this scenario. Thus, zero values were recorded for these algorithms. As shown in Table, the empirical scoring algorithm recorded the highest sensitivity of 97% while Fine Gaussian SVM has the highest specificity of 99% with the lowest classification accuracy of 57% (just 1% ahead of the empirical scoring algorithm). Overall, Coarse Gaussian SVM and Ensemble Bagged Trees achieved the highest accuracy of 78% in this scenario. In this scenario, Coarse Gaussian SVM is the best performing model with a 10% increase in specificity and sensitivity over Ensemble Bagged Trees.

Table 5.2: Scenario 1 modelling results

ID	Model Name	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
1	Fine Tree	142	67	59	112	0.65	0.68	0.67
2	Medium Tree	144	65	53	118	0.69	0.69	0.69
3	Coarse Tree	139	70	71	100	0.58	0.67	0.63
4	Linear Discriminant	161	48	42	129	0.75	0.77	0.76
5	Quadratic Discriminant	0	0	0	0	0	0	0
6	Logistic Regression	163	46	46	125	0.73	0.78	0.76
7	Gaussian Naïve Bayes	0	0	0	0	0	0	0
8	Kernel Naïve Bayes	177	32	57	114	0.67	0.85	0.77
9	Linear SVM	171	38	50	121	0.71	0.82	0.77
10	Quadratic SVM	165	44	51	120	0.7	0.79	0.75
11	Cubic SVM	156	53	58	113	0.66	0.75	0.71
12	Fine Gaussian SVM	207	2	160	11	0.06	**0.99**	0.57
13	Medium Gaussian SVM	174	35	52	119	0.7	**0.83**	0.77
14	Coarse Gaussian SVM	172	37	45	126	0.74	**0.82**	**0.78**
15	Fine KNN	148	61	51	120	0.7	0.71	0.71

16	Medium KNN	161	48	43	128	0.75	0.77	0.76
17	Coarse KNN	121	88	20	151	**0.88**	0.58	0.72
18	Cosine KNN	169	40	59	112	0.65	**0.81**	0.74
19	Cubic KNN	161	48	44	127	0.74	0.77	0.76
20	Weighted KNN	152	57	32	139	**0.81**	0.73	0.77
21	Ensemble Boosted Trees	152	57	44	127	0.74	0.73	0.73
22	Ensemble Bagged Trees	170	39	46	125	0.73	**0.81**	**0.78**
23	Ensemble Subspace Discriminant	167	42	45	126	0.74	0.8	0.77
24	Ensemble Subspace KNN	169	40	64	107	0.63	0.81	0.73
25	Ensemble RUSBoosted Trees	143	66	46	125	0.73	0.68	0.71
26	EMPIRICAL SCORING ALGORITHM	47	162	5	166	**0.97**	0.23	0.56

5.2.2 Results of machine-learning models for Scenario 2

This scenario utilized the transformed sub-dimensions of the data collection instrument as input parameters. Specifically, the modeling results were recorded for each of the machine learning algorithms based on the sub-dimensions A1, A2, A3, B1, B2, B3, and B4 as input parameters, while clinicalStatus was the output parameter predicted. Similar to the findings in Scenario 1, the training session failed for Quadratic Discriminant and Gaussian Naïve Bayes algorithms. Fine KNN achieved the highest sensitivity of 85%. Other variants of KNN (i.e. Medium, Cosine, and Cubic KNN) achieved classification accuracy equal to the empirical scoring algorithm (i.e. 56%) and the highest specificity of 87% with a very low sensitivity of 17%, each. The Medium Gaussian SVM algorithm achieved the highest classification accuracy of 70%. Noteworthy, in the Weighted KNN model the lowest classification accuracy of 54%, lower than that of the empirical scoring algorithm (i.e. 56%), was recorded. The results in Scenario 2 are further explained with the help of

Table 5.31*5.3.*

Table 5.31: Scenario 2 modelling results

ID	Model Name	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
1	Fine Tree	149	60	58	113	0.66	0.71	0.69
2	Medium Tree	149	60	58	113	0.66	0.71	0.69
3	Coarse Tree	140	69	58	113	0.66	0.67	0.67
4	Linear Discriminant	153	56	64	107	0.63	0.73	0.68
5	Quadratic Discriminant	0	0	0	0	0	0	0
6	Logistic Regression	160	49	71	100	0.58	0.77	0.68
7	Gaussian Naïve Bayes	0	0	0	0	0	0	0
8	Kernel Naïve Bayes	163	46	73	98	0.57	0.78	0.69
9	Linear SVM	141	68	63	108	0.63	0.67	0.66
10	Quadratic SVM	150	59	58	113	0.66	0.72	0.69
11	Cubic SVM	150	59	58	113	0.66	0.72	0.69
12	Fine Gaussian SVM	150	59	59	112	0.65	0.72	0.69
13	Medium Gaussian SVM	155	54	59	112	0.65	0.74	**0.70**
14	Coarse Gaussian SVM	140	69	57	114	0.67	0.67	0.67
15	Fine KNN	87	122	26	145	**0.85**	0.42	0.61
16	Medium KNN	182	27	142	29	0.17	**0.87**	0.56
17	Coarse KNN	138	71	54	117	0.68	0.66	0.67
18	Cosine KNN	182	27	142	29	0.17	**0.87**	0.56
19	Cubic KNN	182	27	142	29	0.17	**0.87**	0.56
20	Weighted KNN	176	33	142	29	0.17	0.84	0.54
21	Ensemble Boosted Trees	148	61	59	112	0.65	0.71	0.68
22	Ensemble Bagged Trees	147	62	62	109	0.64	0.7	0.67

23	Ensemble Subspace Discriminant	158	51	65	106	0.62	0.76	0.69
24	Ensemble Subspace KNN	90	119	29	142	**0.83**	0.43	0.61
25	Ensemble RUSBoosted Trees	149	60	58	113	0.66	0.71	0.69
26	EMPIRICAL SCORING ALGORITHM	47	162	5	166	0.97	0.23	0.56

5.2.3 Results of machine-learning models for Scenario 3

This data modeling scenario contains the highest number of input parameters compared to other scenarios experimented on. In this scenario, apart from the uncategorized sub-items in Q18 (i.e. Q18A, Q18B, Q18C, Q18D, Q18E) and Q24 (i.e. Q24A, Q24B), the demographic parameter of severity level was also incorporated into the input parameters. Thus, the input parameters utilized in this scenario were Q4, Q5, Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q16, Q17, Q18A, Q18B, Q18C, Q18D, Q18E, Q19, Q20, Q21, Q22, Q23, Q24A, Q24B, Q25, and Q26 (severity level). Equally, clinicalStatus was the only output parameter considered. As shown in Table **5.42**5.4, besides the empirical scoring algorithm, Coarse KNN achieved the highest sensitivity of 92%. Fine Gaussian SVM achieved the highest specificity of 100% with the lowest sensitivity of 1% and classification accuracy equal to that of the empirical scoring algorithms (i.e. 56%). Overall, Kernel Naïve Bayes appeared to be the best performing algorithm with the highest accuracy of 88%, specificity of 95%, and sensitivity of 81%.

Table 5.42: Scenario 3 modelling results

ID	Model Name	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
1	Fine Tree	141	68	56	115	0.67	0.67	0.67
2	Medium Tree	141	68	49	122	0.71	0.67	0.69
3	Coarse Tree	138	71	60	111	0.65	0.66	0.66
4	Linear Discriminant	167	42	33	138	0.81	0.8	0.80
5	Quadratic Discriminant	166	43	69	102	0.6	0.79	0.71
6	Logistic Regression	175	34	31	140	0.82	0.84	0.83
7	Gaussian Naïve Bayes	180	29	24	147	0.86	0.86	0.86
8	Kernel Naïve Bayes	198	11	33	138	0.81	**0.95**	**0.88**

9	Linear SVM	185	24	34	137	0.8	0.89	0.85
10	Quadratic SVM	178	31	32	139	0.81	0.85	0.83
11	Cubic SVM	175	34	33	138	0.81	0.84	0.82
12	Fine Gaussian SVM	209	0	169	2	0.01	**1.00**	0.56
13	Medium Gaussian SVM	188	21	31	140	0.82	0.90	0.86
14	Coarse Gaussian SVM	185	24	30	141	0.82	0.89	0.86
15	Fine KNN	153	56	38	133	0.78	0.73	0.75
16	Medium KNN	159	50	27	144	0.84	0.76	0.80
17	Coarse KNN	150	59	14	157	**0.92**	0.72	0.81
18	Cosine KNN	177	32	41	130	0.76	0.85	0.81
19	Cubic KNN	159	50	26	145	0.85	0.76	0.80
20	Weighted KNN	154	55	25	146	0.85	0.74	0.79
21	Ensemble Boosted Trees	172	37	43	128	0.75	0.82	0.79
22	Ensemble Bagged Trees	166	43	45	126	0.74	0.79	0.77
23	Ensemble Subspace Discriminant	173	36	30	141	0.82	0.83	0.83
24	Ensemble Subspace KNN	177	32	32	139	0.81	0.85	0.83
25	Ensemble RUSBoosted Trees	154	55	50	121	0.71	0.74	0.72
26	EMPIRICAL SCORING ALGORITHM	47	162	5	166	0.97	0.23	0.56

5.2.4 Results of machine-learning models for Scenario 4

This is closely related to the data modeling approach employed in Scenario 2 in terms of the input parameters utilized. This scenario maintained the transformed sub-dimensions as inputs to the multiple machine learning models with an addition of the demographic parameter of severity level. Thus, the inputs passed to the algorithms are A1, A2, A3, B1, B2, B3, B4, and Q26 (i.e. severity level) while clinicalStatus serves as the only output. The models' performance is lower than Scenario 3 with Coarse KNN achieving the highest sensitivity of 82% while Medium and Cosine KNNs achieved the highest specificity of 78% each. Table **5.5*3***5.5 described the experimental findings of this data modeling scenario.

Table 5.53: Scenario 4 modelling results

ID	Model Name	TN	FP	FN	TP	Sensitivity	Specificity	Accuracy
1	Fine Tree	138	71	53	118	0.69	0.66	0.67
2	Medium Tree	147	62	47	124	0.73	0.7	**0.71**
3	Coarse Tree	135	74	60	111	0.65	0.65	0.65
4	Linear Discriminant	147	62	53	118	0.69	0.7	0.7
5	Quadratic Discriminant	0	0	0	0	0	0	0
6	Logistic Regression	150	59	53	118	0.69	0.72	**0.71**
7	Gaussian Naïve Bayes	0	0	0	0	0	0	0
8	Kernel Naïve Bayes	153	56	61	110	0.64	0.73	0.69
9	Linear SVM	147	62	62	109	0.64	0.7	0.67
10	Quadratic SVM	144	65	54	117	0.68	0.69	0.69
11	Cubic SVM	139	70	56	115	0.67	0.67	0.67
12	Fine Gaussian SVM	153	56	72	99	0.58	0.73	0.66
13	Medium Gaussian SVM	152	57	64	107	0.63	0.73	0.68
14	Coarse Gaussian SVM	146	63	57	114	0.67	0.7	0.68
15	Fine KNN	146	63	90	81	0.47	0.7	0.6
16	Medium KNN	163	46	76	95	0.56	**0.78**	0.68
17	Coarse KNN	110	99	31	140	**0.82**	0.53	0.66
18	Cosine KNN	164	45	81	90	0.53	**0.78**	0.67
19	Cubic KNN	161	48	80	91	0.53	0.77	0.66
20	Weighted KNN	150	59	63	108	0.63	0.72	0.68
21	Ensemble Boosted Trees	144	65	45	126	0.74	0.69	**0.71**
22	Ensemble Bagged Trees	156	53	64	107	0.63	0.75	0.69
23	Ensemble Subspace Discriminant	151	58	59	112	0.65	0.72	0.69
24	Ensemble Subspace KNN	137	72	83	88	0.51	0.66	0.59
25	Ensemble RUSBoosted Trees	139	70	47	124	0.73	0.67	0.69
26	EMPIRICAL SCORING ALGORITHM	47	162	5	166	0.97	0.23	0.56

Overall, variants of KNN and SVM are the best performing models in all the scenarios while the empirical model has achieved better metrics than some of the machine learning models especially in Scenarios 2 and 4. The machine learning models achieved higher performances in Scenarios 1 and 3, both of which have the highest number of none-categorized input parameters.

5.3 Comparative Performance of the Machine-Learning Models and the Empirical Scoring Algorithm

The machine learning algorithms considered in the present study are not the most sophisticated employed in other classification applications but have proved their merits in terms of predictive performance and efficiency. Thus, Figure 5.1-5.7 were provided to visualize the comparative performance of the multiple machine learning models and the empirical scoring algorithm based on the different evaluation metrics adopted.

5.3.1 Comparative results based on number of true positives

True positive is the number of ASD cases that were correctly classified by the algorithms under consideration. Figure 5.1 depicted the comparative performance of the machine learning models and the empirical scoring algorithm in terms of the true positives achieved. As shown in the figure, the empirical scoring algorithm (model_ID = 26) achieved the highest number of true positives compared to the machine learning models. This implies that the empirical scoring algorithm has achieved higher performance in the correct identification of the true ASD cases and has surpassed the machine learning models in the correct classification of the true ASD cases. This might be because of the similarity between the tools used in the clinical diagnosis of the cases and CDHSEBA. Unfortunately, the present study has not captured the history of the specific tools used in the previous clinical diagnosis.

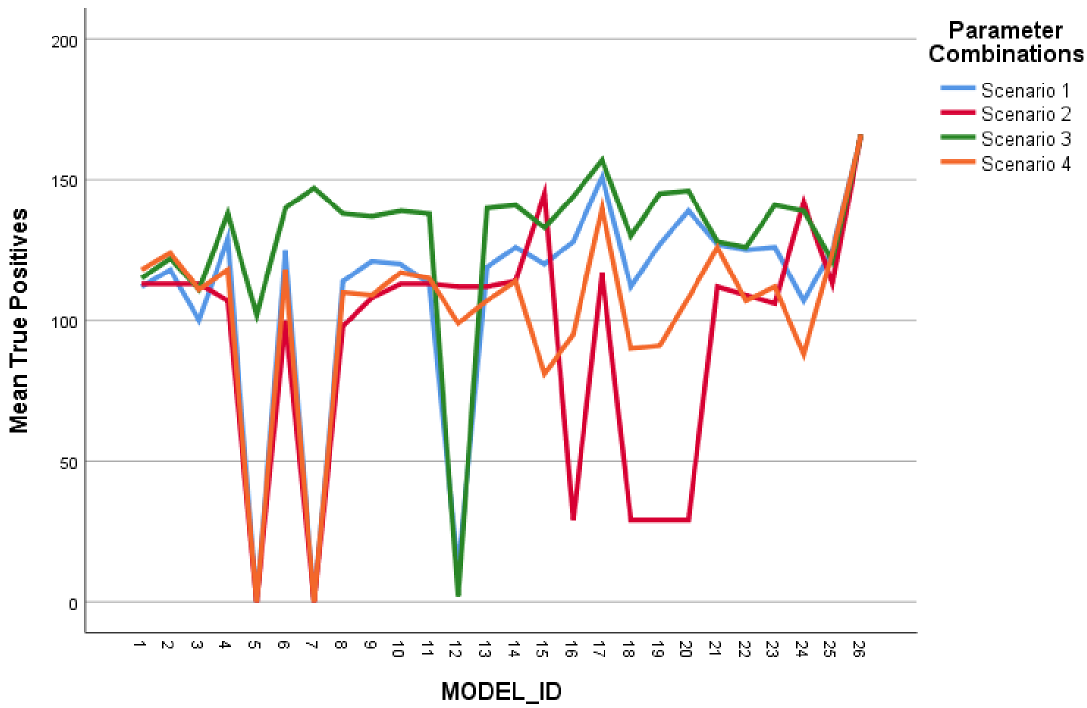


Figure 5.1: Comparative results of the true positives based on the various scenarios

5.3.2 Comparative results based on number of false positives

False positive is one of the values that serve as an indicator of case misclassification. However, the cost of misclassifying a non-autistic person as autistic is lower because further diagnostic tests could correct the error. In medical diagnosis, false negatives rates bear a higher cost than false positives. Thus, the false positives rate is should be given paramount importance while building models for the medical diagnosis of ASD (Alahmari, 2020). The comparatives results of the false positives for the experimental runs of the present study is presented with the help of Figure 5.2.

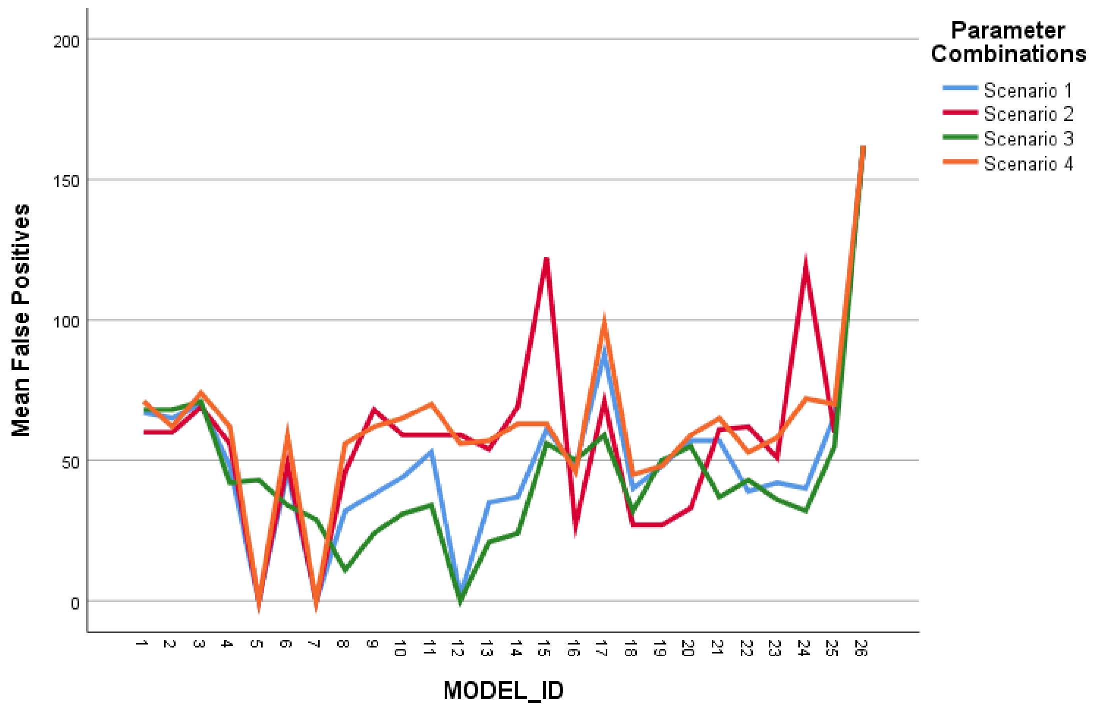


Figure 5.29: Comparative results of the false positives based on the various scenarios

5.3.3 Comparative results based on number of true negatives

True negative is the number of non-ASD cases that were correctly classified by the algorithms under consideration. Figure ***10***5.3 depicted the comparative performance of the machine learning models and the empirical scoring algorithm in terms of the true negatives achieved. As shown in the figure, the empirical scoring algorithm (model_ID = 26) recorded the least number of true negatives when compared to the various machine learning models. This implies that the empirical scoring algorithm has performed poorly in the correct identification of the true non-ASD. On the other hand, Fine Gaussian SVM achieved the highest performance in terms of true negatives under Scenarios 1 and 3 with 207 and 209 correct classifications of non-ASD cases, respectively.

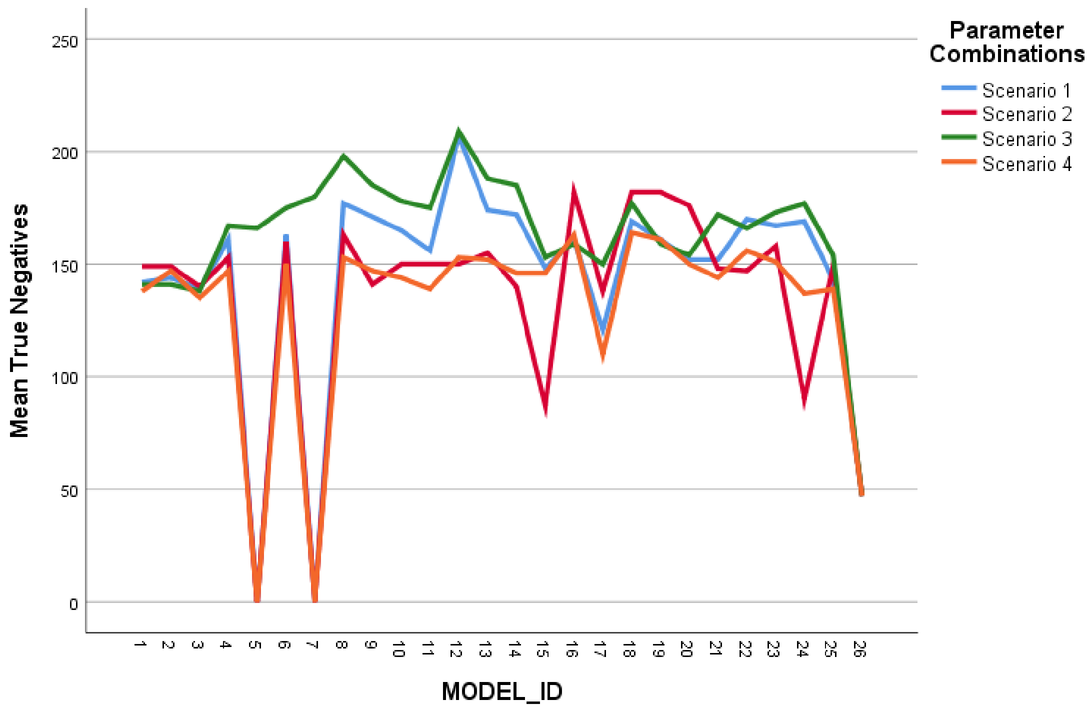


Figure 10: Comparative results of the true negatives based on the various scenarios

5.3.4 Comparative results based on number of false negatives

False negative is another important measure of misclassification. Noteworthy, the cost of misclassifying a person with autism as non-autistic (false negative) is higher than that of false positives because it will result in delayed intervention with possible consequences on the patients and their families. The comparative performance of machine-learning models and the empirical scoring algorithms in terms of the number of false negative cases, as shown in Figure5.4, indicated that the empirical scoring algorithm achieved the lowest number of ASD cases that were misclassified as non-ASD cases.

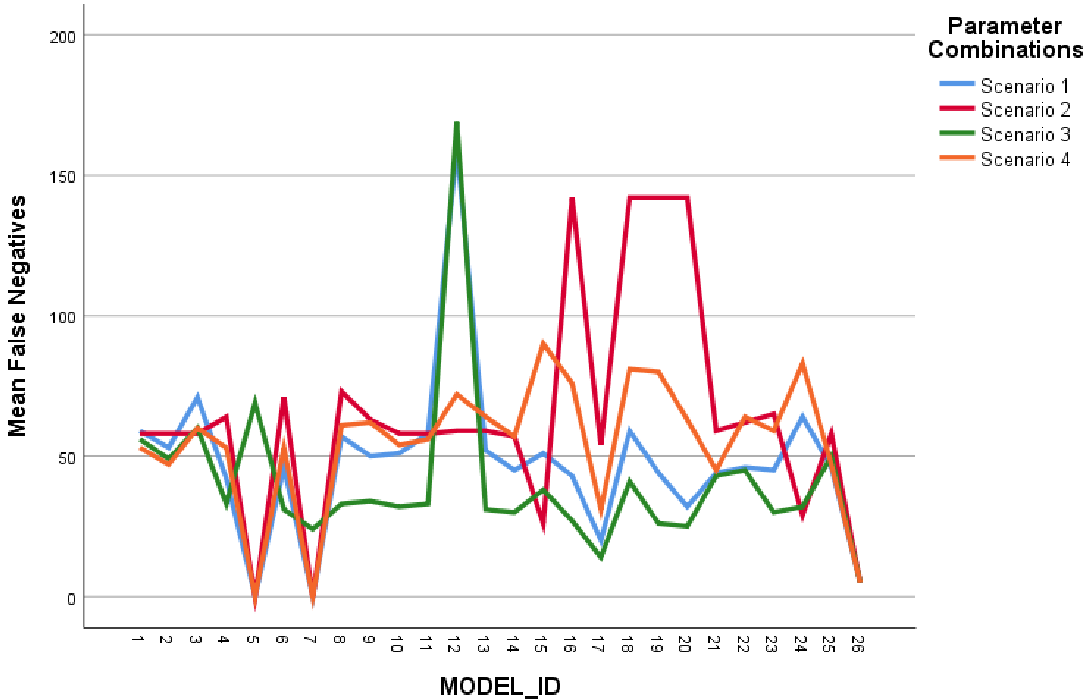


Figure 5.4: Comparative results of the false negatives based on the various scenarios

5.3.5 Comparative results based on sensitivity

Sensitivity is also known as true positives rate, it is the measure of the proposition of ASD cases that are correctly classified by the techniques under consideration. Figure **5.5*11***5.5 depicted the comparative performance of the machine learning models and the empirical scoring algorithm in terms of the true positive rates. As shown in the figure, the empirical scoring algorithm (model_ID = 26) achieved the highest sensitivity compared to the machine learning models. This implies that the empirical scoring algorithm surpassed the machine learning models in the correct classification of true ASD cases. This might be because of the similarity between the tools used in the clinical diagnosis of the cases and CDHSEBA.

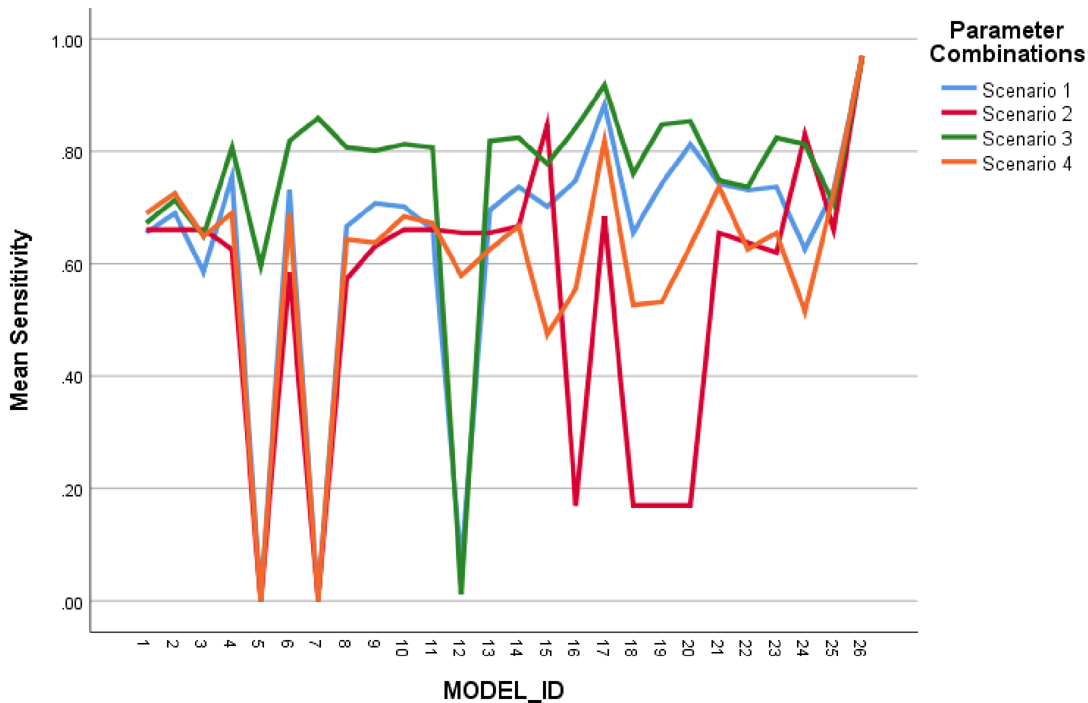


Figure 5.511: Comparative results of the models' sensitivity based on the various scenarios

5.3.6 Comparative results based on specificity

Specificity is also known as true negatives rate, it is the measure of the proposition of non-ASD cases that are correctly classified by the techniques under consideration. Figure **5.6*12***5.6 depicted the comparative performance of the machine-learning models and the empirical scoring algorithm in terms of the true negatives rates. As shown in the figure, the empirical scoring algorithm (model_ID = 26) achieved the lowest specificity compared to the machine-learning models. This implies that the empirical scoring algorithm lagged behind the machine-learning models in the correct classification of true non-ASD cases.

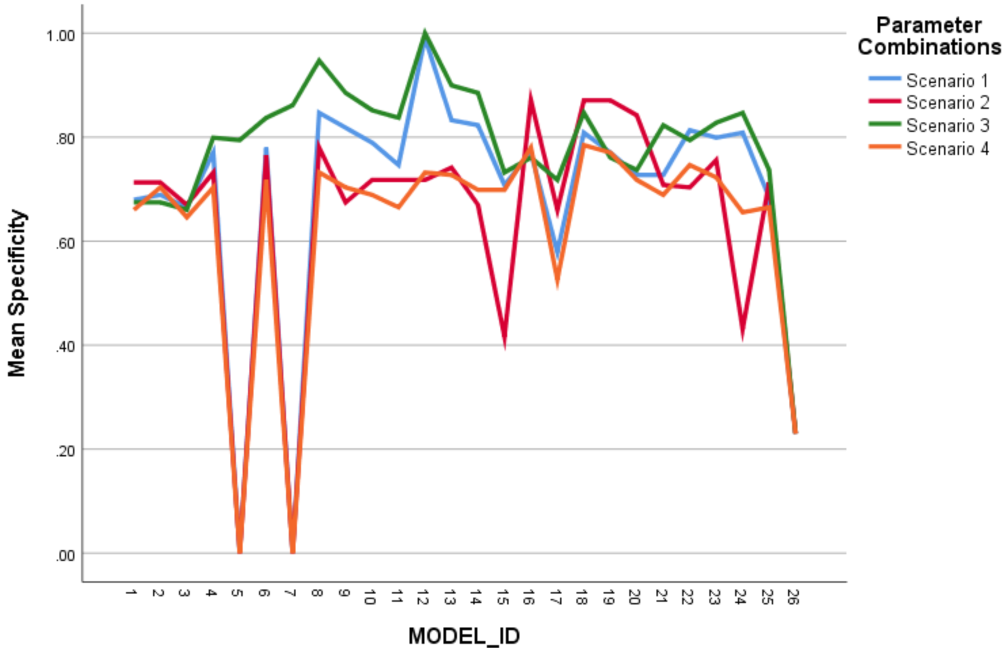


Figure 5.612: Comparative results of the models' specificity based on the various scenarios

5.3.7 Comparative results based on classification accuracy

Classification accuracy is the most commonly used and comprehensive criterion for performance evaluation. Comparative results of the empirical algorithm of the ASD screening instrument and multiple machine learning models based on classification accuracy is visually depicted with the help of Figure **5.7*13***5.7. The results indicated that most of the machine-learning models derived from the third experimental scenario (i.e. Scenario 3) achieved higher classification accuracies with the exeption of Fine Gaussian SVM (model_ID = 12) which achieved classification accuracy equal to the empirical scoring algorithm (accuracy = 0.56) as shown in the figure. On the other hand, Kernel Naïve Bayes is the best perfroming model with classification accuracy of 88%, followed by Gaussian Naïve Bayes (model_ID = 7), Medium Gaussian SVM (model_ID = 13), Coarse Gaussian SVM (model_ID = 14) with classification accuracies of 86% each and Linear SVM (model_ID = 9) that achieved classification accuracy of 85%.

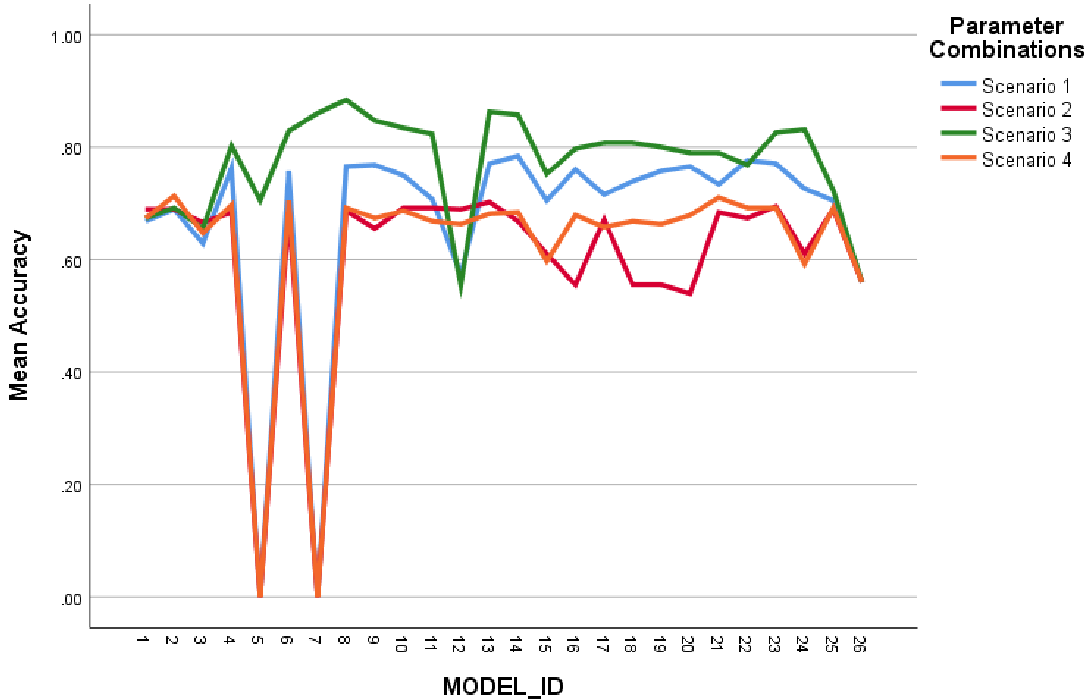


Figure 5.713: Comparative results of the models' accuracies based on the various scenarios

5.4 Discussion

The results have indicated so many revelations especially in terms of the impact that dimensionality reduction has on model performance and result interpretation. Firstly, the results obtained from the two experimental scenarios that utilized the untransformed input parameters (i.e. Scenario 1 and 3) have indicated better performance of the machine learning models in the predictive process based on the high evaluation metrics recorded. Similar to the approach and findings of Levy et al. (2017) and Usta et al. (2019), the inclusion of demographic parameter severity level improve the performance of the machine learning models. For instance, in Scenario 1 Fine Gaussian SVM achieved the highest specificity of 99%, with the inclusion of severity level to the inputs the same algorithm yielded a model specificity of 100%. This implies that apart from the main questionnaire items, demographic factors have a significant influence on the models' performance. Furthermore, the handcrafted rules in the empirical scoring algorithm do not consider demographic factors in the numerical quantification of ASD symptoms as well as the final classification. Thus, the machine-learning approach has proved its merit in determining other influential factors that affect the predictive performance of the models. Another comparative analysis in terms of classification accuracy has verified the assertion that demographic factors influence the models' performance. Specifically, in Scenario 1, Coarse Gaussian SVM and Ensemble Bagged Trees achieved the highest accuracy of 78%. However, with the inclusion of the demographic factor of severity level, in Scenario 3, the Kernel Naïve Bayes classifier (with an accuracy of 88%, specificity of 95%, and sensitivity of 81%) achieved an increase of 10% in the classification accuracy.

Comparative analysis between the original and transformed data provide an insight into the impact of dimensionality reduction on the models' performance and the interpretation of the result. Even though the approach followed in the dimensionality reduction is based on the experts' knowledge, the results differ in respect of the evaluation metrics. Specifically, the models' performances differ between the original and the transformed data. For instance, the highest classification accuracy of 78% recorded in Scenario 1 declined by 8%, and 7% after the data transformations in Scenario 2 and 4, respectively. This is in line with the assertion made by Curtis et al. (2016); Feng et al. (2014); Lapteacru (2016); and Wiesen (2006) on the statistical irrelevancies that exist between original and transformed data. However, unlike data-centric approaches, the present approach preserved the clinical validity of the transformed data, despite a reduction in the models' performance. This can be seen in the comparative performance maintained by the empirical scoring algorithm across all four experimental scenarios. For instance, in Scenario 2, variants of KNN (i.e. Medium, Cosine, and Cubic KNN) achieved classification accuracy equal to that of the empirical scoring algorithm (i.e. 56%) while the Weighted KNN model recorded the lowest classification accuracy of 54%, lower than that of the empirical scoring algorithm. Similarly, Fine Gaussian SVM despite achieving the highest specificity of 100% its classification accuracy is equal to that of the empirical scoring algorithm (i.e. 56%). Other instances that could prove the worthiness of the transformation approach in preserving the clinical validity of the screening instrument can be seen in the individual evaluation metrics highlighted in Figure5.1-5.7, in which the empirical scoring algorithm achieved the staggering high of 97% and very low false natives (FN=5). Despite high FP of 160 recorded from the empirical scoring algorithm, its clinical value is preserved because studies indicated that the false positives rate should be given paramount importance while building models for the medical diagnosis of ASD (Alahmari, 2020). This is because the cost of misclassifying a non-autistic person (FP) as autistic is lower because further diagnostic tests could correct the error. Moreover, in medical diagnosis, false negatives rates bear a higher cost than false positives. However, one of the critical implications in using the empirical scoring algorithm is that, with the high false positives and true positives rates, it will lead to outrageous figures on the cases at-risk of ASD. Overall, while the empirical scoring algorithms achieved outstanding performance in the correct classification of true ASD cases (sensitivity), the best performing machine-learning models outperformed the empirical method in the correct classification of non-ASD cases but both have achieved considerable classification accuracies.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

This is the last chapter of the thesis and it concludes the research work with summarized findings, and recommendations for future studies.

6.1 Conclusion

Assessment of the behavioral symptoms of autism spectrum disorder in form of screening is a common preliminary stage toward identifying people at-risk of ASD and a crucial approach to speeding up diagnostic referrals to extensively assess the symptoms under clinical settings. Nonetheless, scoring autistic traits in the current screening instruments such as the autism spectrum quotient (AQ), and autism diagnostic interview revised (ADI-R) among others rely solely handcrafted rules that have been associated with subjective interpretations. Thus, the tradeoff in ASD screening, and diagnostics studies is on improving the speed of the assessment processes and providing accurate and objective decisions for early assessment to accurate services and treatments. Studies have indicated the merit of automated models based on machine-learning techniques that build accurate assessments systems from retrospective cases and controls. Recently, machine-learning models for behavioral assessment of ASD have been broadly built based on a variety of pre-processed input data. Commonly utilized inputs include the items of ADOS, ADI-R, and AQ. Notably, the major aim of the previous studies was on quick and accurate screening and diagnosis of ASD. However, to achieve quickness, various data selection and transformation techniques were utilized despite evidence on the insufficiency of the reduced items and the inability of the transformation techniques in preserving the clinical validity of the assessment instruments. Specifically, studies have demonstrated inconsistencies in the findings of standard tests between original and transformed data. Similarly, none of the previous studies probes the sufficiency of the reduced parameters against the basis upon which clinicians diagnose ASD. Consequently, clinical validity and real-life applicability of the ML models are at stake despite the high evaluation metrics recorded by the previous studies. The performance of the models was based on the common evaluation metrics of accuracy, specificity, and sensitivity among others. On the other hand, popular machine learning algorithms utilized by the previous studies include the variants of SVM, decision tree, and KNN. In essence, the multitudes of challenges toward quick and accurate assessment of ASD are yet to be addressed by the previous machine learning approaches. In the present study, machine-learning application in the behavioral assessment of autism spectrum disorder was approached using a novel procedure that comprises few behavioral features and preserves the clinical validity of the assessment instrument. Therefore, the proposed approach maintains one of the core aim of ensuring speed and preserving the clinical validity of the machine-learning models. Consequently, comparative results were evaluated between the empirical algorithm of the ASD screening instrument and multiple machine learning models. The study findings revealed that the ML model based on Kernel Naïve Bayes is the best performing model with classification accuracy of 88%.

6.2 Recommendations

Despite several studies that demonstrated the machine-learning approach to ASD assessment, future studies should establish the clinical relevance of the data-centric approaches and readjust the scientific use of the assessment instruments. Accordingly, future studies should explore the best practices of scale development and feature reduction in line with the professional basis of ASD diagnosis in categorizing and evaluating the clinical validity of the robust ML models. Moreover, vital recommendations based on the findings of the present study can be approached following the different experimental scenarios utilized. Specifically, embedding the best performing ML models in any ASD assessment app could be approached based on the parameters utilized in the four data scenarios. In the first scenario, the ML-enable ASD assessment app will have at most 30 input parameters. Although, this scenario did not streamline the parameters the cost of implementation will be cheaper than that of the commonly used instruments such as SRS and ADOS that have 65 and 93 items, respectively. Comparative analysis on the performance of the superior ML model against the empirical scoring algorithm indicated that among the key benefits of implementing the ML model is its outstanding 72% increase in true negatives rate over the 23% recorded with the empirical algorithm. Similarly, implementing the superior ML model of scenario 3 could translate to the same benefits realizable in scenario 1. However, implementing ML models with fewer input parameters translates to a reduction in the cost of the physical gadgets required as well as improvement in the speed of administering the assessment tool. Specifically, implementing the superior models in scenario 2 or scenario 4 will provide an ML-embedded ASD screening app with at most eight input parameters with an overhead of implementing the empirical feature transformation rules. Besides, comparative analysis between the ML models and the empirical scoring algorithm indicated that the best performing ML model in Scenario 2 (i.e. Medium Gaussian SVM) achieved 14% increase in the classification accuracy over the empirical scoring algorithms. Equally, despite having fewer items, the best performing model in scenario 4 outperformed the empirical scoring algorithm with increased accuracy and sensitivity of 15% and 47%, respectively. Another vital recommendation is concerning the present dimensions of the data collection instrument utilized in the present study. Future studies would look at the possibility of redesigning the data collection instrument and improving its scientific robustness as a behavioral scale. Recommendable approaches to categorizing and establishing valid dimensions from CDHSEBA include principal component analysis. Furthermore, future studies would implement the enhanced instruments with more complex and robust algorithms as well as some of the optimization techniques demonstrated in the previous studies. Noteworthy, the visibility in the clinical validity of the proposed approach will provide clinicians with trust on the worthiness of the evaluation metrics recorded.

REFERENCES

Abdelhamid, N., Padmavathy, A., Peebles, D., Thabtah, F., & Goulder-Horobin, D. (2020). Data imbalance in autism pre-diagnosis classification systems: an experimental study. *Journal of Information and Knowledge Management*, *19*(1), 1–16. https://doi.org/10.1142/S0219649220400146

Achenie, L. E. K., Scarpa, A., Factor, R. S., Wang, T., Robins, D. L., & McCrickard, D. S. (2019). A machine learning strategy for autism screening in toddlers. *Journal of Developmental & Behavioral Pediatrics*, *40*(5), 369–376. https://doi.org/10.1097/DBP.0000000000000668

Akter, T., Shahriare Satu, M., Khan, M. I., Ali, M. H., Uddin, S., Lio, P., … Moni, M. A. (2019). Machine learning-based models for early stage detection of autism spectrum disorders. *IEEE Access*, *7*, 166509–166527. https://doi.org/10.1109/ACCESS.2019.2952609

Alahmari, F. (2020). A comparison of resampling techniques for medical data using machine learning. *Journal of Information and Knowledge Management*, *19*(1), 1–13. https://doi.org/10.1142/S021964922040016X

Alcañiz Raya, M., Chicchi Giglioli, I. A., Marín-Morales, J., Higuera-Trujillo, J. L., Olmos, E., Minissi, M. E., … Abad, L. (2020). Application of supervised machine learning for behavioral biomarkers of autism spectrum disorder based on electrodermal activity and virtual reality. *Frontiers in Human Neuroscience*, *14*(April). https://doi.org/10.3389/fnhum.2020.00090

Alhaj, T. A., Siraj, M. M., Zainal, A., Elshoush, H. T., & Elhaj, F. (2016). Feature selection using information gain for improved structural-based alert correlation. *PLOS ONE*, *11*(11), e0166017. https://doi.org/10.1371/journal.pone.0166017

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)* (5th ed.). 5th ed. Washington, DC: American Psychiatric Association.

Baadel, S., Thabtah, F., & Lu, J. (2020). A clustering approach for autistic trait classification. *Informatics for Health and Social Care*, *45*(3), 309–326. https://doi.org/10.1080/17538157.2019.1687482

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., … Dowling, N.

F. (2018). Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. *MMWR. Surveillance Summaries*, *67*(6), 1–23. https://doi.org/10.15585/mmwr.ss6706a1

Bakare, M., & Munir, K. (2011). Autism spectrum disorders in Africa. In M. Mohammad-Reza (Ed.), *A Comprehensive Book on Autism Spectrum Disorders* (1st ed.). Retrieved from http://www.intechopen.com/books/a-comprehensive-book-on-autism-spectrum-disorders/autism-spectrum-disorders-in-africa

Bartolotta, T., & Rizzolo, D. (2019). Recognizing autism spectrum disorder. *Journal of the American Academy of Physician Assistants*, *32*(8), 22–26. https://doi.org/10.1097/01.jaa.0000569776.76198.e1

Batista, G., & Silva, D. F. (2009). How k-nearest neighbor parameters affect its performance. *Argentine Symposium on Artificial Intelligence*, (2009), 95–106.

Becker, M. M., Wagner, M. B., Bosa, C. A., Schmidt, C., Longo, D., Papaleo, C., & Riesgo, R. S. (2012). Translation and validation of autism diagnostic interview-revised (ADI-R) for autism diagnosis in Brazil. *Arquivos de Neuro-Psiquiatria*, *70*(3), 185–190. https://doi.org/10.1590/S0004-282X2012000300006

Bellesheim, K. R., Cole, L., Coury, D. L., Yin, L., Levy, S. E., Guinnee, M. A., … Sohl, K. (2018). Family-driven goals to improve care for children with autism spectrum disorder. *Pediatrics*, *142*(3). https://doi.org/10.1542/peds.2017-3225

Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *57*(8), 927–937. https://doi.org/10.1111/jcpp.12559

Brunner, R. J., & Kim, E. J. (2016). Teaching data science. *Procedia Computer Science*, *80*, 1947–1956. https://doi.org/10.1016/j.procs.2016.05.513

Cambodia, K. (2019). Autism Test (ASD). Retrieved January 5, 2020, from Apps on Google Play website: https://play.google.com/store/apps/details?id=com.kbee.autismtest&hl=enen

Campbell, K., Carpenter, K. L. H., Espinosa, S., Hashemi, J., Qiu, Q., Tepper, M., … Dawson, G.

(2017). Use of a digital modified checklist for autism in toddlers – revised with follow-up to improve quality of screening for autism. *Journal of Pediatrics*, *183*, 133-139.e1. https://doi.org/10.1016/j.jpeds.2017.01.021

Case-Smith, J., Weaver, L. L., & Fristad, M. A. (2015). A systematic review of sensory processing interventions for children with autism spectrum disorders. *Autism*, *19*(2), 133–148. https://doi.org/10.1177/1362361313517762

Cavus, N., Lawan, A. A., Ibrahim, Z., Dahiru, A., Tahir, S., Abdulrazak, U. I., & Hussaini, A. (2021). A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder. *Journal of Personalized Medicine*, *11*(4), 299. https://doi.org/10.3390/jpm11040299

Chan, W., Smith, L. E., Hong, J., Greenberg, J. S., & Mailick, M. R. (2017). Validating the social responsiveness scale for adults with autism. *Autism Research*, *10*(10), 1663–1671. https://doi.org/10.1002/aur.1813

Chauhan, A., Sahu, J., Jaiswal, N., Kumar, K., Agarwal, A., Kaur, J., … Singh, M. (2019). Prevalence of autism spectrum disorder in Indian children: A systematic review and meta-analysis. *Neurology India*, *67*(1), 100. https://doi.org/10.4103/0028-3886.253970

Chojnicka, I., & Pisula, E. (2017). Adaptation and validation of the ADOS-2, Polish version. *Frontiers in Psychology*, *8*(NOV), 1–14. https://doi.org/10.3389/fpsyg.2017.01916

Consurgo. (2015). Autism Test. Retrieved January 5, 2020, from Apps on Google Play website: https://play.google.com/store/apps/details?id=com.consurgo.autismtest&hl=enen

Curtis, A., Smith, T., Ziganshin, B., & Elefteriades, J. (2016). The mystery of the z-score. *AORTA*, *04*(04), 124–130. https://doi.org/10.12945/j.aorta.2016.16.014

Dahiya, A. V., McDonnell, C., DeLucia, E., & Scarpa, A. (2020). A systematic review of remote telehealth assessments for early signs of autism spectrum disorder: Video and mobile applications. *Practice Innovations*, *5*(2), 150–164. https://doi.org/10.1037/pri0000121

Duda, M., Haber, N., Daniels, J., & Wall, D. (2017). Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Translational Psychiatry*, *7*(5), 2–8. https://doi.org/10.1038/tp.2017.86

Duda, M., Kosmicki, J., & Wall, D. (2015). Testing the accuracy of an observation-based classifier for rapid detection of autism risk. *Translational Psychiatry*, *5*(June), e556. https://doi.org/10.1038/tp.2015.51

Duda, M., Ma, R., Haber, N., & Wall, D. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational Psychiatry*, *6*(2), 1–5. https://doi.org/10.1038/tp.2015.221

Durkin, M. S., Elsabbagh, M., Barbaro, J., Gladstone, M., Happe, F., Hoekstra, R. A., … Shih, A. (2015). Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide. *Autism Research*, *8*(5), 473–476. https://doi.org/10.1002/aur.1575

Egger, H. L., Dawson, G., Hashemi, J., Carpenter, K. L. H., Espinosa, S., Campbell, K., … Sapiro, G. (2018). Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. *Npj Digital Medicine*, *1*(1), 20. https://doi.org/10.1038/s41746-018-0024-6

Ertuğrul, Ö. F., & Tağluk, M. E. (2017). A novel version of k nearest neighbor: Dependent nearest neighbor. *Applied Soft Computing*, *55*, 480–490. https://doi.org/10.1016/j.asoc.2017.02.020

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child and Adolescent Psychiatry*, *22*(6), 329–340. https://doi.org/10.1007/s00787-013-0375-0

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., & Tu, X. M. (2014). Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, *26*(2), 105–109. https://doi.org/10.3969/j.issn.1002-0829.2014.02

Fu, C. H. Y., & Costafreda, S. G. (2013). Neuroimaging-based biomarkers in psychiatry: Clinical opportunities of a paradigm shift. *Canadian Journal of Psychiatry*, *58*(9), 499–508. https://doi.org/10.1177/070674371305800904

Ghafouri-Fard, S., Taheri, M., Omrani, M. D., Daaee, A., Mohammad-Rahimi, H., & Kazazi, H. (2019). Application of single-nucleotide polymorphisms in the diagnosis of autism spectrum disorders: a preliminary study with artificial neural networks. *Journal of Molecular*

Neuroscience, *68*(4), 515–521. https://doi.org/10.1007/s12031-019-01311-1

Goel, N., Grover, B., Gupta, D., Khanna, A., & Sharma, M. (2020). Modified grasshopper optimization algorithm for detection of autism spectrum disorder. *Physical Communication*, *41*, 101115. https://doi.org/10.1016/j.phycom.2020.101115

Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., … Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*, *144*(4). https://doi.org/10.1542/peds.2018-3963

Hashemi, J., Dawson, G., Carpenter, K. L. H., Campbell, K., Qiu, Q., Espinosa, S., … Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing*, *3045*(AUGUST), 1–12. https://doi.org/10.1109/TAFFC.2018.2868196

Jack, A. (2018). Neuroimaging in neurodevelopmental disorders: Focus on resting-state fMRI analysis of intrinsic functional brain connectivity. *Current Opinion in Neurology*, *31*(2), 140–148. https://doi.org/10.1097/WCO.0000000000000536

Jadhav, S. D., & Channe, H. P. (2016). Comparative study of K-NN, naive bayes and decision tree classification techniques. *International Journal of Science and Research (IJSR)*, *5*(1), 1842–1845. https://doi.org/10.21275/v5i1.NOV153131

Jiang, L., Cai, Z., Wang, D., & Jiang, S. (2007). Survey of improving K-nearest-neighbor for classification. *Proceedings - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007*, *1*, 679–683. https://doi.org/10.1109/FSKD.2007.552

Kosmicki, J., Sochat, V., Duda, M., & Wall, D. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry*, *5*(2), 1–7. https://doi.org/10.1038/tp.2015.7

Küpper, C., Stroth, S., Wolff, N., Hauck, F., Kliewer, N., Schad-Hansjosten, T., … Roepke, S. (2020). Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-020-61607-w

La Trobe, U. (2017). ASDetect. Retrieved January 5, 2020, from Apps on Google Play website:

https://play.google.com/store/apps/details?id=au.edu.latrobe.asdetect&hl=en

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care and Pain*, *8*(6), 221–223. https://doi.org/10.1093/bjaceaccp/mkn041

Lapteacru, I. (2016). On the consistency of the z-score to measure the bank risk. *SSRN Electronic Journal*, 1–33. https://doi.org/10.2139/ssrn.2787567

Levy, S., Duda, M., Haber, N., & Wall, D. (2017). Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism. *Molecular Autism*, *8*(1), 1–17. https://doi.org/10.1186/s13229-017-0180-6

Liu, G., Li, Y., Zhang, W., & Zhang, L. (2020). A brief review of artificial intelligence applications and algorithms for psychiatric disorders. *Engineering*, *6*(4), 462–467. https://doi.org/10.1016/j.eng.2019.06.008

Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, *9*(8), 888–898. https://doi.org/10.1002/aur.1615

Low, D. M., Bentley, K. H., & Ghosh, S. S. (2020). Automated assessment of psychiatric disorders using speech: A systematic review. *Laryngoscope Investigative Otolaryngology*, *5*(1), 96–116. https://doi.org/10.1002/lio2.354

Lubis, A. R., Lubis, M., & Khowarizmi, A.-. (2020). Optimization of distance formula in K-Nearest Neighbor method. *Bulletin of Electrical Engineering and Informatics*, *9*(1), 326–338. https://doi.org/10.11591/eei.v9i1.1464

Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, *91*, 216–231. https://doi.org/10.1016/j.patcog.2019.02.023

Matson, J. L., & Konst, M. J. (2014). Early intervention for autism: who provides treatment and in what settings. *Research in Autism Spectrum Disorders*, *8*(11), 1585–1590. https://doi.org/10.1016/j.rasd.2014.08.007

Medda, J. E., Cholemkery, H., & Freitag, C. M. (2019). Sensitivity and specificity of the ADOS-2 algorithm in a large german sample. *Journal of Autism and Developmental Disorders*, *49*(2), 750–761. https://doi.org/10.1007/s10803-018-3750-3

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moon, S. J., Hwang, J., Kana, R., Torous, J., & Kim, J. W. (2019). Accuracy of machine learning algorithms for the diagnosis of autism spectrum disorder: Systematic review and meta-analysis of brain magnetic resonance imaging studies. *Journal of Medical Internet Research*, *21*(12). https://doi.org/10.2196/14108

Murphy, C. M., Ellie Wilson, C., Robertson, D. M., Ecker, C., Daly, E. M., Hammond, N., … McAlonan, G. M. (2016). Autism spectrum disorder in adults: diagnosis, management, and health services development. *Neuropsychiatric Disease and Treatment*, *12*, 1669–1686. https://doi.org/10.2147/NDT.S65455

Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *Proceedings - 2011 IEEE Control and System Graduate Research Colloquium, ICSGRC 2011*, 37–42. https://doi.org/10.1109/ICSGRC.2011.5991826

Nazneen, N., Rozga, A., Smith, C. J., Oberleitner, R., Abowd, G. D., & Arriaga, R. I. (2015). A novel system for supporting autism diagnosis using home videos: iterative development and evaluation of system design. *JMIR MHealth and UHealth*, *3*(2), e68. https://doi.org/10.2196/mhealth.4393

Øien, R. A., Candpsych, S. S., Volkmar, F. R., Shic, F., Cicchetti, D. V., Nordahl-Hansen, A., … Chawarska, K. (2018). Clinical features of children with autism who passed 18-month screening. *Pediatrics*, *141*(6). https://doi.org/10.1542/peds.2017-3596

Patra, S., & Arun, P. (2011). Use of Indian scale for assessment of autism in child guidance clinic: An experience. *Indian Journal of Psychological Medicine*, *33*(2), 217. https://doi.org/10.4103/0253-7176.92043

Penner, M., Anagnostou, E., Andoni, L. Y., & Ungar, W. J. (2018). Systematic review of clinical guidance documents for autism spectrum disorder diagnostic assessment in select regions. *Autism*, *22*(5), 517–527. https://doi.org/10.1177/1362361316685879

Peterson, L. (2009). K-nearest neighbor. *Scholarpedia*, *4*(2), 1883. https://doi.org/10.4249/scholarpedia.1883

Plitt, M., Barnes, K. A., & Martin, A. (2015). Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage: Clinical*, *7*, 359–366. https://doi.org/10.1016/j.nicl.2014.12.013

Poyarkov, A., Drutsa, A., Khalyavin, A., Gusev, G., & Serdyukov, P. (2016). Boosted decision tree regression adjustment for variance reduction in online controlled experiments. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 235–244. https://doi.org/10.1145/2939672.2939688

Pradhan, A. (2012). Support Vector Machine- a survey. *International Journal of Emerging Technology and Advanced Engineering*, *2*(8), 82–85.

Pratama, T. G., Hartanto, R., & Setiawan, N. A. (2019). Machine learning algorithm for improving performance on 3 AQ-screening classification. *Communications in Science and Technology*, *4*(2), 44–49. https://doi.org/10.21924/cst.4.2.2019.118

Puerto, E., Aguilar, J., López, C., & Chávez, D. (2019). Using multilayer fuzzy cognitive maps to diagnose autism spectrum disorder. *Applied Soft Computing Journal*, *75*, 58–71. https://doi.org/10.1016/j.asoc.2018.10.034

Rish, I. (2001). An empirical study of the naive bayes classifier. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Vol. 3, pp. 41–46.

Roobaert, D., Karakoulas, G., & Chawla, N. V. (2006). Information gain, correlation and support vector machines. In *Feature Extraction* (Vol. 207, pp. 463–470). https://doi.org/10.1007/978-3-540-35488-8_23

Rudra, A., Banerjee, S., Singhal, N., Barua, M., Mukerji, S., & Chakrabarti, B. (2014). Translation and usability of autism screening and diagnostic tools for autism spectrum conditions in India. *Autism Research*, *7*(5), 598–607. https://doi.org/10.1002/aur.1404

Ruparelia, K., Abubakar, A., Badoe, E., Bakare, M., Visser, K., Chugani, D. C., … Newton, C. R. (2016). *Autism Spectrum Disorders in Africa : Current Challenges in Identification , Assessment , and Treatment : A Report on the International Child Neurology Association Meeting on ASD in Africa*. *31*(8), 1018–1026. https://doi.org/10.1177/0883073816635748

Sarabadani, S., Schudlo, L. C., Samadani, A. A., & Kushski, A. (2020). Physiological detection of affective states in children with autism spectrum disorder. *IEEE Transactions on Affective Computing*, *11*(4), 588–600. https://doi.org/10.1109/TAFFC.2018.2820049

Schein, A. I., & Ungar, L. H. (2007). Active learning for logistic regression: an evaluation. *Machine Learning*, *68*(3), 235–265. https://doi.org/10.1007/s10994-007-5019-5

Sekaran, K., & Sudha, M. (2021). Predicting autism spectrum disorder from associative genetic markers of phenotypic groups using machine learning. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), 3257–3270. https://doi.org/10.1007/s12652-020-02155-z

Shahamiri, S. R., & Thabtah, F. (2020). Autism AI: a new autism screening system based on artificial intelligence. *Cognitive Computation*, *12*(4), 766–777. https://doi.org/10.1007/s12559-020-09743-3

Singh, G., Kumar, B., Gaur, L., & Tyagi, A. (2019). Comparison between multinomial and bernoulli naïve bayes for text classification. *2019 International Conference on Automation, Computational and Technology Management, ICACTM 2019*, 593–596. https://doi.org/10.1109/ICACTM.2019.8776800

Song, D.-Y., Kim, S. Y., Bong, G., Kim, J. M., & Yoo, H. J. (2019). The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: a literature review. *Journal of the Korean Academy of Child and Adolescent Psychiatry*, *30*(4), 145–152. https://doi.org/10.5765/jkacap.190027

Su, J., & Zhang, H. (2006). A fast decision tree learning algorithm. *Proceedings of the National Conference on Artificial Intelligence*, *1*(Quinlan 1993), 500–505.

Surén, P., Saasen-Havdahl, A., Bresnahan, M., Hirtz, D., Hornig, M., Lord, C., … Stoltenberg, C. (2019). Sensitivity and specificity of early screening for autism. *BJPsych Open*, *5*(3), 1–8. https://doi.org/10.1192/bjo.2019.34

Suresh Kumar, R., & Renugadevi, M. (2019). Differential evolution tuned support vector machine for autistic spectrum disorder diagnosis. *International Journal of Recent Technology and Engineering*, *8*(2), 3861–3870. https://doi.org/10.35940/ijrte.B3063.078219

Susmaga, R. (2004). Confusion matrix visualization. In *Intelligent Information Processing and Web Mining* (pp. 107–116). https://doi.org/10.1007/978-3-540-39985-8_12

Suthaharan, S. (2016). Machine learning models and algorithms for big data classification. In *IEEJ Trans. EIS*. https://doi.org/10.1007/978-1-4899-7641-3

Thabtah, F. (2018). Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*, *00*(00), 1–20. https://doi.org/10.1080/17538157.2017.1399132

Thabtah, F. (2019). An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Informatics Journal*, *25*(4), 1739–1755. https://doi.org/10.1177/1460458218796636

Thabtah, F., Abdelhamid, N., & Peebles, D. (2019). A machine learning autism classification based on logistic regression analysis. *Health Information Science and Systems*, *7*(1). https://doi.org/10.1007/s13755-019-0073-5

Thabtah, F., Kamalov, F., & Rajab, K. (2018). A new computational intelligence approach to detect autistic features for autism screening. *International Journal of Medical Informatics*, *117*(June), 112–124. https://doi.org/10.1016/j.ijmedinf.2018.06.009

Thabtah, F., & Peebles, D. (2020). A new machine learning model based on induction of rules for autism detection. *Health Informatics Journal*, *26*(1), 264–286. https://doi.org/10.1177/1460458218824711

Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Tollet, T. (2019). Autism Quiz: AQ-10 quotient test. Retrieved January 5, 2020, from Apps on Google Play website: https://play.google.com/store/apps/details?id=com.theo.aq10

Torres, E. B., Rai, R., Mistry, S., & Gupta, B. (2020). Hidden aspects of the research ADOS are bound to affect autism science. *Neural Computation*, *32*(3), 515–561. https://doi.org/10.1162/neco_a_01263

Trevethan, R. (2017). Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Frontiers in Public Health*, *5*(11), 1–7. https://doi.org/10.3389/fpubh.2017.00307

Usta, M. B., Karabekiroglu, K., Sahin, B., Aydin, M., Bozkurt, A., Karaosman, T., … Ürer, E. (2019). Use of machine learning methods in prediction of short-term outcome in autism spectrum disorders. *Psychiatry and Clinical Psychopharmacology*, *29*(3), 320–325. https://doi.org/10.1080/24750573.2018.1545334

Wall, D., Dally, R., Luyster, R., Jung, J. Y., & DeLuca, T. (2012). Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS ONE*, *7*(8). https://doi.org/10.1371/journal.pone.0043855

Wall, D., Kosmicki, J., Deluca, T., Harstad, E., & Fusaro, V. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*, *2*(4), e100-8. https://doi.org/10.1038/tp.2012.10

Wang, H., Shi, Y., Zhou, X., Zhou, Q., Shao, S., & Bouguettaya, A. (2010). Web service classification using support vector machine. *Proceedings - International Conference on Tools with Artificial Intelligence*, *ICTAI*, *1*, 3–6. https://doi.org/10.1109/ICTAI.2010.9

Wang, T., Qin, Z., Jin, Z., & Zhang, S. (2010). Handling over-fitting in test cost-sensitive decision tree learning by feature selection, smoothing and pruning. *Journal of Systems and Software*, *83*(7), 1137–1147. https://doi.org/10.1016/j.jss.2010.01.002

Ward, S. L., Sullivan, K. A., & Gilmore, L. (2018). Combining parent and clinician ratings of behavioural indicators of autism spectrum disorder improves diagnostic classification. *Early Child Development and Care*, *188*(6), 748–758. https://doi.org/10.1080/03004430.2016.1235567

Webb, G. I. (2016). Naïve Bayes. In *Encyclopedia of Machine Learning and Data Mining* (pp. 1–2). https://doi.org/10.1007/978-1-4899-7502-7_581-1

Wiesen, J. (2006). Benefits, drawbacks, and pitfalls of z-score weighting. *30th Annual IPMAAC*

Conference, 1–41. Las Vegas.

Wingfield, B., Miller, S., Yogarajah, P., Kerr, D., Gardiner, B., Seneviratne, S., … Coleman, S. (2020). A predictive model for paediatric autism screening. *Health Informatics Journal*. https://doi.org/10.1177/1460458219887823

Xie, X. (2019). CHAT23-autism screening scale. Retrieved January 5, 2020, from App Store website: https://apps.apple.com/us/app/chat23-autism-screening-scale/id1195518734

Xu, S. (2018). Bayesian naïve bayes classifiers to text classification. *Journal of Information Science*, *44*(1), 48–59. https://doi.org/10.1177/0165551516677946

Yuen, T., Carter, M. T., Szatmari, P., & Ungar, W. J. (2018). Cost-Effectiveness of universal or high-risk screening compared to surveillance monitoring in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *48*(9), 2968–2979. https://doi.org/10.1007/s10803-018-3571-4

Zhang, X., Qin, F., Chen, Z., Gao, L., Qiu, G., & Lu, S. (2020). Fast screening for children's developmental language disorders via comprehensive speech ability evaluation—using a novel deep learning framework. *Annals of Translational Medicine*, *8*(11), 707–707. https://doi.org/10.21037/atm-19-3097

Zhou, X., Zhang, X., & Wang, B. (2016). Online support vector machine: a survey. *Advances in Intelligent Systems and Computing*, *382*, 269–278. https://doi.org/10.1007/978-3-662-47926-1_26

APPENDICES

APPENDIX I

PERMISSION TO USE QUESTIONNAIRE

Childhood Neuropsychiatric Disorders Initiatives (CNDI)

E46, Goshen Estate, Premier Layout,
Enugu, Enugu State, Nigeria.

CNDI

Department of Computer Information Systems,

Near East University,

Near East Boulevard,

ZIP: 99138 Nicosia

TRNC Mersin 10 – Turkey

Dear Abdulmalik Ahmad Lawan,

RE: CHILD DEVELOPMENT QUESTIONNAIRE FOR HOUSEHOLD SURVEY TO ESTIMATE BURDEN OF ASD

Thank you for your interest in the above-named questionnaire.

You are permitted to use this questionnaire for your study as requested in as much as the source of paper questionnaire you adapted to electronic version is duly acknowledged.

Kindly share with us the final findings of your study after completing it.

Thank you once again.

With Best Wishes,

Dr. Muideeen Owolabi Bakare

Chairman & Chief Consultant Psychiatrist

Address – No. E46, Goshen Estate, Premier Layout, Enugu, Enugu State, Nigeria

Email – cndinigeria@gmail.com; mobakare2000@yahoo.com

Phone Numbers - +2347030970079 or +2348052210933

Mailing Address:
c/o Dr. Muideen O. Bakare
Child & Adolescent Unit, Federal
Neuropsychiatric Hospital, New Haven,
Enugu, Enugu State, Nigeria

Website:
www.cndinitiatives.org
Tel:
+2347030970079, +2348052210933.
Email:
cndinigeria@gmail.com

APPENDIX II

ETHICS APPROVAL

YAKIN DOĞU ÜNİVERSİTESİ

BİLİMSEL ARAŞTIRMALAR ETİK KURULU

15.01.2020

Dear Abdulmalik Ahmad Lawan

Your application titled **"ICT-based Screening for Autism SpectrumDisorder"** with the application number YDÜ/FB/2020/82 has been evaluated by the Scientific Research Ethics Committee and granted approval. You can start your research on the condition that you will abide by the information provided in your application form.

Assoc. Prof. Dr. Direnç Kanol

Rapporteur of the Scientific Research Ethics Committee

Direnc Kanol

Note: If you need to provide an official letter to an institution with the signature of the Head of NEU Scientific Research Ethics Committee, please apply to the secretariat of the ethics committee by showing this documen

APPENDIX III

CDHSEBA

CHILD DEVELOPMENT FOR HOUSEHOLD SURVEY TO ESTIMATE BURDEN OF AUTISM SPECTRUM DISORDER (CDHSEBA)

Please, describe the child's behavior based on the following questions. Your response will remain confidential and will only be used for research purpose.

Thank you.

Abdulmalik Ahmad Lawan

aaltofa2000@gmail.com

1. Your relationship with the child

   a. Daughter/Son b. Brother/Sister c. Neighbor d. Cousin e. Others ………………….

2. About the child

   a. Gender: Male Female

   b. Is the child between 2 and 18 years old today? Yes No

3. Going back to the first 3 years of the child's life, was there anything that seriously worried you or anyone else about his ...	Yes	No
a) Language and communication development?	Yes	No
b) Relationship with peers?	Yes	No
c) Gross Motor Development and use of hands and limbs?	Yes	No
d) Odd or repetitive behaviour?	Yes	No
e) Ability to learn and do new things – things such as puzzles or helping get dressed?	Yes	No

A1: DEFICITS IN SOCIO-EMOTIONAL RECIPROCITY

4. The child does not speak at all (he or she can't make himself or herself understood in words; he or she can't say any recognizable words)?	Yes	No
5. The child does not speak normally for his/her age?	Yes	No
8. The child cannot communicate with you by using gestures? E.g. pointing with the index finger, nodding/shaking head for yes/no etc.	Yes	No
9. The child does not initiate a conversation with you?	Yes	No

A2: DEFICITS IN NON-VERBAL COMMUNICATION

10. The child does not smile back when you smile at him/her?	Yes	No
11. The child does not maintain eye contact when talking to people?	Yes	No
12. The child does not show the typical range of facial expressions? E.g. he/she doesn't smile when happy? He/she doesn't show sadness when unhappy? He/she doesn't express surprise when something unexpected happens?	Yes	No

A3: DEFICITS IN DEVELOPING, MAINTAINING AND UNDERSTANDING RELATIONSHIPS

13. The child does not participate in imaginative role-playing with other kids interactively. Like (cooking play)/dolls/clay/telephone/toy gun/motor car OR 'teacher-student', 'thief-police', 'mother-child', etc.	Yes	No
14. Does the child appear to be in his/her own world, no matter what he/she is doing (even when with other children)?	Yes	No
15. Does the child prefer to play alone rather than joining his peers?	Yes	No

B1: STEREOTYPED MOVEMENTS, LANGUAGE OR USE OF SPEECH

6. Does the child often repeat the same word or phrase repeatedly in the same manner?	Yes	No
7. Does the child repeat what you say? Copy your speech or the speech of others.	Yes	No
16. Does the child have interests that are not typical for children his or her age, like an interest in objects like fans, light switches, radios, etc.?	Yes	No
17. Does the child have any repetitive behavior? For instance, arranging toys or household objects in a specific manner repeatedly.	Yes	No
18. Does the child keep on repeating any of the following?		
a) Flapping hands (moving hands up and down)	Yes	No
b) Hand wringing (as if squeezing clothes)	Yes	No
c) Toe-walking (walking on tip-toe)	Yes	No
d) Swinging or spinning his/her body	Yes	No
e) Making unusual finger or hand movements near his/her face.	Yes	No

B2: INSISTENCE ON SAMENESS AND INFLEXIBILITY THOUGHT

19. Does the child insist on sameness and actively resist any change in his/ her routines? For example, insisting on the same dress/asking for the same place to sit while eating/insisting on no change in the arrangement of the toys or household items. Bathing or getting dressed at certain time and when unable to do so for some particular reason, does get very upset?	Yes	No

B3: HIGHLY RESTRICTED, FIXATED INTERESTS AND ABNORMAL INTENSITY IN FOCUS

20. Has the child memorized unusual facts like schedules, history facts, or other sorts of facts that preoccupy him or her daily?	Yes	No
21. Is the child 'too obsessed' with certain activities or interests beyond what you would expect for a child of his/her age?	Yes	No
22. Does the child have excessive interest in odd or unusual things/activities that other children do not have? E.g. collecting sweet wrappers, nylon bags, piece of rope, pulling thread and rubber band etc.	Yes	No
23. Does the child prefer to play with a particular part of a toy/object rather than the whole toy/object? For example, when playing with a toy car, only want to play with the tyres and not the rest of the car.	Yes	No

B4: SENSITIVITY TO SENSORY INPUT

24A. Does the child do anything to hurt or harm him/herself? E.g. banging his/hers head on objects, biting him/herself, piercing him/herself with sharp objects?	Yes	No
24B. Is the child hypersensitive or under sensitive to certain sensory inputs? i.e. is indifferent to pain? Overly upset by certain sounds or too sensitive to light?	Yes	No
25. Does the child show an unusual interest in certain sensory aspects of the environment? E.g. excessive touching or smelling of objects?	Yes	No

26. If the child presents with any difficulties in the areas we have mentioned above please tell us the extent to which these difficulties interfere with their day-to-day functioning at home and in school

a) Not all

b) Causes minor interferences

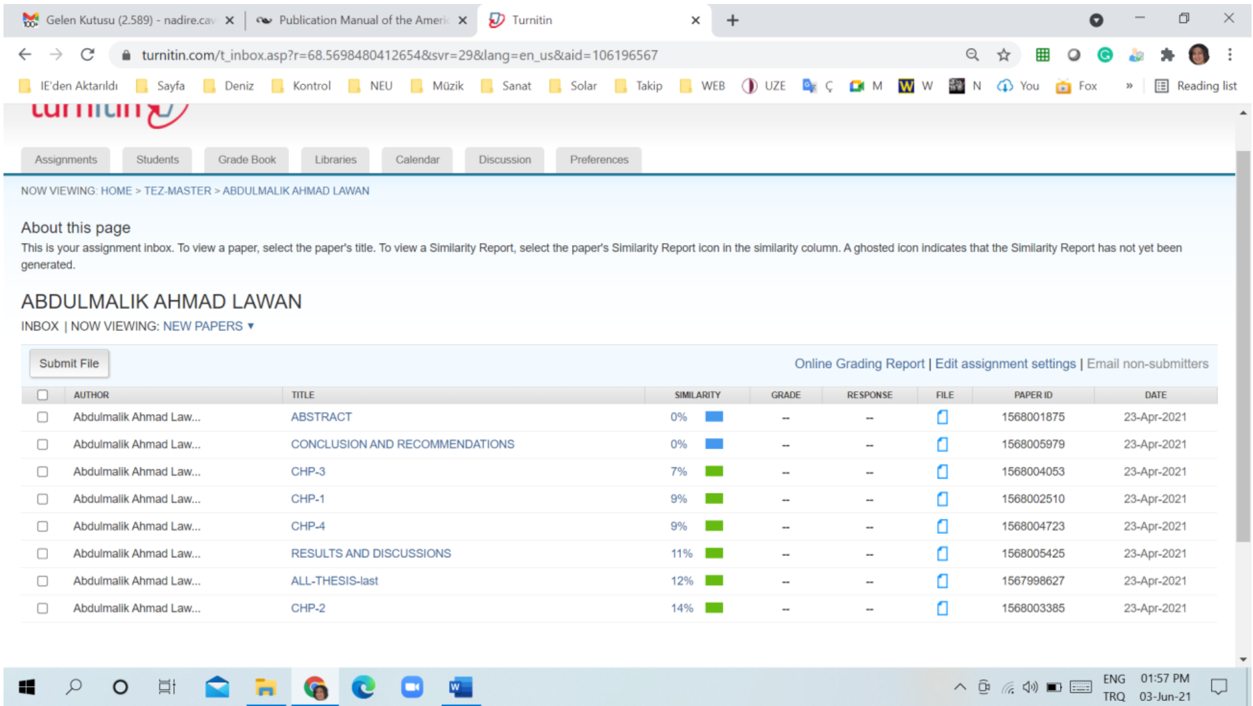
c) Causes major interferences

d) Symptoms described above makes it impossible for the child to function in the above settings.

27. Is the child previously diagnosed with ASD?	Yes	No

APPENDIX IV

SIMILARITY REPORT



NOW VIEWING: HOME > TEZ-MASTER > ABDULMALIK AHMAD LAWAN

About this page

This is your assignment inbox. To view a paper, select the paper's title. To view a Similarity Report, select the paper's Similarity Report icon in the similarity column. A ghosted icon indicates that the Similarity Report has not yet been generated.

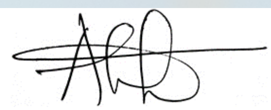
ABDULMALIK AHMAD LAWAN

INBOX | NOW VIEWING: NEW PAPERS ▾

Submit File

Online Grading Report | Edit assignment settings | Email non-submitters

AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
Abdulmalik Ahmad Law...	ABSTRACT	0%	--	--		1568001875	23-Apr-2021
Abdulmalik Ahmad Law...	CONCLUSION AND RECOMMENDATIONS	0%	--	--		1568005979	23-Apr-2021
Abdulmalik Ahmad Law...	CHP-3	7%	--	--		1568004053	23-Apr-2021
Abdulmalik Ahmad Law...	CHP-1	9%	--	--		1568002510	23-Apr-2021
Abdulmalik Ahmad Law...	CHP-4	9%	--	--		1568004723	23-Apr-2021
Abdulmalik Ahmad Law...	RESULTS AND DISCUSSIONS	11%	--	--		1568005425	23-Apr-2021
Abdulmalik Ahmad Law...	ALL-THESIS-last	12%	--	--		1567998627	23-Apr-2021
Abdulmalik Ahmad Law...	CHP-2	14%	--	--		1568003385	23-Apr-2021



Abdulmalik Ahmed Lawan

APPENDIX V

CURRICULUM VITAE

1. PERSONAL DATA

Name: Abdulmalik Ahmad Lawan

Date of Birth: 15TH October, 1989

Place of Birth: Hotoro

Residence: 449 Salen Zana Link, Hotoro

State of Origin: Kano State

Nationality: Nigerian

Marital Status: Married

Gender: Male

Languages Spoken: Hausa, English, Arabic, Russian, Turkish

Contact Address: Department of Computer Science, KUST **Tel:** +2347066498622

Email: aaltofa2000@gmail.com || aalawan@kustwudil.edu.ng

Next of Kin: Abdulaziz Ahmad Lawan

2. AREA OF SPECIALIZATION

Human Computer Interaction

3. INSTITUTIONS ATTENDED

2018-Date- Near East University, Cyprus

Qualification: PhD. Computer Information Systems (**In View**)

2015- Near East University, Cyprus

Qualification: MSc. Computer Information Systems **(First class** honor division**)**

2011- Kano University of Science and Technology, Wudil

Qualification: BSc. Computer Science **(First class** honor division**)**

2009- Gateway Computer Training Institute, Kano

Qualification: Certificate in Information Technology (CIT)

2007- Science College Dawakin Kudu, Kano

Qualification: Senior School Certificate Examination (SSCE)

2004- Government Day Sec. Sch. Hotoro North, Kano

Qualification: Junior School Certificate Examination (JSCE)

2001- Hotoro South Special Primary School, Kano state

Qualification: Primary School Certificate

4. ACADEMIC CAREER

2015-Date – Department of Computer Science

Kano University of Science and Technology, Wudil, **Position:** Assistant Lecturer

2013-2014 – Khadija Memorial College, Kano

No. 50 CBN Quarters Hotoro GRA, **Position:** Classroom Teacher

2013 – National Youth Service Corps

Government College Birnin Kudu, Jigawa State, **Position:** Classroom Teacher

5. COURSES TAUGHT AT UNIVERSITY

1. Introduction to Computer Science
2. Introduction to Web Design
3. Special Topics in Computer Science (Research Methodology)
4. Human Computer Interface
5. Data Structures and Algorithm (Summer Semester)
6. Coding System and Information Theory (Summer Semester)
7. System Modelling and Simulation (Summer Semester)

6. PROJECT SUPERVISION

1. Secured Organizational Messaging System: An Implementation of Private and Public Key Cryptography - A Case Study of Federal Medical Centre (FMC) Azare, Bauchi State - Nigeria
2. Java Implementation of Machine Translation using Transfer Approach - A Case Study of English-Igala Noun Phrases
3. Computer Base Test System, Population Information System
4. Online Commodity Sales Application System (a case study of multipurpose cooperative society KUST, Wudil)

5. Design and Implementation of Online Scholarship Application and Screening System (a case study of Jigawa state scholarship board)
6. Design and Implementation of NYSC Reporting System Based on ISO10002 (a case study of directorate of salary and pension, Jigawa state)
7. Design and Implementation of E-Learning System (a case study of faculty of computing and mathematical science)
8. A New Watermarking Algorithm Based on Discrete Wavelet-Chirp Z Transform-Singular Value Decomposition-QR Decomposition
9. Staff Induction Training System (a case study of MTN, Nigeria) (etc.)

7. PUBLICATIONS

7.1 Journal articles

Cavus, N., **Lawan, A. A.**, Ibrahim, Z., Dahiru, A., Tahir, S., Abdulrazak, U. I., & Hussaini, A. (2021). A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder. *Journal of Personalized Medicine, 11*(4), 299. https://doi.org/10.3390/jpm11040299

Cavus, N., Sani, A. S., Haruna, Y., & **Lawan, A. A.** (2021). Efficacy of social networking sites for sustainable education in the era of COVID-19: A systematic review. *Sustainability*, *13*(2). https://doi.org/10.3390/su13020808

Shehu, N. S., Tahir, S., Abdullahi, D., & **Lawan, A. A.** (2021). Examining the differences in autism quotient scores based on ethnicity. *Malaysian Journal of Medical Research (MJMR), 5*(1), 1-5.

Abdulkadir, M., **Lawan, A. A.**, & Mamman, S. (2021). Peer-to-peer approach for distributed privacy-preserving deep learning. *International Journal of Computer, 40*(1), 91–108.

Ahmad, A. L., & Cavus, N. (2018). The efficacy of employing social media for educational practice in the unrest regions of the world. *Journal of Learning and Teaching in Digital Age, 3*(2), 22-26.

Ahmad, A. L., & Cavus, N. (2019). Motives behind preference of internet communication tools among university students. *Journal of Learning and Teaching in Digital Age, 4*(1), 41-45.

7.2 Book/ Book Chapters

Zakari, A., **Lawan, A. A.**, & Bekaroo, G. (2017). Towards improving the security of low-interaction honeypots: Insights from a comparative analysis. *In Lecture Notes in Electrical Engineering* (Vol. 416). https://doi.org/10.1007/978-3-319-52171-8_28

Zakari, Abubakar, **Lawan, A. A.**, & Bekaroo, G. (2017). A Hybrid Three-Phased Approach in Requirement Elicitation. *In Lecture Notes in Electrical Engineering* (vol. 416, pp. 331–340). https://doi.org/10.1007/978-3-319-52171-8_30

Ahmad, A. L., & Hussain, A. (2018). General Computer Operations for Introductory Computer Labs (1st ed.). Mauritius: Scholar's Press.

7.3 Conference Presentations

Lawan, A. A., Abdi, A. S., Abuhassan, A. A., & Khalid, M. S. (2019). what is difficult in learning programming language based on problem-solving skills? *In Proceedings of the International Conference on Advanced Science and Engineering* (pp. 18-22). IEEE. https://doi.org/10.1109/ICOASE.2019.8723740

Ahmad, A. L., & Cavus, N. (2015). Efficacy of employing social media for educational practices in the unrest regions of the world. *Paper presented at the Cyprus International Conference on Educational Research*, 19-21 March 2015 Girne American University, Cyprus.

Ahmad, A. L., & Cavus, N. (2015).Motives behind preference of internet communication tools among university students. *Paper presented at the International Computer & Instructional Technologies Symposium*, 23-24 May 2015, Sandıklı-Afyonkarahisar, Turkey.

7.4 Workshop, Seminar Participations and others

2015- Participation in Model Checking Contest with a Model of "Internet Open Trading Protocol (IOTP model) using Petri Net" (published at Model Checking Contest MCC2015, France)

2016- Departmental Seminar presentation on "What is & What is not Big Data??" at the Department of Computer Science, KUST

2011- Conference attended on North-West ICT4NEED at KUST, Wudil

2011- Conference attended on **"**Developing Our Nation's Economy through Information Technology (DONEIT)" at Federal University of Technology Minna

2010- Computer Operation during "Student Industrial Working Experience Scheme (SIWES)" at Abubakar Rimi TV (ARTV) Kano.

2015- Researched MSc. Thesis on "Investigating Computer Technology Acceptance and Readiness of Students: A Case Study in Northwestern Nigeria"

2011- Researched BSc. Project on "Implementation of Computer Security using Cryptography"

8. SPECIAL TRAINING AND AWARD

2018- Training on "Life-Long Learning Skills for Academic Development" at Near East University, Nicosia Cyprus

2012- Training on "School Records and Classroom Management" at Government College Birnin Kudu, Jigawa State.

2012- Extensive Training on "Assessment of the Core Competence in Employability Service Sector (ACCESS)" at NestTrust Nigeria, Kano.

2003- Trained Peer Educator during HIV/AIDS Peer Educators workshop at Government Day Secondary School Hotoro North, Kano.

2012- Award of Best Computer Graduating Student 2011/2012, KUST, Wudil

2010- Award of 1st Position at Nigerian Computer Society (NCS): Bi-Annual Olatunji Odegbami Competition.

2011- Award of 1st Position at North-West Zone Ict4needs: Computer Programming Contest

9. ADMINISTRATIVE DUTIES

2018- Departmental Examinations Officer, Computer Science, KUST

2016/2017- Students Advisor/ Level Coordinator, Computer Science, KUST

2018- Secretary Departmental A&PC, Computer Science, KUST

2018- Departmental Financial Secretary, Computer Science, KUST

2016- Preparation of proposal for mounting MSc Computer Science, KUST

2016- Preparation of proposal for mounting BSc Information Technology, KUST

2016- Member Undergraduate Project Standardization Committee, Computer Science

2016- Member Undergraduate Curriculum Review Committee, Computer Science, KUST

2016- Member Departmental Committee for Research and Publications, KUST

2016- Preparation of Lab Manual for CSC1301: Introduction to Computer Science **(etc.)**

10. COMMUNITY SERVICES

2008-2010 Voluntary Physics/ Biology Teacher at GGSS Hotoro South, Kano State

2015- Voluntary Basic Science Teacher at GJSS Hotoro Walawai, Kano State

2013-2018 Math Teacher at Attarbiyya International College during annual 3-month SSCE extra lesson organized by Almajlisul Islami Hotoro

2008- To-date Member in many community-based development committees, Hotoro **(etc.)**

11. SKILLS

- Skillful in computer operations, internet and software development.
- Proficient in programming with C/C++ and Java etc.
- Skillful in working with software packages esp. SPSS, MATLAB etc.
- Ability to work under pressure and with little supervision.
- Good cognitive, communication and interpersonal skills.

12. REFEREES

Prof. Dr. Abubakar Musa

The Deputy Vice Chancellor,

Kano University of Science and Technology, Wudil.

Email: ammaikaratu@kustwudil.edu.ng

Prof. Dr. Nadire Cavus

The Director Graduate School of Applied Sciences,

Near East University, Cyprus.

Email: nadire.cavus@neu.edu.tr