Chapter One

Introduction

1.1 Background to the Study

Autism, or autism spectrum disorder (ASD), is a developmental disorder that presents in early childhood[1]. It has the potential to impact an individual's social, communication, relationship, and self-regulation capabilities[2]. Typically, the onset of this condition occurs within the initial three years and persists throughout the individual's lifespan. The aetiology of autism remains uncertain, although it is believed to be influenced by genetic, environmental, and prenatal factors such as stress, inflammation, toxins, alcohol, air pollution, autoimmune disorders, the child's emotional state, and environmental factors during the first two years of life[3].

Autism is classified into four levels, namely mild, moderate, severe, and very severe. Despite the increasing prevalence of autism among children, there remains a significant gap in research pertaining to the aetiology and impact of this condition[4]. Autism is characterized by challenges in communication within both the remote and local networks of the brain. These networks have an impact on the development of the brain, and the associated illness becomes apparent during early childhood, with diagnoses typically occurring between 24 and 36 months[5].

Individuals with autism spectrum disorder exhibit unique and diverse behaviors, however, they all have an effect on social cohesion. Some children with autism may exhibit a preference for solitude or distance even when in the company of their family. Additionally, repetitive behaviors such as frequent hand-waving and other unusual gestures may be observed[5].

In some cases, these children may experience sudden regression in their language and communication skills. This condition is the result of various factors, such as social, health,

psychological, genetic, and environmental influences on the developing brain[6]. Autism presents itself in diverse manners. The patient exhibits a high degree of shyness in contrast to their peers and expresses a desire for isolation[14]. Individuals experience feelings of hopelessness and a lack of leadership due to mental abnormalities in their brain. In addition, they exhibit outbursts of emotional distress and may inadvertently harm themselves, their siblings, or their peers without any discernible provocation. Performance ratings tend to be higher than verbal ratings[4].

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that can impact an individual's language, communication, cognitive abilities, and social behaviour[7]. This condition is categorized as a psychiatric disorder and is characterized by atypical sensory perception as a fundamental trait. Typical features of autism spectrum disorder (ASD) include enduring difficulties with social interactions and stereotyped patterns of behavior, as well as repetitive actions and thoughts. Additionally, individuals with ASD may exhibit imaginative thinking, which is often accompanied by a gradual deterioration in their ability to communicate effectively[8].

According to data from the World Health Organization, approximately 1 in 160 children is diagnosed with autism. Furthermore, the prevalence of ASD is rapidly increasing, with 0.63% of children being diagnosed with ASD in 2022[9]. The symptoms of autism spectrum disorder (ASD) typically manifest within the initial five years of an individual's life. Autism Spectrum Disorder (ASD) is a condition that typically emerges in childhood and can persist into adulthood and adolescence. Autism Spectrum Disorder (ASD) is a significant neurodevelopmental condition that can incur substantial costs for treatment and management.

The National Health Service (NHS) has reported a 39% rise in the suspected prevalence of autism spectrum disorder (ASD) during the period of April 2021 to March 2022[9]. The initial

documentation of the prevalence of autism spectrum disorder (ASD) in Nigeria was a quasi-epidemiological study conducted in 1978. This study focused on children with intellectual disability and included five other sub-Saharan African countries, namely Ghana, Kenya, Zimbabwe, Zambia, and South Africa[10]. A study conducted in sub-Saharan African countries revealed a prevalence of ASD of approximately 1 in 145 among children with intellectual disabilities. A prevalence rate of 0.8% for ASD was recorded among 393 children attending a mental hospital clinic in Southeastern Nigeria over a one-year period[10].

The recorded prevalence of Autism Spectrum Disorder (ASD) among 44 children with intellectual disabilities in a privately-owned special school for children with neurodevelopmental disabilities in Southeastern Nigeria was 11.4%[11]. Furthermore, a study conducted on a community-based sample of 85 children with neurodevelopmental issues in Lagos State, Southwestern Nigeria, revealed a prevalence rate of 34.5% for ASD[10].

While there are presently no pharmaceutical interventions available to manage this condition, professionals have identified alternative methods to alleviate or address it. The following techniques are among those that have been identified:

i. Functional Analysis: This approach entails assessing the child's environment to identify the antecedents and outcomes of the child's conduct. A professional conducts interviews with their parents and maintains records of their living environment. These interviews serve to identify the factors that impact individuals and subsequently address them accordingly[12].

ii. Selecting Target Behaviors: This approach involves the identification of unique behaviors exhibited by individuals, including but not limited to linguistic, auditory, or aggressive tendencies. The professional identifies the specific behaviors that significantly impact the child with autism and utilizes specialized interventions to modify these behaviors.

iii. Teaching Procedures: In this approach, a professional expert educates the child's behaviors and subsequently determines various methods or procedures to be employed for their treatment[13].

iv. Machine Learning Classification: This involves the use of machine learning algorithms to predict and classify autism disorder early.

Individuals diagnosed with autism spectrum disorder (ASD) may exhibit exceptional intellectual and artistic abilities. The presence of these exceptions has posed significant challenges in the clinical assessment and diagnosis process. Recently, there has been an increase in social and computational research conducted on ASD. In the field of computation, deep learning and machine learning algorithms have been widely adopted due to their ability to enhance diagnostic precision while simultaneously reducing diagnostic costs and timeframes.

Machine learning is a subfield of Artificial Intelligence (AI) that involves the development of computer algorithms capable of accessing data and using it to improve the performance without being explicitly programmed. The main goal is to enable automated machine learning without requiring human intervention. Furthermore, it enables the computers to adapt their actions accordingly. The process of learning begins with the observation of data, wherein the program is provided with examples, direct experience, and instructions to enable the machine to identify patterns in data and make informed decisions[14].

The field of Machine Learning utilizes various algorithms to address data-related challenges. Data scientists often emphasize that there is no universally applicable algorithm that can effectively address every problem. The selection of an appropriate algorithm is contingent upon the nature of the problem to be solved, the quantity of variables involved, and the most suitable model for the task at hand. There exist four primary classifications of machine

learning techniques: Supervised Learning, Unsupervised Learning, Semi-supervised Learning, and Reinforcement Learning[15,16].

Supervised learning is a machine learning technique that involves learning a function that can map input data to output data based on a set of example input-output pairs[17]. The process entails deducing a function from labelled training data, which comprises a collection of training examples. Supervised machine learning algorithms require external assistance in order to operate effectively. The dataset is partitioned into training and testing sets[17]. The training dataset contains an output variable that requires prediction or classification. Algorithms typically acquire patterns from the training dataset and subsequently utilize them to predict or classify data in the test dataset. Supervised learning algorithms can be classified into two distinct categories, namely classification and regression[18].

In contrast to supervised learning, unsupervised learning does not involve correct answers or a teacher[19]. Instead, algorithms are tasked with independently discovering and presenting the underlying structure within the data. Unsupervised learning algorithms typically extract a limited number of features from the data. Upon the introduction of new data, the system utilizes the previously acquired features to classify the data into its respective class. This technique is primarily utilizing for the purposes of clustering and feature reduction. Unsupervised learning is characterized by the absence of a pre-defined outcome. The system endeavors to extract valuable insights from vast quantities of data. The algorithms can be classified into two categories: clustering and association[20].

Semi-supervised machine learning is a hybrid approach that integrates both supervised and unsupervised machine learning techniques[21]. In the fields of machine learning and data mining, utilizing existing unlabeled data can prove to be advantageous in situations where obtaining labelled data is a laborious task. Supervised machine learning methods typically

involve training machine learning algorithms on a labelled dataset, where each record contains the corresponding outcome information. Some of the algorithms used in semi-supervised learning include Support Vector Machines (SVM) and Generative Models[22].

Reinforcement learning is a machine learning field that focuses on the action's software agents should take in an environment to optimizes a cumulative reward[22]. Reinforcement learning is considered as one of the fundamental paradigms of machine learning, alongside supervised and unsupervised learning.

1.2 Statement of the Problem

Prior research has employed accuracy as the metric of measurement. The accuracy metric takes into account both positive and negative classifications to provide an overall evaluation. Prior analyses of machine learning algorithms indicate that models such as Support Vector Machines (SVM), K-Nearest Neighbours (KNN), Random Forest (RF), Gaussian Naive Bayes (GNB), Multi-Layer Perceptron (MLP), Extreme Gradient Boosting (XGBoost), Bagging, and Stacking exhibit the most favourable outcomes in terms of performance[23,24]. The metric of sensitivity, also known as the true positive rate (TPR), provides a superior means of evaluating model performance by accurately calculating the instances that have been correctly labelled. This will facilitate accurate diagnosis of Autism Spectrum Disorder (ASD), which is a key focus of the present study. Consequently, it is necessary to conduct an enquiry into their fundamental structure. Therefore, an in-depth examination and exploration of the aforementioned models is relevant. The purpose of this study is to conduct unbiased comparisons by utilising the optimal model, which refers to the model with the best hyperparameters. The assessment criteria will be computed, and the sensitivity (recall), area under the Precision-Recall curve, and area under the ROC curve will be used as the scoring

metrics for comparison. This is because accuracy may not be the most appropriate performance metric.

1.3 Aim and Objectives of the Study

The study aims to perform a comparative analysis of machine classification algorithms including support vector machine (SVN), gaussian naïve bayes (GNB), k-nearest neighbours (KNN), and multilayer perceotron (MLP) for the prediction of autism in children.

The specific objectives were to:

i. Perform feature selection to remove redundant features, improve classification and pre-process data to remove missing values and normalize data;
ii. apply the top four classification algorithms from pre-existing literature: Support Vector Machine, Gaussian Naïve Bayes, K-Nearest Neighbours, and Multilayer Perceptron to the dataset and evaluate their respective performances;
iii. improve model performance using cross validation (GridsearchCv Randomised SearchCV) and ensemble models like Random Forest, XGboost, Bagging and Stacking to the optimal models;
iv. Comparatively analyse model's performance metrics like sensitivity, specificity, ROC, F1 and employ sensitivity (recall).

1.4 Research Questions

1. What classification technique provides optimal overall performance amongst the selected machine learning classifiers?
2. Does performance improve upon the tuning of hyperparameters?
3. Is performance enhanced in ensemble techniques?
4. Which ensemble technique proved to be the best in performance?

1.5 Significance of the Study

The result of this study will develop a highly accurate machine learning model that uses physical characteristics of children to predict the presence of ASD. Also, this study will

identify the physical characteristics that are the most significant indicators of ASD in children which will enhance better prediction there is no known cure. The outcome of this research will enhance the existing methods of predicting ASD by providing more accurate classifications. Academically, the study will contribute to the body of knowledge and proffer solutions to issues regarding autism among children especially at the infant stage. The findings of this study will also serve as a reference for computer science students, lecturers, and researchers, as well as serve as a catalyst for further on the subject. Additionally, findings may result in the development of new theories regarding using artificial intelligence for Health care. The study would also be significant to knowledge when published.

1.6 Scope of the Study

The study will use the autism spectrum disorder for toddler's dataset obtained from a public dataset online (University of California, Irvine repository) to investigate the performance of the selected machine learning models through a comparative analysis. Also, the criteria for assigning the class variable will be based on a subject's score in the Diagnostic and Statistical Manual of Mental Health Disorders (DSM-V), a revamped version of the DSM-IV, which takes into consideration at least two of the predefined characteristics. Exploratory data analysis will be done, eight nonparametric models will be compared in total, with four models– Support Vector Machine (SVM), K-Nearest Neighbours (KNN), Gaussian Naïve Bayes (GNB), and Multilayer Perceptron (MLP) and the other four, ensemble methods - Random Forest (RF), Extreme Gradient Boost (XGBoost), Bootstrap Aggregating (Bagging), and stacking. Evaluation metrics will be computed, and the scoring metric in the comparison will be sensitivity (recall), area under the Precision-Recall curve, and area under the ROC curve.

1.7 Limitation of Study

The study's limitations, which are important for contextualizing the findings and guiding future research, include:

i. The study's results are based on a specific dataset. This can limit the generalizability of the findings, as the data may not represent the broader autism spectrum or include sufficient demographic diversity.
ii. The use of synthetic data for demonstration purposes may not capture the complexity and variability of real-world data, which could result in models that do not perform as well when applied to actual patient data.
iii. While the models performed well on the given dataset, they may not necessarily generalize to other datasets or clinical environments without further tuning and validation.
iv. Despite precautions, there is always a risk that the models, particularly those that achieved very high accuracy, may be overfitted to the particular characteristics of the training data.
v. The exclusion of highly predictive features like "Q_Chat_10" to avoid overfitting may also result in the loss of potentially valuable information that could contribute to the predictive power of the models.

1.8 Operational Definition of Terms

Algorithm: Algorithms typically acquire patterns from the training dataset and subsequently utilizes them to predict or classify data in the test dataset.

Autism: Autism, or autism spectrum disorder (ASD), is a developmental disorder that presents in early childhood. It has the potential to impact an individual's social, communication, relationship, and self-regulation capabilities.

Autism Spectrum Disorder (ASD): autism spectrum disorder (ASD) is a neurodevelopmental condition that can impact an individual's language, communication, cognitive abilities, and social behavior

Machine Learning: Machine learning is a subfield of Artificial Intelligence (AI) that involves the development of computer algorithms capable of accessing data and using it to improve their performance without being explicitly programmed.

Reinforcement Machine Learning: Reinforcement learning is a machine learning field that focuses on the action's software agents should take in an environment to optimize a cumulative reward

Semi-supervised Learning: Semi-supervised machine learning is a hybrid approach that integrates both supervised and unsupervised machine learning techniques. In the fields of machine learning and data mining, utilizing existing unlabeled data can prove to be advantageous in situations where obtaining labelled data is a laborious task

Supervised Learning: Supervised learning is a machine learning technique that involves learning a function that can map input data to output data based on a set of example input-output pairs

Unsupervised Learning: Unsupervised learning algorithms typically extract a limited number of features from the data. Upon the introduction of new data, the system utilizes the previously acquired features to classify the data into its respective class.

Endnotes

1. M.G Logrieco, L Casula, G.N Ciuffreda, R.L Novello, M Spinelli, F Lionetti, I Nicolì, M Fasolo, V Giovanni. *Risk and protective factors of quality of life for children with autism spectrum disorder and their families during the COVID-19 lockdown*. An Italian study. Research in developmental disabilities. 1(120), 2022,104130

2. E Feige, R Mattingly, T Pitts, A.F Smith. *Autism spectrum disorder: Investigating predictive adaptive behavior skill deficits in young children*. Autism research and treatment. 31, 2021.

3. Y.K Harmanşa, N Bedikyan, O Erbaş. *Autism and Cholesterol*. **Journal of Experimental and Basic Medical Sciences**. 2(2), 2021, 247.

4. H. Hadoush, M Alafeef, E Abdulhay. *Brain complexity in children with mild and severe autism spectrum disorders: Analysis of multiscale entropy in EEG*. **Brain Topography**. 30(32), 2019, 914.

5. S.L Hyman, S.E Levy, S.M Myers, D.Z Kuo, S Apkon, L.F Davidson, K.A Ellerbeck, J.E Foster, G.H Noritz, M.O Leppert, B.S Saunders. *Identification, evaluation, and management of children with autism spectrum disorder*. **Pediatrics**. 145(1), 2020.

6. S. Bölte, S Girdler, P.B Marschik. *The contribution of environmental exposure to the etiology of autism spectrum disorder*. **Cellular and Molecular Life Sciences**. 15(76), 2019, 97.

7. M.M Hassan, S.A Taher. *Analysis and classification of autism data using machine learning algorithms*. **Science Journal of University of Zakho**.10(4), 2022, 206-12.

8. A.S Heinsfeld, A.R Franco, R.C Craddock, A Buchweitz. and F Meneguzzi. *Identification of autism spectrum disorder using deep learning and the ABIDE dataset*. *NeuroImage: Clinical*, *17*, 2018 16-23.

9. C. Depastas, A Kalaitzaki. *The epidemiology of autism spectrum disorder and factors contributing to the increase in its prevalence*. Archives of Hellenic Medicine/Arheia Ellenikes Iatrikes. 39(3), 2022.

10. M.O Bakare, O.G Taiwo, M.A Bello-Mojeed, K.M Munir. *Autism spectrum disorders in Nigeria: A scoping review of literature and opinion on future research and social policy directions*. **Journal of Health Care for the Poor and Underserved**.30(3): 2019, 899.

11. L. Franz, N Chambers, M von Isenburg, P.J de Vries. *Autism spectrum disorder in sub-saharan africa: A comprehensive scoping review.* **Autism Research**. 10(5), 2017, 723-49.

12. S.J Rogers, P Yoder, A Estes, Z Warren, J McEachin, J Munson, M Rocha, J Greenson, L Wallace, E Gardner, G Dawson. *A multisite randomized controlled trial comparing the effects of intervention intensity and intervention style on outcomes for young children with autism*. **Journal of the American Academy of Child & Adolescent Psychiatry**. 60(6): 2021, 710-22.

13. J.B McCauley, R Elias, C Lord. *Trajectories of co-occurring psychopathology symptoms in autism from late childhood to adulthood.* **Development and psychopathology**. 32(4), 2020, 1287-302.

14. K. Danysz, S Cicirello, E Mingle, B Assuncao, N Tetarenko, R Mockute, D Abatemarco, M Widdowson, S Desai. *Artificial intelligence and the future of the drug safety professional*. **Drug safety**. 5(42), 2019, 491-7.

15. J. Bagherzadeh, H Asil. *A review of various semi-supervised learning models with a deep learning and memory approach*. **Iran Journal of Computer Science**. 2019 Jun 1;2:65-8.

16. K.K Jha, R Jha, A.K Jha, M.A Hassan, S.K Yadav, T Mahesh. *A brief comparison on machine learning algorithms based on various applications: a comprehensive survey*. In2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS) 2021 Dec 16 (pp. 1-5). IEEE.

17. P.C Sen, M Hajra, M Ghosh. *Supervised classification algorithms in machine learning: A survey and review.* InEmerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018 2020 (pp. 99-111). Springer Singapore

18. B. Mahesh. *Machine learning algorithms-a review*. **International Journal of Science and Research (IJSR)**.[Internet]. 2020 Jan;9:381-6.

19. R. Hou, Y Kong, B Cai, H Liu. *Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning*. **Neural Computing and Applications**. 2020 May;32:5399-407.

20. N.N Pise, P Kulkarni. *A survey of semi-supervised learning methods*. In2008 International conference on computational intelligence and security 2008 Dec 13 (Vol. 2, pp. 30-34). IEEE.

21. A. Ligthart, C Catal, B Tekinerdogan. *Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification*. **Applied Soft Computing**. 2021 Mar 1;101:107023.

22. X. Shen, C Yin, X Hou. *Self-attention for deep reinforcement learning*. InProceedings of the 2019 4th International Conference on Mathematics and Artificial Intelligence 2019 Apr 12 (pp. 71-75)

23. X.A .Bi, Y Wang, Q Shu, Q Sun, Q Xu. *Classification of autism spectrum disorder using random support vector machine cluster*. Frontiers in genetics. 2018 Feb 6;9:18.

24. R. Sujatha, S.L Aarthy, J Chatterjee, A Alaboudi, N.Z Jhanjhi. *A machine learning way to classify autism spectrum disorder.* **International Journal of Emerging Technologies in Learning (iJET)**. 2021 Mar 30;16(6):182-200

Chapter Two

Literature Review

2.1 Conceptual Review

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by diverse symptoms affecting communication, behavior, and social interactions. Early and accurate diagnosis is crucial for effective intervention, yet traditional diagnostic methods can be subjective and time-consuming. Machine learning (ML) techniques offer promising advancements in ASD prediction by analyzing behavioral and physiological data. This conceptual review explores various ML classification techniques, including Support Vector Machines (SVM), Random Forest (RF), k-Nearest Neighbors (k-NN), and Artificial Neural Networks (ANN), assessing their effectiveness in predicting ASD in children. The review emphasizes the importance of data pre-processing, feature selection, and model training in optimizing ML performance, and highlights the challenges and limitations of using ML for early ASD detection. By comparing these techniques, the review aims to identify the most

effective approaches for early and accurate ASD diagnosis, ultimately contributing to better outcomes for children and their families.

2.1.1 Autism Spectrum Disorder (ASD)

Autism spectrum disorder (ASD) is a cognitive and psychiatric condition with a very variable range of symptoms, courses, and causes[1]. Children with ASD may exhibit a variety of traits, including behavioural deviations, interpersonal communication difficulties, and social impairments[2]. Although confined, iterative behaviour, or activities are present alongside chronic challenges in social communication and interaction. These are not the only characteristics of ASD, and it is crucial to not disregard the dysfunctional motor traits linked to ASD because they are extremely common (79%), and can significantly affect quality of life, and the development of social skills[3,4].

According to the 5th edition of Diagnostic and Statistical Manual of Mental Disorders, individuals with ASD have limited areas of interest, repetitive behaviors that limit their communication and social interaction[3]. If this disease is diagnosed early, special education can start early with an opportunity to reduce the negative effects of the disease. ASD does not yet have a known clinical treatment in the world. According to the Centers for Disease Control and Prevention, one in 59 children in the United States has been diagnosed with ASD[6]. Given the high prevalence of ASD, high levels of comorbidity, as well as the cost and consequences of delays for treatment, the diagnosis of ASD should be made as soon as possible. The diagnosis of ASD with a high degree of accuracy is of paramount importance. The symptoms of autism are very wide ranging[6]. The perpetual continuation of a wide variety of symptoms raises suspicions of autism in the individual.

Currently, autism is diagnosed on the basis of:

i. the parental interview,

ii. scales such as the Autism Diagnostic Observation Scale (ADOS), Autism Observation Scale for Infants, Social Orienting Continuum and Response Scale, and Early Social and Communication Scale, and

iii. clinical observations[7].

This highly subjective diagnostic process may vary based on the physician's experience, the behavior of the patient, the knowledge of parents and answers on the scale. The diagnostic process can take a long time depending on the physician and the parents, and there is always the possibility of misdiagnosis due to these differences. This shows that scientific advances and support are still needed for the diagnosis of autism[8].

Clinical findings or biomarkers are needed for early diagnosis of autism[9]. However, clinical biomarkers alone will not be sufficient for early diagnosis. Authors advocate the research ideology that early diagnosis and early special education can be achieved by using computer-aided methods based on EEG signals and/or imaging, machine learning (ML), deep learning (DL) and other advanced computing technologies. Early detection and treatment of ASD is possible, but this is only possible through clinical diagnosis[9]. This diagnosis can then guide the development of individualized treatment plans to enhance the health and social balance of ASD-affected children and their immediate families and careers. Unfortunately, an ASD diagnosis can be time-consuming and expensive. The recent rise in ASD cases worldwide has prompted medical professionals and scientists to look for improved screening techniques[10].

2.1.2 Autism Spectrum Disorder (ASD) in Children

ASD is revealed in childhood, usually in the first five years of life and lasts until maturity[11]. These symptoms first show in childhood, but it takes about 3 years after symptoms are noticed before an ASD diagnosis is made[11].

It is customary for trained professionals with extensive experience to handle the difficult decision-making procedures involved in accurately screening and assessing children with autism. Even while the medical community still finds it difficult to detect and gauge the gravity of childhood autism, the prevalence of this illness has intensified in recent times. This is mainly because of different causes; Early onset of the condition, which makes it challenging to diagnose, lack of knowledge about the illness and its signs and its resemblance to various neurological conditions in nature[12].

Availability and access to early intervention programs for children with ASD are dependent on an official ASD diagnosis for them[13]. Using standardized tests of mental perception, linguistic, and adaptability, as well as looking out for indicative behaviours of ASD are promoted. Best practices in ASD diagnosis include information from caregivers which may include reports of history, progression and present symptoms[13]. While there are generally approved diagnostic tools like the revised Edition of the Autism Diagnostic Observation Schedule (ADOS), to facilitate eliciting and observing ASD symptoms, they were made to be used with standardized objects or toys in in-person interviews in quiet environments[14].

2.1.3 Detection of Autism Spectrum Disorder (ASD)

Autism detection is a challenging and time-demanding task[15]. Numerous behavioural and physiological strategies have been utilized to successfully and accurately detect autism in children at an early age. The current job done by scientific research centres on discovering suitable answers, treatments and predictive indications, that inform parents of their children's behaviour and psychological status is laudable[15]. Especially in youngsters, diverse mental diseases like ADHD and ASD are very challenging to detect[16].

A current method of diagnosing mental illness in psychiatry (DSM-5/ICD-10), which is solely dependent on observations of symptoms and is potentially inaccurate[17]. No quantitative test

exists yet that a doctor may recommend to a patient that will result in an accurate diagnosis[16]. For other illnesses including diabetes, HIV, and hepatitis C, such quantitative and conclusive tests are standard procedure. The absence of a clinical test and seemingly overlapping symptoms in children make the diagnosis of mental health illnesses a challenging procedure[19]. Children with ASD experience social disabilities such as communicative challenges and behavioural deviations throughout their entire life. This condition affects more than 1% of children, and identifying it early can be helpful[16.] The incidence of ASD in men as compared to women was four times higher, and some demographic characteristics including gender and colour difference between people with ASD and healthy people.

Various approaches to ASD diagnosis have been investigated, including behaviour observation, and the extraction and analysis of discriminant patterns from demographic and brain data[16]. Two general techniques for detecting ASD are:

1. Brain Imaging

There are various regions in human brain[21]. Even while each region is responsible for its own functions, there are useful connections between them that make up the brain network. Brain imaging data can yield useful biomarkers through quantitative analysis, leading to more precise diagnoses of ADHD, MCI, ASD, and Alzheimer's[21]. For the diagnosis of these cognitive illnesses, machine learning algorithms have been integrated with magnetic resonance imaging (MRI) and functional magnetic resonance imaging (fMRI)[16]. The ABIDE initiative, (Autism Brain Imaging Data Exchange) which offers structural and functional brain imaging datasets gathered from numerous facilities around the world, has recently attracted a lot of applause for its work on ASD detection using fMRI data. Based on ABIDE data, other studies and techniques have been created [23,24,25].

Numerous researches have examined regression and classification methods like SVM and RF[24,25,26]. Using the SVM, 79.17% accuracy was achieved in a work to study the impact of various frequencies for creating functional brain networks[27]. In a work that classified 1,035 individuals with 70% accuracy using a deep learning approach (505 ASD and 530 controls)[28]. The authors asserted that this strategy enhanced the cutting-edge method. They used an approach that regarded different pairwise correlation coefficients as characteristics. To extract lower dimensional data, stacked autoencoders were initially trained. Following autoencoder training, the weights from the autoencoders were added to an MLP classifier to fine-tune the classification process. 17 sites in the ABIDE dataset were categorized independently, and an accuracy of 52% was achieved. It was established that insufficient training samples was the cause of the poor performance on individual sites.

Typically, research employing machine learning to diagnose ASDs have only taken into account a portion of the ABIDE dataset or have included information other than fMRI data in their models[16]. There are just a few studies, that examined all 1,035 subjects in the ABIDE dataset using solely fMRI data without making any assumptions about their demographics. An extensive review of literature indicates that Heinsfeld's method is the most exotic for diagnosing ASD with the entire ABIDE dataset.

2. Eye Tracking Technologies

Despite the late identification of clinical and physiological traits, some key behavioural traits are very capable of identifying autism and its severity[15]. Due to its speed, low cost, ease of analysis, and applicability to all ages, eye-tracking technology is undoubtedly significant in the indication of ASD[15]. The eye movement tracking method creates, monitors, and records points while also computing eye movement via these locations. Eye responses significantly affect how people respond to verbal and visual stimuli as biomarkers of ASD. Additionally,

research revealed that a connection between clinical assessment and early ASD identification by eye movement tracking[30]. Additionally, eye-tracking diagnosis aids in the quick identification of children with ASD (Constantino et al., 2017).

2.1.4 Machine Learning

Machine learning is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to learn and make predictions or decisions without being explicitly programmed[31]. Machine learning algorithms require large amounts of data to learn from. This data can be structured (e.g., tables) or unstructured (e.g., text, images, audio). In machine learning Features are the variables or characteristics extracted from the data that the algorithm uses to make predictions. Feature engineering involves selecting and transforming these variables to improve the algorithm's performance[31]. There are various machine learning algorithms, including supervised learning (where the model is trained on labeled data), unsupervised learning (where the model identifies patterns in unlabeled data), and reinforcement learning (where agents learn by interacting with an environment[32].

During the training phase, the machine learning model learns patterns and relationships in the data[32]. It adjusts its internal parameters to minimize the difference between its predictions and the actual outcomes in the training data. Machine learning has a wide range of applications, including image and speech recognition, natural language processing, recommendation systems, autonomous vehicles, fraud detection, and medical diagnosis, among many others. It continues to advance rapidly and has the potential to transform various industries and aspects of our daily lives.

2.1.4.1 Types of Machine Learning

Machine learning can be categorized into several types or paradigms, depending on the learning approach and the nature of the data. The main types of machine learning are:

1. Supervised Machine Learning

In supervised learning, the algorithm is trained on a labeled dataset, where each input is associated with the correct output or target[33]. The goal is to learn a mapping from inputs to outputs. Examples include classification (assigning inputs to predefined categories) and regression (predicting a continuous numeric value)[35]. Supervised learning is one of the fundamental paradigms in machine learning. In supervised learning, an algorithm learns from a labeled dataset, where each input example is associated with a corresponding target or output[33]. The goal is to learn a mapping or relationship between the inputs and their corresponding outputs so that the algorithm can make predictions or classify new, unseen data

Supervised learning can be applied to different types of problems, depending on the nature of the output variable:

Classification: In classification problems, the output variable is categorical, meaning it belongs to one of a finite number of classes or categories[36].

Regression: In regression problems, the output variable is continuous, meaning it can take any real-numbered value within a certain range[37]. Supervised learning is a versatile and widely used approach in various domains, including natural language processing, computer vision, healthcare, finance, and more. Its success depends on the quality and quantity of labeled data and the choice of an appropriate model for the task at hand.

There are several types of supervised learning algorithms, each designed for different types of problems and data[38].

1. **Regression Algorithms**: These algorithms are used when the target variable is continuous or numeric[39]. The goal is to predict a real-valued output. Common regression algorithms include:

i. Linear Regression,

ii. Ridge Regression,

iii. Lasso Regression,

iv. Support Vector Regression,

v. Decision Trees for Regression

Classification Algorithms: These algorithms are used when the target variable is categorical or discrete, and the goal is to classify data points into predefined categories or classes[40]. Common classification algorithms include: Logistic Regression, Decision Trees for Classification, Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Naive Bayes, Neural Networks (Deep Learning)[41].

Ensemble Methods: Ensemble methods combine the predictions of multiple base models to improve overall performance. Some popular ensemble methods include: Bagging (Bootstrap Aggregating), Boosting (e.g., AdaBoost, Gradient Boosting), Stacking[43].

Instance-Based Learning: Instance-based learning algorithms make predictions based on similarity to existing labeled instances in the training data. k-Nearest Neighbors (k-NN) is a prime example of an instance-based learning algorithm[44].

Classification Algorithms

Naive Bayes: Naive Bayes is a classification algorithm based on Bayes' theorem with a strong independence assumption between features. It's particularly effective for text classification tasks[45].

Support Vector Machines (SVM): SVM is a versatile classification algorithm that finds a hyperplane that best separates data into different classes. It can be used for both binary and multiclass classification.

Neural Networks: Neural networks, including deep neural networks, are used for complex tasks such as image recognition, natural language processing, and speech recognition. They consist of interconnected layers of artificial neurons (perceptrons).

Decision Trees: Decision trees are used for both classification and regression tasks. They create a tree-like structure of decision rules based on features to make predictions.

Random Forest: Random Forest is an ensemble learning method that combines multiple decision trees to improve accuracy and reduce overfitting.

Gradient Boosting: Gradient boosting algorithms like XGBoost and LightGBM are powerful techniques for regression and classification[46]. They build an ensemble of decision trees sequentially, with each tree correcting the errors of the previous ones.

Logistic Regression: Logistic regression is used for binary classification tasks and models the probability of a data point belonging to a particular class.

Ordinal Regression: Ordinal regression is used when the target variable has ordered categories. It's a specialized form of regression for ordinal data.

2. Unsupervised Machine Learning

Unsupervised machine learning is a type of machine learning where the algorithm is trained on unlabeled data, meaning that there are no explicit labels or target values provided for the input data[47]. Common tasks include clustering (grouping similar data points together) and dimensionality reduction (reducing the number of features while preserving important information)[47]. In contrast to supervised learning, where the algorithm learns to make

predictions based on labeled examples, unsupervised learning algorithms seek to find patterns, structures, or relationships within the data on their own[47].

There are several types of supervised learning algorithms. They include;

i. Clustering Algorithms: Clustering algorithms group similar data points together into clusters or clusters of clusters (hierarchical clustering)[48]. Common clustering algorithms include: K-Means Clustering which partitions data into 'K' clusters based on similarity, Hierarchical Clustering builds a tree-like structure of clusters to represent data hierarchy. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) identifies clusters based on the density of data points, Agglomerative Clustering which is type of hierarchical clustering that starts with individual data points as clusters and merges them iteratively[48].

ii. Dimensionality Reduction Algorithms: Dimensionality reduction techniques aim to reduce the number of features (variables) while preserving important information[49]. Common methods include: Principal Component Analysis (PCA) which finds orthogonal axes that maximize variance, t-Distributed Stochastic Neighbor Embedding (t-SNE) used for visualization and reduces dimensionality while preserving pairwise similarities, autoencoders are Neural networks that learn a compact representation of data in their hidden layers[49].

iii. Association Rule Mining: Association rule mining identifies interesting relationships or patterns between variables in large datasets. Common algorithms include: Apriori Algorithm (Used in market basket analysis to find associations between items), FP-Growth (Frequent Pattern Growth) which is efficient for discovering frequent patterns[50].

iv. Generative Models: Generative models aim to learn the underlying probability distribution of data and can generate new data samples that resemble the training data[51]. Common generative models include: Variational Autoencoders (VAEs) which ombines

autoencoders with probabilistic models, Generative Adversarial Networks (GANs) which comprises a generator and discriminator to generate realistic data[52].

v. Anomaly Detection Algorithms: Anomaly detection identifies data points that deviate significantly from the majority of the data. Algorithms used for anomaly detection include: One-Class SVM which trains a model on normal data and identifies outliers, Isolation Forest (Constructs random decision trees to isolate anomalies).

vi. Density Estimation Algorithms: Density estimation models the probability distribution of data points in a dataset[53]. Common methods include: Kernel Density Estimation (KDE) estimates the probability density function using kernel functions, Gaussian Mixture Models (GMM) which represents data as a mixture of Gaussian distributions.

vii. Self-Organizing Maps (SOMs): SOMs are neural network-based algorithms used for clustering and visualization[54].

viii. Latent Dirichlet Allocation (LDA): LDA is a probabilistic model used for topic modeling in text data[55].

ix. Independent Component Analysis (ICA): ICA separates multivariate data into additive, independent components. The choice of the appropriate unsupervised learning algorithm depends on the specific task and the nature of the data[56]. These algorithms are used in various domains, including data analysis, image processing, natural language processing, and more, to gain insights from unlabeled data or to preprocess data for supervised learning tasks.

3. Semi-Supervised Learning

Semi-supervised learning is a machine learning paradigm that combines elements of both supervised and unsupervised learning. In semi-supervised learning, the training dataset contains a mixture of labeled and unlabeled data[57]. Semi-supervised learning has been applied in various domains, including natural language processing, computer vision, and speech

recognition[58]. It's particularly valuable when obtaining large amounts of labeled data is challenging or costly but you still want to improve the performance machine learning models. There are different approaches and algorithms within semi-supervised learning:

Self-Training: Self-training is a straightforward and intuitive approach. It involves training a model on the available labeled data and then using that model to make predictions on the unlabeled data[59]. The predicted labels on the unlabeled data become pseudo-labels, and the model is retrained iteratively with both the original labeled data and the newly pseudo-labeled data.

Co-Training: Co-training is applicable when there are multiple views or sources of data. It involves training separate models on different subsets or views of the features and using the agreement or disagreement between the models' predictions as a form of confidence measure[60]. Data points with high agreement can be considered as pseudo-labeled data.

Multi-View Learning: Multi-view learning is similar to co-training but focuses on multiple sets of features or representations of the data. Each view provides additional information that can be used to improve the model's performance on the unlabeled data[61].

Multi-Instance Learning: In multi-instance learning, the data points are organized into bags, and each bag contains multiple instances. Labels are assigned to entire bags rather than individual instances[62]. This approach is often used when only bag-level labels are available, and instances within a bag are assumed to share the same label.

Semi-Supervised Variational Autoencoders (VAEs): Variational autoencoders can be extended to semi-supervised settings. The encoder network learns both the data representations and the class labels of the labeled data, while the decoder aims to reconstruct the input data[63]. VAEs encourage the latent space to be disentangled between classes, making them suitable for semi-supervised classification.

Semi-Supervised Graph-Based Methods: These methods leverage the graph structure of data to propagate labels from labeled to unlabeled data points. Label Propagation, Label Spreading, and Graph Convolutional Networks (GCNs) are examples of techniques that use graph-based semi-supervised learning[64].

Generative Models for Semi-Supervised Learning: Generative models like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) can be trained in a semi-supervised manner[65]. The generator generates data samples, and the discriminator distinguishes between real and generated data, with some labeled data used for training the discriminator.

Transfer Learning for Semi-Supervised Learning: Transfer learning methods, where a model pre-trained on a large dataset is fine-tuned on a smaller labeled dataset, can be used in semi-supervised learning[66]. The pre-trained model provides a strong feature extractor, which can be helpful in leveraging unlabeled data effectively.

4. Reinforcement Learning

Reinforcement learning involves an agent that interacts with an environment and learns to make a sequence of decisions (actions) to maximize a cumulative reward[67]. This type of learning is commonly used in tasks such as game playing, robotic control, and autonomous systems. There are several algorithms and approaches in reinforcement learning.

1. Value-Based Methods

Q-Learning: Q-learning is a model-free algorithm that learns the optimal action-value function (Q-function) by iteratively updating Q-values using the Bellman equation.

Deep Q-Networks (DQN): DQN combines Q-learning with deep neural networks to handle high-dimensional state spaces, making it suitable for complex tasks68.

2. Policy-Based Methods:

Policy Gradient Methods: These algorithms learn a parameterized policy directly, optimizing it to maximize the expected cumulative reward. Examples include: REINFORCE (Monte Carlo Policy Gradient), Proximal Policy Optimization (PPO), Trust Region Policy Optimization (TRPO)[66].

3. Model-Based Methods:

Monte Carlo Tree Search (MCTS): MCTS is a model-based algorithm often used in games and planning problems. It builds a tree of possible actions and their outcomes to make decisions[70].

Model Predictive Control (MPC): MPC is used for continuous control tasks and involves solving an optimization problem at each time step to find the best action[70].

4. Exploration Strategies

Epsilon-Greedy: A simple exploration strategy where the agent selects the greedy action (exploitation) most of the time but occasionally explores by choosing a random action[71].

Upper Confidence Bound (UCB): UCB-based algorithms select actions based on the estimated uncertainty or confidence intervals of their values[72].

Thompson Sampling: A Bayesian exploration strategy that maintains a probability distribution over action values and selects actions according to their sampled values.

Deep Deterministic Policy Gradients (DDPG): DDPG is an algorithm designed for continuous action spaces. It combines the actor-critic framework with deep neural networks[73].

Hindsight Experience Replay (HER): HER is a technique used in reinforcement learning for learning from failures. It replays and re-labels past experiences to help the agent learn from unsuccessful attempts[73].

Inverse Reinforcement Learning (IRL): IRL is used to learn the reward function from expert demonstrations, allowing the agent to learn tasks by imitating expert behavior[74].

Multi-Agent Reinforcement Learning (MARL): MARL deals with scenarios where multiple agents interact in the same environment. Algorithms like Multi-Agent Deep Deterministic Policy Gradients (MADDPG) and Multi-Agent Proximal Policy Optimization (MAPPO) are used in multi-agent settings[75].

2.1.5 Classification of ASD Using Machine Learning

ML models can be used both to confirm a diagnosis and make early diagnosis of ASD. In most ML methods used for the diagnosis of ASD, data must be preprocessed before training and testing the model which includes various classification algorithms like GBoost, Support Vector Regression, Extreme Learning Machines (SVM), Gaussian Naive Bayes (GNB), Multilayer Perceptron, Esemble Technique and K-Nearest Neighbour (KNN)[76]. A data set is created by identifying features to help recognize ASD and converting these features into numerical values and data matrices. The feature extraction or feature selection process depends on the nature of the data to be used[76]. For example, if brain imaging, sMRI or fMRI, is used, structural and, structural (1) gray matter (GM) readings derived from cortical

thickness (CT) (2) GM density values from voxel-based morphometry (Ashburner and Friston 2000), (3) microstructural changes in the white matter (WM) from diffusion-weighted magnetic resonance imaging (DWI) (fractional anisotropy [FA]). (4) connectivity matrixes or (5) parameters obtained from network analysis and resting state fMRI (rs-fMRI) [77].

2.2 Methodological Framework

In this section, the supervised classification algorithm used in this study was presented

2.2.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is commonly utilised for tasks involving regression or two-group classification, with a primary focus on classification applications[78]. This approach involves the representation of individual data points in a multi-dimensional space, where each dimension corresponds to the value of a specific attribute. In this context, the process of categorization involves identifying the demarcation line that effectively separates the two distinct classes[78]. An ideal hyperplane can be defined as a linear decision function that exhibits the most distinct boundary between vectors belonging to different groups. In the event that there is a requirement for an additional feature, the Support Vector Machine (SVM) employs the kernel algorithm to transform a low-dimensional input space into a higher-dimensional space[79]. In essence, it converts seemingly irreconcilable problems into ones that can be integrated. If there are no errors in this separation, then the expected value of the error can be expressed as:

$$E[\Pr(error)] \leq \frac{E[number\ of\ support\ vectors]}{[number\ of\ training\ vectors]} \quad (2.1)$$

The decision function will be given as

$$D(x) = w\emptyset(x) + b \quad (2.2)$$

which is the best line that integrates the training data, w and b are parameters of the SVM, and $\emptyset(x)$ is the function which transforms the data into the new M dimension

There exist two distinct categories of Support Vector Machines (SVM): linear SVM, which is applicable when the data can be represented in a linear fashion[80]. This suggests that the data points exhibit a high degree of linearity, allowing them to be effectively fitted by a single straight line. The second method is non-linear Support Vector Machine (SVM), which is employed when the data cannot be separated by a linear boundaryinear SVM, which is applicable when the data can be represented in a linear fashion[80].

In such cases, sophisticated techniques, known as kernel tricks, are utilized to enhance the SVM's performance[79]. The selection of a kernel is a crucial aspect in data analysis, as it is contingent upon the specific characteristics of the dataset under consideration. If the problem at hand is linear in nature, it is necessary to utilize a linear kernel function. Commencing with the supposition that the data exhibits linearity is deemed advantageous, afterwards proceeding to evaluate other kernels in order to assess their performance measures. Support Vector Machines (SVM) exhibit superior performance when applied to datasets with a linear structure, and their effectiveness is further boosted when dealing with data that exists in high-dimensional spaces[81]. Support Vector Machines (SVM) provide a high degree of robustness due to its non-sensitivity towards outliers.

In a work that utilized the support vector machine (SVM) classification algorithm to analyse three distinct datasets[82]. Each dataset had a sample of 29 Chinese children ranging in age from 4 to 11 years. A total of six photos were shown to the participants, with three depicting Chinese faces categorised as "same race" and the remaining three featuring Caucasian faces labelled as "other races". The analysis of the entire set of faces in the sample space yielded an accuracy rate of 88.51%. The recall or sensitivity rate, which measures the proportion of actual positive cases correctly identified, was found to be 93.10%. The area under the receiver operating characteristic curve (AUC), a metric used to evaluate the performance of a

classification model, was determined to be 89.63%[82]. Additionally, the specificity rate, indicating the proportion of actual negative cases correctly identified, was calculated to be 86.21%. In order to improve the efficacy of their model, the researchers exclusively utilized facial data from individuals of different racial backgrounds. As a result, the model exhibited enhanced performance metrics, including an accuracy rate of 90.80%, sensitivity of 96.55%, AUC of 94.41%, and specificity of 87.93%[82]. In comparison, when the model was trained on facial data from individuals of the same race, the performance metrics were lower, with an accuracy rate of 81.61%, specificity of 84.48%, sensitivity of 75.86%, and AUC of 82.40%[82].

2.2.1.1 Hyperparameters of SVM

Hyperparameters are features employed in models to improve the predictive and performance power of the models or, to speed up computation time. SVMs have several hyperparameters that can be tuned to optimize their performance. They include;

Kernel: SVMs can use different types of kernels, such as linear, polynomial, radial basis function (RBF), and sigmoid. The choice of kernel can significantly impact the model's performance[83].

C (Regularization Parameter): The C parameter, often referred to as the regularization parameter, controls the trade-off between maximizing the margin and minimizing classification errors[84]. A smaller C encourages a larger margin but may allow some misclassification, while a larger C aims to classify all training points correctly and may result in a smaller margin.

Gamma (Kernel Coefficient): Gamma is a parameter for the RBF kernel and defines how far the influence of a single training example reaches[85]. A small gamma implies a Gaussian with a large variance, while a large gamma creates a Gaussian with a small variance. Tuning gamma is crucial for controlling overfitting[85].

Kernel Coefficient (coef0): This is another parameter used in the polynomial and sigmoid kernels. It influences the model's behavior, particularly in the sigmoid kernel[86].

Class Weights: SVMs allows assigning different weights to different classes. This is useful when dealing with imbalanced datasets, where one class has significantly fewer samples than the others[87].

Decision Function Shape (decision_function_shape): In multi-class classification, SVMs can use either "ovr" (one-vs-rest) or "ovo" (one-vs-one) decision function shapes[88]. This parameter determines the shape to be used.

2.2.2 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a simple and widely used machine learning algorithm for classification and regression tasks[89]. KNN uses a distance metric, typically Euclidean distance, to measure the similarity between data points. It calculates the distance between the new data point (the one you want to classify or predict) and all data points in the training dataset. KNN then selects the K-nearest data points (neighbors) based on the calculated distances[89]. These neighbors are the training examples that are most similar to the new data point.

For classification tasks, KNN predicts the class label of the new data point by taking a majority vote among its K-nearest neighbors. In regression tasks, it predicts the target value by averaging the target values of its K-nearest neighbors[89].

The K-nearest neighbours (KNN) strategy involves estimating the conditional distribution of variable Y given variable X, and afterwards assigning an observation to the class with the highest probability, also known as the modal class[90]. To forecast the result for an observation X=x, the observations that are in close proximity to x are identified. Subsequently, X is assigned to the category that encompasses these observations[90]. The impact of the initial measurement values of the predictors on the distance results necessitates the scaling and

centering of all predictor values. This process aims to mitigate any bias towards predictors with larger scales and ensure that all predictors are equally considered when computing Euclidean distance.

KNN is a non-parametric method, which requires that all observations in the experiment be higher than the number of chosen predictors. Assuming a training set D made up of X_i, $i \in [1, |D|]$ samples, where F described the number of features and assumed normality[91]. Each output was labelled with class label $y_j \in Y$. The aim was to classify an unknown variable q. For every $x_i \in D$, the distance between q and x_i was computed thus:

$$d(q, x_i) = \sum_{f \in F} w_f \, \partial(q_f, x_{if}) \tag{2.3}$$

and the equation, which measured distance in both discrete and continuous cases was written as:

$$\partial(q_f, x_{if}) = \begin{cases} 0 & f \ discrete \ and \ q_f = x_{if} \\ 1 & f \ discrete \ \ and \ q_f \neq x_{if} \\ |q_f - x_{if}| & f \ discrete \end{cases} \tag{2.4}$$

upon which the k nearest neighbours are chosen. The class of q is ideally selected by assigning it to the majority class among the chosen nearest neighbours[91]. It will often make sense to assign more weights are assigned to the nearest neighbours, and a voting system is implemented where the neighbours decide the class of q, based on the inverse of their distance from it.

$$Vote(y_i) = \sum_{c=1}^{k} \frac{1}{d(q, x_c)^n} 1(y_j, y_c) \tag{2.5}$$

$1(y_j, y_c)$ results in 1, if the class labels match and 0, if otherwise. n, is usually 1, even though greater values can be applied to minimise the effect of further distant neighbours.

2.2.2.1 Hyperparameters of KNN

n neighbours: the number of neighbours(groups) selected to classify a data point

weights: assigned using a kernel function. Weights assign more value to nearby points and less, to the points farther away.

Distance : Minkowski, euclidean, manhattan etc. Calculates the distance between a point, and the distance between points in the training set.

2.2.3 Gaussian Naive Bayes (GNB)

Naive Bayes (NB) refers to a set of supervised learning models used for predictive purposes. They are simple and effective models which learn the probabilities of certain features belonging to a given group[92]. Gaussian Naive Bayes (GNB) is a probabilistic machine learning algorithm used for classification tasks. It's particularly useful when dealing with continuous or real-valued features. GNB is a variant of the Naive Bayes algorithm, which is based on Bayes' theorem and assumes that features are conditionally independent given the class label. In the case of GNB, it assumes that the likelihood of the features follows a Gaussian (normal) distribution[92].

Bayes' Theorem: Bayes' theorem is the foundation of Naive Bayes algorithms. It relates the conditional probability of an event A given an event B to the conditional probability of event B given event A.

$$P(A|B) = \frac{P(B|A)*P(A)}{P(B)} \quad (2.6)$$

In the context of classification, we want to find the probability of a class (C) given a set of features (X):

$$P(C|X) = \frac{P(X|C)*P(C)}{P(X)} \quad (2.7)$$

Where:

$P(C|X)$ is the posterior probability of class C given features X.

$P(X|C)$ is the likelihood of the features given class C.

$P(C)$ is the prior probability of class C.

P(X) is the marginal likelihood of features X.

1. Gaussian Distribution

GNB assumes that the likelihood P(X|C) for each feature follows a Gaussian (normal) distribution. The Gaussian probability density function (PDF) for a continuous random variable X with mean μ and variance σ^2 is[92]:

$$P(X|C) = \frac{1}{\sqrt{2\pi\sigma^2}} . e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.8)$$

In the case of GNB, we calculate this distribution separately for each feature, assuming that they are conditionally independent given the class label.

2. Naive Bayes Assumption

The "naive" part of GNB comes from the assumption that features are conditionally independent given the class label. This simplifies the calculation of the joint likelihood P(X|C)as follows[92]:

$$P(X|C)=P(X_1|C).\ P(X_2|C)\ldots\ldots..P(X_n|C) \quad (2.9)$$

Where:

X_1 represents individual features.

$(X_1|C)$ is the Gaussian probability density function for each feature, assuming a Gaussian distribution.

Classification

To classify a new data point with features X, GNB calculates the posterior probability for each class and selects the class with the highest posterior probability:

$$y = \text{argmax}_c\ P(C = c|X) \quad (2.10)$$

Where y is the predicted class label, and c iterates over all possible class labels.

GNB uses Bayes' theorem and the Gaussian distribution to estimate the likelihood of features given each class, assuming feature independence. It then computes the posterior probabilities for each class and selects the class with the highest probability as the predicted class label for a new data point.

2.2.3.1 Hyperparameters of a GNB Model

It is pertinent to find an optimal combination of parameters whose sole aim is to minimise loss and improve the efficiency of the model.

var_smoothing: is used to smoothen the likelihood curve and provide stability for sample points far from the mean.

param_grid: This is a collection of string parameters and their possible combinations and settings, which makes parameter searching easy given any combination of settings.

verbose: refers to the number of messages. Verbose is usually set to 1

cv: this is the parameter that spearheads cross-validation, also called an iterator. Common settings are 3, 5 and 10.

n_jobs: This is the largest number of works carried out simultaneously; it is commonly set to -1 which means that all CPUs are used.

2.2.4 Random Forest (RF)

Random Forest (RF) is an ensemble learning technique used for both classification and regression tasks[93]. It is based on the idea of creating multiple decision trees during training and then combining their predictions to improve accuracy and reduce overfitting[93]. Random Forest is built upon the foundation of decision trees. Decision trees are a hierarchical structure of nodes that split the data based on feature values to make predictions. Each node represents a decision or a test on a feature, and each leaf node represents the final prediction. Random

Forest builds an ensemble of decision trees. Each tree is trained on a random subset of the training data (bootstrapped samples) and a random subset of the features (feature bagging)94. This randomness ensures that the trees are diverse and reduces overfitting.

For classification tasks, each tree in the Random Forest "votes" for a class, and the class with the most votes becomes the final prediction. In the case of regression, each tree makes a prediction, and the final prediction is the average of all the individual tree predictions[95]. Random Forest can handle missing values in the dataset. When making a prediction for a data point with missing values, it can use the predictions from multiple trees and combine them appropriately. Random Forest can also calculate the out-of-bag error during training. It uses the data points that were not included in the bootstrap samples for each tree to estimate the model's performance without the need for a separate validation set[95].

The RF algorithm is very efficient, as it handles datasets that contain continuous variables, as well as categorical variables robustly. An RF classifier contains subsets of various tree classifiers $\{h(x, \Theta_k), k = 1,2, ...\}$ where the Θ_k are independently and identically distributed random vectors with each tree being able to specify the modal class at input x [95]. The performance index, which solely approximates the confidence interval (CI) of the RF model is given as

$$mg(x, y) = av_k I(h_k(x, \Theta_k) = y) - \max_{j \neq y} av_k I(h_k(x, \Theta_k) = j) \quad (2.11)$$

where $I(.)$ denotes an indicator function, and $av(.)$, the average value. It is observed that as the margin increases, the confidence level also increases. The generalisation error becomes

$$PE^* = P_{x,y}(mg(x, y) < 0), \quad (2.12)$$

where $P(.)$ denotes probability. With an increase in trees for all sequences Θ_k, PE^* converges to

$$P_{x,y}(P_{\Theta}(h(x,\Theta)=y) - \max_{j\neq y} P_{\Theta}(h(x,\Theta)=j<0) \quad (2.13)$$

Convergence of this generalisation error proves that the RF model does not overfit as more trees are introduced. The upper bound for the generalisation error is given as

$$PE^{*} \leq \frac{\bar{\rho}(1-s^{2})}{s^{2}} \ , \quad (2.14)$$

where $\bar{\rho}$ is the average correlation value, s is the strength of each tree in the model. An increased strength of individual trees and a low correlation between them produces more accurate prediction results.

2.2.4.1 Hyperparameters of an RF Model

Number of Estimators (n_estimators): This hyperparameter controls the number of decision trees in the random forest ensemble. A higher number of estimators can lead to better performance but may also increase computation time[96].

i. Criterion: RF uses a criterion to measure the quality of a split when building individual decision trees.
ii. Maximum Depth of Trees (max_depth): This hyperparameter limits the maximum depth of the individual decision trees in the ensemble[96]. It helps control overfitting. Setting it too high can lead to overfitting, while setting it too low can result in underfitting.
iii. Minimum Samples per Leaf (min_samples_leaf): It specifies the minimum number of samples required to be in a leaf node[97]. This hyperparameter can also help prevent overfitting. A larger value results in a simpler model.
iv. Minimum Samples per Split (min_samples_split): It specifies the minimum number of samples required to split an internal node. Like `min_samples_leaf`, it helps control the tree's complexity[97].

v. Maximum Features (max_features): This hyperparameter determines the maximum number of features that are considered when making a split decision in a tree[97]. It can be specified as a percentage or a fixed number. Lower values can lead to more diverse trees and reduce overfitting.

vi. Bootstrap Sampling (bootstrap): This binary hyperparameter determines whether to use bootstrap sampling when building each decision tree[96]. Bootstrap sampling involves randomly selecting data points with replacement from the training set. Setting it to `True` (the default) enables bootstrap sampling, while setting it to `False` disables it.

vii. Class Weights (class_weight): For imbalanced datasets, you can assign different weights to classes[97]. This gives more importance to minority classes during training.

viii. Out-of-Bag (OOB) Score (oob_score): If set to `True`, the RF model calculates an out-of-bag score during training, providing an estimate of the model's performance without the need for a separate validation set.

ix. Warm Start (warm_start): This binary hyperparameter allows you to train RF incrementally. When set to `True`, you can add more trees to the existing RF model using the `n_estimators` parameter without starting from scratch[97].

x. Feature Importance (feature_importances): RF can provide information about feature importance, which can help identify which features are more influential in making predictions.

2.2.5 Multilayer Perceptron (MLP)

A Multilayer Perceptron (MLP) is a type of artificial neural network used for various machine learning tasks, including classification, regression, and pattern recognition[98]. It consists of multiple layers of interconnected neurons (also called nodes or units), including an input layer,

one or more hidden layers, and an output layer. Each connection between neurons has a weight associated with it, and each neuron has an activation function that determines its output based on the weighted sum of its inputs.

Equations and components of an MLP include[98]:

1. Input Layer

The input layer consists of neurons that represent the features of the input data. If we have n features, it can represent the input as a vector $X = [x_1, x_2, \ldots\ldots, x_n]$. The input neurons simply pass their values forward to the next layer without any processing[99].

2. Hidden Layers

The hidden layers are where the processing occurs. Each neuron in a hidden layer receives input from the neurons in the previous layer, applies a weighted sum, and passes the result through an activation function. Let's denote the weighted sum at neuron j in hidden layer i as zij. It is calculated as[99]:

$$z_{ij} = \sum_{K=1}^{N} w_{ijk} \cdot a_{i-1,k} + b_{ij} \qquad (2.15)$$

Where:

w_{ijk} is the weight of the connection between neuron k in layer i-1 and neuron j in layer i.

$a_{i-1,k}$ is the output (activation) of neuron k in layer i-1.

b_{ij} is the bias associated with neuron j in layer i.

N is the number of neurons in layer i-1.

After calculating z_{ij}, it is passed through an activation function, denoted as σ:

$a_{ij} = \sigma\ (z_{ij})$

Common activation functions include the sigmoid function, the hyperbolic tangent function, and the rectified linear unit (ReLU) function.

Output Layer

The output layer processes the information from the final hidden layer and produces the network's output. The output layer can have one or more neurons, depending on the task (e.g., binary classification, multi-class classification, regression)[100]. Similar to the hidden layers, each neuron in the output layer calculates a weighted sum and applies an activation function. The final output of the network is typically represented as a vector Y for multi-class classification or regression:

$Y = [y_{1,} y_{2,}\ldots\ldots ym]$ (2.16)

Where $y_{1,}$ is the output of neuron i in the output layer.

3. Training

MLPs are trained using techniques like backpropagation and gradient descent to minimize a loss or cost function[100]. The goal is to adjust the weights and biases to make the network's output as close as possible to the true target values (ground truth).

4. Loss Function

The loss function, denoted as L, quantifies the error between the network's predictions and the true target values. Common loss functions include mean squared error (MSE) for regression and cross-entropy loss for classification[99].

MLP is a feedforward neural network with multiple layers, including input, hidden, and output layers. It processes input data through weighted connections, applies activation functions, and produces output. The network is trained to minimize a loss function using optimization algorithms like gradient descent. The specific equations and activation functions used depend on the architecture and parameters of the MLP.

2.2.5.1 Hyperparameters of a Multilayer Perceptron

activation function: specifies saturation region, e.g. ReLU, ELU, Tanh

Optimization function: is used to modify the weights and learning rates to minimise error e.g. stochastic gradient descent method (sgd), and the batch gradient descent method (bgd)

Learning rate: this rate governs how quickly the model adjusts to the problem. Higher learning rates produce faster changes and require fewer epochs for training e.g., inverse Scaling.

Number of iterations: number of repetitions in the model execution.

2.2.6 Ensemble Techniques

Ensemble techniques are the methods employed when classification algorithms (base learners) perform poorly in their defined tasks. This technique performs by merging the predictions of various models, retraining these predictions, and in turn producing an optimal predictive model[101]. Ensemble techniques combine multiple models to improve predictive performance. These techniques are based on the idea that aggregating the predictions of several models can often lead to better results than using a single model[102]. Common ensemble methods are bagging, boosting, stacking, and pasting[103].

1. Bagging (Bootstrap Aggregating): Bagging is a technique that involves training multiple base models independently and then combining their predictions. It reduces the variance of the model, making it less prone to overfitting[104]. Bagging is the most straightforward ensemble method. It fits each base classifier on randomised subsets of the original data, drawn with replacement, and then piles up the separate predictions (by averaging or voting) to produce a final classifier[105]. It helps in the reduction of variance in a classification model, by adding randomness to its building. Bagging is an important concept in machine learning because it avoids overfitting data. It is commonly applied to decision trees but can also be used on other classifiers.

The equation for Bagging can be represented as follows: Suppose there is dataset (D) consisting of N examples, and you want to train M base models. Each base model is trained on a subset of the data, which is obtained by resampling D with replacement (bootstrap

samples)[105]. The prediction of each base model is combined by taking a majority vote (for classification) or averaging (for regression):

For classification (Majority Vote):

$$y = \text{mode}(y_1, y_2, y_3, .., , y_M) \quad (2.16)$$

For regression (Averaging):

$$y = \frac{1}{M}\sum_{i=1}^{M} y_i \quad (2.17)$$

Where

y is the final prediction.

y_i is the prediction of the ith base model.

2. Boosting

Boosting is a technique that combines multiple weak learners to create a strong learner. It focuses on improving model accuracy by giving more weight to examples that were previously misclassified[106]. It is an ensemble method that aims to create a better classification model from a combination of serial weak classifiers. In the boosting method, weights are assigned to the training datasets at every iteration[107]. After every iteration, higher weights are assigned to mislabelled observations and lesser weights to the correctly labelled ones[106]. Boosting algorithms aim to find the optimal weights and order of combining base models to minimize the prediction error. The weights are adjusted such that more weight is given to misclassified examples, which helps the ensemble focus on the most challenging cases

The equation for Boosting can be represented as follows: Suppose a dataset (D) consisting of N examples, and you want to train M base models. At each iteration, a new base model (h_m) is trained, and its weight (α_m) is determined based on the model's accuracy. The final prediction is a weighted sum of the base models' predictions:

$$y = \sum_{m=1}^{M} \alpha_m h_m(x) \quad (2.18)$$

Boosting is important in dealing with variance and bias. Examples of boosting techniques are extreme gradient boosting (XGboost), Adaptive boosting (Adaboost), gradient boosting, light gradient boosting machine (Light GBM), and CATboost (Category boosting).

Adaptive Boosting: AdaBoost creates a strong learner through multiple iterations of adding a weak learner in each cycle[107]. The weight vector is also tweaked to account for previously misclassified sample points. Hence, the resulting classifier has higher accuracy. Adaboost is not robust to outliers and noise. The hyperparameters in Adaboost are learning_rate, n_estimators, and base_estimator.

Gradient Boosting: Gradient boosting is an improvement of Adaboost, which aims to minimize the loss function by using the gradient descent optimization method while combining weak learners[108]. The loss function estimates the best model depending on the problem task. Weak learners are added based on the additive model component. Gradient boosting is notably improved as it inculcates subsampling, which encourages randomness of the model, shrinkage reduces the impact of each added learner, and the size of added steps, thus penalising consecutive iteration[108].

Extreme Gradient Boost (XGboost)**:** XGBoost is an improvement of the gradient boost designed to improve speed and performance. This technique employs regularized learning for smoothing and shrinkage to reduce the impact of each tree and make room for future trees to aid improvement, and feature subsampling to prevent over-fitting. All these features, speed up the run time of the algorithm. Parameters of XGboost which can be tuned by the user are eta, gamma, max_depth, seed, eval_metric etc[109].

Light Gradient Boosting Machine: LightGBm uses two gradient-based methods: exclusive feature bundling (EFB) and gradient-based onside sampling (GOSS)[110]. GOSS operates by excluding the portion of the dataset with relatively small gradients and then uses the remaining data to compute the overall information gain. The EFB uses the mutually exclusive features in the dataset, and non-zero values simultaneously to minimise the number of features. This enhances the overall accuracy of the split point[110].

Stacking**:** As opposed to the previous methods which use homogenous weak models, stacking employs heterogeneous weak learners. Learning occurs simultaneously, and then they are combined by training a meta-learner, which then makes predictions using the various models' predictions[106].

2.2.7 Model Evaluation

It is crucial to critically evaluate a classifier's performance and there are several methods and metrics implemented to achieve this. Machine learning models are decision functions because they provide a response or action that supports the decision-making process.

A table of these events is called a confusion matrix. And is shown below

Table 2.1 A Confusion Matrix Layout

	Predicted Positives	Predicted Negatives

Actual Positives	True Positive (TP)	False Negative (FN)
Actual Negatives	False Positive (FP)	True Negative (TN)

Source: [107]

False positives and false negatives are called errors, and the probability of error is the total of the false-positive rate and false-negative rate weighted by their prior probabilities[111]. From the events in the table above, functions like accuracy, specificity, precision, F1, and recall can be calculated.

2.2.7.1 Cross-Validation

Cross-validation is a resampling technique used in machine learning and statistics to assess the performance and generalization ability of a predictive model[112]. It helps in estimating how well a model will perform on unseen data. Cross-validation involves partitioning the dataset into subsets and repeatedly training and evaluating the model on different combinations of these subsets. Cross-validation is especially important in cases where the dataset is limited, and it helps ensure that the model's performance evaluation is robust and not heavily dependent on a particular random data split. It's a valuable tool for model selection, hyperparameter tuning, and assessing a model's expected performance on unseen data[113].

Cross-validation aims to overcome selection bias or overfitting issues and to generate insights into how a model will generalize an independent dataset[113]. Cross-validation refers to the process where these machine learning models are trained on fixed or random subsets of the input data and then tested on the complementary part of the dataset. In summary, cross-validation tests the model's propensity to predict fresh instances that were absent in its estimation (i.e., a newly generated dataset). Cross-validation is performed by taking away a

part of the data set, and keeping it until the final evaluation takes place. This portion is called the validation set[113].

There are different types of cross-validation techniques, but the commonest, easy-to-perform methods are k-fold CV, repeated k-fold CV, leave one out CV, Jackknife CV, Stratified k-fold, and leave p out CV, Monte Carlo (shuffle-split), and time series (rolling cross CV)[114]. In the *k*-fold CV method, the train set is divided into *k*-equal subsets and the model is trained using k−1 of the folds as training data (one part kept till final estimation)[114]. The dataset is divided into k equally-sized or nearly equal-sized subsets or "folds." Each fold contains a portion of the data. The model is trained on k-1 of the folds (training set) and tested on the remaining fold (testing set). This process is repeated k times, with each fold serving as the testing set exactly once[114].

For each iteration of training and testing, a performance metric (e.g., accuracy, mean squared error) is calculated based on the model's predictions on the testing set. The performance metrics from all k iterations are averaged to provide an overall evaluation of the model's performance[114]. The resultant model is evaluated on this reserved subset. Performance metrics reported by the *k*-fold method are the averages of the estimated values computed in the loop. Not carrying out these operations in the loop results in “data leakage”-which means taking a peak into the test set, and this leads to model bias. Other validation techniques are said to follow this baseline procedure[114].

The key benefits of k-fold cross-validation are: It provides a more reliable estimate of a model's performance because it assesses its ability to generalize across different data subsets, It helps identify issues such as overfitting (model performs well on training data but poorly on test data) and underfitting (model performs poorly on both training and test data), It allows for

better utilization of available data, as each data point is used for testing in one of the k iterations.

2.2.7.2 Accuracy

Accuracy depicts the extent to which an algorithm correctly predicts both positive and negative cases. It is a proportion often specified as a percentage[115].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2.19)$$

while classification error is the exact opposite.

$$Classification\ error = \frac{FP+FN}{TP+TN+FP+FN} \quad (2.20)$$

Researchers argue that solely depending on accuracy values tends to be misleading as accuracy is not a good measurement of a model's performance. They rather agree that using Receiver Operator Characteristic (ROC) curve explicitly specifies the number of correctly and incorrectly predicted samples when solving binary problems[116]. However, if the distribution is significantly skewed, ROC curve may give an excessively positive impression of a model's effectiveness.

2.2.7.3 Balanced Accuracy

Balanced Accuracy was developed in a bid to correctly classify samples and minimise classification error. The weights are evenly distributed to each class label[117]. Balanced accuracy is a performance metric used in machine learning to evaluate the effectiveness of a classification model, particularly in cases where the dataset is imbalanced. It is a modified version of traditional accuracy that accounts for the distribution of classes in the dataset. Balanced accuracy is especially useful when the number of instances in different classes varies significantly, making traditional accuracy less informative.

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \quad (2.21)$$

Sensitivity (True Positive Rate or Recall) is the proportion of actual positive cases that the model correctly identifies as positive[118]. It's calculated as $\frac{TP}{TP+FN}$ where TP is the number of true positives, and FN is the number of false negatives.

Specificity is the proportion of actual negative cases that the model correctly identifies as negative[119]. It's calculated as $\frac{TN}{TN+FN}$ where TN is the number of true negatives, and FP is the number of false positives.

Balanced accuracy combines both sensitivity and specificity by taking their average. This metric provides a more balanced view of a classifier's performance, especially in scenarios where the classes are imbalanced. It helps prevent overly optimistic accuracy estimates that can result from models that predict the majority class almost all the time. Balanced accuracy ranges from 0 to 1, with higher values indicating better model performance. A balanced accuracy of 1 implies perfect classification, while 0 means the model is no better than random guessing[119]. It is particularly useful when assessing models for tasks with imbalanced classes, such as fraud detection, disease diagnosis, or rare event prediction. It has a robust performance metric when the cost of false positives and false negatives differs significantly. In such cases, you may want to optimize the model to maximize balanced accuracy rather than traditional accuracy. Balanced accuracy is valuable when comparing different classifiers, especially if the classes are imbalanced, as it provides a more fair and informative assessment of their performance[119].

2.2.7.4 Precision

Precision simply refers to the positive predictive value (PPV) of a model and is calculated as the number of true positives divided by the sum of the true positives and false positives. It is often referred to as confidence level[119]. Precision is a commonly used performance metric in machine learning and statistics, particularly in the context of classification tasks. It measures

the accuracy of positive predictions made by a model and is particularly relevant when dealing with imbalanced datasets or situations where the cost of false positives is high.

The precision of a model is calculated using the following formula:

$$Precision = \frac{TP}{TP+FP} \quad (2.22)$$

Where TP (True Positives) is the number of instances that were correctly predicted as positive by the model.

FP (False Positives) is the number of instances that were incorrectly predicted as positive when they were actually negative. Precision ranges from 0 to 1, with 1 indicating perfect precision and 0 indicating no precision at all. Precision is particularly important in scenarios where false positives are costly or can have negative consequences. It is often used in combination with other metrics like recall, F1-score, and the receiver operating characteristic (ROC) curve to provide a comprehensive assessment of a model's performance[118]. High precision means that when the model predicts a positive class, it is usually correct.

2.2.7.5 Recall (Sensitivity)

Recall is given as the total number of relevant instances that were retrieved, i.e., the true positive rate (TPR), hit rate or sensitivity. Recall is computed as the incidence of true positives divided by the total prediction (sum of the true positives and the false negatives)[118]. Recall is a more efficient metric in this study because it depicts correct classification and diagnosis (correctly classifying an ASD child, is more important than incorrectly classifying a neurotypical child). With a high sensitivity value, it is less likely for a screening/diagnostic test to return false positives.

$$Recall = \frac{TP}{TP+FN} \quad (2.23)$$

2.2.7.6 Specificity

Specificity is a performance metric used in classification tasks, particularly in the field of machine learning and medical diagnostics. It measures the ability of a model to correctly identify negative instances, or true negatives (TN), and is essential when evaluating models, especially in situations where false positives have significant consequences, such as medical testing. It's calculated as $\frac{\mathbf{TN}}{\mathbf{TN+FN}}$ where TN is the number of true negatives, and FP is the number of false positives.

Specificity is a performance metric that assesses a model's ability to correctly identify negative instances, making it particularly important in scenarios where false positives have significant consequences. It provides a key component in evaluating a model's overall performance in classification tasks.

2.2.7.7 F1 Score

The F1 score function is designed and interpreted as a weight of recall average and precision[120]. The F1 score is a widely used performance metric in machine learning and statistics, especially for binary classification tasks. It combines both precision and recall (sensitivity) into a single metric, making it useful when you want to balance the trade-off between precision and recall. The F1 score is calculated using the following formula:

$$F_1 = 2*\frac{Precision*Recall}{Precision+Recall} \quad (2.24)$$

Where:

Precision is the ratio of true positives (TP) to the sum of true positives and false positives (FP), and it measures the accuracy of positive predictions

Recall (also known as sensitivity or true positive rate) is the ratio of true positives to the sum of true positives and false negatives (FN), and it measures the ability to correctly identify positive instances. The F1 score combines these two metrics into a single value that balances precision and recall. It's particularly useful in cases where false positives and false negatives have different consequences or where you want to find a good trade-off between precision and recall[120].

2.2.7.8 Receiver Operating Characteristic (ROC) Curve

The ROC is a graph of specificity (FPR) against sensitivity (TPR). The curve shows both the correctly labelled positive sample points and the mislabelled negative ones. The area under the curve, popularly termed AUC or AUROC, is the estimator of the classification performance[121]. Excellent performance is a TPR of 1, and an FPR of 0. Hence a good model is one in which all labelled points fall in the upper left section of the curve.

$$AUC = \int_0^1 ROC(t)dt \qquad (2.25)$$

2.2.7.9 Precision–Recall (PR) Curve

The PR curve is a graph of recall against precision and provides an alternative for the evaluation of a model's classification performance. The Precision-Recall (PR) curve is a graphical representation used to evaluate the performance of binary classification models, especially in situations where imbalanced datasets or the trade-off between precision and recall are critical considerations[122]. It provides a visual representation of how a model's precision and recall change at different classification thresholds. The area under the PR curve is often used as a single value to summarize the model's performance. A higher AUC-PR indicates better model performance[122]. The Precision-Recall (PR) curve is a valuable tool for assessing and comparing the performance of binary classification models, especially when precision, recall, and the trade-off between them are important considerations.

2.3 Review of Empirical Studies

In order to classify clinical datasets used to predict the early indications of autism in children and adolescents, a study assessed the efficacy of several machine-learning algorithms and pre-processing techniques. In their work, KNN, RF, and logistic regression were tested together with a few other machine learning approaches for conducting classification across datasets. Their study proved that basic pre-processing procedures, proper data encoding, and various classification algorithms, including Random Forest, logistic regression, KNN, and Random Forest, produce high standards when compared with those produced by state-of-the-art systems[123].

In a work where the authors attempt to fill this gap by developing a novel discriminative few shot learning method to analyze hour-long video data and exploring the fusion of facial dynamics for the trait classification of ASD. Leveraging well-established computer vision tools from spatio-temporal feature extraction and marginal fisher analysis to few-shot learning and scene-level fusion, they constructed a three-category system to classify an individual into Autism, Autism Spectrum, and Non-Spectrum. Experimental results are reported to demonstrate the potential of the proposed automatic ASD trait classification system (achieving 91.72% accuracy on the Caltech ADOS video dataset) and the benefits of few-shot learning and scene-level fusion strategy by extensive ablation studies[124].

In another study that presented a new contrastive learning-based framework for decision tree-based classification of actions, including human-human interactions (HHI) and human-object interactions (HOI). The key idea is to translate the original multi-class action recognition into a series of binary classification tasks on a pre-constructed decision tree. Under the new framework of contrastive learning, they present the design of an interaction adjacent matrix (IAM) with skeleton graphs as the backbone for modeling various action-related attributes

such as periodicity and symmetry. Through the construction of various pretext tasks, the authors obtained a series of binary classification nodes on the decision tree that can be combined to support higher-level recognition tasks. Experimental justification for the potential of their approach in real-world applications ranges from interaction recognition to symmetry detection[125].

In another work that provided a complete overview of autistic children's diagnosis and treatment options. By, using various modalities, autistic diagnosis and therapeutic strategies are provided. Following that, a multimodal sensing technique for autism assessment is implemented. The research concludes with several suggestions regarding future research and difficulties. This paper displays the summarization as per current research in several fields such as autism diagnosis. Intervention theory and the applications with majorly focusing on ICF -based results[126].

A study devises a Jellyfish Search Optimization with Deep Learning Driven ASD Detection and Classification (JSODL-ASDDC) model. The goal of the JSODL-ASDDC algorithm is to identify the different stages of ASD with the help of biomedical data. The proposed JSODLASDDC model initially performs min-max data normalization approach to scale the data into uniform range. In addition, the JSODL-ASDDC model involves JSO based feature selection (JFSO-FS) process to choose optimal feature subsets. Moreover, Gated Recurrent Unit (GRU) based classification model is utilized for the recognition and classification of ASD. Furthermore, the Bacterial Foraging Optimization (BFO) assisted parameter tuning process gets executed to enhance the efficacy of the GRU system. The experimental assessment of the JSODL-ASDDC model is investigated against distinct datasets. The experimental outcomes highlighted the enhanced performances of the JSODL-ASDDC algorithm over recent approaches[127].

In a similar study attempts to build a suitable prediction model enabled by ML technology to predict ASD and dyslexia for people of any age. This work seeks to examine the possible use of Random Forest, SVM with linear kernel, SVM with polynomial kernel, SVM with rbf kernel, SVM with sigmoid kernel, XGBoost, Decision Tree, Logistic Regression, Naïve Bayes, and KNN to forecast and assess ASD and dyslexia difficulties in children, adolescents and adults. Using real data set collected from individuals with and without autistic traits, the proposed model and the AQ-10 screening tool were assessed. The data for dyslexia is made up of 3644 cases with 197 properties, 196 of which are independent variables and one is a dependent variable. The data for autism consists of 704 cases with 22 characteristics, 21 independent variables, and 1 dependent variable with binary values (YES or NO). The results of the research showed that, in terms of accuracy, precision, F1 score, and recall, the recommended prediction model gave better results for the data set[128].

Another study aims to analyze and make a comparison on which prediction model that gives a high accuracy after the feature selection. The importance of attributes is investigated using correlation and the predictive models are constructed for the detection of this disorder in children. The dataset consists of 1054 instances and each instance includes 19 attributes. Experimental results clearly show that using feature selection with 10 attributes can lead the impact of accuracy with predictive model of Random Forest (RF) returns the highest accuracy with 94.78%. The findings also indicated that the number of questions in screening tools can be reduced and give an impact with the good results[129].

Another study aims to propose a data mining architecture that combines behavioural and clinical characteristics with demographic data and to provide a quick, acceptable and easy way to support the ASD diagnosis. this can be performed by conducting a comparison study to determine the efficacy of four possible classifiers: logistic regression (LR), sequential

minimum optimization (SMO), naïve Bayes, and instance-based technique based on k-neighbors (IBK). These classifiers have been performed with Waikato Environment for Knowledge Analysis (WEKA) tools to distinguish autistic adults from healthy, normal subjects. The results showed that, with 99.71%, SMO classification accuracy was 99.71, which exceeded the accuracy of other classifiers. The proposed architecture allows for early detection of ASD, distinguishing between ASD and healthy control subjects[130].

In this study, a model has been proposed for the early prediction and analysis of ASD problems in children, adolescents, and adults. Principal Component Analysis (PCA) is used for the feature reduction method and then various machine learning algorithms have been applied to three ASD datasets relating to Children, adolescents, and adults. 10-fold cross-validation is used to evaluate the performance of applied machine learning algorithms. The result produces from their proposed model shows that Logistic Regression shows outperform comparing to all other benchmark machine learning algorithms in terms of accuracy, precision, recall, and F1-score during the prediction of ASD cases[131].

In a study that develops a recommender model with multi-classifiers to enhance precision in the prediction of ASD. Various machine learning algorithms are experimented to assess the model's performance. The authors showed that Decision Trees and Random Forests exhibit improved performance if analyzed with other algorithms with respect to accuracy, precision, recall, and F1-score as evaluation metrics[132].

In a proposed work that is based on an ASD detection mechanism that uses a convolutional neural network and a particle swarm optimization algorithm (PSO-CNN). Initial pre-processing removes missing information from the dataset. To increase the effectiveness of the suggested, we analyse the four underlying techniques, including the SVM, NB, LR, and PSOCNN with four different dataset types, including ASD Screening Data for Children,

Adults, Children, and Adolescents are used for the ASD Prediction analysis. Outcomes strongly indicate that PSO-CNN based prediction models perform more accurately and efficiently on all of these datasets of 99.1% after using various ML approaches with handling missing information[133].

Another study proposes a feature selection algorithm (which is based on feature correlation and ranking) for early ASD detection on the clinical ASD dataset. The performance of the feature selection algorithm is compared with different machine learning algorithms (LR, GBC, AdaBoost, and DT). The result shows that 5 out of the 30 features with a Logistic Regression model are sufficient to detect Autism with 98.18% accuracy, 98.16% sensitivity, and 98.16% precision. The result also shows that the Gradient Boost achieves 98.18% accuracy with 5 features, and the AdaBoost achieves 97.10% accuracy with 5 features[134].

A similar work revealed that multiple SVMs were employed in their study to categorize a set of ASD patients, alongside the neurotypical cohort, having claimed that the accuracy of a single SVM was found to be low[135].

Another work applied KNN, SVM, RF, SGD, AdaBoost, Naïve Bayes, and CN2 Induction Rule to four ASD datasets from the UCI ML repository to determine the best model. After testing, they obtained 99.7% by SGD for the adult set, 97.2% by RF in the child data, and 99.8% by AdaBoost for the toddler set[136].

Further, a similar study used restrictive kinematic features to study the diagnosis and categorization of ASD (RKF). 23 kids with normal development (TD) and 20 kids with high-functioning autism were enlisted. They were given a motor task to complete that called for the execution of the most varied movement. Five machine learning classifiers: SVM, LDA, DT, RF, and KNN, were trained and tested. With four kinematic features, KNN with (k = 1)

produced the highest evaluation metrics (acc: 88.37%, spec: 91.3%, recall: 85%, AUC: 0.8815)[137].

A study conducted proof-of-concept research in their study in which a supervised machine-learning system was created to segregate accurately between fifteen (15) preschoolers with ASD and 15 neurotypical kids using the straightforward reach-to-drop task[138].

To facilitate early identification of ASD, a study sought to categorize ASD data in a rapid, simple, and straightforward manner. For kids, teens, and adults, three ASD datasets were employed. They employed KNN, RF and SVM methods to categorize the ASD data. To test the classification methods, 100 random samples of the data were chosen. Their results were evaluated using average values. The effectiveness of SVM and RF as ASD classification techniques is demonstrated. For all of the aforementioned datasets, the RF technique specifically categorized the data with a 100% accuracy rate[139].

This research introduces a framework called ASD-DiagNet, which aims to identify individuals with Autism Spectrum Disorder (ASD) from healthy individuals using exclusively fMRI data. The researchers devised and executed a collaborative learning algorithm employing an autoencoder and a single layer perceptron (SLP), leading to enhanced feature extraction quality and optimized model parameters. In addition, they devised and executed a data augmentation method, utilizing linear interpolation on existing feature vectors, enabling the generation of synthetic datasets essential for the training of machine learning models. The suggested methodology is assessed using a publicly available dataset obtained from Autism Brain Imaging Data Exchange. This dataset consists of 1,035 individuals who were gathered from 17 distinct brain imaging centres. Their machine learning approach surpasses other cutting-edge methods used in 10 imaging centres, with a classification accuracy gain of up to 28%, with a maximum accuracy of 82%. The machine learning strategy described in this

work not only produces higher quality results but also offers significant advantages in terms of execution time, with a reduction from 7 hours to just 40 minutes compared to other methods[140].

In this study, three artificial-intelligence techniques were developed, namely, machine learning, deep learning, and a hybrid technique between them, for early diagnosis of autism. The first technique, neural networks [feedforward neural networks (FFNNs) and artificial neural networks (ANNs)], is based on feature classification extracted by a hybrid method between local binary pattern (LBP) and grey level co-occurrence matrix (GLCM) algorithms. This technique achieved a high accuracy of 99.8% for FFNNs and ANNs. The second technique used a pre-trained convolutional neural network (CNN) model, such as GoogleNet and ResNet-18, on the basis of deep feature map extraction. The GoogleNet and ResNet-18 models achieved high performances of 93.6% and 97.6%, respectively. The third technique used the hybrid method between deep learning (GoogleNet and ResNet-18) and machine learning (SVM), called GoogleNet + SVM and ResNet-18 + SVM. This technique depends on two blocks. The first block used CNN to extract deep feature maps, whilst the second block used SVM to classify the features extracted from the first block. This technique proved its high diagnostic ability, achieving accuracies of 95.5% and 94.5% for GoogleNet + SVM and ResNet-18 + SVM, respectively[141].

To improve the precision and time required for diagnosis, machine learning techniques are being used to complement the conventional methods. The authors have applied models such as Support Vector Machines (SVM), Random Forest Classifier (RFC), Naïve Bayes (NB), Logistic Regression (LR), and KNN to our dataset and constructed predictive models based on the outcome. The main objective of their paper is to thus determine if the child is susceptible to ASD in its nascent stages, which would help streamline the diagnosis process.

Based on our results, Logistic Regression gives the highest accuracy for our selected dataset[142].

In this study, the authors obtained slightly earlier detected ASD datasets pertaining to children and highly processed the dataset as needed. Various ML approaches were applied to the collected dataset and compared their performance based on accuracy, precision, recall, f-measure, log loss, kappa statistics, and MCC. They found that Random Forest and XGBoost provide the best performance with 97.70% accuracy. Model Interpretation methods were applied to interpret the models how they predict positive or negative class. In addition to that important feature and the impact of features on each class been found by shapely value. The impact of features on ASD is found to find the risk factors, which are mostly related or responsible for ASD. The study's findings indicate that, when properly tuned, machine learning approaches can offer accurate forecasts of ASD status. According to the findings, the suggested model has the ability to diagnose ASD in its early phases[143].

Another study used FL technique for autism detection by training two different ML classifiers including logistic regression and support vector machine locally for classification of ASD factors and detection of ASD in children and adults. Due to FL, results obtained from these classifiers have been transmitted to central server where meta classifier is trained to determine which approach is most accurate in the detection of ASD in children and adults. Four different ASD patient datasets, each containing more than 600 records of effected children and adults have been obtained from different repository for features extraction. The proposed model predicted ASD with 98% accuracy (in children) and 81% accuracy (in adults)[144].

In this paper, the authors proposed a novel hybrid ensemble model which is the ensemble of ResNet101 and a Bidirectional Gated recurrent unit (Bi-GRU) with a Weighted Average Ensemble (WAE). The preprocessing is initially carried out to remove undesirable aspects

from the input signal, such as noise, computational burden, etc. A convolutional neural network-based ResNet model is utilized for the classification and detection of input data. Bi-GRU neural network learns data from two different sources such as forward and backward to provide extremely precise predictions. We integrate ResNet and Bidirectional Gated recurrent unit (Bi-GRU) to produce accurate classification results and this technique is optimized by using the Chaotic Henry Gas Solubility Optimization (CHGSO) algorithm. To enhance the efficiency of the proposed method we compare the four underlying techniques such as the Support vector machine (SVM), K-Nearest Neighbor (KNN), Modified Grasshopper Optimization Algorithm- Random forest (MGOA-RF), and Deep Neural Network (DNN), and the hybrid model is evaluated with dataset EEG microstates dataset. Accuracy, precision, sensitivity, specificity, F1-score, and MCC are utilized for assessing the classification performance more accurately. Our Hybrid ensemble model attains Sensitivity of 98%, 99% higher Specificity, 98% F1-Score, MCC of 99%, Accuracy of 98%, and Precision of 99%[145].

In another study, the authors adopted different machine learning algorithms including neural network-based classifiers for effective ASD classification and detection. They performed extensive experiments on three ASD datasets (Adult, Adolescent, and Child) to evaluate the models. The experimental results demonstrate that the tree-based classifiers: Random Forest and Decision Tree classifiers outperformed other classifiers by achieving the highest accuracy, F-beta, recall and precision scores. The experimental results demonstrate the robustness and effectiveness of the models. Moreover, we applied SHAP methods to investigate the important features in ASD prediction for all three datasets[146].

Several machine learning models, including logistic regression, k-nearest neighbours, Naïve Bayes, support vector machine, decision trees, random forest, AdaBoost, XGBoost, and deep

neural network, were used in a study to analyse a publically available dataset on children with Autism Spectrum Disorder (ASD). The models were chosen for their capacity to provide model explainability. The support vector machine, logistic regression, and AdaBoost models achieved the highest performance, with accuracy rates ranging from 99% to 100%. Feature significance determined the crucial features required for accurate ASD prediction. The best-performing models exhibited these traits, which were also consistent with clinical assessment of ASD. LIME facilitated the visualization of the logical processes utilized by the models in order to generate their predictions, hence enhancing the comprehension of clinicians. This is necessary for the possible integration of Explainable Artificial Intelligence (XAI) tools into regular clinical practice. Their research results are encouraging and have provided valuable insights in the creation of efficient and dependable automated algorithms that can assist clinicians in accurately detecting autism in children. Utilizing XAI techniques for early and timely assessment has the potential to decrease the number of patients who need to undergo a lengthy and complex diagnostic process. Implementing timely intervention strategies and providing effective therapy guarantee optimal healthcare for our patients[147].

A related study examines the assessment of Transfer Learning (TL) techniques employed in non-clinical analysis. All utilized transfer learning methods are highly recommended and widely recognized for their effectiveness in extracting facial features. We utilized an open-source dataset obtained from Zenodo, comprising of photos depicting children's faces focusing on a specific region of interest within each image. The analysis employs established transfer learning models, including VGG16, VGG19, InceptionV3, AlexNet, and ResNet50, to assess their performance in comparison to the CNN model. All models undergo thorough hyperparameter tuning and are subsequently employed for predicting autistic features, rather than for data classification. InceptionV3 demonstrated superior performance compared to all other models, with prediction accuracies of 87.99% and 84.33% on the validation and test

datasets, respectively. These accuracies are notably greater than those achieved by other models proposed. Furthermore, this is the sole endeavour that has employed a transfer learning model on the dataset provided by Zenodo[148].

A different research conducted an analysis on various machine learning models, including logistic regression, k-nearest neighbours, Naïve Bayes, support vector machine, decision trees, random forest, AdaBoost, XGBoost, and deep neural network. These models were utilized to examine publically accessible datasets of individuals with Autism Spectrum Disorder (ASD) across different age groups, including children, adolescents, and adults. The kid dataset yielded the most successful models through AdaBoost, support vector machine, and logistic regression, achieving an accuracy of 95–99%. Meanwhile, logistic regression, random forest, and support vector machine emerged as the top-performing models for the teenage dataset, achieving an accuracy of 95–100%. The AdaBoost, support vector machine, and logistic regression algorithms achieved exceptional performance on the adult dataset, with accuracy rates ranging from 99% to 100%[149].

A study developed a machine learning (ML) architecture that is capable of effectively analyzing autistic children's datasets and accurately classifying and identifying ASD traits. The authors considered the ASD screening dataset of children in this study. They utilized the SMOTE method to balance the dataset, followed by feature transformation and selection methods. Then, they utilized several classification techniques in conjunction with a hyperparameter optimization approach. The AdaBoost method yielded the best results among the classifiers. They employed ML and statistical approaches to identify the most crucial characteristics for the rapid recognition of ASD patients[150].

This study developed various hybrid systems to diagnose facial feature images for an ASD dataset by combining convolutional neural network (CNN) features. The first approach

utilized pre-trained VGG16, ResNet101, and MobileNet models. The second approach employed a hybrid technique that combined CNN models (VGG16, ResNet101, and MobileNet) with XGBoost and RF algorithms. The third strategy involved diagnosing ASD using XGBoost and an RF based on features of VGG-16-ResNet101, ResNet101-MobileNet, and VGG16-MobileNet models. Notably, the hybrid RF algorithm that utilized features from the VGG16-MobileNet models demonstrated superior performance, reached an AUC of 99.25%, an accuracy of 98.8%, a precision of 98.9%, a sensitivity of 99%, and a specificity of 99.1%[151].

The aim of a study was to detect ASD at an early stage to improve brain development and increase the awareness of parents and caregivers about ASD. Machine learning methods are now used to predict the spectrum of autism. This study provides a comprehensive assessment of documents that use machine learning to predict ASD, as well as data analysis and classification algorithms. This work aims to classify and study the different methods of Machine Learning, as well as to explain the nature of ASD and to evaluate performance and demonstrate research potential using different criteria. This publication serves as a roadmap for imminent researchers who want to work on the topic of ASD prediction using machine learning[152].

In this study, the author presents a useful framework for comparing several Machine Learning (ML) approaches to ASD diagnosis at an early age. The suggested framework incorporates an effective prediction method into the overall design by using a variety of Feature Scaling (FS) techniques, including minmax scalar, principal component analysis, and intuitive visual representation analysis. Our suggested design using the IPD model shows that the overall verification of the features, various functional qualities reflects the adult autism spectrum disorder categorization in its entirety. In our IPD model, we use a chi-squared and probability

density functional hybrid to do the mathematical analysis. The structure is meant to identify and categorize patients according to their observable characteristics. At the end of the day, we can see that the 95.9% accuracy is far higher than the Machine Learning method[153].

In a paper whose objective of this paper is to train and validate various models (Decision Tree, Random Forest, SVM, KNN and Gated Recurrent Units or GRUs), and conduct a comparative analysis based on various performance metrics. We were able to achieve a 100% training and validation accuracy with both, Random Forest and GRUs, after rigorous tuning of the models. KNN classifier yielded the poorest results[154].

In a study that uses the data mining technique in the process of autism detection, which provides multiple beneficial impacts with high accuracy as it identifies the essential genes and gene sequences in a gene expression microarray dataset. For optimally selecting the genes, the Artificial Bee Colony (ABC) Algorithm is utilized in this study. In contrast, the feature selection process is carried out by five different algorithms: tabu search, correlation, information gain ratio, simulated annealing, and chi-square. The proposed work utilizes a hybrid Extreme Learning Machine (ELM) algorithm based Adaptive Neuro-Fuzzy Inference System (ANFIS) in the classification process, significantly assisting in attaining high-accuracy results. The entire work is validated through Java. The obtained outcomes have specified that the introduced approach provides efficient results with an optimal precision value of 89%, an accuracy of 93%, and a recall value of 87%[155].

In another study aims to create a classification model that can predict the likelihood of ASD with the greatest degree of precision. To investigate the potential for predicting and analyzing ASD traits in the Toddler, Child, Adolescent, and adult age groups, we used several supervised Machine Learning (ML) models. These include Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Nave Bayes (NB), Logistic Regression

(LR), and Random Forest (RF). Four publicly available, distinctive non-clinical ASD screening datasets from Kaggle and the UCI machine learning library are used to test these models. The first dataset includes 1054 instances and 19 toddler-related features. The remaining ones consist of 21 traits and 292, 104, and 704 cases involving children, adolescents, and adults, respectively. After implementing different ML approaches over the pre-processing datasets, the results showed that the DT, LR, and RF classifiers are the dominant models. These dominated models achieve the highest prediction accuracy, among other studied models, of about 100% for all the utilized datasets[156].

The objective of this study is to examine the eye-tracking data collected from computer games in order to categorise youngsters as either autistic or neurotypical. This study involved the selection of 20 typically developing children from Namjoo public elementary school and 20 children with autism from the Daily Rehabilitation Clinic for Children with Special Disorders in Tehran. The data were analyzed using a multilayer perceptron neural network, employing a backpropagation learning technique and a scaled conjugate gradient training algorithm. The evaluation of the classification findings was conducted using the confusion matrix. This study involved analyzing the graphs from the game and the gaze fixation data of children in order to conduct a person-to-person evaluation and identify appropriate features for classifying children as either autism or typical. The situation in which all features were considered achieved the maximum output precision and accuracy, with values of 76.6% and 73.8% respectively[157].

A study also aims to bridge a gap in understanding by bringing together the results of recent studies and technologies that use machine learning based approaches for ASD screening in infants and children younger than 18 months. Individuals on the autism spectrum have restricted interests and behaviors and struggle to communicate and interact socially with

others. The prevalence of ASD has increased in recent years. The potential for innovative methods like machine learning to be integrated into established therapeutic practices is quite encouraging. How computers can be taught to recognize patterns in data is the focus of machine learning research. Artificially intelligent technologies can identify signs and symptoms, organize data, make diagnoses, and forecast outcomes. Machine learning is an area of artificial intelligence that focuses on teaching computers new behaviors by observing how they interact with the world. These days, there are a wide variety of machine learning methods available. In this piece, we look at the existing literature on the frequency of ASD in the general community. The main interest of this paper is the academic literature were sought by searching numerous databases[158].

Another study aims to automate the diagnosis process, compare and identify the best classifiers, and find out the most significant traits for supporting detection. We attempted to investigate the use of Logistic Regression, Decision Tree, Naive Bayes, Random Forest, Support Vector Machine, Linear Discriminant Analysis, K-Nearest Neighbors, XGBoost Classifier, Multi-Layer Perceptron for predicting if the person is susceptible to ASD. For the analysis, ASD data for children , children, adolescents, and adults were employed, and the best classifier techniques were determined using measures like recall, precision, F-measures, etc. Feature engineering was performed to identify the most impactful and effective features of the datasets. We identified that LR and MLP models provide the best results in adult and toddler datasets. In case of adolescent data, LR classifier showed the highest performance whereas LR and XGBoost algorithms worked well for child dataset. The results obtained justify that ML and DL techniques showed promise in developing an improved system for supporting the detection of Autism[159].

A study presents a review of the recent literature on different ML approaches for the classification and identification of ASD in children and the study of the characteristics of ASD. The concept of ASD and the techniques used to discover it are explained; at the end of the paper. The most prominent techniques used in ML to identify children with ASD were discussed in detail. Finally, the main objective of our study is to focus on whether the child is at risk of developing ASD spectrum disorder in its early stages and highlight the importance of early diagnosis using ML methods to reduce the symptoms of the disease[160].

Another related study aims to develop a parental autism screening tool termed the Indian Autism Grading Tool (IAGT) for early screening of autism. Data are collected using the Indian Autism Parental Questionnaire and assigned with grades. This dataset is employed to test five supervised machine learning models, which compare classification performance based on accuracy, precision and recall. The most effective model should be used to implement the autism screening application. MLR is known to be more robust and to support fewer data sets, so it can be employed for the implementation of ML-powered mobile applications. MLR achieves the overall accuracy of 97.85%, which equates to 0.72%, 2.37%, 0.84% and 1.54% better than SVM, DT, KNN and GNB respectively. The proposed tool is developed in both Tamil and English. The pilot study is conducted with 30 children and the predictability of the tool is compared with the clinician. Therefore, the tool consistently achieves the same level of accuracy as clinicians[161].

In a proposed system that can handle time-related NL queries in the healthcare domain. A temporal database has been created where various Autism Spectrum Disorders (ASD) real-life problem statements have been stored with time. The proposed system is able to extract data from a temporal database using time-related NL queries with aggregation. Time tick has been assigned for handling temporal queries whereas aggregation methods have been used for

handling computational type queries. Auxiliary verbs are used as an indication of whether the NL query is a temporal query. The proposed system has the capacity to produce structured query language (SQL) from temporal and computational types of queries which are frequently occurring in NL. This system will help the parents of an autistic child to enhance their knowledge about autism after getting many real-life autism problems according to time[162].

A study was conducted to analyse autism spectrum disorder at its onset, with the aim of enhancing brain development and ensuring adequate attention to this disorder by parents and guardians. ML algorithms currently have a crucial function in identifying ASD. The objective of this study is to identify and analyse several machine learning approaches used in the field of autism research. It also seeks to evaluate the characteristics of autism and assess the performance of these techniques using many metrics. Additionally, the study will explore potential avenues for future research in this area. The survey demonstrates that the Random Forest algorithm has favourable results in comparison to other algorithms[162].

A similar study aims to estimate ASD (autism spectrum disorder) at a sooner possible time and increase more accuracy than the previous research and reduce medical costs. The authors want to predict and distinguish between autistic and non-autistic children by using a machine learning approach. Firstly, they have gathered data from the surveillance side as much as possible. They also set some particular questions and try to find maximum accurate answers to all questions. Furthermore, supervised learning algorithms are applied to diagnosis whether children meet the symptoms for ASD. Among all applied algorithms KNN and Random Forest shows maximum accuracy and speed to diagnosis. Above all, their final goal is to create an online tool that can provide machine learning-based analysis to a user to detect autism at an early age precisely[164].

This work puts forward the method for the detection of ASD utilizing Machine Learning (ML) with the features extracted from sMRI (Structural Magnetic Resonance Imaging). Surface morphometric and volumetric morphometric features have been utilized for training the machine learning models. The cross-validation approach has been used to avoid overfitting problem occurred during training and testing steps. Machine learning models such as Random Forest (RF), Extra Trees (ET), Linear Support Vector Machine (SVM), Non - Linear SVM, and K- Nearest Neighbors (KNN) have been used for classification between ASD and controls. To evaluate the performance of classification, accuracy, precision, recall, and ROC-AUC score values have been considered[165].

This work proposes newness toward the approach of ensembling the same model by combining the same DCNN model with different optimizers. The numbers of subjects considered for this work are 484 ASD and 491 Controls. The proposed ensemble model of DCNN with Adam and Nadam optimizer has achieved the accuracy of 77.58%, 77.66%, and 81.35% on the data division ratio of 70:30, 80:20, and 90:10 respectively. Experimental results validate the superior performance of the proposed model compared to the sMRI-based state-of-the-art approaches for the detection of ASD[166].

Another study proposes a method for detecting Autism Spectrum Disorder (ASD) using facial photographs. This method combines histograms of oriented gradients (HOG) and local binary pattern (LBP) characteristics in a hybrid approach. The dataset employed in this study was obtained from Kaggle and consists of 2940 facial photographs, with an equal distribution between children diagnosed with autism and those without autism. Four distinct machine learning methods, including support vector machine (SVM), random forest (RF), K-nearest neighbours (KNN), and decision tree (DT), have been trained using a hybrid combination of

extracted data to detect ASD. The SVM classifier has attained the highest classification performance, with an accuracy rate of 77.96%[167].

This research presents an analysis of machine learning techniques to identify a specific set of circumstances that are collectively predictive of autistic disorder. Furthermore, this article presents an autism management system that utilises knowledge fusion of collected features from mobile apps used by parents for monitoring. This study work examined approximately 65 characteristics of children across various behavioural categories in order to enhance the precision of the predictive machine learning model. In addition, we have implemented several machine learning algorithms, including Decision Tree (DT), Multinomial Naive Bayes (MNB), Random Forest (RF), Adaboost, Multilayer Perceptron (MLP), K-Nearest Neighbour (KNN), Support Vector Machine (SVM), and Logistic Regression (LR). These algorithms were applied to a dataset consisting of 500 children in order to compare their performance with a weighted voting approach. The goal was to achieve a higher accuracy rate of 97%. The results showed that the Weighted Voting Classifier (WVC) outperformed the other state-of-the-art classifiers[168].

In this study, the authors explored the toddler dataset. The selection of the important features in medical applications such as ASD screening therefore plays a key role in model development, since that directly affects classification accuracy. This study examines several feature selections approaches. We use Generalized Learning Vector Quantization (GLVQ) and propose a method, Modified Generalized Learning Vector Quantization (MGLVQ) for selecting important features. They applied knearest neighbor (kNN), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Multilayer Perceptron (MLP) classification machine learning algorithms. The performance of classifiers is evaluated by evaluation methods such as accuracy, Fβ and area

under the curve (AUROC). The results of all classifiers have been examined, and the performance of the MLP in different evaluation methods is excellent. Moreover, we illustrate the MGLVQ technique explores important features that impact the primary diagnostic procedure of ASD traits in children . As a result, the proposed machine learning model could be effective in detecting ASD at an earlier stage[169].

The objective of this study is to build an online application that can accurately identify ASD at an early stage using machine learning.in this paper there is an attempt to explore the possibility to use Naïve Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network and Convolutional Neural Network for predicting and analysis of ASD problems in a child, adolescents, and adults. The proposed techniques are evaluated on publicly available three different non-clinically ASD datasets. First dataset related to ASD screening in children has 292 instances and 21 attributes. Second dataset related to ASD screening Adult subjects contains a total of 704 instances and 21 attributes. Third dataset related to ASD screening in Adolescent subjects comprises of 104 instances and 21 attributes. After applying various machine learning techniques and handling missing values, results strongly suggest that CNN based prediction models work better on all these datasets with higher accuracy of 99.53%, 98.30%, 96.88% for Autistic Spectrum Disorder Screening in Data for Adult, Children, and Adolescents respectively[170].

In this paper, some of the research works in the field of application of AI, ML, and IoT in autism were reviewed. State-of-the-art articles were collected and around 58 articles were selected which have significant contribution in this field. The selected research works were analyzed, represented, and compared. Finally, incorporation of the autism facilities in smart city environment is described, some research gaps and challenges were pointed out, and recommendations were provided for further research work[171].

This study uses several deep convolutional neural network (CNN)-based transfer learning approaches to detect autistic children using the facial image. An empirical study is conducted to select the best optimizer and set of hyperparameters to achieve better prediction accuracy using the CNN model. After training and validating with the optimized setting, the modified Xception model demonstrates the best performance by achieving an accuracy of 95% on the test set, whereas the VGG19, ResNet50V2, MobileNetV2, and EfficientNetB0 achieved 86.5%, 94%, 92%, and 85.8%, accuracy, respectively. Our preliminary computational results demonstrate that our transfer learning approaches outperformed existing methods. Our modified model can be employed to assist doctors and practitioners in validating their initial screening to detect children with ASD disease[172].

The research describes an effective deep learning-based, data-centric approach for diagnosing autism spectrum disorder from facial images. To classify ASD and non-ASD subjects, this method requires training a convolutional neural network using the facial image dataset. As a part of the data-centric approach, this research applies pre-processing and synthesizing of the training dataset. The trained model is subsequently evaluated on an independent test set in order to assess the performance matrices of various data-centric approaches. The results reveal that the proposed method that simultaneously applies the pre-processing and augmentation approach on the training dataset outperforms the recent works, achieving excellent 98.9% prediction accuracy, sensitivity, and specificity while having 99.9% AUC. This work enhances the clarity and comprehensibility of the algorithm by integrating explainable AI techniques, providing clinicians with valuable and interpretable insights into the decision-making process of the ASD diagnosis model[173].

Another study suggests the detection of ASD in high-functioning adults with the contribution of Transfer Learning. Decision Trees, Logistic Regression and Transfer Learning were

applied on a dataset consisting of high-functioning ASD adults and controls, who looked for information within web pages. A high classification accuracy was achieved regarding a Browse (80.50%) and a Search (81%) task showing that our method could be considered a promising tool regarding automatic ASD detection. Limitations and suggestions for future research are also included (Kollias KF, Syriopoulou-Delli CK, Sarigiannidis P, Fragulis GF. Autism detection in High-Functioning Adults with the application of Eye-Tracking technology and Machine Learning[174].

In another study that designs an Equilibrium Optimizer with Deep Learning Model for Autism Spectral Disorder Classification (EODL-ASDC) technique. The presented EODL-ASDC technique mainly focuses on the identification and classification of ASD. To attain this, the presented EODL-ASDC technique exploits the deep belief network (DBN) system to perform the classification procedure. In addition, the EO algorithm is employed for the optimal hyperparameter tuning of the DBN approach. To demonstrate the enhanced ASD classification result of the EODL-ASDC approach, an extensive range of experimental evaluates was executed. The experimental results demonstrate the improvements of the EODL-ASDC technique over other approaches[175].

2.4 Summary of Gap in Literature Reviewed

This section gave an in-depth review of various concepts used in this study and was organised into four sub-headings conceptual review, theoretical review/framework, review of empirical studies related to the research topic and conceptual model. The conceptual review explained in depth the concepts of the study. These concepts are Autism spectrum disorder (ASD), Autism spectrum disorder (ASD) in children, detection of Autism spectrum disorder and machine learning. It also richly gave insights into sub-concepts such as types of machine learning, and classification of Autism spectrum disorder using machine learning which

includes various classification algorithms like GBoost, Support Vector Regression, Extreme Learning Machines (SVM), Gaussian Naive Bayes (GNB), Multilayer Perceptron, Esemble Technique and K-Nearest Neighbour (KNN). Other sub-concept which includes analysis and interpretation of ML model, model optimization and data balancing were also explained. In the methodological review, the main classification algorithms used in this study were fully explained. They are GBoost, Support Vector Regression, Extreme Learning Machines (SVM), Gaussian Naive Bayes (GNB), Multilayer Perceptron, Ensemble technique and K-Nearest Neighbour (KNN) including their hyper parameters

In the review of empirical studies, several studies on classifications of Autism using machine learning algorithms were presented. The studies show that many empirical research works similar to the topic under study have been carried out. However, past empirical studies using other algorithms showed lower accuracy, lower f1 scores and precision. Also, from previous comparative analyses on ML algorithms it appears that models like SVM, KNN, RF, GNB, MLP, XGBoost, Bagging and Stacking produce the highest tallies of best performances[40,50,160,165].

An inquisition into their underlying framework has been ignited, hence, a thorough exploration of the aforementioned models is pertinent. With the motive of unbiased comparisons, the optimal model, i.e. the model with the best hyperparameters will be employed. Evaluation metrics will be computed, and the scoring metric in the comparison will be sensitivity (recall), area under the Precision-Recall curve, and area under the ROC curve, since using accuracy as the performance metric may not be best practice.

Endnotes

1. A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, I. Castiglioni. *Use of machine learning to identify children with autism and their motor abnormalities.* **Journal of Autism and Developmental Disorders**. 2015 Jul;45:2146-56.

2. T. Eslami, V. Mirjalili, A. Fong, A.R. Laird, F. Saeed. *ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data*. Frontiers in Neuroinformatics. 2019 Nov 27;13:70.

3. Division of Research, American Psychiatric Association. *Highlights of changes from dsm-iv to dsm-5: Somatic symptom and related disorders.* Focus. 2013 Oct;11(4):525-7.

4. O.A. Saffar. "*The effectiveness of computer-based sight words to improve reading skill in autism*." Discuss. Way. 2019;10(14).

5. D.L. Christensen, M.J. Maenner, D. Bilder, J.N. Constantino, J. Daniels, M.S. Durkin, R.T. Fitzgerald, M. Kurzius-Spencer, S.D. Pettygrove, C. Robinson, J. Shenouda. P*revalence and characteristics of autism spectrum disorder among children aged 4 years—early autism and developmental disabilities monitoring network, seven sites, United States, 2010, 2012, and 2014.* MMWR Surveillance Summaries. 2019 Apr 4;68(2):1.

6. S. Jacob, J.J. Wolff, M.S. Steinbach, C.B. Doyle, V. Kumar, J.T. Elison. *Neurodevelopmental heterogeneity and computational approaches for understanding autism*. Translational Psychiatry. 2019 Feb 4;9(1):63.

7. H.S. Nogay, H. Adeli. *Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging.* Reviews in the Neurosciences. 2020 Nov 18;31(8):825-41.

8. D.S American Psychiatric Association, American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5. Washington, DC: American Psychiatric Association; 2013 May 22.*

9. C. Bridgemohan, D.M. Cochran, Y.J. Howe, K. Pawlowski, A.W. Zimmerman, G.M. Anderson, R. Choueiri, L. Sices, K.J. Miller, M. Ultmann, J. Helt. *Investigating potential biomarkers in autism spectrum disorder*. Frontiers in Integrative Neuroscience. 2019 Aug 2;13:31.

10. S. Jaffer, I. Abdulazez, N. Al-Qazzaz, T. Yousif. *Data Mining for Autism Spectrum Disorder detection among Adults*. **Al-Nahrain Journal for Engineering Sciences.** 2022 Dec 23;25(4):142-51.

11. C.E. Blackmore, E.L. Woodhouse, N. Gillan, E. Wilson, K.L. Ashwood, V. Stoencheva, A. Nolan, G.M. McAlonan, D.M. Robertson, S. Whitwell, Q. Deeley. *Adults with autism spectrum disorder and the criminal justice system: An investigation of prevalence of contact with the criminal justice system, risk factors and sex differences in a specialist assessment service.* Autism. 2022 Nov;26(8):2098-107.

12. C.S. Kanimozhiselvi, D. Jayaprakash, K.S. Kalaivani. *Grading autism children using machine learning techniques*. **International Journal of Applied Engineering Research**. 2019;14(5):1186-8.

13. S. Broder Fingert, A. Carter, K. Pierce, W.L. Stone, A. Wetherby, C. Scheldrick, C. Smith, E. Bacon, S.N. James, L. Ibanez, E. Feinberg. *Implementing systems-based innovations to improve access to early screening, diagnosis, and treatment services for children with autism spectrum disorder: An Autism Spectrum Disorder Pediatric, Early Detection, Engagement, and Services network study*. Autism. 2019 Apr;23(3):653-64.

14. N.I. Berger, A.L. Wainer, J. Kuhn, K. Bearss, S. Attar, A.S. Carter, L.V. Ibanez, B.R. *IngersCharacterizing available tools for synchronous virtual assessment of children with suspected autism spectrum disorder: A brief report*. **Journal of Autism and Developmental Disorders**. 2021 Feb 19:1-2.

15. I.A. Ahmed, E.M. Senan, T.H. Rassem, M.A. Ali, H.S.A. Shatnawi, S.M. Alwazer, M. Alshahrani. Eye *Tracking-Based Diagnosis and Early Detection of Autism Spectrum Disorder Using Machine Learning and Deep Learning Techniques.* Electronics, 2022:11(4), p.530.

16. T. Eslami, J.S. Raiker, F. Saeed, 2021. *Explainable and scalable machine learning algorithms for detection of autism spectrum disorder using fMRI data*. In Neural engineering techniques for autism spectrum disorder (pp. 39-54). Academic Press.

17. C. Tang, W. Zheng, Y. Zong, N. Qiu, C. Lu, X. Zhang, X. Ke, C. Guan. *Automatic identification of high-risk autism spectrum disorder: a feasibility study using video and audio data under the still-face paradigm*. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020 Sep 29;28(11):2401-10.

18. National Collaborating Centre for Mental Health (UK), *Attention deficit hyperactivity disorder: diagnosis and management of ADHD in children, young people and adults. Leicester (UK): British Psychological Society, 2018. National Institute for Health and Care Excellence: Clinical Guidelines)*. http://www.ncbi.nlm.nih.gov/books/NBK493361/. (consulted in March 2018).

19. J. Baio, L. Wiggins, D.L. Christensen, M.J. Maenner, J. Daniels, Z. Warren, M. Kurzius-Spencer, W. Zahorodny, C.R. Rosenberg, T. White, M.S. Durkin. CDC. *Prevalence of autism spectrum disorder among children aged 8 years-Autism and Developmental Monitoring Network, 11 Sites, United States.* MMWR Surveillance Summaries, 67 (6), 1-23.

20. N. Zhang, M. Ruan, S. Wang, L. Paul, X. Li. *Discriminative few shot learning of facial dynamics in interview videos for autism trait classification*. IEEE Transactions on Affective Computing. 2022 May 30.

21. X.A. Bi, Y. Wang, Q. Shu, Q. Sun, Q. Xu. *Classification of autism spectrum disorder using random support vector machine cluster*. Frontiers in Genetics. 2018 Feb 6;9:18.

22. T. Iidaka. *Resting state functional magnetic resonance imaging and neural network classified autism and control. Cortex*. 2015 Feb 1;63:55-67.

23. A. Abraham, MP Milham, A Di Martino, RC Craddock, D Samaras, B Thirion, G Varoquaux. *Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example*. NeuroImage. 2017 Feb 15;147:736-45.

24. S. Itani, D Thanou. *Combining anatomical and functional networks for neuropathology identification: A case study on autism spectrum disorder*. Medical image analysis. 2021 Apr 1;69:101986.

25. V. Subbaraju, M.B. Suresh, S. Sundaram, S. Narasimhan. *Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging: A spatial filtering approach.* Medical image analysis. 2017 Jan 1;35:375-89.

26. A.J. Fredo, A. Jahedi, M. Reiter, R.A. Müller. *Diagnostic classification of autism using resting-state fMRI data and conditional random forest*. Age. 2018 Jul;12(2.76):6-41.

27. H. Chen, X. Duan, F. Liu, F. Lu, X. Ma, Y. Zhang, L.Q. Uddin, H. Chen. *Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity—a multi-center study*. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2016 Jan 4;64:1-9.

28. A.S. Heinsfeld, A.R. Franco, R.C. Craddock, A. Buchweitz, F. Meneguzzi. *Identification of autism spectrum disorder using deep learning and the ABIDE dataset.* NeuroImage: Clinical. 2018 Jan 1;17:16-23.

29. E. Loth, T. Charman, L. Mason, J. Tillmann, E.J. Jones, C. Wooldridge, J. Ahmad, B. Auyeung, C. Brogna, S. Ambrosino, T. Banaschewski. *The EU-AIMS Longitudinal European Autism Project (LEAP): design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. Molecular autism*. 2017 Dec;8(1):1-9.

30. M.K. Kwon, A. Moore, C.C. Barnes, D. Cha, K. Pierce. *Typical levels of eye-region fixation in children with autism spectrum disorder across multiple contexts*. **Journal of the American Academy of Child & Adolescent Psychiatry**. 2019 Oct 1;58(10):1004-15.

31. A.M. Bur, M. Shew, J. New. *Artificial intelligence for the otolaryngologist: a state of the art review*. Otolaryngology–Head and Neck Surgery. 2019 Apr;160(4):603-11.

32. R. Gupta. *A survey on machine learning approaches and its techniques*. In2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) 2020 Feb 22 (pp. 1-6). IEEE.

33. N. Al-Azzam, I. Shatnawi. *Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer.* Annals of Medicine and Surgery. 2021 Feb 1;62:53-64.

34. M.A. El Mrabet, K. El Makkaoui, A. Faize. *Supervised machine learning: A survey*. In2021 4th International Conference on Advanced Communication Technologies and Networking (CommNet) 2021 Dec 3 (pp. 1-10). IEEE.

35. K.K. Hiran, R.K. Jain, K. Lakhwani, R. Doshi. *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples (English Edition).* BPB Publications; 2021 Sep 16.

36. P.C. Sen, M. Hajra, M. Ghosh. *Supervised classification algorithms in machine learning: A survey and review.* InEmerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018 2020 (pp. 99-111). Springer Singapore.

37. N.P. Tigga, S. Garg. *Prediction of type 2 diabetes using machine learning classification methods*. Procedia Computer Science. 2020 Jan 1;167:706-16.

38. S. Uddin, A. Khan, M.E. Hossain, M.A. Moni. *Comparing different supervised machine learning algorithms for disease prediction.* BMC medical informatics and decision making. 2019 Dec;19(1):1-6.

39. E. Hamza, M. Azizbek. *Regression based on decision tree algorithm*. Universum: технические науки. 2022(6-6 (99)):55-8.

40. S. Venkatesan. *Design an intrusion detection system based on feature selection using ML algorithms.* Mathematical Statistician and Engineering Applications. 2023 Feb 18;72(1):702-10.

41. P. Palimkar, R.N. Shaw, A. Ghosh. *Machine learning technique to prognosis diabetes disease: Random forest classifier approach*. InAdvanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021 2022 (pp. 219-244). Springer Singapore.

42. M. Zounemat-Kermani, O. Batelaan, M. Fadaee, R. Hinkelmann. *Ensemble machine learning paradigms in hydrology: A review*. **Journal of Hydrology**. 2021 Jul 1;598:126266.

43. S. Jain, D.S. Chauhan. *Instance-based learning of marker proteins of carcinoma cells for cell death/survival*. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization. 2020 May 3;8(3):313-22.

44. N. Boyko, K. Boksho. *Application of the naive bayesian classifier in work on sentimental analysis of medical data*. InIDDM 2020 Nov 19 (pp. 230-239).

45. H. Jafarzadeh, M. Mahdianpari, E. Gill, F. Mohammadimanesh, S. Homayouni. *Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation.* Remote Sensing. 2021 Nov 2;13(21):4405.

46. S. Sah. *Machine learning: A review of learning types*. Preprints 2020, 2020070230. https://doi.org/10.20944/preprints202007.0230.v1.

47. S. Sharma, N. Batra. *Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering*. In2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon) 2019 Feb 14 (pp. 568-573). IEEE.

48. M. Alkhayrat, M. Aljnidi, K. Aljoumaa. *A comparative dimensionality reduction study in telecom customer segmentation using deep learning and PCA*. **Journal of Big Data**. 2020 Dec;7:1-23.

49. D. Dwiputra, A.M. Widodo, H. Akbar, G. Firmansyah. *Evaluating the performance of association rules in apriori and fp-growth algorithms: Market basket analysis to discover rules of item combinations.* **Journal of World Science**. 2023 Aug 30;2(8):1229-48.

50. L. Ruthotto, E. Haber. *An introduction to deep generative modeling. GAMM-Mitteilungen.* 2021 Jun;44(2):e202100008. https://doi.org/10.1002/gamm.202100008.

51. *B. Ahmad, J. Sun, Q. You, V. Palade, Z. Mao. Brain tumor classification using a combination of variational autoencoders and generative adversarial networks*. Biomedicines. 2022 Jan 21;10(2):223.

52. M. Wahbah, B. Mohandes, T.H. EL-Fouly, M.S. El Moursi. *Unbiased cross-validation kernel density estimation for wind and PV probabilistic modelling.* Energy Conversion and Management. 2022 Aug 15;266:115811.

53. U.C. Jaleel, N.V. Chethala, A. Thottapillil, J.R. KR, S.K. Kukkupuni, P. Kulkarni, A. Safeeda, A.T. Manuel, S. EPA. *Artificial neural network based self organizing maps analysis for clinical trials of Indian systems of medicine*. Available at SSRN: https://ssrn.com/abstract=4468365 or http://dx.doi.org/10.2139/ssrn.4468365

54. V. Garbhapu, P. Bodapati. *A comparative analysis of Latent Semantic analysis and Latent Dirichlet allocation topic modeling methods using Bible data*. **Indian Journal of Science and Technology**. 2020 Dec 15;13(44):4474-82.

55. R. Guo, Z. Luo, M. Li. *A survey of optimization methods for independent vector analysis in audio source separation.* Sensors. 2023 Jan 2;23(1):493.

56. J. Yan, X. Wang. *Unsupervised and semi-supervised learning: the next frontier in machine learning for plant systems biology*. **The Plant Journal**. 2022 Sep;111(6):1527-38.

57. J. Bharadiya. A comprehensive survey of deep learning techniques natural language processing. **European Journal of Technology**. 2023 May 23;7(1):58-66.

58. L. Yang, W. Zhuo, L. Qi, Y. Shi, Y. Gao. St++: *Make self-training work better for semi-supervised semantic segmentation*. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022 (pp. 4268-4277).

59. A. Kumar, J Yadav. *A review of feature set partitioning methods for multi-view ensemble learning.* Information Fusion. 2023 Aug 1:101959.

60. H. Sadr, M.M. Pedram, M. Teshnehlab. *Multi-view deep network: a deep model based on learning features from heterogeneous neural networks for sentiment analysis.* IEEE access. 2020 May 4;8:86984-97.

61. S. Huang, Z. Liu, W. Jin, Y. Mu. *Bag dissimilarity regularized multi-instance learning*. Pattern Recognition. 2022 Jun 1;126:108583.

62. R. Zemouri. *Semi-supervised adversarial variational autoencoder*. Machine Learning and Knowledge Extraction. 2020 Sep 6;2(3):20.

63. W. Lin, Z. Gao, B. Li. *Shoestring: Graph-based semi-supervised learning with severely limited labeled data*. arXiv preprint arXiv:1910.12976. 2019 Oct 28.

64. T. Kumarage, S. Ranathunga, C. Kuruppu, N. De Silva, M. Ranawaka. *Generative adversarial networks (GAN) based anomaly detection in industrial software systems*. In2019 Moratuwa Engineering Research Conference (MERCon) 2019 Jul 3 (pp. 43-48). IEEE.

65. K. Nguyen, Y. Nguyen, B. Le. *Semi-supervising learning, transfer learning, and knowledge distillation with SimCLR*. arXiv preprint arXiv:2108.00587. 2021 Aug 2.

66. T.T. Davarakis, G. Palaiokrassas, A. Litke, T. Varvarigou. *Reinforcement learning with smart contracts on blockchains*. Future Generation Computer Systems. 2023 Nov 1;148:550-63.

67. K.A. ElDahshan, H. Farouk, E. Mofreh. *Deep reinforcement learning based video games: A review*. In2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) 2022 May 8 (pp. 302-309). IEEE.

68. M.A. Chadi, H. Mousannif. *Understanding reinforcement learning algorithms: The progress from basic q-learning to proximal policy optimization*. arXiv preprint arXiv:2304.00026. 2023 Mar 31.

69. W. Li, Y. Liu, Y. Ma, K. Xu, J. Qiu, Z. Gan. *A self-learning Monte Carlo tree search algorithm for robot path planning*. Frontiers in Neurorobotics. 2023;17.

70. S.S. Wagner, P. Arndt, J. Robine, S. Harmeling. *Cyclophobic reinforcement learning*. arXiv preprint arXiv:2308.15911. 2023 Aug 30.

71. W. Zhang, Y. Song, X. Liu, Q. Shangguan, K. An. *A novel action decision method of deep reinforcement learning based on a neural network and confidence bound. Applied Intelligence*. 2023 May 29:1-3.

72. C. Wu, J. Ruan, H. Cui, B. Zhang, T. Li, K. Zhang. *The application of machine learning based energy management strategy in multi-mode plug-in hybrid electric vehicle, part i: Twin delayed deep deterministic policy gradient algorithm design for hybrid mode*. Energy. 2023 Jan 1;262:125084.

73. M. Hwang, W.C. Jiang, Y.J. Chen. *A critical state identification approach to inverse reinforcement learning for autonomous systems*. **International Journal of Machine Learning and Cybernetics**. 2022 May;13(5):1409-23.

74. Z. Feng, M. Huang, D. Wu, E.Q. Wu, C. Yuen. *Multi-agent reinforcement learning with policy clipping and average evaluation for uav-assisted communication markov game*. IEEE Transactions on Intelligent Transportation Systems. 2023 Jul 28.

75. A.S. Albahri, R.A. Hamid, A.A. Zaidan, O.S. Albahri. *Early automated prediction model for the diagnosis and detection of children with autism spectrum disorders based on effective sociodemographic and family characteristic features*. Neural Computing and Applications. 2023 Jan;35(1):921-47.

76. H.S. Nogay, H. Adeli. *Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging.* Reviews in the Neurosciences. 2020 Nov 18;31(8):825-41.

77. P. Chhajer, M. Shah, A. Kshirsagar. *The applications of artificial neural networks, support vector machines, and long–short term memory for stock market prediction*. **Decision Analytics Journal**. 2022 Mar 1;2:100015.

78. I.S. Al-Mejibli, J.K. Alwan, H. Abd Dhafar. *The effect of gamma value on support vector machine performance with different kernels*. **International Journal of Electrical and Computer Engineering**. 2020 Oct 1;10(5):5497.

79. M. Bansal, A. Goyal, A. Choudhary. *A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning*. **Decision Analytics Journal**. 2022 Jun 1;3:100071.

80. Y. Lan, Y. Zhang, W. Lin. *Diagnosis algorithms for indirect bridge health monitoring via an optimized AdaBoost-linear SVM. Engineering Structures*. 2023 Jan 15;275:115239.

81. M. Manoj, J.I. Praveen. *A hybrid approach to support the detection of autism spectrum disorder (asd) through machine learning and deep learning techniques*. In2023 12th International Conference on Advanced Computing (ICoAC) 2023 Aug 17 (pp. 1-7). IEEE.

82. M.A. Almaiah, O. Almomani, A. Alsaaidah, S. Al-Otaibi, N. Bani-Hani, A.K. Hwaitat, A. Al-Zahrani, A. Lutfi, A.B. Awad, T.H. Aldhyani. P*erformance investigation of principal component analysis for intrusion detection system using different support vector machine kernels*. Electronics. 2022 Nov 1;11(21):3571.

83. M. Shamsi, S. Beheshti. *Separability and scatteredness (s&s) ratio-based efficient svm regularization parameter, kernel, and kernel parameter selection*. arXiv preprint arXiv:2305.10219. 2023 May 17.

84. Z. Ye, S. Guo, D. Chen, H. Wang, S. Li. *Drilling formation perception by supervised learning: Model evaluation and parameter analysis*. **Journal of Natural Gas Science and Engineering**. 2021 Jun 1;90:103923.

85. C.L. Pessoa, V.H. Peres Silva, R. Stefani. *Prediction of the self-healing properties of concrete modified with bacteria and fibers using machine learning*. **Asian Journal of Civil Engineering**. 2023 Aug 30:1-0.

86. A. Taherkhani, G. Cosma, T.M. McGinnity. AdaBoost-CNN: *An adaptive boosting algorithm for convolutional neural networks to classify multi-class imbalanced datasets using transfer learning*. Neurocomputing. 2020 Sep 3;404:351-66.

87. N.B. Jenkins. *Topic-based classification of MAUDE adverse problem reports through multi-application machine learning.* 2023 (Doctoral dissertation, The George Washington University).

88. M. Bansal, A. Goyal, A. Choudhary. *A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning*. **Decision Analytics Journal**. 2022 Jun 1;3:100071.

89. S. Mangalathu, S.H. Hwang, E. Choi, J.S. Jeon. *Rapid seismic damage evaluation of bridge portfolios using machine learning techniques*. Engineering Structures. 2019 Dec 15;201:109785.

90. P. Cunningham, S.J. Delany. *k-Nearest neighbour classifiers-A Tutorial. ACM computing surveys (CSUR)*. 2021 Jul 13;54(6):1-25.

91. L. Ali, S.U. Khan, N.A. Golilarz, I. Yakubu, I. Qasim, A. Noor, R. Nour. *A feature-driven decision support system for heart failure prediction based on statistical model and Gaussian naive bayes*. Computational and Mathematical Methods in Medicine. 2019 Nov;2019.

92. S.K. Kiangala, Z. Wang. *An effective adaptive customization framework for small manufacturing plants using extreme gradient boosting-XGBoost and random forest ensemble learning algorithms in an Industry 4.0 environment*. Machine Learning with Applications. 2021 Jun 15;4:100024.

93. A.A. Akinyelu, A.O. Adewumi. *Classification of phishing email using random forest machine learning technique*. **Journal of Applied Mathematics**. 2014 Apr 3;2014.

94. R.A. Disha, S. Waheed. *Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique*. Cybersecurity 2022 Jan 4;5(1):1.

95. P. Probst, M.N. Wright, A.L. Boulesteix. *Hyperparameters and tuning strategies for random forest. Wiley Interdisciplinary Reviews: data mining and knowledge discovery*. 2019 May;9(3):e1301.

96. M. Daviran, A. Maghsoudi, R. Ghezelbash, B. Pradhan. *A new strategy for spatial predictive mapping of mineral prospectivity: Automated hyperparameter tuning of random forest approach*. Computers & Geosciences. 2021 Mar 1;148:104688.

97. A.K. Sahoo, C. Pradhan, H. Das. *Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making.* Nature inspired computing for data science. 2020:201-12.

98. H. Taud, J.F. Mas. *Multilayer perceptron (MLP). Geomatic approaches for modeling land change scenarios*. 2018:451-5.

99. E. Bisong, E. Bisong. *The multilayer perceptron (mlp). Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*. 2019:401-5.

100. M. Abdar, M. Zomorodi-Moghadam, X. Zhou, R. Gururajan, X. Tao, P.D. Barua, R. Gururajan. *A new nested ensemble technique for automated diagnosis of breast cancer*. Pattern Recognition Letters. 2020 Apr 1;132:123-31.

101. O. Sagi, L. Rokach. *Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2018 Jul;8(4):e1249.

102. M. Hosni, I. Abnane, A. Idri, J.M. de Gea, J.L. Alemán. *Reviewing ensemble classification methods in breast cancer. Computer methods and programs in biomedicine*. 2019 Aug 1;177:89-112.

103. I.D. Mienye, Y. Sun. *A survey of ensemble learning: Concepts, algorithms, applications, and prospects*. IEEE Access. 2022 Sep 16;10:99129-49.

104. T. Le Minh, P.T. Truyen, T. Van Phong, A. Jaafari, M. Amiri, N. Van Duong, N. Van Bien, D.M. Duc, I. Prakash, B.T. Pham. *Ensemble models based on radial basis function network for landslide susceptibility mapping*. Environmental Science and Pollution Research. 2023 Aug 23:1-9.

105. A. Asselman, M. Khaldi, S. Aammou. *Enhancing the prediction of student performance based on the machine learning XGBoost algorithm*. Interactive Learning Environments. 2023 Aug 18;31(6):3360-79.

106. J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, M. Asadpour. *Boosting methods for multi-class imbalanced data classification: an experimental review*. **Journal of Big Data**. 2020 Dec;7:1-47.

107. X. Yang, Y. Wang, R. Byrne, G. Schneider, S. Yang. *Concepts of artificial intelligence for computer-assisted drug discovery*. Chemical reviews. 2019 Jul 11;119(18):10520-94.

108. C. Bentéjac, A. Csörgő, G. Martínez-Muñoz. *A comparative analysis of gradient boosting algorithms*. Artificial Intelligence Review. 2021 Mar;54:1937-67.

109. M. Hajihosseinlou, A. Maghsoudi, R. Ghezelbash. A novel scheme for mapping of MVT-Type Pb–Zn Prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm. Natural Resources Research. 2023 Aug 12:1-22.

110. T.D. Lyon. *Child maltreatment, the law, and two types of error*. Child maltreatment. 2023 Aug:10775595231176454.

111. J.M. Kernbach, V.E. Staartjes. *Foundations of machine learning-based clinical prediction modeling: Part ii—generalization and overfitting*. Machine Learning in Clinical Neuroscience: Foundations and Applications. 2022:15-21.

112. P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, A. Brenning. *Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data*. Ecological Modelling. 2019 Aug 24;406:109-20.

113. S. García, J. Luengo, F. Herrera. *Data preprocessing in data mining*. Cham, Switzerland: Springer International Publishing; 2015.

114. M.E. Lokanan, K. Sharma. *Fraud prediction using machine learning: The case of investment advisors in Canada*. Machine Learning with Applications. 2022 Jun 15;8:100269.

115. J. Hagenauer, M. Helbich. *A comparative study of machine learning classifiers for modeling travel mode choice.* Expert Systems with Applications. 2017 Jul 15;78:273-82.

116. J.M. Johnson, T.M. Khoshgoftaar. *Survey on deep learning with class imbalance*. **Journal of Big Data**. 2019 Dec;6(1):1-54.

117. G.S. Handelman, H.K. Kok, R.V. Chandra, A.H. Razavi, S. Huang, M. Brooks, M.J. Lee, H. Asadi. *Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods*. **American Journal of Roentgenology**. 2019 Jan;212(1):38-43.

118. R.A. Poldrack, G. Huckins, G. Varoquaux. *Establishment of best practices for evidence for prediction: a review*. JAMA psychiatry. 2020 May 1;77(5):534-40.

119. C. Goutte, E. Gaussier. *A probabilistic interpretation of precision, recall and F-score, with implication for evaluation*. InEuropean conference on information retrieval 2005 Mar 21 (pp. 345-359). Berlin, Heidelberg: Springer Berlin Heidelberg.

120. J. Fan, S. Upadhye, A. Worster. *Understanding receiver operating characteristic (ROC) curves*. **Canadian Journal of Emergency Medicine**. 2006 Jan;8(1):19-20.

121. L. Simon, R. Webster, J. Rabin. *Revisiting precision recall definition for generative modeling*. InInternational Conference on Machine Learning 2019 May 24 (pp. 5799-5808). PMLR.

122. D.V. G. Devika, R. Chinnaiyan. *Optimized machine learning classification approaches for prediction of autism spectrum disorder*. Ann Autism Dev Disord. 2020; 1 (1). 2020;1001.

123. N. Zhang, M. Ruan, S. Wang, L. Paul, X. Li. *Discriminative few shot learning of facial dynamics in interview videos for autism trait classification*. IEEE Transactions on Affective Computing. 2022 May 30.

124. M. Ruan, X. Yu, N. Zhang, C. Hu, S. Wang, X. Li. *Video-based contrastive learning on decision trees: from action recognition to autism diagnosis*. InProceedings of the 14th Conference on ACM Multimedia Systems 2023 Jun 7 (pp. 289-300.

125. S. Lylath, L.B. Rananavare. *Efficient approach for autism detection using deep learning techniques: A survey*. In 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT) 2022 Oct 13 (pp. 1-6). IEEE.

126. S.R. Sree, I. Kaur, A. Tikhonov, E.L. Lydia, A.A. Thabit, Z.H. Kareem, Y.K. Yousif, A. Alkhayyat. *Jellyfish search optimization with deep learning driven autism spectrum disorder classification*. Cmc-Computers Materials & Continua. 2023 Jan 1;74(1):2195-209.

127. S. Shilaskar, S. Bhatlawande, S. Deshmukh, H. Dhande. *Prediction of autism and dyslexia using machine learning and clinical data balancing*. In 2023 International Conference on Advances in Intelligent Computing and Applications (AICAPS) 2023 Feb 1 (pp. 1-11). IEEE.

128. A.H. Noruzman, N.A. Ghani, N.S. Zulkifli. *A comparative study on autism among children using machine learning classification*. InProceedings of International Conference on Emerging Technologies and Intelligent Systems: ICETIS 2021 Volume 2 2022 (pp. 131-140). Springer International Publishing.

129. S. Jaffer, I. Abdulazez, N. Al-Qazzaz, T. Yousif. *Data mining for autism spectrum disorder detection among adults*. **Al-Nahrain Journal for Engineering Sciences.** 2022 Dec 23;25(4):142-51.

130. T.Y. Rashme, L. Islam, A.A. Prova, S. Jahan. *Autism screening disorder: Early prediction*. In2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON) 2021 Sep 24 (pp. 1-6). IEEE.

131. A.V. Shinde, D.D. Patil. *A multi-classifier-based recommender system for early autism spectrum disorder detection using machine learning*. Healthcare Analytics. 2023 Dec 1;4:100211.

132. A.H. Noruzman, N.A. Ghani, N.S. Zulkifli. *A comparative study on autism among children using machine learning classification*. InProceedings of International Conference on Emerging Technologies and Intelligent Systems: ICETIS 2021 Volume 2 2022 (pp. 131-140). Springer International Publishing.

133. U. Singh, S. Shukla, M.M. Gore. *An improved feature selection algorithm for autism detection*. In2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) 2022 Dec 2 (pp. 1-8). IEEE.

134. X.A. Bi, Y. Wang, Q. Shu, Q. Sun, Q. Xu. *Classification of autism spectrum disorder using random support vector machine cluster*. Frontiers in genetics. 2018 Feb 6;9:18.

135. R. Sujatha, S.L. Aarthy, J. Chatterjee, A. Alaboudi, N.Z. Jhanjhi. *A machine learning way to classify autism spectrum disorder*. **International Journal of Emerging Technologies in Learning (iJET)**. 2021 Mar 30;16(6):182-200.

136. Z. Zhao, X. Zhang, W. Li, X. Hu, X. Qu, X. Cao, Y. Liu, J. Lu. *Applying machine learning to identify autism with restricted kinematic features*. IEEE Access. 2019 Oct 30;7:157614-22.

137. A. Crippa, C. Salvatore, P. Perego, S. Forti, M. Nobile, M. Molteni, I. Castiglioni. *Use of machine learning to identify children with autism and their motor abnormalities*. **Journal of autism and developmental disorders**. 2015 Jul;45:2146-56.

138. U. Erkan, D.N. Thanh. *Autism spectrum disorder detection with machine learning methods*. Current Psychiatry Research and Reviews Formerly: Current Psychiatry Reviews. 2019 Dec 1;15(4):297-308.

139. T. Eslami, V. Mirjalili, A. Fong, A.R. Laird, F. Saeed. *ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data*. Frontiers in neuroinformatics. 2019 Nov 27;13:70.

140. I.A. Ahmed, E.M. Senan, T.H. Rassem, M.A. Ali, H.S. Shatnawi, S.M. Alwazer, S.M. Alshahrani. *Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques*. Electronics. 2022 Feb 10;11(4):530.

141. K. Vakadkar, D. Purkayastha, D. Krishnan. *Detection of autism spectrum disorder in children using machine learning techniques*. SN Computer Science. 2021 Sep;2:1-9.

142. A.A. Hemu, R.B. Mim, M.M. Ali, M. Nayer, K. Ahmed, F.M. Bui. *Identification of significant risk factors and impact for ASD prediction among children using machine learning approach*. In2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) 2022 Apr 21 (pp. 1-6). IEEE.

143. A.A. Hemu, R.B. Mim, M.M. Ali, M. Nayer, K. Ahmed, F.M. Bui. *Identification of significant risk factors and impact for ASD prediction among children using machine learning approach*. In2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) 2022 Apr 21 (pp. 1-6). IEEE.

144. S. Loganathan, C. Geetha, A.R. Nazaren, M.H. Fernandez. *Autism spectrum disorder detection and classification using chaotic optimization based Bi-GRU network: An weighted average ensemble model. Expert Systems with Applications*. 2023 Nov 15;230:120613.

145. M. Masum, I. Nur, M.H. Faruk, M. Adnan, H. Shahriar. *A comparative study of machine learning-based autism spectrum disorder detection with feature importance analysis*. InCOMPSAC 2022: Computer Software and Applications Conference 2022 (Vol. 3).

146. M.S. Magboo, V.P. Magboo. *Explainable AI for autism classification in children*. InAgents and Multi-Agent Systems: Technologies and Applications 2022: Proceedings of 16th KES International Conference, KES-AMSTA 2022, June 2022 2022 Aug 23 (pp. 195-205). Singapore: Springer Nature Singapore.

147. R.V. Bidwe, S. Mishra, S. Bajaj. *Performance evaluation of transfer learning models for ASD prediction using non-clinical analysis*. InProceedings of the 2023 Fifteenth International Conference on Contemporary Computing 2023 Aug 3 (pp. 474-483).

148. V.P. Magboo, M.S. Magboo. *Classification models for autism spectrum disorder.* InInternational Conference on Artificial Intelligence and Data Science 2021 Dec 17 (pp. 452-464). Cham: Springer Nature Switzerland.

149. M.J. Uddin, M.M. Ahamad, P.K. Sarker, S. Aktar, N. Alotaibi, S.A. Alyami, M.A. Kabir, M.A. Moni. *An integrated statistical and clinically applicable machine learning framework for the detection of autism spectrum disorder*. Computers. 2023 Apr 30;12(5):92.

150. B. Awaji, E.M. Senan, F. Olayah, E.A. Alshari, M. Alsulami, H.A. Abosaq, J. Alqahtani, P. Janrao. *Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined cnn features*. Diagnostics. 2023 Sep 14;13(18):2948.

151. V. Kavitha, R. Siva. *Review of machine learning algorithms for autism spectrum disorder prediction.* In2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) 2022 Dec 13 (pp. 608-613). IEEE.

152. S. Sandeep, A. Singh, A. Khamparia. Deep Learning (DL) *based improved probabilistic dense model (IPD) for autism spectrum disorders (ASD) classification analysis*. **Journal of Data Acquisition and Processing**. 2023;38(3):1872. ISSN: 1004-9037 https://sjcjycl.cn/ DOI: 10.5281/zenodo.98549407.

153. S. Lodha, N. Lodha, H. Malani, P. Devashetti, A. Rajguru. *Early diagnosis of autism using machine learning techniques and gated recurrent units*. In2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS) 2022 Mar 25 (Vol. 1, pp. 1152-1158). IEEE. DOI: 10.1109/ICACCS54159.2022.9785287.

154. M. Pushpa, M. Sorna Mageswari. *Early stage autism detection using ANFIS and extreme learning machine algorithm.* **Journal of Intelligent & Fuzzy Systems**, vol. 45, no. 3, pp. 4371-4382, 2023. DOI: 10.3233/JIFS-231608.

155. D.D. Khudhur, S.D. Khudhur. *The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups*. Measurement: Sensors. 2023 Jun 1;27:100774.

156. S. Aminoleslami, K. Maghooli, N. Sammaknejad, S. Haghipour, V. Sadeghi-Firoozabadi. *Classification of autistic and normal children using analysis of eye-tracking data from computer games.* Signal, Image and Video Processing. 2023 Nov;17(8):4357-65.

157. K. Garg, N.N. Das, G. Aggrawal. *A review on: Autism spectrum disorder detection by machine learning using small video*. In2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT) 2023 Jan 19 (pp. 1-8). IEEE. DOI: 10.1109/ICCT56969.2023.10076139.

158. M. Manoj, J.I. Praveen. A hybrid approach to support the detection of autism spectrum disorder (ASD) through machine learning and deep learning techniques. In2023 12th International Conference on Advanced Computing (ICoAC) 2023 Aug 17 (pp. 1-7). IEEE. DOI: 10.1109/ICoAC59537.2023.102499620.

159. A.D. Sallibi, K.M. Alheeti. *Machine learning methods to detect autism among children*. In2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT) 2023 Jul 4 (pp. 224-228). IEEE. DOI: 10.1109/AICCIT57614.2023.10218312.

160. K. Selvi, D. Jayaprakash, S. Poonguzhali. *Early diagnosis of autism using indian autism grading tool*. **Journal of Intelligent & Fuzzy Systems**. 2023 Jan 1(Preprint):1-5. DOI: 10.3233/JIFS-221087.

161. P. Mukherjee, C. Chakraborty, N. Banerji, K.P. Mandal, B. Chakraborty. *Time-related natural language query handling to extract autism spectrum disorder information from temporal database*. InInternational Conference on Advanced Computational and Communication Paradigms 2023 Feb 16 (pp. 133-145). Singapore: Springer Nature Singapore.

162. M. Swedha, A. Devendran. *An analysis of the use of machine learning in the diagnosis of autism spectrum disorder*. InComputational Intelligence for Clinical Diagnosis 2023 Jun 6 (pp. 177-189). Cham: Springer International Publishing.

163. S. Islam, T. Akter, S. Zakir, S. Sabreen, M.I. Hossain. *Autism spectrum disorder detection in children for early diagnosis using machine learning*. In2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) 2020 Dec 16 (pp. 1-6). IEEE. DOI: 10.1109/CSDE50874.2020.9411531.

164. M. Mishra, U.C. Pati. *Autism detection using surface and volumetric morphometric feature of sMRI with Machine learning approach*. InInternational Conference on Advanced Network Technologies and Intelligent Computing 2021 Dec 17 (pp. 625-633). Cham: Springer International Publishing.

165. M. Mishra, U.C. Pati. *A classification framework for Autism Spectrum Disorder detection using sMRI: Optimizer based ensemble of deep convolution neural network with on-the-fly data augmentation*. Biomedical Signal Processing and Control. 2023 Jul 1;84:104686.

166. A. Khanna, M. Mishra, U.C. Pati. *A hybrid feature based approach of facial images for the detection of autism spectrum disorder*. InInternational Conference on Data Analytics and Insights 2023 May 11 (pp. 389-399). Singapore: Springer Nature Singapore.

167. M.M. Rahman, S.A. Mamun. *Prior prediction and management of autism in child through behavioral analysis using machine learning approach*. InRhythms in Healthcare 2022 Sep 14 (pp. 63-77). Singapore: Springer Nature Singapore.

168. M. Bala, M.H. Ali. *Check for early stage prediction model of autism spectrum disorder traits of children* . InComputer, Communication, and Signal Processing. AI, Knowledge Engineering and IoT for Smart Systems: 7th IFIP TC 12 International Conference, ICCCSP 2023, Chennai, India, January 4–6, 2023, Revised Selected Papers 2023 (p. 3). Springer Nature.

169. S. Raj, S. Masood. *Analysis and detection of autism spectrum disorder using machine learning techniques*. Procedia Computer Science, 167, 994-1004, https://doi.org/10.1016/j.procs.2020.03.399.

170. T. Ghosh, M.H. Al Banna, M.S. Rahman, M.S. Kaiser, M. Mahmud, A.S. Hosen, G.H. Cho. *Artificial intelligence and internet of things in screening and management of autism spectrum disorder*. Sustainable Cities and Society. 2021 Nov 1;74:103189. https://doi.org/10.1016/j.scs.2021.103189.

171. M.S. Alam, M.M. Rashid, R. Roy, A.R. Faizabadi, K.D. Gupta, M.M. Ahsan. *Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach*. Bioengineering. 2022 Nov 18;9(11):710. Bioengineering 2022, 9(11), 710; https://doi.org/10.3390/bioengineering9110710.

172. M.S. Alam, M.M. Rashid, R. Roy, A.R. Faizabadi, K.D. Gupta, M.M. Ahsan. *Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach*. Bioengineering. 2022 Nov 18;9(11):710. Technologies 2023, 11(5), 115; https://doi.org/10.3390/technologies11050115.

173. K.F. Kollias, C.K. Syriopoulou-Delli, P. Sarigiannidis, G.F. Fragulis. *Autism detection in High-Functioning Adults with the application of Eye-Tracking technology and Machine Learning*. In2022 11th International Conference on Modern Circuits and Systems Technologies (MOCAST) 2022 Jun 8 (pp. 1-4). IEEE.

174. A. Praveena, N. Senthamilarasi, T.S. Karthik, S.K. Abirami, S. Das. *Equilibrium optimizer with deep learning model for autism spectral disorder classification*. In2022 International Conference on Automation, Computing and Renewable Systems (ICACRS) 2022 Dec 13 (pp. 582-587). IEEE. DOI: 10.1109/ICACRS55517.2022.10029197.

175. N. Japkowicz, M. Shah. *Performance evaluation in machine learning*. Machine Learning in Radiation Oncology: Theory and Applications. 2015:41-56.

Chapter Three

Methodology

3.1 Research Approach

This research focused on the comparison of acclaimed best classifiers centred on evidence-based literature. Parameters of interest upon execution (classification) were classification accuracy, sensitivity, and specificity. The classification techniques employed in this study were Gaussian Naive Bayes, Support Vector Machine, K-Nearest neighbour, Random Forest model, and Multilayer perceptron. The ensemble methods employed are bagging (bootstrap aggregating), boosting (XGboost), and stacking.

3.2 Requirement Specification

All techniques are deployed using Jupyter notebook of the python programming language. Appropriate libraries such as Data Visualization Libraries (Matplotlib for creating various types of graphs and visualizations and Seaborn), Data Manipulation and Analysis Libraries

(Pandas for data manipulation and analysis and NumPy for numerical computations) and Machine Learning Libraries (Scikit-Learn Metrics functions for calculating accuracy, precision, recall, F1-scorefunctions for calculating accuracy, precision, recall, F1-score) were installed when needed.

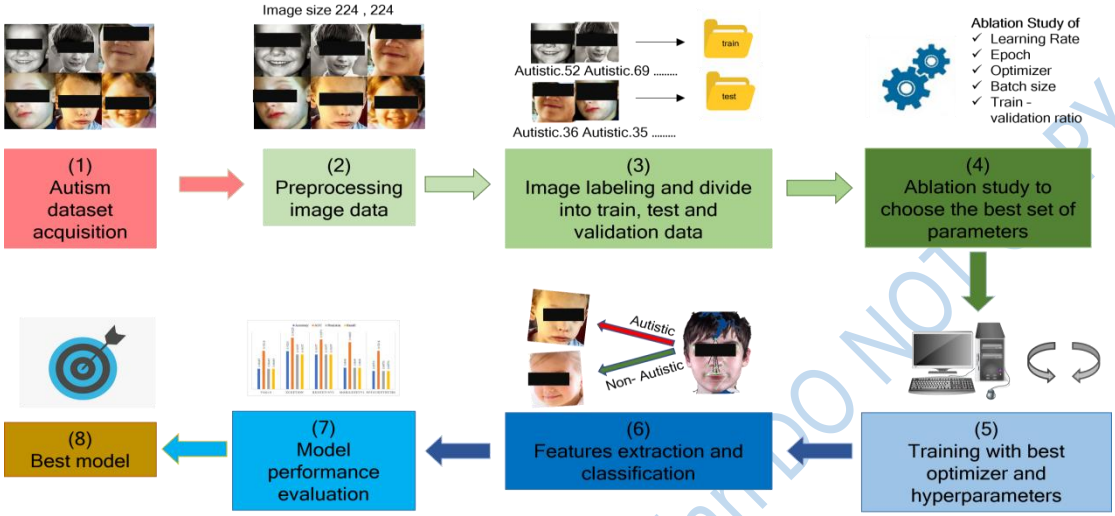

Figure 3.1. Design Framework[1].

3.3 Research Design

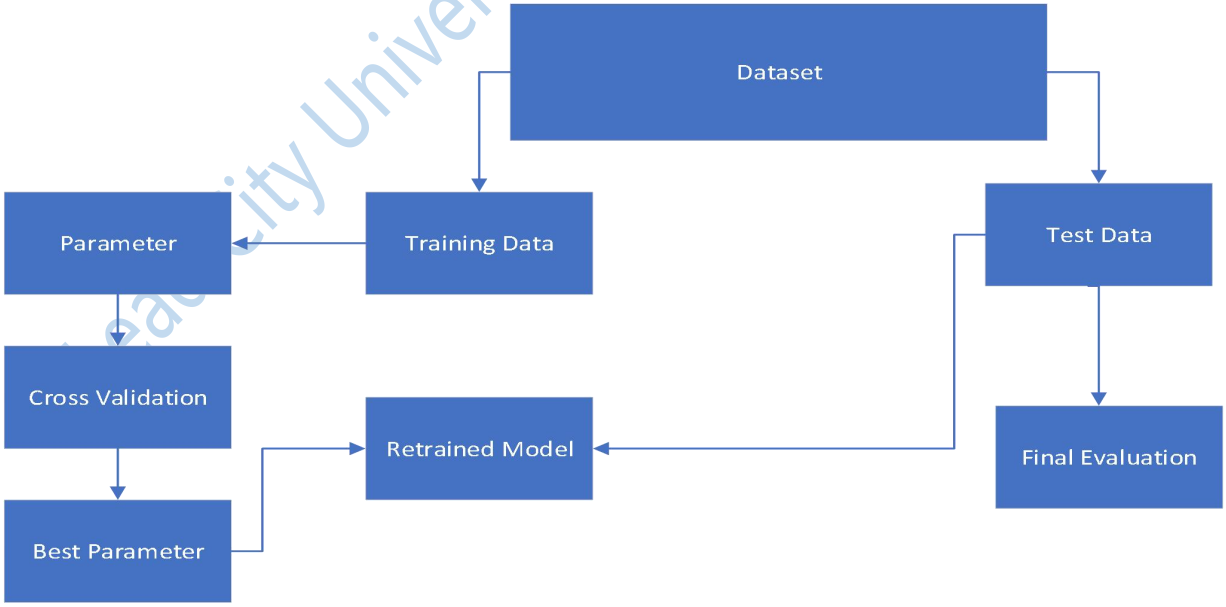

Figure 3.2: Conceptual Framework
Source: Researchers Design, 2024

3.3.1 Data Collection

Various databases and repositories like Kaggle.com, UCI machine learning repository, data.world, Mendeley, and National Database for Autism Research (NDAR) was systematically searched for suitable and viable data, and the Autism dataset for children was chosen from UCI.

The dataset represents a screening mechanism for the diagnosis of autism spectrum disorder in children, and it contains very important features to be used in the classification of ASD outcomes. Ten behavioural characteristics were assessed and recorded based. These said attributes had been proven to be strong determinants of ASD in the behavioural study of ASD. The data is of a continuous nature, majorly nominal /categorical (binary classification) hence it is suitable for classification/prediction purposes.

Eighteen (18) attributes and a total of 1054 records were captured including the class variable. No missing values were encountered. Class variable was assigned based on the subject's total count score after undergoing the screening process using an ASD test application. The questions that captured the independent variables are specified below:

Q1 – Does the child turn to look at you when called by their given name?

Q2 – Is it easy to make eye contact with them?

Q3 – Do they point to show their need for a thing?

Q4 – Do they point to show or share their interest with you?

Q5 – Do they like to play pretend? (talking on the phone, caring for dolls)

Q6 – Do they follow the direction of where you're looking?

Q7 – Do they show signs of sympathy? (wanting to comfort you or a family member when they are visibly upset, e.g hugging, stroking of hair)

Q8 – Description of subject's first words

Q9 – Do they use basic gestures like waving?

Q10 – Do they stare fixatedly at nothing without clear purpose?

A1 to A10 are binary variables since their outputs were classed as "0" or "1". For questions A1 to A9, if the response was "Never", "Rarely", or "Sometimes", "1" was assigned, if "Always" or "Usually", then "0" was recorded. For A10, "1" was recorded when the answer was "Sometimes", "Usually", or "Always", and "0" for "Rarely" and "Never".

Other features (collected from the ASD test app) are:

Age – specified in months (Numeric)

Count score by Q-Chat-10 – Between 1 and 10 (Numeric)

Ethnicity -String variable

Jaundiced when born – Yes/No (Boolean)

Any family member with a history of ASD – Yes/No (Boolean)

Who is taking the test – String

Why the test is being undertaken (String)

Class Variable – Automatically assigned by the ASD test app

The class variable was assigned based on the criteria that if a subject's score was lower than or equal to 3, then they had no ASD traits, if higher than 3, then it means ASD traits had been detected.

3.3.2 Data Preprocessing

The dataset was pre-processed to minimise noise and handle missing values. Data will also be normalised and transformed to stabilise the data and maximize accuracy, and feature selection will be done based on relevance analysis i.e., removing redundant features. Pre-processing data to remove missing values and normalize it is crucial for creating accurate and efficient machine learning models. These steps help in ensuring that the model receives high-quality input data, devoid of anomalies like missing values or vastly different scales among features, which can significantly skew or misguide the learning process.

Identification of Missing Values: Use functions to identify missing values in your dataset. In Python's pandas library, ‘isnull()’ can help identify these.

Handling Missing Values:

i. Drop Rows/Columns: If a column/row contains a high percentage of missing values, it might be practical to drop it using `dropna()`.

ii. Imputation: Replace missing values using Scikit-learn's `SimpleImputer` can automate this process.

Normalizing Data: Normalization adjusts the scale of your data attributes so that they range over a common scale without distorting differences in the ranges of values. There are two common methods:

i. Min-Max Scaling: This technique rescales the data to a fixed range, usually 0 to 1. The formula is ‘(X - min) / (max - min)’. It's useful when you need specific bounded values.

ii. Standardization (Z-score normalization): This technique rescales data so that it has a mean of 0 and a standard deviation of 1.

3.3.3 Data Splitting

Data splitting, also known as data partitioning. It involves dividing a dataset into distinct subsets for various purposes, primarily for model training, validation, and testing. The main goals of data splitting are to assess a model's performance, prevent overfitting, and ensure that the model generalizes well to unseen data. The training set will made up of 70% of the data. The four classification algorithms Support Vector Machine, Gaussian Naïve Bayes, K-Nearest Neighbours, and Multilayer Perceptron will be used to train the dataset and evaluate their

respective performances. The test set will be used to assess the model's generalization performance. It is made up of 30% of the data.

3.3.4 Evaluating the Performance of the Models

Evaluating the performance Support Vector Machine (SVM), Gaussian Naïve Bayes (GNB), K-Nearest Neighbours (KNN), and Multilayer Perceptron (MLP) involves several key steps and metrics. These steps ensure that the evaluation is comprehensive, covering aspects like accuracy, precision, recall, and sometimes more sophisticated metrics like the F1 score.

For each model,

i. Initialize the Model: Create an instance of the model.

ii. Train the Model: Fit the model to the training data.

iii. Predict: Use the trained model to make predictions on the test set.

Step 3: Evaluate the Models

Use various metrics to evaluate model performance. Common metrics include:

i. Accuracy: The proportion of correctly predicted observations to the total observations. Good for balanced classes but not as informative for imbalanced datasets.

ii. Precision: The ratio of correctly predicted positive observations to the total predicted positives. Important when the cost of a false positive is high.

iii. Recall (Sensitivity): The ratio of correctly predicted positive observations to the all observations in actual class. Crucial when the cost of a false negative is high.

iv. F1 Score: The weighted average of Precision and Recall. Useful when you need a balance between Precision and Recall and there is an uneven class distribution.

v. Confusion Matrix: A table showing correct predictions and types of incorrect predictions.

1. Support Vector Machine (SVM)

The SVM was employed for either regression or two-group classification problems, though it is mainly used in classification. In this method, each data point is plotted in n-dimensional space, with each coordinate representing the value of a given feature. Here, classification is done by locating the line that clearly segregates both classes. If there is no error in this separation, then, the Expected value of error is given as

$$E[\Pr(error)] \leq \frac{E[number\ of\ support\ vectors]}{[number\ of\ training\ vectors]} \quad 3.1$$

The decision function was given as

$$D(x) = w\emptyset(x) + b \quad 3.2$$

which is the best line that integrates the training data, w and b are parameters of the SVM, and $\emptyset(x)$ is the function which transforms the data into the new M dimension

$$\frac{D(x)}{||w||} \quad 3.3$$

represents the line, which is the distance between item x and the hyperplane.

The parameters of the linear decision function that will maximize M are:

$$w = \sum_k a_k y_k x_k \quad 3.4$$

$$b = (y_k - w * x_k) \quad 3.5$$

The principal function of the above training algorithms is to solve the equation quadratically.

$$J = \left(\frac{1}{2}\right)||w|| \quad 3.6$$

2. Gaussian Naïve Bayes

Suppose a vector $(x_1, x_2, ...x_n)$ represents the n features of X. Let θ represent the class label of X. The naïve theorem describes the conditional probability of observing X given the class label θ, $P(\theta|X)$ as a product of several simpler probabilities as shown below:

$$P(\theta|f_1, f_2, ...f_n) = \frac{P(\theta)P(f_1, f_2, ..., f_n|\theta)}{P(f_1, f_2, ...f_n)} \quad 3.7$$

Where $(f_1, f_2, \ldots f_n)$ represents the features of vector X. Following the assumption of independence, the probability can be expressed as;

$$P(\theta|f_1, f_2, \ldots, f_{i-1}, f_{i+1}, \ldots f_n) = P(f_i|\theta) \quad 3.8$$

For all i,

$$P(\theta|f_1, f_2, \ldots f_n) = P(\theta)\frac{\prod_{i=1}^{n} P(f_i|\theta)}{P(f_1, f_2, \ldots f_n)} \quad 3.9$$

Assuming every feature is constant, the classification rule follows below:

$$P(\theta|f_1, f_2, \ldots f_n) \alpha\, P(\theta)\prod_{i=1}^{n} P(f_i|\theta), \quad 3.10$$

And,

$$\hat{\theta} = arg\max_{\theta} P(\theta)\prod_{i=1}^{n} P(f_i|\theta) \quad 3.11$$

It is worth noting that for the GNB model, the probability of feature occurrence follows a Gaussian distribution

$$P(f_i|\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta}}\exp\left(\frac{-(f_i-\mu_\theta)}{2\sigma_\theta^2}\right) \quad 3.12$$

whose parameters σ_θ and μ_θ are computed using maximum likelihood.

3. K Nearest Neighbours

Assuming a training set D made up of X_i, $i \in [1, |D|]$ samples, where F described the number of features and assumed normality. Each output was labelled with class label $y_j \in Y$. The aim was to classify an unknown variable q. For every $x_i \in D$, the distance between q and x_i was computed thus:

$$d(q, x_i) = \sum_{f\in F} w_f\, \partial(q_f, x_{if}) \quad 3.13$$

$$\partial(q_f, x_{if}) = \begin{cases} 0 & f\ discrete\ and\ q_f = x_{if} \\ 1 & f\ discrete\ and\ q_f \neq x_{if} \\ |q_f - x_{if}| & f\ discrete \end{cases}$$

upon which the k nearest neighbours are chosen

4. Multilayer Perceptron

The fully connected network is then expressed as:

$h_i^1 = \emptyset^1\left(\sum_j w_{ij}^1 x_j + b_i^1\right)$

$h_i^2 = \emptyset^2\left(\sum_j w_{ij}^2 h_j^1 + b_i^2\right)$

$y_i = \emptyset^3\left(\sum_j w_{ij}^3 h_j^2 + b_i^3\right)$

The activation function is then represented as a vector of units, and weights as a matrix. The resulting vector is:

$\boldsymbol{h}^1 = \emptyset^1(\boldsymbol{W}^1\boldsymbol{X} + \boldsymbol{b}^1)$

$\boldsymbol{h}^2 = \emptyset^2(\boldsymbol{W}^2\boldsymbol{h}^1 + \boldsymbol{b}^2)$

$\boldsymbol{y} = \emptyset^3(\boldsymbol{W}^3\boldsymbol{h}^2 + \boldsymbol{b}^3)$

Finally, transpose the vectors into a single matrix H, for computational ease.

$\boldsymbol{H}^1 = \emptyset^1(\boldsymbol{X}\boldsymbol{W}^{(1)T} + \boldsymbol{1}\boldsymbol{b}^{(1)T})$

$\boldsymbol{H}^2 = \emptyset^2(\boldsymbol{H}^1\boldsymbol{W}^{(2)T} + \boldsymbol{1}\boldsymbol{b}^{(2)T})$

$\boldsymbol{Y} = \emptyset^3(\boldsymbol{H}^1\boldsymbol{W}^{(3)T} + \boldsymbol{1}\boldsymbol{b}^{(3)T})$

3.3.5 Cross-Validation

Cross-validation is very important in data resampling as it aims to avoid overfitting and evaluate the generalizability of the classification models[2]. A 10-fold cross-validation method will be applied to the selected models, and the gridsearchCV function will be employed to tune all hyperparameters to provide the optimal model with the best combination of parameters. GridsearchCv is a search method which provides a thorough procedure while looking for an estimator over all the values in the parameter sample space[3].

3.3.6 Model Retraining

The optimal parameters will then be used to retrain the model. Also, Ensemble models like Random Forest, XGboost, Bagging and Stacking will also be fitted to the optimal models, and then all comparisons will be done based on the results. Instead of a single train/test split, use

cross-validation to get a more reliable estimate of the model's performance using K-fold cross-validation.

While cross-validation and ensemble methods are distinct concepts, they can be used together in the model development process. Cross-validation can be used to evaluate the performance of ensemble methods just like it is used for any other machine learning model. This helps in understanding how well the ensemble method generalizes to unseen data. Also, during the training of some ensemble models, like in boosting, cross-validation can be used to determine the number of iterations or to prevent overfitting by stopping the training process early if the cross-validation score starts to decrease.

In stacking, cross-validation predictions from the base models can be used to train the final model to ensure that the meta-model is not overfitted to the training data. Cross-validation is a technique for assessing model performance, and ensemble methods are strategies for building models.

3.3.7 Model Evaluation

Evaluation will be done based on chosen metrics; accuracy, precision, auroc score, specificity, ROC, and F1 and sensitivity (recall). The confusion matrix will also plotted to aid visualization of the classification results. Grouped bar charts will also be plotted using seaborn and matplotlib, to enhance understanding and visualisation.

Endnotes

1. M.S Alam, MM Rashid, R Roy, AR Faizabadi, KD Gupta, MM Ahsan. *Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach*. Bioengineering. 2022 Nov 18;9(11):710.

2. T.T Wong, PY Yeh. *Reliable accuracy estimates from k-fold cross validation*. IEEE Transactions on Knowledge and Data Engineering. 2019 Apr 25;32(8):1586-94.

3. G.S Ranjan, AK Verma, S Radhika. *K-nearest neighbors and grid search cv based real time fault monitoring system for industries*. In2019 IEEE 5th international conference for convergence in technology (I2CT) 2019 Mar 29 (pp. 1-5). IEEE.

Chapter Four

Result and Discussion of Findings

Introduction

This chapter presents the results and discussion of findings which was based on the data and analysis with respect to the objectives (research questions) of the study.

4.1 Result on Data Preprocessing

4.1.1 Data Importation and Correlational Analysis

All necessary libraries for the execution of the experiment were imported, and the autism data set was imported into the python environment as a *.csv* file. The data were processed and a correlational analysis was done to describe link or relationship between two or several input variables. The correlation coefficient which is the numerical value of this relationship determines the strength or weakness of the association.

From figure 4.1, the heatmap represents correlations, with numbers ranging from -1 to 1, which are typical of correlation matrices. In a correlation matrix heatmap, A `1` or close to `1` (red on this heatmap) indicates a strong positive correlation. A `-1` or close to -1 (dark blue

on this heatmap) indicates a strong negative correlation. A `0` or close to `0` (white or light blue/purple in this heatmap) indicates no correlation.

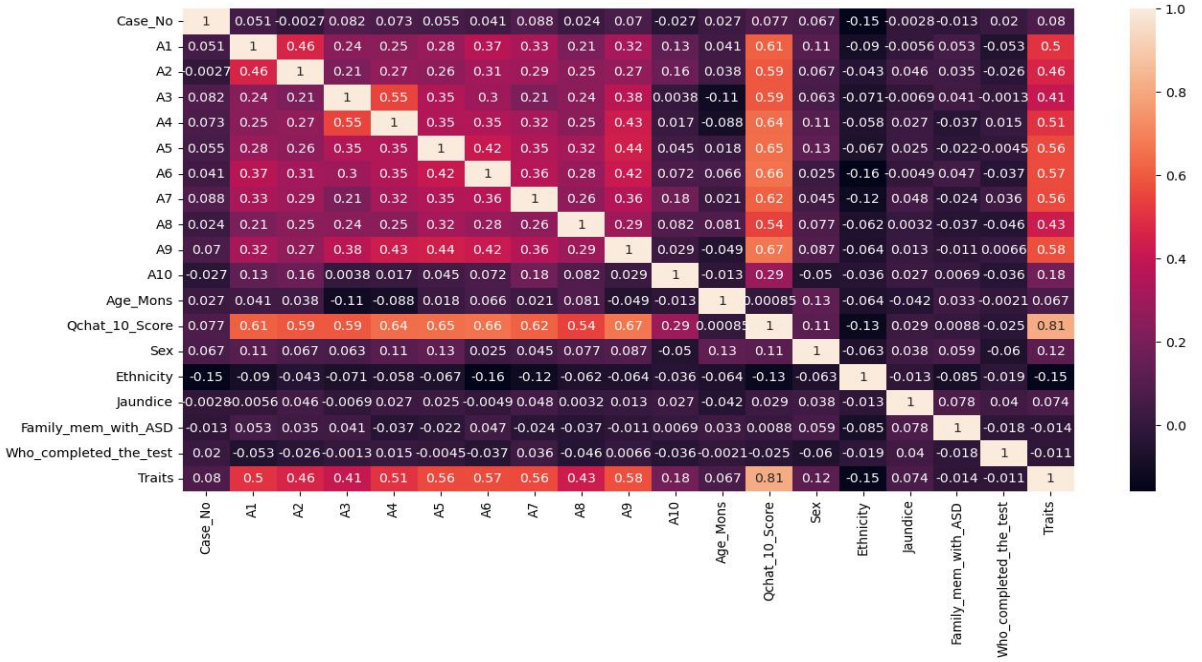

Figure 4.1 First Correlation Heat Map Research Result, 2024

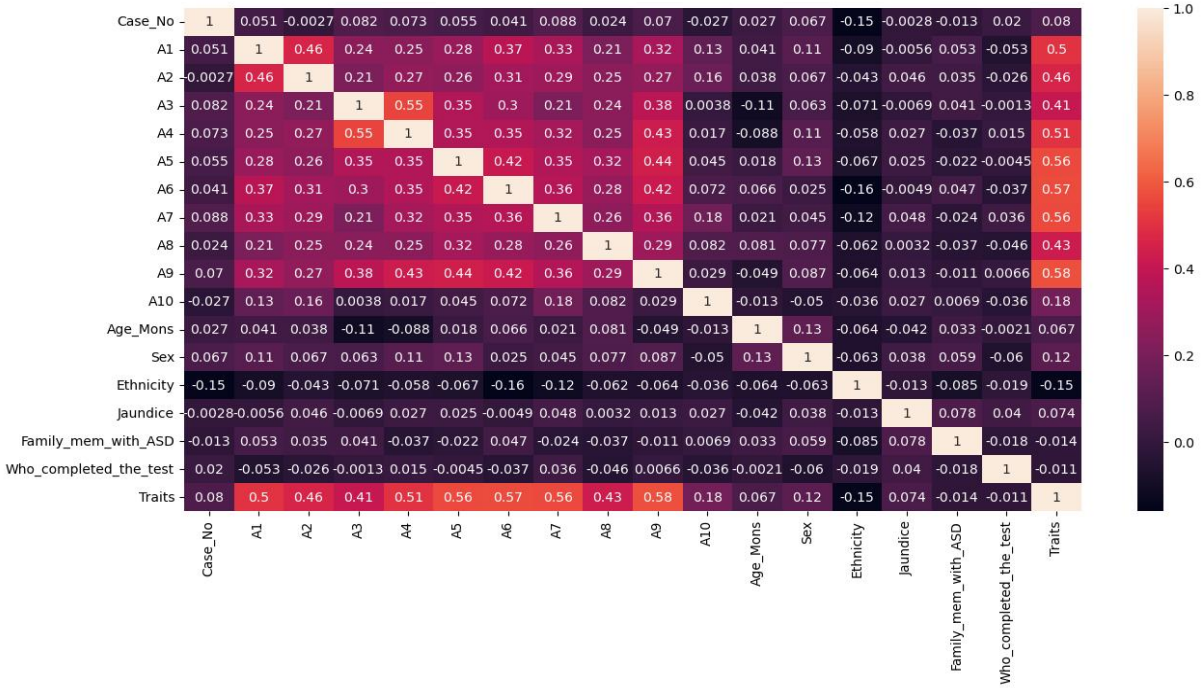

Figure 4.2 Second Correlation Heat Map Research Result, 2024

From figure 4.1 and 4.2, the correlation function, *.corr()*, was applied to the preprocessed data, and column 12, denoted by "Q_Chat_10 was observed to have the highest correlation coefficients. The 12th row and column from the top have predominantly red cells, which indicates high positive correlation coefficients. In the context of predictive modeling or classification, if a feature (like "Q_Chat_10") is used to define the class labels, then it is essentially a perfect predictor of the class. Including this feature in the model would lead to overfitting. Overfitting occurs when the model learns not only the genuine patterns in the training data but also the noise. This makes the model perform very well on the training data but poorly on unseen data because it fails to generalize. It was pertinent to discard the (Q-Chat-10) column since it was the basis for assigning the class label. Retaining it would lead to model overfitting, a grave instance which ought to be avoided.

4.1.2 Data Partitioning

Values for X and y were assigned using the .iloc[] function. This returns the dataset as an array, and it was then partitioned into train and test sets, using the *train_test_split* command, with test_size = 0.2, and random_state = 1. The resulting size of X_train was 13488, X_test = 3376, y_train = 843, and y_test = 211.

4.2 Cross Validation

Cross-validation was used for data resampling as it aims to avoid overfitting and evaluate the generalizability of the classification models[1]. For this project, a 10-fold cross-validation method was applied to the selected models, and the gridsearchCV function was employed to tune all hyperparameters (except in RF where RandomisedSearchCV was implemented); to provide the optimal model with the best combination of parameters. GridsearchCv is a search method which provides a thorough procedure while looking for an estimator over all the values in the parameter sample space[2]. The chosen parameters are those that maximise the

score of the data that was separated in the cross-validation. RandomisedSearchCV, parameter values are sampled at random and only a given number of settings are sampled from the chosen distributions[3]. The number of sampled parameter settings is given by n_iter.

4.3 Application of Classification Algorithm

The optimal parameters of the four classification algorithms Support Vector Machine, Gaussian Naïve Bayes, K-Nearest Neighbours, and Multilayer Perceptron were used to retrain the model and then evaluation was done based on chosen metrics; accuracy, precision, recall, F1, auroc score. The confusion matrix was also plotted to aid visualization of the classification results. Ensemble models like Random Forest, XGboost, Bagging and Stacking were also fitted to the optimal models, and then all comparisons are done based on the results. Grouped bar charts were plotted using seaborn and matplotlib, to enhance understanding and visualisation. All results and outputs are specified as tables and figures where necessary and were also explained in detail.

4.4 Classification Results

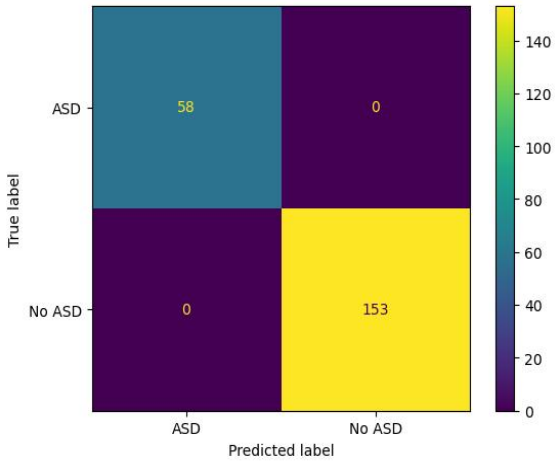

Figure 4.3 SVM Confusion Matrix

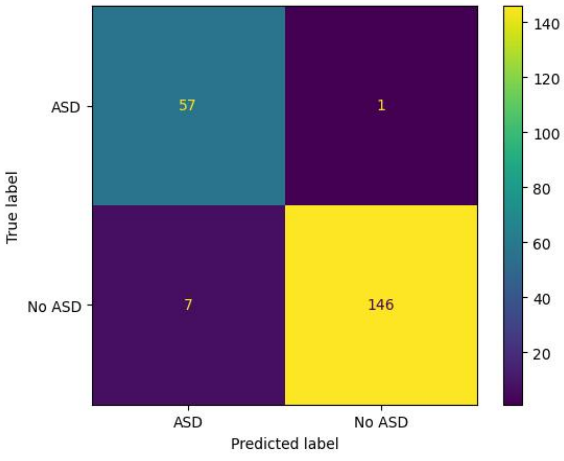

Figure 4.4 KNN Confusion Matrix

From Figure 4.3, the confusion matrix for the Support Vector Machine (SVM) classifier on the autism dataset presents the following information:

True Positives (TP): 58 cases were correctly predicted as having ASD (Autism Spectrum Disorder).

True Negatives (TN): 153 cases were correctly predicted as not having ASD.

False Positives (FP): 0 cases were incorrectly predicted as having ASD when they did not.

False Negatives (FN): 0 cases were incorrectly predicted as not having ASD when they actually did.

The absence of false positives and false negatives suggests that the model has achieved a perfect separation of the classes in the test set. The SVM model correctly classified all the children. No false positives or false negatives were returned.

From Figure 4.4 the confusion matrix for the K-Nearest Neighbors (KNN) classifier presents the following information:

True Positives (TP): 57 cases were correctly predicted as having ASD.

True Negatives (TN): 146 cases were correctly predicted as not having ASD.

False Positives (FP): 1 case was incorrectly predicted as having ASD when it did not.

False Negatives (FN): 7 cases were incorrectly predicted as not having ASD when they actually did.

The high numbers of TP and TN indicate that the KNN classifier is highly accurate. Most individuals are being correctly classified in both categories. However, there are more false negatives (7) than false positives (1). In a medical context, this could be a significant concern because it means the model fails to identify several individuals who actually have ASD. The presence of false negatives affects the recall, which is a measure of the model's ability to find all the relevant cases within a dataset. The single false positive affects precision, the measure of the model's ability to identify only the relevant data points. Since there's only one false.

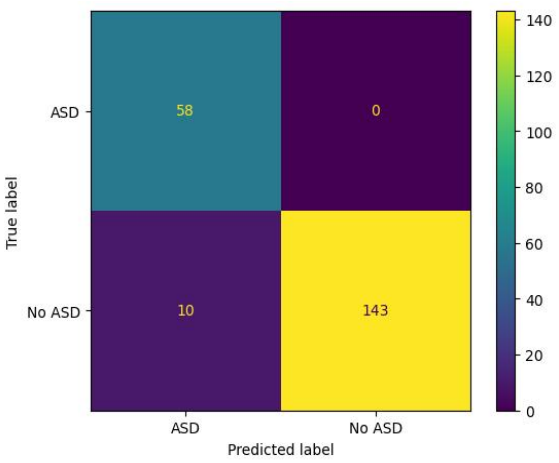

Figure 4.5 GNB Confusion Matrix

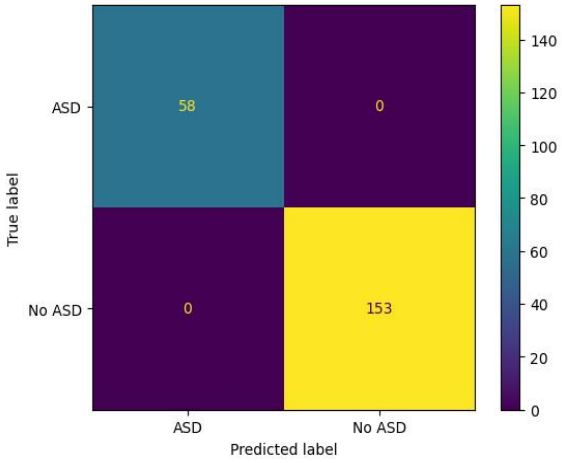

Figure 4.6 MLP Confusion Matrix

The confusion matrix for the Gaussian Naïve Bayes (GNB) classifier on figure 4.5 provides the following information:

True Positives (TP): 58 cases were correctly predicted as having ASD.

True Negatives (TN): 143 cases were correctly predicted as not having ASD.

False Positives (FP): 0 cases were incorrectly predicted as having ASD when they did not.

False Negatives (FN): 10 cases were incorrectly predicted as not having ASD when they actually did.

The GNB classifier correctly identified all cases that it predicted as ASD. This means there were no false positives; however, there were false negatives. The absence of false positives indicates a high level of precision, while the presence of false negatives indicates a lower recall. In the context of autism diagnosis, the false negatives are particularly critical as they represent individuals who have ASD but were not identified by the model. The number of false negatives (10) is higher compared to the previous KNN model (7). The number of true negatives is high, suggesting that the classifier is quite accurate and specific in identifying those without ASD. The accuracy and specificity are important, but they should not overshadow the need for high sensitivity in medical diagnostic tests. The GNB classifier's

performance suggests it is quite specific but less sensitive. For clinical application, it's vital to reduce the number of false negatives to avoid missing ASD diagnoses.

The confusion matrix for the Multilayer Perceptron (MLP) classifier in figure 4.6 shows that

True Positives (TP): 58 cases were correctly identified as ASD;

True Negatives (TN): 153 cases were correctly identified as not having ASD;

False Positives (FP): 0 cases were wrongly identified as having ASD when they did not.

False Negatives (FN): 0 cases were wrongly identified as not having ASD when they did.

The MLP has classified all the test cases correctly, as indicated by the lack of false positives and false negatives. This suggests a very high level of both precision and recall.

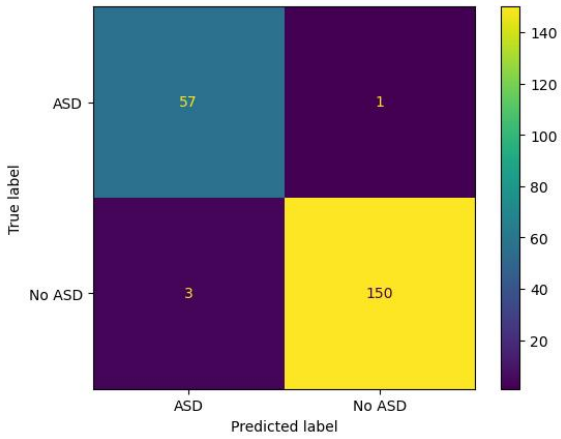

Figure 4.7 RF Confusion Matrix

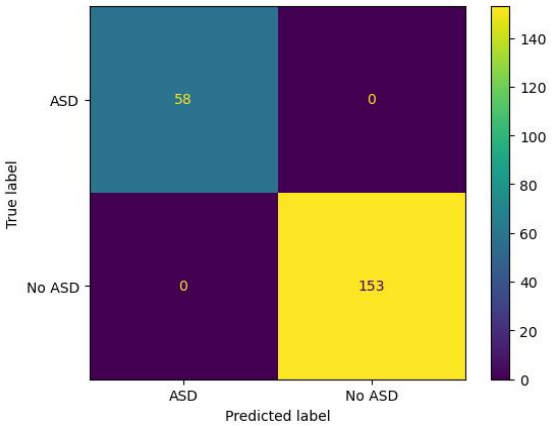

Figure 4.8 XGB Confusion Matrix

In the RF model, True Positives (TP): 57 cases were correctly predicted as ASD.

True Negatives (TN): 150 cases were correctly predicted as not having ASD.

False Positives (FP): 1 case was incorrectly predicted as ASD when it was not.

False Negatives (FN): 3 cases were incorrectly predicted as not having ASD when they did.

The Random Forest classifier has a high number of true positives and true negatives, which indicates a high level of accuracy overall. With only one false positive and three false

negatives, the RF model shows a balance between precision and recall. While precision is almost perfect, indicating that the model is reliable when it predicts ASD, the presence of false negatives means that it sometimes misses ASD cases. The three false negatives are important in a clinical setting because these represent individuals with ASD who have been missed by the classifier. This can be critical, as early detection of ASD can be beneficial for intervention. When compared to the previously evaluated models, the RF classifier offers a balance similar to the KNN, with a slight increase in false negatives but one fewer false positive.

XGBoost (XGB) classifier confusion matrix shows: True Positives (TP): 58 cases were correctly predicted as ASD.

True Negatives (TN): 153 cases were correctly predicted as not having ASD.

False Positives (FP): 0 cases were incorrectly predicted as ASD.

False Negatives (FN): 0 cases were incorrectly predicted as not having ASD.

Similar to the previously discussed MLP model, the XGB classifier has achieved perfect classification with zero false positives and zero false negatives. This indicates very high precision and recall.

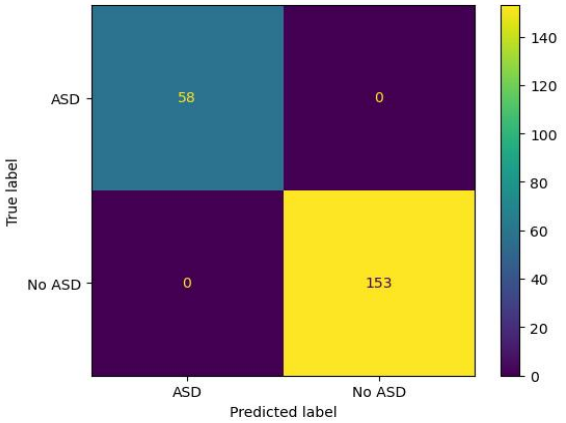

Figure 4.9 Bagging with SVM

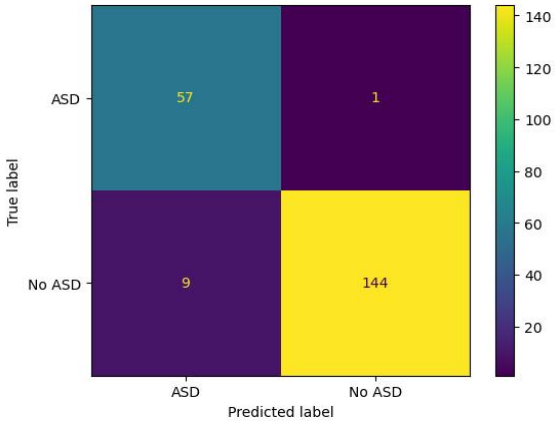

Figure 4.10 Bagging with KNN

Bagging did not affect the SVM classification, while it slightly improved KNN by reducing the false positives to 7. Bagging also improved the GNB slightly by reducing the false positives to 9, but in the case of MLP, Bagging reduced the classification accuracy by classifying a child with ASD as not having ASD.

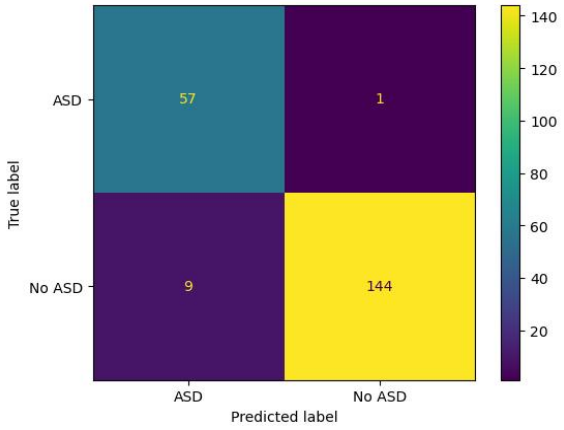

Figure 4.11 Bagging with GNB

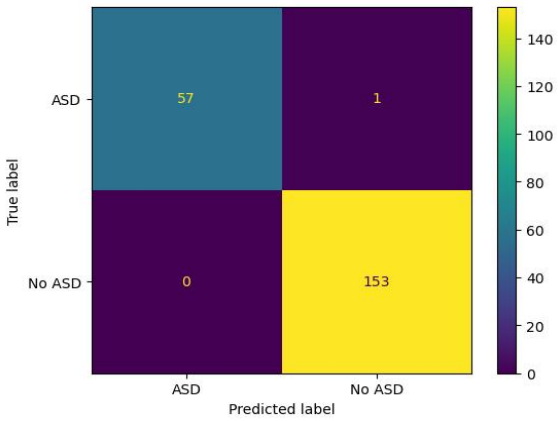

Figure 4.12 Bagging with MLP

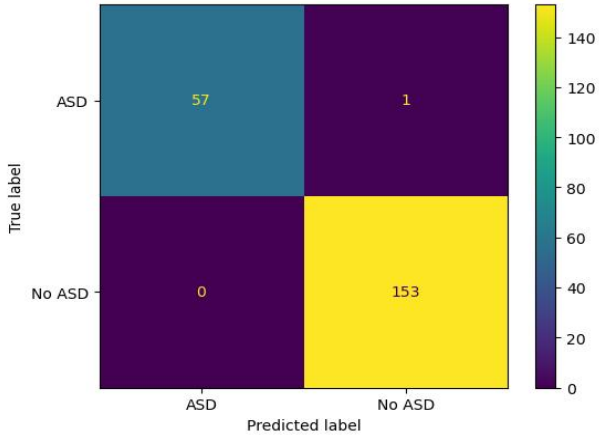

Figure 4.13 Stacking Classifier

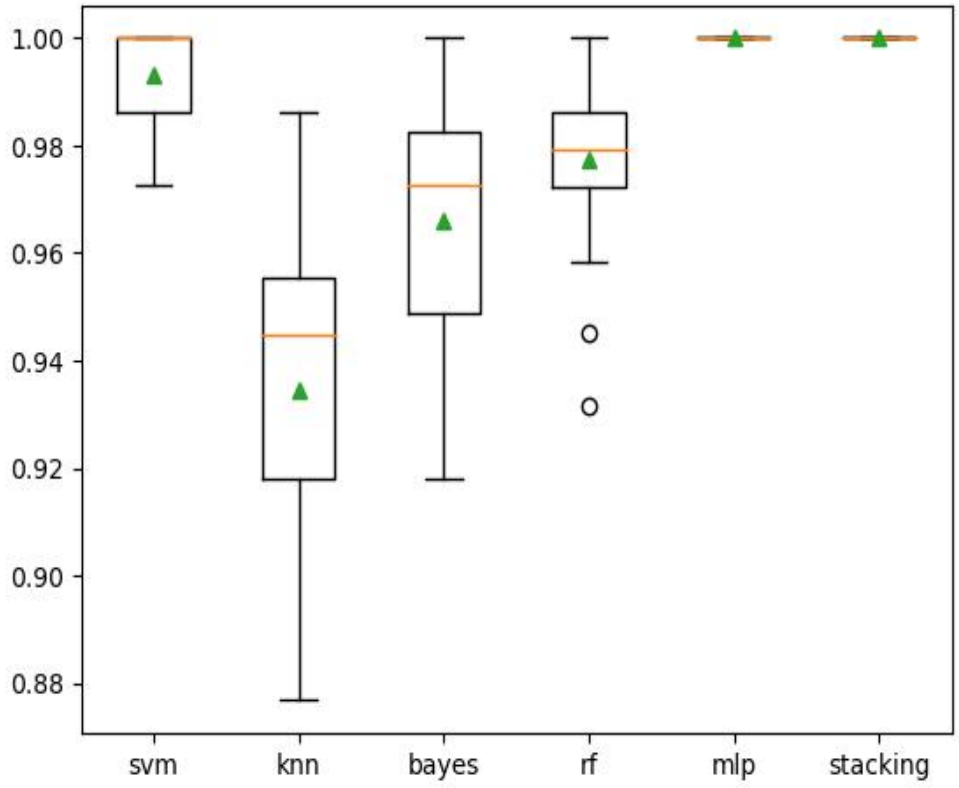

Figure 4.14 Box plot comparing the Stacking Classifier to Base Learners
Research Result, 2024

The stacking classifier correctly classed 57 children with ASD, and 153 children with no traits of ASD, while one child was wrongly classified as not having ASD. There was no significant difference between the stacking classifier and the MLP. The mean and standard deviations of recall, the scoring parameter for the model comparison were less than 0.05 and are quite negligible.

4.5 Discussion of Findings

Exploratory data analysis was first performed to understand the data set and its properties better. The dataset recorded interviews regarding 1054 children, out of which 70% were male and 30% were female. 73% of the children had previously had jaundice. 83.8% did not have family members with ASD traits, while 16.2% had family members with ASD. 32% of the sampled children were white European, 22% were Asian, 18% were Middle Eastern, 6% were south Asian and 5% were black. Ethnicities such as native Indian, Latino, Pacifica, mixed and others accounted for 17%. The average age in months was 27.87 with a standard deviation of 7.98. The youngest sampled child was 12 months old while the oldest was 36 months old.

It is important to specify the runtime that elapsed during hyperparameter tuning, as this proved to be a constraint in the case of the Random Forest model, which was very slow and cumbersome using the GridSearchCV method which exhausts all possible combinations of the parameter space. RandomizedSearchCV was used as a worthy alternative, as it reduced the search time by searching the sample space randomly. Table 4.1 shows the runtime values of all the models. Table 4.2 shows the evaluation metrics achieved from the exploration of the classification algorithms. These are optimised values, given that the correct procedures were adhered to at every stage. The base classifiers were applied independently, while the ensemble techniques were dependent on some of these base models.

It was noticed that accuracy equalled recall in all models, and the average precision returned the same values as the area under the Precision-Recall curve for the ensemble models. The F1 score returned the lowest metric in all sampled models. All base models performed brilliantly, with SVM being the best and most precise in this experiment with a result of 1, amongst all classifiers even before hyperparameter tuning.

After hyper tuning, SVM and MLP ranked highest with all metrics returning scores of 100%. KNN returned 96.2% accuracy and recall and GNB trailed at 95.26% for both metrics. For the ensemble methods, XGB and Bagging with SVM performed best at 100% for all metrics. RF produced an accuracy and recall of 98.1% while Bagging with KNN and GNB had both metrics at 95.2%. The classification performance of Bagging with MLP and the stacking classifier were equivalent across all metrics though, producing accuracy and recall of 99.5%.

Having very minimal false positives is the bane of precision, and a positive predictive value of 100 is the true algorithm on which it relies. In this case, SVM and MLP had no false positives, same with XGB, Bagging with SVM, and Bagging with MLP but their scores are varied with respect to the false negatives, as both sum up to the denominator in the determination of precision. While SVM and MLP achieved 100% F1-score, KNN and GNB were 96.3% and 95.4% respectively. XGB and Bagging with SVM attained an F1 score of 100%, RF achieved 98.1% and Bagging with KNN, GNB and MLP returned 95.3%, 95.3%, and 99.5% respectively. The F1 score is an averaged score estimated by the harmonic mean of precision and recall.

For the area under the ROC curve, SVM, MLP, XGB, and Bagging with SVM returned a value of 100%, KNN and GNB produced values of 96.85% and 96.73% respectively and RF attained 98.1%. All the models will be classified as having performed excellently, and all points will be situated at the top left area of the ROC curve. Average precision and area under the Precision-Recall Curve returned equal scores for all models, with SVM, MLP, XGB, Bagging with SVM and MLP at 100%, KNN at 99.5%, GNB at 99.98%, RF with 99.91%, bagging with KNN at 99.5% and Bagging with GBN at 99.95%. These results corroborate their excellent classification performance.

Table 4.1 Model Runtime for Hyperparameter Tuning

MODEL	RUN TIME (in seconds)
SVM	10.8
KNN	4.98
GNB	7.48
MLP	658
RF (Using GridsearchCV)	172,800
RF (Using RandomizedSearchCV)	41.7
XGB	2.45

Source: Research Result, 2024

Classifier	Model Parameters	Optimal Model Parameters
SVM	{"C":[0.1,1,10], "kernel":['linear', 'rbf', 'sigmoid','poly'], "gamma":['scale', 'auto'], "random_state":[15]}	{'C': 1, 'gamma': 'scale', 'kernel': 'linear', 'random_state': 15}
KNN	{'n_neighbors' : [5,7,9,11,13,15], 'weights' : ['uniform','distance'], 'metric': ['minkowski','euclidean','manhattan']}	{'metric': 'manhattan', 'n_neighbors': 15, 'weights': 'uniform'}
GNB	{'var_smoothing': np.logspace(0,-9, num=100)}	{'priors': None, 'var_smoothing': 0.004328761281083057}
RF	{'n_estimators': n_estimators, 'max_features': max_features, 'max_depth': max_depth, 'min_samples_split': min_samples_split, 'min_samples_leaf': min_samples_leaf, 'bootstrap': bootstrap}	{'n_estimators': 200, 'min_samples_split': 2, 'min_samples_leaf': 2, 'max_features': 'auto', 'max_depth': 80, 'bootstrap': False}
MLP	{'hidden_layer_sizes': [(10,30,10),(20,)], 'activation': ['tanh', 'relu'], 'solver': ['sgd', 'adam'], 'alpha': [0.0001, 0.05], 'learning_rate': ['constant','adaptive'],}	{'n_estimators': 200, 'min_samples_split': 2, 'min_samples_leaf': 2, 'max_features': 'auto', 'max_depth': 80, 'bootstrap': False}

Figure 4.15: Optimal Model Parameters after Tuning of Hyperparameters
Source: Research Result, 2024

	models	accuracy	precision	recall	f1score	roc_auc_score	average_precision	auc_precision_recall
0	SVM	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1	KNN	0.962085	0.965002	0.962085	0.962638	0.968503	0.994473	0.995329
2	GNB	0.952607	0.959576	0.952607	0.953687	0.967320	0.999829	0.999828
3	MLP	0.995261	0.995291	0.995261	0.995248	0.991379	0.999958	0.999957
4	RF	0.976303	0.977141	0.976303	0.976485	0.978307	0.999097	0.999094
5	XGB	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
6	BAG_SVM	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
7	BAG_KNN	0.952607	0.957515	0.952607	0.953499	0.961968	0.995718	0.995705
8	BAG_GNB	0.952607	0.957515	0.952607	0.953499	0.961968	0.999566	0.999564
9	BAG_MLP	0.995261	0.995291	0.995261	0.995248	0.991379	0.999829	0.999828
10	STACK	0.995261	0.995291	0.995261	0.995248	0.991379	0.999829	0.999828

Figure 4.16 Evaluation Metrics of Each Model
Source: Research Result, 2024

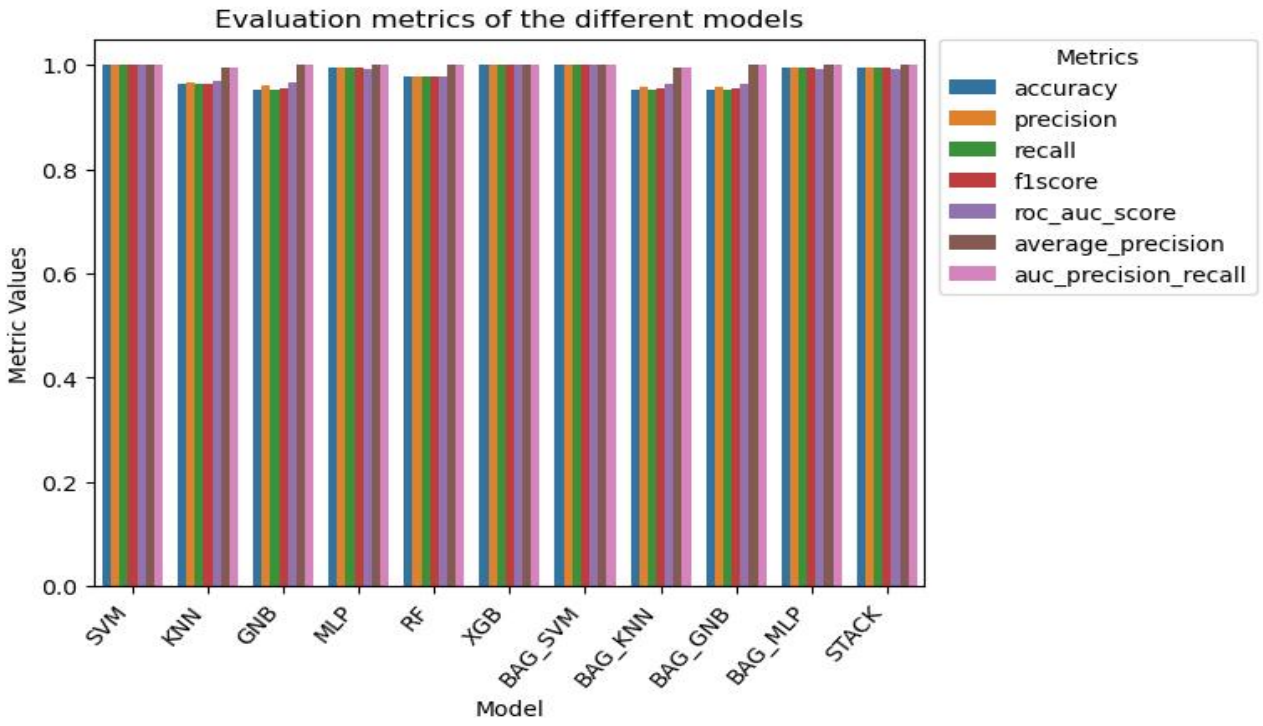

Fig 4.17 Grouped Bar Chart Comparing Evaluation Metrics of Classifiers
Source: Research Result, 2024

Endnotes

1. J.M Kernbach, VE Staartjes. *Foundations of machine learning-based clinical prediction modeling: Part II—Generalization and overfitting.* Machine Learning in Clinical Neuroscience: Foundations and Applications. 2022:15-21.

2. Y. Zhao, W Zhang, X Liu. *Grid search with a weighted error function: Hyper-parameter optimization for financial time series forecasting*. Applied Soft Computing. 2024 Feb 3:111362.

3. S. Kumar, H Tiwari, M Jaiswal. *Diabetes prediction using optimisation techniques with machine learning algorithms.* **International Journal of Electronic Healthcare**. 2023;13(2):158-68.

Chapter Five

Conclusion

This chapter discusses a summary of the findings of the research, as well as the conclusion, recommendations, contributions to knowledge, and areas where additional research is needed.

5.1 Summary of Findings

The research developed an optimal classification/predictive model for use in the assessment, screening, detection, and diagnosis of autism spectrum disorder (ASD). From the findings, "Q_Chat_10" was found to have the highest correlation with the class labels, suggesting it was a perfect predictor. However, to avoid overfitting, this feature was excluded from the model training. The dataset was partitioned using the `.iloc[]` function and split into training and test sets with a 20% test size. A 10-fold cross-validation was applied to prevent overfitting and ensure the models' generalizability. GridSearchCV and RandomizedSearchCV were used for hyperparameter tuning. Four primary algorithms (Support Vector Machine, Gaussian Naïve Bayes, K-Nearest Neighbours, Multilayer Perceptron) were optimized and evaluated on metrics like accuracy, precision, recall, F1, and AUROC score. Furthermore, ensemble models (Random Forest, XGBoost, Bagging, and Stacking) were applied, and their performances were compared.

Also, confusion matrices for each model revealed that SVM and MLP have perfect classification with no false positives or negatives, KNN has high accuracy but with more false negatives than false positives, which is a concern in a medical context. GNB has no false positives but several false negatives, indicating high precision but lower recall. RF showed a balance between precision and recall with few misclassifications, XGB achieved perfect classification, akin to SVM and MLP. It was noted that Bagging did not affect SVM's classification but improved KNN by reducing false positives. It also slightly improved GNB. Stacking Classifier performed comparably to MLP, with only one instance of

misclassification. The differences in recall and precision across models were marginal and deemed negligible.

Further, the dataset was characterized by a diverse population in terms of gender, ethnicity, and age. A large percentage of children had jaundice, and a minority had family members with ASD. The time taken for hyperparameter tuning varied across models, with Random Forest being notably slow with GridSearchCV, prompting the use of RandomizedSearchCV as an alternative. Accuracy and recall were equivalent across models, with base classifiers performing well, especially SVM which had a perfect score pre- and post-hyperparameter tuning. Ensemble methods like XGB and Bagging with SVM also reached 100% in all metrics. F1 scores were generally high but lower than other metrics. All models performed excellently on the ROC curve, indicating effective classification abilities. The area under the Precision-Recall Curve matched the average precision, further confirming the models' classification performance.

Finally, the study found that SVM and MLP were the most precise models in the experiment, with perfect scores across various metrics. The ensemble models also performed exceptionally well, with the stacking classifier showing no significant difference compared to the individual models. The study highlighted the importance of feature selection and model validation techniques such as cross-validation to ensure generalizability and avoid overfitting.

5.2 Conclusion

Motivated by the age-long debate of whether accuracy is the best metric for model evaluation in machine learning models, It was pertinent to thoroughly investigate the machine learning classification algorithms that were selected. The selection was solely based on empirical evidence of their high performances and adjudged robustness. The global pandemic of neurodivergence at this time further fanned this flame. Exploratory data analysis revealed

that ASD is more prevalent in males than females. It can be argued that this was merely a factor of the available population dataset. No concrete evidence was extrapolated from ethnicity. No subject instance was dropped in terms of outliers, as every sampled case contained useful information for the success and completion of this diagnostic experiment, hence, was taken into account.

Hyperparameter tuning is a methodology which should be lauded as it enhanced the model performances as well as fine-tuned the experimental outcomes. The SVM proved to be the best classifier at the initial fitting to the data, as it returned all evaluation metrics as 100%. These values raised utmost concerns about overfitting of the model. After a critical appraisal of all methods applied, parameter hyper tuning seemed to be a principal contribution to overfitting, but proper mitigative approaches like cross-validation had been applied to checkmate this menace. For example, during manual parameter tuning, it was observed that all evaluation metrics dropped drastically when the kernel in SVC was changed to poly, sigmoid or rbf. The resultant performance values were in the range of 60%. This explicitly shows that the dataset is strongly linear. For the ensemble methods, there was no significant difference in metrics, as the parameters of the model that was fit had already been optimised. It was also deduced that the decision boundary was strictly simple and linear, and it was not cost-effective to have built complex models for the classification task. Another method employed to mitigate overfitting is the loss function which measures the "error" or "departure " from the true values of the outcome. For every model, the train and test accuracy score was computed and the absolute deviations were recorded. It was observed that the loss function ranged from 0 to 2.4 (which was recorded by the RF model). Given that models are usually not expected to perform this synchronously, there might still be some predictive bias.

Given that the models were fit to a clinical dataset, the 95% - 100% results on each model are a step in the right direction for the screening and diagnosis of Autism Spectrum Disorder. In

the medical field, only a 1% chance of error is allowed in design experiments, hence, it suffices to say that the DSM-V interview method is robust and efficient in the diagnosis of Autism Spectrum Disorder. Conclusively, classification models will perform optimally if the data collection method is well-tailored to capture important and deterministic features, the underlying structure of the data is properly understood, and appropriate techniques and algorithms are employed.

5.3 Recommendations

Based on the findings the following recommendations were made

i. conduct thorough feature selection to eliminate features that could lead to overfitting, such as "Q_Chat_10" in this study. Use techniques like cross-validation consistently to assess the generalizability of models to unseen data.

ii. Also, hyperparameter tuning for all models, using methods like GridSearchCV and RandomizedSearchCV, as they can significantly improve model performance. However, consider the trade-off between computational cost and performance gain, particularly for complex models like Random Forest.

iii. When deploying models in a clinical setting, prioritize not only accuracy but also interpretability. Models that provide insight into their decision-making process are more valuable to clinicians.

5.4 Contribution to Knowledge

This study's contributes to knowledge in the following ways:

i. Highlighting the significance of careful feature selection in model training has underlined the potential pitfalls of including highly predictive but potentially overfitting features. The decision to exclude "Q_Chat_10" due to its high correlation with the target

variable serves as a valuable case study for proper data preparation in predictive modeling.

ii. By employing both GridSearchCV and RandomizedSearchCV for hyperparameter tuning, the study provides empirical evidence on the effectiveness of these methods. This work reinforces the notion that proper hyperparameter tuning is essential for achieving optimal model performance.

iii. The comprehensive evaluation of traditional classifiers alongside ensemble methods offers a robust benchmark for future research. It presents a detailed comparison of the performance metrics (accuracy, precision, recall, F1, and AUROC scores) across models, establishing a precedent for the assessment of classification models in medical datasets.

iv. The study contributes insights into the utility of ensemble methods in enhancing predictive accuracy. The successful application and performance of stacking classifiers, in particular, illustrate the potential of ensemble methods in clinical predictive analytics.

v. By discussing the performance of various models, including those with perfect classification abilities, this study contributes to the understanding of the balance needed between model accuracy, interpretability, and practical deployment in a clinical setting.

5.5 Suggested Area of Further Studies

Based on the findings, the following are suggested for future research;

i. Future studies could explore longitudinal data tracking individuals over time to better understand how predictive models can adapt to changes in behavioral and medical indicators of autism.

ii. While the study's models have shown high accuracy in a controlled environment, further research is needed to validate these models in real-world clinical settings with diverse and larger populations.

iii. There is a growing need for models that are not only accurate but also interpretable. Research into explainable AI could make machine learning models more transparent and their decisions more understandable to clinicians.

iv. Exploring how to effectively integrate different types of data (genetic, neuroimaging, behavioral, etc.) could lead to more robust and comprehensive models for ASD prediction and diagnosis.

v. Comparing machine learning approaches with traditional statistical methods or other non-machine learning approaches used in medical diagnostics could provide insights into the unique benefits and limitations of these advanced techniques.