Towards Machine Learning-based Autism Spectrum Detection and Management Using Multimodal Data

by

Sabbir Ahmed (ID: 172516)
Md. Farhad Hossain (ID: 172506)
Silvia Binte Nur (ID:172463)

A Research Project Report submitted to the
Institute of Information Technology
in partial fulfillment of the requirements for the degree of
Bachelor of Science in
Information Technology

Supervisor: Prof. M Shamim Kaiser, PhD

Institute of Information Technology
Jahangirnagar University
Savar, Dhaka-1342

March 2022

DECLARATION

We hereby declare that this thesis is based on the results found by ourselves. Materials of work found by other researcher are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Sabbir Ahmed	Md. Farhad Hossain	Silvia Binte Nur
Roll:172516	Roll:172506	Roll:172463

CERTIFICATE

This is to certify that the thesis entitled **Towards Machine Learning-based Autism Spectrum Detection and Management Using Multimodal Data** has been prepared and submitted by **Sabbir Ahmed**, **Md. Farhad Hossain** and **Silvia Binte Nur** in partial fulfilment of the requirement for the degree of Bachelor of Science (honors) in Information Technology on March 28, 2022.

Dr. M. Shamim Kaiser
Supervisor

Accepted and approved in fulfilment of the requirement for the degree Bachelor of Science (honors) in Information Technology.

Prof. Dr. Mohammad Abu Yousuf	Dr. Mohammad Shahidul Islam	Dr. Rashed Mazumder
Chairman	Member	Member

Prof. Dr. Md. Hasanul Kabir
External Member

ACKNOWLEDGEMENTS

We feel pleased to have the opportunity of expressing our heartfelt thanks and gratitude to those who rendered their cooperation in making this report.

This thesis is performed under the supervision of Dr. M. Shamim Kaiser, professor, Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. During the work, he has supplied us with several books, journals, and materials related to the present investigation. Without his help, kind support, and the generous period he has given, we could not perform the project work successfully in due time. First and foremost, we wish to acknowledge our profound and sincere gratitude to him for his guidance, valuable suggestions, encouragement, and cordial cooperation.

We express our utmost gratitude to the Director of IIT, Jahangirnagar University, Savar, Dhaka, for his support to complete the work within the time frame. Moreover, we would also like to thank the other faculty members of IIT who have helped us directly or indirectly by providing their valuable support in completing this work.

We express our gratitude to all other sources from where we have found helpful. We are indebted to those who have helped us directly or indirectly in completing this work.

Last but not least, we would like to thank all the staff of IIT, Jahangirnagar University, and our friends and family who have helped us by giving their encouragement and cooperation throughout the work.

ABSTRACT

The human body's sensory processing system is capable of gathering and integrating information via sensory organs. Children with autism spectrum disorder(ASD) have been found to have a sensory impairment. Individuals with ASD are prone to hyper/hyposensitivity, which might cause changes in information management, and affect cognitive impairment and social reactions to everyday events. This report proposed a two-stage strategy consisting of a questionnaire-based classification of ASD and a multimodal screening method for ASD screening that incorporates all possible input factors such as facial imaging, eye tracking, motor movement analysis, behavioral analysis, heart rate, and audio data. In the first stage, a questionnaire is designed, based on ASD symptoms found in previous studies with 82 questions. Then, a dataset is built by the use of the questionnaire and performing a survey. Following that, video and audio data from ASD patients are captured, as well as physiological data from ASD patients via a smartwatch (Fitbit) for screening and monitoring purposes. The final dataset is constructed by extracting the relevant multimodal features. Finally, different machine learning classifiers for type detection and screening are constructed and compared to different architectures as well as single modality classifiers. The proposed multimodal neural network with selective dropout attained a testing accuracy of 99.98 percent.

Keywords: Autism Spectrum Disorder, Machine Learning, Neural Network Classifiers, Multimodal Classification, Questionnaire, Multimodal Data.

LIST OF ABBREVIATIONS

AD	Autistic disorder
ADOS	Autism Diagnostic Observation Schedule
AS	Asperger's Syndrome
ASD	Autism spectrum disorder
ANN	Artificial Neural Network
CDD	Childhood Disintegrative Disorder
CNN	Convolutional Neural Network
DL	Deep Learning
FC	Feature Concatenation
FCA	Feature Concatenation with Attention
FCSD	Feature Concatenation with Selective Dropout
HC	Hidden Concatenation
LSTM	Long Short Term Memory
ML	Machine learning
OC	Output Concatenation
PDD-NOS	Pervasive Developmental Disorder-Not Otherwise Specified.
RNN	Recurrent Neural Network

LIST OF FIGURES

Figure

LIST OF TABLES

Table

TABLE OF CONTENTS

CHAPTER I

Introduction

1.1 Overview

In this section, we have discussed the impact of Autism Spectrum Disorder (ASD) on global health and the lives of patients and their families. In the following sections, we have explained what ASD is and how it differs from other neurological conditions. ASD must be detected early in a patient's life because there is no accurate lab test for its detection. Finally, we have outlined the shortcomings in existing research. Finally, we have discussed sensing and IoT technology in the detection of ASD, machine learning (ML) detection, multimodal ASD screening, and an autism health monitoring system.

1.2 Background

1.2.1 Effect of ASD in Global Health

According to the Centers for Disease Control (CDC) and Prevention of the USA, 17 % of children aged three to seventeen were diagnosed with a developmental disability between 2009 and 2017 [1]. ASD is a group of complicated development in social contact, speech, and non-verbal expression, and restricted/repetitive behavior that entails ongoing difficulties [2]. The constellation of both communication and physical rigidity impairments with genetic complications are early manifestations of ASD. According to Lord et al. [3]. Two deficiencies characterize ASD: 1) social and communication, 2) highly restricted and repetitive pattern of sensing - motor behavior. In each person, the causes of ASD and the seriousness of the symptoms vary. ASD has been diagnosed in every 1 in 270 people in the world [4]. In the United States alone, 1 out of every 54 children has been diagnosed with ASD. Hence, early detection of ASD creates awareness both in the family and socially, enables better care and less negligence for diagnosed individuals, and results in overall better psychological growth.

1.2.2 ASD and Its Type

The ASD is a complicated, irreversible developmental disorder that typically manifests in early childhood and has a detrimental effect on an individual's overall abilities, self-control relationships, and communication. ASD is a neurodegenerative disease defined by occurrence of repetitive and restricted activities, behavior, or interests and impairments in social communication, as per the fifth edition of the diagnostic and statistical manual of mental disorders [2]. Deficits in nonverbal communication for social interaction, social-emotional reciprocity, relationship development, maintenance, and comprehension are the primary indicators of impairments in social communication. Multiple symptoms are indicative of restricted and repetitive patterns of activities, interests, or behavior. These symptoms include rigid adherence to schedules, insisted sameness, repetitive motor or body movements, ritualistic behavior, hyper or hypoactivity, unusual interest in sensory aspects of the environment, and an uncertain response to sensory inputs. A great deal of study has been done on ASD, its varieties, symptoms, and identification. Faras et al [5] has classified autism as a Pervasive Developmental Disorder (PDD) and categorized ASD as Autistic Disorders (AD), Asperger's Syndrome (AS), Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS), Childhood Disintegrative Disorder (CDD), and Rett Syndrome (RS).

1.2.3 The Necessity of Early Intervention

Toddlers who got early intensive therapy not only had functional gains but also required fewer services than those who received "treatment as usual" in a two to three-year follow-up of a clinical study, resulting in overall cost savings [6]. As a result, early intervention—and, by implication, early diagnosis—have the potential to enhance function while lowering societal costs [7]. Early diagnosis as well as treatment for children with impairments have been found to enhance child outcomes for AD [8] and for autism. Developmental surveillance which is also known as developmental monitoring by the American Academy of Pediatrics. It is promoted by the World Health Organization [9] as a procedure for the early diagnosis of developmental disabilities, particularly in low middle-income countries. Structured screening for AD or neurodevelopmental problems, as part of a progressive perspective to diagnosis and treatment, is one of the proposed techniques for monitoring children's development [10]. In comparison to relying solely on clinical judgments, information from high income countries suggests that providing screening instruments during regular health care examinations may yield to a more fast and effective assessment of children in need [11]. Regular screening for autism or DD at health-care visits is a simple and effective way to detect the condition early, while also allowing for referral for additional diagnosis and intervention if necessary. Despite its significance, however, early diagnosis remains a hurdle in both high and low middle-income countries [12]. Early in life, when developmental changes are fast, domains overlap, and early symptoms are typically modest, identification is difficult [13].

1.2.4 Deficits in Traditional Medical Approaches

Even though there may be few visible physical impairments, people with ASD suffer from significant psychological sickness. Since there are no physical attributes quantifiable in lab tests, ASD diagnosis has been quite difficult until now. Impairment among individuals varies for different ages; children are more prone to the intellectual disability with inferior motor control, whereas adult patients suffer from atypical, selective, and repetitive interaction in case of communication. Thorough development history of the child is normally gathered through parents; and recent behavior of the child as well as monitoring of the child in conjunction with group and unfamiliar people, are two crucial components of the children's ASD screen procedure. On the other hand, ASD screening of adults primarily relies on the behavioral pattern of development from childhood, screening tests, and questionnaires [3]. Doctors analyze communication, social, and behavioral development data to make a decision. Since people with ASD who come in for an autism evaluation are more likely to have co-incided mental health difficulties, any mental health evaluation technique must be able to identify such symptom and behavior irregularities from autism-related abnormalities; thus creating anomalies and ASD screening results. The Diagnostic and Statistical Manual of Mental Disorders (DSM 5) [2] and Autism Diagnostic Observation Schedule (ADOS) [14], the two most often used manuals, have made a difference in detecting ASD. DSM-5 defined two key domains of ASD to assess impairment: (i) communication and social interaction, and (ii) restricted interests and repetitive behaviors. On the other hand, ADOS evaluation utilizes planned social circumstances to generate target responses and interpersonal interactions divided into four modules. These modules are suited to people depending on their language and stage of development to guarantee that a varied range of behavioral events is covered. Nonetheless, the psychometric features of each method are restricted, dependent on outdated diagnostic standards, various behaviors, restrictions on present operation, and age.

1.2.5 Sensing and IoT Technology in ASD Detection

Various sensing technology embedded with wearable devices, such as microphones, heart rate, motion, accelerometer, and pulse oximeter, provide much better insight into the ailments and symptoms of a patient. Similarly, video, simulation, eye tracking, and virtual reality-based scenarios contribute to the accurate assessment of ASD diagnosis rather than questionnaire-based one [15]. These devices can collect critical information about patients and then employ various ML and DL algorithms to determine suitable responses. Preliminary detection of ASD based on IoT devices might perhaps assist professionals. Moreover, IoT and ML systems allow for mass primary diagnosis with little or no health professional involvement.

1.2.6 Effect of ML in Neurological Diseases

Complex characteristics and symptoms of developmental and cognitive disorders add complications to classifying in clinical decision making as well as deterministic computational methods. ML algorithms have been utilized broadly to solve developmental disorders, specifically ASD. Hyde et al. [16] addressed the effectiveness of utilizing ML for autism identification and reviewed several detection methods. These methods include the detection of behavioral and neuroimaging data, behavioral and developmental data, genetic data, and electronic health records. Reviewed methods include classifiers like Support Vector Machine (SVM), Decision Tree (DT), Alternating Decision Tree (AD Tree), Neural Networks (NN), Bayesian Network (BN) Random Forest (RF), Logistic Regression (LR), Random Tree (RT), Naive Bayes (NB) and more.

ASD diagnosis is a difficult procedure due to the disorder's developmental and widespread character. Numerous signs of ASD, including as delayed expressive communicative competence, poor interpersonal responsiveness, behavioral issues, and repetitive/restricted behavior, are also present in other diseases and syndromes. Intellectual handicap hinders the different diagnosis even more, particularly in children, whose interpersonal interaction impairments produced by severe cognitive impairment must be separated from those caused by ASD. A variety of official and informal autism screening techniques are used by professionals. These might be anything from casual observations to official evaluations. The M-CHAT (Modified Checklist for Autism in

Children) is a common 20-question assessment designed for toddlers aged between sixteen to thirty months [17]. The Ages and Stages Questionnaire [18] is a broad developmental screening method that focuses on developmental issues at distinctive ages. The ASQ consists of 40 questions on reciprocal social interaction (such as willingness to participate with other children, trying to offer comfort to others, and social smiling), language and communication (consisting interpersonal conversation, the use of traditional gestures, and generalizing utterances), and stereotyped or repetitive behavioral schemes. The ASQ also contains a question concerning self-injurious conduct and another about the individual's present language functioning [18]. ADOS is a series of semi-structured observation based questions in which examiner assesses the child's answers to a variety of familiar and unusual circumstances, looking for ASD-related behaviors. ADOS consists of several subtasks whose goal is to assess social capabilities, such as eye contact and social smiling [14]. The ADI-R is a similar questionnaire for caregivers ot family members to diagnose ASD by interpreting their observations about their children's everyday activities.[19] Because the ADI-R assessment is dependent on these reports, it does not allow for direct observation of children's social behaviors. Instead of relying solely on psychoanalysis or other theoretical assumptions, CARS is based on direct behavioral observation. The CARS is particularly helpful for research and administrative classification, as well as for generating a comprehensive explanation of a child's unusual behavior [20].

1.2.7 Multimodal ASD Screening and Type Detection

Though individual screening systems like questionnaires, eye tracking, neuroimaging, genetic data, electronic health records provide adequate results, the combination of these methods yields greater accuracy. A multimodal screening system consists of each screening as mentioned earlier methods, then summarizes each prediction into a final prediction. Multiple ML classifiers run parallelly with a final classifier that takes each of the first stage classifiers as input. In this manner, several ASD symptoms can be analyzed for diagnosis [21].

1.2.8 Health Monitoring System for ASD

People, particularly ASD patients, confront a variety of emotional situations in their daily lives, which in terms affect the physiological growth of individuals. All aspects of these physiological and psychological activities are hard to be cataloged by health professionals. Hence, a physiological outcome monitoring system that records continuous communication and behavioral changes produce intuition of the patient's well-being. Again these systems assess health professionals to monitor the growth in different contexts. Heart Rate Reactivity (HRR), Heart Rate Variability (HRV), Respiration Rate (RR), blood volume pulse, oxygen saturation, body temperature, electrodermal signals of skin, blood pressure, and many more autonomous sensors can create a ubiquitous environment around the ASD patient and provides insights about their daily lifestyles and emotions [22]. Merging the cloud, IoT, and ML technology, these insights allow for autonomous emergency notification to health professionals and caregivers. The following are some popular techniques for controlling and monitoring ASD systems: at the outset, ASD symptoms must be discovered, which may be done in a variety of ways, such as utilizing questionnaires to identify the most prevalent ASD signals. Once this step is complete, we may use IOT-based devices that use different sensor motors or wearable devices to track an individual's everyday actions. We can monitor an individual's behavior and how he or she responds to diverse circumstances by deploying these technologies. We can develop a system to monitor or ease their behavioral cycle after we've identified it, and we can utilize a variety of apps or web-based platforms to do it. Thus, we can train them and improve their lives by providing diverse approaches, equipment, and other resources.

1.3 Motivation

40% of individuals with ASD are nonverbal, and 31% of children are intellectually disabled. Medical approaches necessitated the involvement of an expert specialist and a significant amount of time. Thus, early intervention is critical for detecting ASD, and ongoing monitoring will considerably improve a child's quality of life. Traditional approaches require an individual to provide uni-type data, which complicates the procedure. Additionally, the overwhelming majority of datasets are based on questionnaires, which may not adequately represent ASD individuals' subconscious behavior and social inadequacies. As a result of a scarcity of training data, only a few deep learning (DL) algorithms have been deployed. Thus, we aim to acquire multimodal data from an individual, implement it using a multimodal model, and compare it to a unimodal model.

1.4 Problem Statement

ASD is a pervasive developmental impairment characterized by persistent difficulties with speech and nonverbal communication, social interaction, as well as restricted behaviours. Each person has very distinctive set of impacts and disabilities. Again the severity of symptoms vary by individual. And in some cases, individuals are not diagnosed before they reach the age of adolescence or adulthood. This delay results in children with ASD not receiving the necessary early intervention. Thus, early detection of ASD allows more efficient intervention.

1.5 Objective and Contribution

1.5.1 Objective

The objective of this research is following:

- To create questionnaire and multimodal data collection strategy
- To collect ASD dataset using questionnaire and multimodality
- To design ML classifiers and multimodal neural networks for ASD classification.
- To evaluate ML classifiers for ASD classification and multimodal neural networks for screening.
- To compare Classifiers for screening and type detection.

1.5.2 Contributions

Our contributions in this research are given below :

- Accumulated all possible symptoms related to ASD found in the current literature, and prepared a questionnaire to screen ASD and type detection using ML from 82 selected questions;
- Listed the most salient signs that distinguish ASD children from non-ASD children;
- Prepared a multimodal dataset with eye scan path, face, behaviour, activity, audio data. A video camera is used to collect eye scan path, facial image and behaviour where as a Wearable device (Fitbit) is used to collect activity, heart rate, calories consumed, metabolic equivalent of task etc.

- Designing a multimodal Neural network that encompasses all factors in the screening;
- On the created dataset, we have evaluated and compared the performance of various multimodal classifiers.

1.6 Research Outline

The remainder of the report is structured as follows: The second section reviews the literature; the third section discusses the proposed methodology,the fourth section contains the experimental analysis and finally, section 5 contains the conclusion

CHAPTER II

Literature Review

2.1 Overview

A brief explanation of the ASD symptoms has been provided in this part, which follows a review of relevant medical studies. Our next topic was a discussion of ASD detection datasets currently available. In the next section, we looked at how ML may be used to identify (ASD) by utilizing Questionnaires, Physiological and Video Data, and Multimodal Data. Finally, we discussed the monitoring system and the software that can be used to diagnose autism spectrum conditions.

2.2 ASD Symptoms

Ample research has been conducted related to ASD, its types, symptoms, and detection. Faras et al. [5] classified autism as a Pervasive Developmental Disorder (PDD) and categorized ASD as Autistic Disorders (AD), Asperger's Syndrome (AS), Childhood Disintegrative Disorder (CDD), Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) and Rett Syndrome (RS). Biomarkers related to cognitive, behavioral, visual, and structural connectivity have demonstrated promise in several clinical screening and diagnostic procedures, like ADOS, DSM-5, Autism Diagnostic Interview-Revised (ADI-R), Developmental, Dimensional and Diagnostic Interview (3di), Social Responsiveness Scale (SRS, SRS-2) [23]. Clinical standards, however, usually require the involvement of multidisciplinary teams in ASD diagnosis, and these processes need substantial amounts of time. Seltzer et al. [24] discovered that patients with ASD show a tendency to query inappropriately, spontaneously imitate, lack interest in people, difficulty sharing meals, and repetitive use of objects. Faras et al.[5] explored red flags indicating ASD in participants and described delays in speaking, repetitive play with toys, and communication difficulties. In addition,

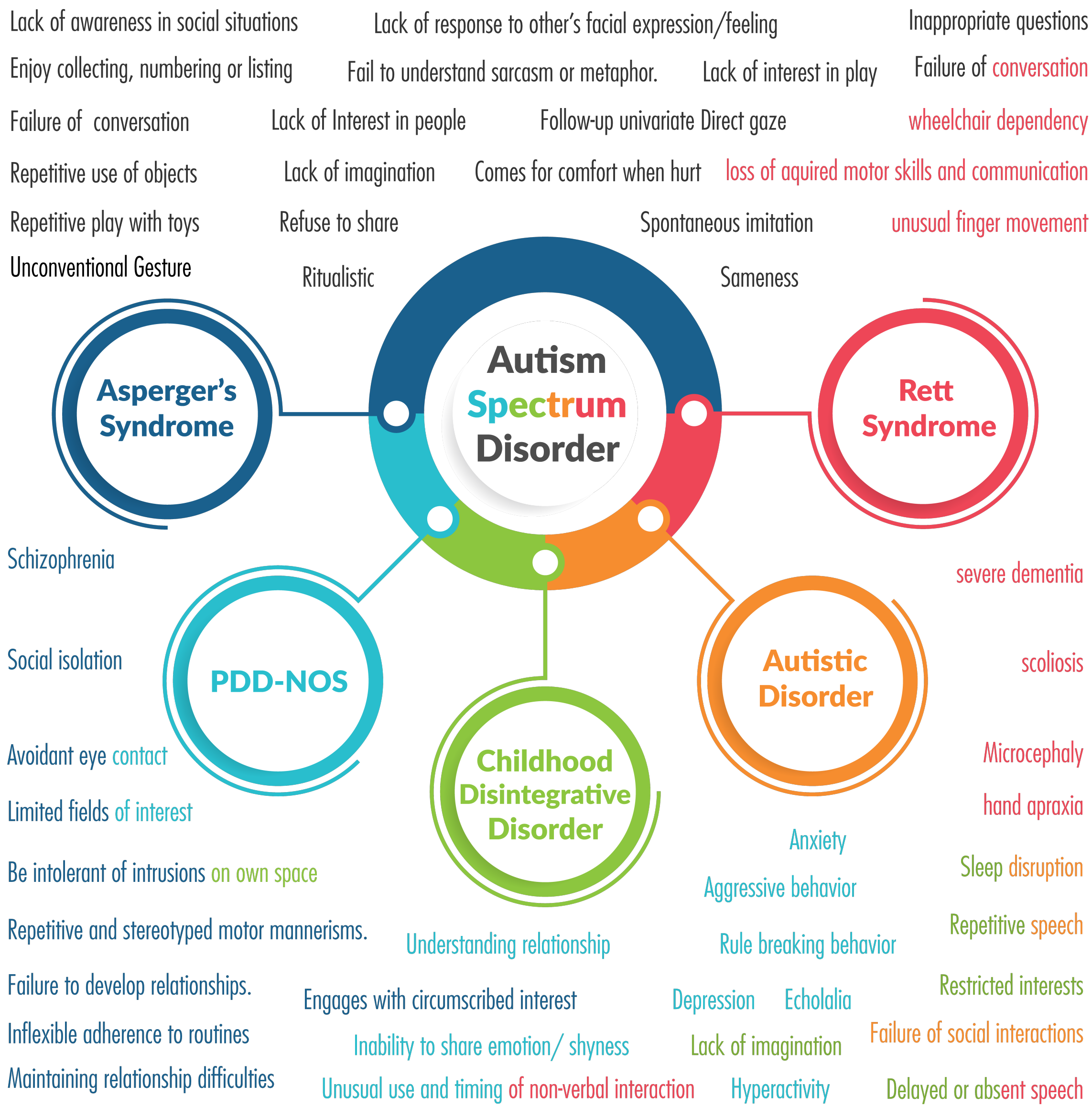

Figure 2.1: Different types of ASD and their symptoms. The black color indicates common ASD symptoms exists in all subtypes. Different colors indicate the corresponding symptoms found in different subtypes.

there was a lack of facial expression, pretend play, imagination, interest in playing near peers on purpose, ability to comprehend sarcasm, and awareness of personal space. Dawson et al. [25] discovered substantial disparities in task assessment performance between children with autism and those who did not have the condition. The symptoms included were social orienting, symbolic play, shared attention, distress responses, and immediate and delayed imitation. Taylor et al. [26] investigated sleep issues related to ASD. Solomon et al. [27] conducted a study comparing autism and internalizing symptoms in boys and girls with ASD. They discovered no differences in ASD symptoms, but found that females with ASD had greater internalizing symptoms than males with ASD. Volkmar et al. [28] developed a questionnaire to assess autistic and non-autistic individuals. They evaluated the variation score for five symptom areas: social/self-help, sensory, body/object use, language, and relating.

2.2.1 Asperger's Symptoms

According to Baskin et al.[29], individuals with AS manifest an inflexible adherence to specific nonfunctional routines, schizophrenia, repetitive and stereotyped motor mannerisms, and limited fields of interest. People with AS engage with others who share similar interests, struggle to maintain and develop good peer relationships, and prefer social isolation. In contrast, children have trouble starting to play with their peers due to a lack of social skills, according to Mirkovic et al.[30]. As a result, individuals choose to engage in activities that require less interaction. Their motions are clunky and erratic. They often have dyspraxia, a motor coordination disorder. Some people may be hypersensitive. On the other hand, they are pretty adept at memorizing and frequently have interests in mathematics, meteorology, music, dinosaurs, card collecting, and so on. A structured interview with the parents of these children was carried out by Eisenmajer et al. [31] to compare AD and Asperger's disorder. In addition, children with AS have been discovered to avoid eye contact, have less willingness to become friends, and participate in long-running pedantic speech patterns and repeated question talks. Machintosh et al.[32] reviewed current published literature and discovered that people with AS are more likely to have hyperactivity disorder, emotional and behavioral disturbances in childhood

2.2.2 PDD-NOS Symptoms

Karabekiroglu et al.[33] researched PDD-NOS symptoms and discovered that the participants exhibit unusual nonverbal movement, lack of eye contact while interacting hyperactivity and hostility, and inappropriate laughter. Initially, they created some preliminary symptom screening scale items: poor social interaction, hyperactivity, aggressiveness, obsessions, impulsiveness, echolalia, stubbornness, language retardation, choosiness, prosody problems, confusing pronouns, and highly interested televisions, etc. Later found that inattentiveness, hyperactivity, poor social skill, not responsiveness were the significantly discriminative symptoms. Snow et al.[34] identified a concern with rule-breaking and aggressive conduct, as well as anxiety and depression among PDD-NOS patients. Jensen et al.[35] investigated PIC (Personality Inventory for Children) components to assess PDD-NOS symptoms. Undiscipline, cognitive development, internalizing symptoms, and social inadequacy were the factors. They also discovered that the people suffer psychosis, delinquency, and a high frequency of odd symptoms. The distinctions between PDD and related disorders were investigated, and categorization systems were developed by Buitelaar et al.[36]. For example, they demonstrated a lack of social or emotional reciprocity, incorrect communication language, uncomfortable discussion, a lack of shared enjoyment, a lack of nonverbal behavior, and so on in patients with PDD-NOS symptoms.

2.2.3 CDD Symptoms

Mehra et al. [37] examined symptoms of CDD and discovered that the participants exhibited limited interest, lack of imagination, sleep problems and decreased motor abilities. Aggression, agitation, and self-injurious behavior are all common in CDD. Case reports for CDD indicate at least two years of seemingly normal growth, with all developmental milestones met. However, CDD has been linked to difficulties with adaptive skills including toileting, self-help, and sleep, as well as emotional and behavioral management. In addition, regression in CDD has been reported as being fast and abrupt. The communication disruption, according to Malhotra et al malhotra1999, is the most noticeable. The loss of speech is often so severe that it resembles the extreme lack of speech development found in autism. Social and self-help abilities are deteriorating at a greater rate than motor skills.

2.2.4 AD Symptoms

Elia et al.[38] observed that the diagnosis of AD can be based on the first REM delay, muscle twitches density and rapid eye movement density. They found that muscle cramps were significantly higher in patients with AD. In addition, they observed patients with AD, patients with fragile X syndrome with mental retardation, and normal subjects. Patients' first REM delay, sleep latency, bedtime, stage transitions, and awake following sleep initiation were all substantially associated. They suspected that people with AD and X syndrome with mental impairment had a neurophysiological sleep pattern. And also, these patterns could be correlated with other psychological parameters which were used to diagnose AD. Repetitive behaviors may not be major characteristics of AD; nevertheless, Militerni et al. [39] discovered that younger subjects demonstrated repetitive motor and sensory behaviors, whereas older youngsters with higher IQ scores demonstrated complex repetitive behaviors. Eisenmajer et al.[31] examined patients' parents and discovered that children with AD were less likely to participate in prosocial activities than kids with AS. Eye contact avoidance was more acute in children with AD than in children with AS. Children with AD were more likely to utilize echolalic speech and less likely to employ pedantic speech patterns and repeated question discussions.

2.2.5 Rett Syndrome Symptoms

Hagberg et al. [40] identified the essential clinical symptoms of RTT and classified it into five categories. At six months to 2 years, the subjects' attained talents tend to fade: hand skill, hand use, communication, inner language, emotional interaction, and so on. Hand movements that are unusual include hand-washing, wringing, clapping, and other expressions such as teeth grinding and body rocking. Rapid deterioration may occur after this stage. The individual is between the ages of 1-3 years throughout this time period. Severe dementia, rapid social regression, apraxia, and seizures can be observed in 75-80 percent of patients. The last stage may include motor handicap, such as being confined to a wheelchair or bed, severe scoliosis, growth retardation, and so on. RTT, according to Kyle et al.[41], is a neurological condition caused by genetic abnormalities. The RTT was classified into four stages: stagnation, rapid regression, pseudo stationery, and motor degeneration. Developmental progress is slowed during the stagnation stage, which includes crawling, sitting, vocalization, and so on. Postural delays appear, and microcephaly (head shrinkage) occurs. Breathing problems, severe microcephaly, and the possibility of seizures are all frequent during the rapid regression stage. Seizures are more likely in the third stage, and the individual may show more interest in others, as well as hand apraxia/dyspraxia. Physical difficulties deteriorate in the final stage, scoliosis is discovered, and wheelchair dependency may ensue. RTT is related to severe mental impairment, cortical atrophy, stereotyped hand movements, and extrapyramidal dysfunction, according to Wenk[42]. According to Wulffaert [43], RTT patients have a short attention span, make nonspeech noises, and laugh for no apparent reason. Nomura et al.[44] discovered that the participants' intellectual ability is substantially impacted from childhood to adulthood.

2.3 ASD Dataset

To analyze biomarkers and traits among ASD and neurotypical people, many researchers have conducted studies among different modalities. Through suitable processes, a lot of surveys have been conducted to acquire data from different people in order to find differences in biomarkers between ASD, neurodegenerative and

Name	Size	Input format
Detect Autism from a facial image	2536 Train, 200 test, 100 validation	Image
Autism Screening	704 data, 21 field	Structure Data
Autism Detection Based on Eye Movement Sequences on the Web: A Scanpath Trend Analysis Approach	-	Structure Data
Autism screening data for toddlers	1054 data	Structure Data
Visualization of Eye-Tracking Scanpaths in Autism Spectrum Disorder: Image Dataset	-	Image
Autistic Spectrum Disorder Screening Data for Children Data Set	292 instances, 21 attributes	Structure Data
A Dataset of Eye Movements for the Children with Autism Spectrum Disorder	300 images	Image
A set of video clips of reach-to-grasp action Request Needed	1837 video sequences	Video
Identification of common genetic risk variants for autism spectrum disorder	18,381 individuals with ASD and 27,969 controls	Genetic
Meta-analysis of GWAS of over 16,000 individual	Meta-analysis of GWAS of over 16,000 individual	Genetic

Table 2.1: Available dataset details

neurotypical patients. By labeling these features manually each of the datasets provides future research scope for researchers of different backgrounds. Since making a dataset is a tedious task, requiring multidisciplinary contribution, a handful of datasets is publicly available. The modalities of data range from structured data to image, audio, and video data. A total of 20 features(10 behavioral features, and ten individual characteristics) are utilized to classify ASD cases in this dataset[45]. A further innovative way to ASD detection using the Scanpath Trend Analysis is the use of an individual eye move from the web[46]. Another dataset featuring displays of eye-tracking scan paths focusing on ASD is here[47]. By employing a mobile app called ASDTests, another structured dataset was built to test autism in toddlers[48]. 300 images of 14 individuals diagnosed with ASD and 14 normal individuals is given here[49]. Another dataset, consisting of video clips of children with ASD performing reach-to-grasp activities, is proposed here[50]

2.4 ASD Detection Using ML

2.4.1 Questionnaire

Goal et al. proposed [51] proposed a modified Grasshopper Optimization Algorithm (GOA) for feature selection in AQ10 questionnaire-based data set, which improved the classification accuracy of RF, LR, NB, and KNN and achieved 100 percent accuracy for child and adolescence dataset. Thabtah and Peebles [52] suggested new Rule-based machine learning(RML) for detecting ASD On several ASD datasets while bagging, boosting, rule induction, and decision trees methods were empirically assessed, with RML outperforming the others. Kupper et al [53] prepared an ADOS questionnaire with a survey conducted in Germany and minimized the feature of ADOS into only five attributes. Then classification has been applied using SVM with an AUC of 82 percent.levi et al [54] proposed sparsity to build regulation techniques to pick features of the ADOS questionnaire, utilizing 17 different supervised learning models and to encode missing data for a grid search styled cross-validation. They had incorporated feature sets along with sex and age to offer a simple classifier that may be developed to differentiate ASD from non-ASD reliably. The ML technique includes LR, Lasso, SVM, ADTree, RF, Ridge, Elastic net, Nearest shrunken centroids, LDA, AdaBoost, Relaxed Lasso, and others. In another work, Pratama et al. [55] applied SVM, RF, ANN into the AQ-10 dataset with 10 fold cross-validation, though no feature selection has been used.

Table 2.2: Review of recent literature

Author	**Methodology**	**ML Model**	**Data Type**	**Dataset Size**	**Evaluat -ion matric**	**Limitation**	**Year**
Carette [56]	An automated technique using on LSTM neural nets that focuses on the saccade portion of eye during reading	LSTM	Eye Trackng	ASD 17,Non 15	Accuracy 83%	Exclusion of Comparison with other NN	2017
Carette [57]	Using color gradients, cohesively represent eye motions into an image-based manner while preserving the dynamic features of eye movement	RF,SVM, LR, NB, ANN	Eye Trackng	ASD 29,Non 30	AUC 90%	NA	2019
Elbattah [58]	a eye tracking scanpath method for ASD classification using K-mean clustering algorithm,and find maximum accuracy with the k value set at four,	Grayscale ,PCA ,t-SNE,Autoencoder Features, K-Means Clustering	Eye Trackng	ASD 29,Non 30	Accuracy 94%	NA	2019
Tao [59]	eye scan path by saliency mapping from CNN generator and discriminator part with encoder, decoder utilizing a CNN-LSTM algorithm for classification.	SalGAN, CNN-LSTM	Eye Trackng	ASD 14,Non 14	Accuracy 74.22%	NA	2019
Goel [51]	Optimization algorithm GOA for accelarating ML algorithm	GOA, , LR, NB, KNN, RF-CART- ID3, BACO	AQ-10, Questionnaire	1100	Accuracy of near 100%	Low convegence speed for GOA	2020
Thabtah [52]and Peebles	On several ASD datasets, bagging, rule induction, boosting, and decision trees methods were empirically assessed, with RML outperforming the others.	RIDOR, Nnge, RIPPER, RML, Bagging, CART, PRISM , C4.5	AQ-10, Q-CHAT, Questionnaire	ASD 189 , Non ASD 515	F1 Score more than 90%	Not applied to any toddlar dataset	2020
Kupper [53]	ADOS questionnaire based data, minimizing features into only five attributes, classification using SVM	recursive feature selection, kohen cappa, SVM	ADOS, Questionnaire	ASD 385, Non ASD 288	AUC of 82%	Affect of diffrent ML algorthim in selected feature is known	2020
Levy [54]	Collected seperate data for child and adult, applied 17 supervised machine learning algorithm , 10 feuture from ADOS and SVM, LDA gained most AUC	LR, Lasso, SVM, ADTree, RF, Ridge, Elastic net, Nearest shrunken centroids, LDA, AdaBoost, Relaxed Lasso	ADOS, Questionnaire		AUC of 95%	Impure source and survey of data	2017
Pratama [55]	SVM, RF, ANN applied into AQ-10 datset with 10 fold cross validation	SVM, RF, ANN	AQ-10, Questionnaire	4189 ASD	Sensitivity of 87.89%	absence of feature selection	2019

2.4.2 Eye-tracking

Carette et al [56] proposed an eye movement tracking autism screening system using LSTM. A total of 17 ASD and 15 neurotypical patients were studied in this article with an accuracy of 83 percent. Elbattah et al. [58] suggested an eye-tracking scan path method for ASD classification using the K-mean clustering algorithm, and finding maximum accuracy with the k value set at four. Similarly, Carette et al. [57] presented a similar methodology that converts eye-tracking scan path into an image then classifies the image for ASD screening. They experimented with RF, SVM, LR, NB, ANN algorithms with a maximum AUC of 90 percent. Tao et al. [59] classified ASD utilizing eye scan path by saliency mapping from CNN generator and discriminator part with encoder, decoder. An image has been given to the patient to capture their eye scan path then utilizing a CNN-LSTM algorithm for classification. Tegmark et al. [60] reviewed 38 eye-tracking and eye scan path-based ASD detection methods. They consider eight factors to measure the effectiveness of each method.

2.4.3 EEG

ASD detection can be done from brain activity also. Lih Oh et al [61]. proposed EEG based features that were selected via Marginal Fisher Analysis and Student's t test. Using SVM classifiers 37 children were classified with 98 percent accuracy. Similar EEG-based classification using transfer learning pre-trained models is utilized by bygin et al. [62]. For feature extraction, local binary pattern and short-term Fourier transform is used to create image formats from EEG. Elizondo et al [63] has done a case study of an ASD patient using EEG signals which basically measures attention. The EEG data was collected when the ASD user engaged in several didactic learning tasks in this case study. The recommended data acquisition for their case study was to first put a headset on the participants, then start EEG tracking and video recording, then give the test subject a worksheet and instructions, and lastly record the entire process until the activity is complete. They discovered that band PSD parameters such Relative Beta Power (RBP), Theta–Beta Ratio (TBR), Relative Theta Power (RTP), Theta–Alpha Ratio (TAR), Relative Alpha Power (RAP), and the TBAR are helpful for attention categorization. The multi-layer perceptron neural network model (MLP-NN) had the greatest performance with these characteristics.Tawhid et al [64] proposed a system where, primarily, Re-referencing, filtering, and normalizing are used to pre-process the raw EEG data. The pre-processed signals are then transformed into two-dimensional spectrogram pictures using the Short-Time Fourier

Transform. The pictures are then analyzed using ML and DL models. Textural features are recovered and distinguished using PCA before being given to six different ML classifiers. DL tests three different convolutional neural network models. For this, Grossi et al employed a seldom used system called TWIST (Training with Input Selection and Testing), a recursive hybrid ML technique that integrates an evolutionary optimization technique Gen-D with a backpropagation neural network. TWIST iterations selected a trustworthy set of characteristics for ML classifier input. A classification strategy for discovering and prioritizing the most valuable parameters to feed the predictive model, is essential for the proper deployment of the ML system, they claim. Their findings reveal that two EEG channels, C3 and C4, provide enough information to allow powerful ML algorithms to detect ASD signatures.

2.4.4 Facial Imaging Based

On the basis of a unique ASD dataset of clinically diagnosed children, Lu et al [65]developed a feasible ASD screening method based on face photographs. They claim that the ASD Children Facial Image Dataset is the only one that is publicly available. They were all sourced via online searches of ASD Facebook groups and other sources, according to the Kaggle dataset's creator. Because of this, there is a concern about the picture quality in the dataset. Accuracy is critical in the categorization of ASD face pictures using DL. The Kaggle dataset contains 89 percent white children and 11 percent children of color. They also clarify why the racial mix in the Kaggle dataset is problematic. They just used the Kaggle dataset to show the consequences of the decision. The Elim Autism Rehabilitation Center's clinically diagnosed ASD children's images were used to train their algorithm to accurately identify ASD, which fills a gap in the field.

2.4.5 Neuroimaging Based

Both Eslami et al and Heinsfeld et al used the ABIDE dataset. It comprises morphological and anatomical brain imaging data from around the world to help researchers better understand autism's neurological basis. To categorize brain illnesses, Eslami et a [66]l introduced the Auto-ASD-Network, which combines DL's ability to identify significant patterns in data with the discriminative power of the SVM classifier. The SVM classifier uses DL model features as input. They use a data augmentation approach called SMOTE to double the quantity of items in the training set to improve generalizability and accommodate for DL's tendency to overfit. Heinsfeld et al [67] claimed the greatest classification accuracy. They achieved a mean classification accuracy of 70% (sensitivity 74%, specificity 63%) in cross-validation folds, and a range of 66-71% in individual folds. The SVM classifier had a mean accuracy of 62% (sensitivity 68%, specificity 62%), whereas the Random Forest classifier had a mean accuracy of 63%. (sensitivity 69percent, specificity 58%).

2.4.6 Body Movement Based

According to Großekathöfer et al[68] using Stereotypical Motor Movement as a marker for ASD detection improved detection accuracy by 5 to 9% compared to existing published results. They also observed that excluding users from classifier training results in good detection accuracy. The chest accelerometer gave the most significant features for categorizing. This illustrates that a single torso sensor can reliably detect both body moving and hand flapping. Sadouk et al [69] employed DL on stereotypical motor movement (SMM) activities. Outperforming earlier research, they constructed temporal and frequency-domain CNN models with model parameters determined by input space. These frameworks adapt to stereotyped behavior patterns of every new atypical individual with only very few labeled SMMs, and (ii) overcome the medical problem of a shortage of marked SMMs per subject. Because the cross-domain transfer learning approach employs a dataset from such a target domain other than the target domain, it does not require SMMs for training. As a consequence, lack of medical target data no longer reduces the range and learning capacity of CNN models.

2.4.7 Video Based

Leblanc et al [70] built a pipeline that generated characteristics and input these into a diagnosing classification model. There was no difference in performance between the LR9 and ADTree7 models. A 10-fold GridSearch cross validation customized for UAR trains all pipeline pieces for each model and each feature imputation methodology. The developed pipeline is then tested on the U tube dataset, which includes various ratings (i.e., most frequent value). Comparing generic and dynamic feature substitution to single and multivariable feature imputation All in all, they discovered that using algorithmic-driven replacement questions and dynamic feature imputation enhanced UAR. Using DL, Li et al [71] propose diagnosing ASD from raw video. Our initial step is to monitor the eye movements in each video. Thei r second step is to turn these monitoring trajectories into angle and length histograms. Last but not least, we categorize these two histograms using a three-layer LSTM network. The LSTM network outperforms other ML methods like SVM by 6.2%. (from 86.4% to 92.6%). Given that this approach was designed to identify ASD, the sensitivity (TPR) and specificity (TNR) are excellent.

2.4.8 Multimodal Based

Alcañiz et al [72] implemented a multimodal virtual reality experience, where 24 children with ASD and 25 children with typical neurodevelopment participated.They tracked changes in their bodily motions in this Virtual Reality. Then it was determined which bodily parts may have a role in the discriminating between the two groups. The purpose of this study was to see whether virtual stimulus condition could better distinguish between the two populations' body regions. To assess body motions in a multimodal VR experience, they employed a ML technique to analyze body movements in three stimuli conditions: visual, auditory, and olfactory.
Sadek et al [73] studied many categories for ASD identification and analyzed various forms of detection systems that employ ML, computer vision, and neural networks to identify autistic individuals. ASD using ML was detected by Rahman et al. (citerahman2020review) who offered numerous methods to speed up the implementation of processing data for ASD detection using ML. Several strategies for finding and processing unbalanced data in such detection methods have also been investigated by the researchers.

2.5 ASD Monitoring and Management

2.5.1 Common Approaches of Management and Monitoring Systems

The following are some popular techniques for controlling and monitoring ASD systems: at the outset, ASD symptoms must be discovered, which may be done in a variety of ways, such as utilizing questionnaires to identify the most prevalent ASD signals. Once this step is complete, we may use IOT-based devices that use different sensor motors or wearable devices to track an individual's everyday actions. We can monitor an individual's behavior and how he or she responds to diverse circumstances by deploying these technologies. We can develop a system to monitor or ease their behavioral cycle after we've identified it, and we can utilize a variety of apps or web-based platforms to do it. Thus, we can train them and improve their lives by providing diverse approaches, equipment, and other resources.

2.5.2 Artificial Intelligence in Autism Management

Since ASD is correlated to physiological changes associated with negative emotions in people, monitoring these emotional changes can provide caregivers with a real-time picture of what these people are going through. An effective monitoring system may raise caregivers' awareness of a person's emotional condition, allowing them to take the appropriate steps to reduce stress symptoms and encourage the individual to use better stress coping skills.

In this section, we'll look at how IoT-based and wearable gadgets might aid autistic individuals by detecting their actions using various sensors, and then we'll look at the applications and websites that can be used to monitor their behavior. Anxiety problems are common in children and adolescents with ASD, with an estimated incidence rate of 40 percent [74]. Various physiological indicators and markers are considered for physiological and emotional evaluation. Heart Rate Variability (HRV) is a useful metric that computes the time intervals between two successive R peaks in an ECG signal obtained by an ECG sensor. Jansen et al. found some evidence of heart rate arousal variability in response to public speaking stresses [75]. The total number of breaths, or respiratory cycles, that occur each minute is known as the respiratory rate (RR). The rate of respiration might alter owing to disease, stress, and other factors. The respiratory center, which is located inside the medulla oblongata of the brain, regulates breathing rate. The rate of respiration has been proved to be an effective stress indicator [76]. The conductivity of the skin is measured by the Galvanic Skin Response (GSR). The GSR may be measured by inserting two electrodes on the skin's surface, one of which injects a small amplitude AC current into the skin and the other of which uses Ohm's Law to calculate the skin's impedance given a certain voltage. GSR has been suggested as a potential stress indicator [77].

Wearable gadgets for physiological and, to a lesser extent, emotional monitoring are widely available on the market. Cabibihan et al. [78] conducted a review of the academic literature on several sensing technologies that might be used for ASD screening and treatments. Eye trackers, movement trackers, physiological activity monitors, tactile sensors, voice prosody and speech detectors, and sleep quality assessment devices were among the sensing technologies studied. The devices' advantages and usefulness in assisting the treatment of various symptoms in people with ASD, as well as their limits, were evaluated. Tang et al. [79] focused on the integration of a realistic environment based on multi-sensor interrogation to assist NT persons in "reading" the emotional state of ASD youngsters. They include four different types of meters into their design in order to collectively perceive users' behavioral patterns, as well as individual and group emotions: Individual physiological data: perspiration, pulse rate; then individual behavioral data: gestures, head, facial expressions, hands, and upper-body movements, motions, and Following that, sociometry with integrated sensors, RGB-B sensors, and low-cost depth sensors is used to gather data. Notenboom et al. [80] used physiological signals to evaluate autistic people's emotions, and they created recommendations based on the target group's user requirements. Because the target population is sensitive to stimuli and has difficulty adjusting, certain design criteria are important. As possible designs, a smartwatch, a patch, and an infrared camera were considered. The recommendations were developed as a result of the examination of these designs. The smartwatch came out on top, followed by the patch. An infrared camera isn't the best option. The principles can be utilized to create a wearable that measures autistic children's physiological signals. Northrup et al. [81] created a combination of tailored features in IoT based wearable compatable with mobile application which analyzes stress reactivity in ASD children with real-time capabilities and sends on spot notifications to a caregiver through a mobile portable device.

2.5.3 Available Softwares and Websites for ASD Management

Let me talk: The free AAC talker [1] software, which is accessible for both Android and iOS, assists in the creation of intelligible phrases by aligning graphics. This row of photographs may be interpreted as a phrase if the images are linked together in a meaningful way. This program, they believe, is appropriate for autism symptoms, AS, and ASD. AAC stands for aligning images (Augmentative and Alternative Communication). LetMeTalk's picture library includes over 9,000 easy-to-understand photos from the library. Again, the built-in camera may be used to add existing photographs from the device or to capture new ones. Considering ASD patients have a hard time interacting with neurotypicals or other peers in society, this type of software can be a huge assistance in expressing their emotions.

Cough Drop : Symbol based AAC [2] is another AAC app. The history of building Coughdrop is quite interesting. Brian Whitmer, a software engineer, and entrepreneur was looking for an effective communication method for his daughter who had been diagnosed with RS when he came up with Coughdrop. He was disappointed by bad design decisions and outdated technology because of his experience in usability, so he teamed with around 30 SLPs, OTs, and IT specialists to create something better.

[1] https://apps.apple.com/us/app/letmetalk

[2] https://www.coughdrop.com/

CoughDrop was the end outcome. Brian soon met Scot Wahlquist, whose son has autism and is nonverbal, and the two collaborated to improve CoughDrop for people with any communication requirement. They believed that too many suppliers were attempting to "lock in" clients through proprietary solutions and expensive costs, and they sought to change that. As a result, CoughDrop is open source and includes open-licensed material such as free symbols and community-generated boards, as well as all of our word sets, which are all provided under a Creative Commons license. This software is accessible on the App Store, Google Play, and Amazon, and it may also be accessed through web browsers and Windows. This software allows you to communicate with friends and family in a personalized way across numerous devices. With a simple interface and enough support and teaching, one may gradually increase one's vocabulary. It operates offline with cloud backup, making switching devices a breeze if something goes wrong. It's also possible to share boards with others and open-license them for usage by anybody. across classes, and make access management easier. Individuals may plan a successful approach using built-in goal-tracking tools and gain recommendations for how to strengthen communication tactics from community experts. License consumption may be readily tracked and data can be viewed across rooms, buildings, or teams. People may easily travel between classrooms, and access constraints can be simplified.

ABA Flashcards and games-emotion [3] is a useful program for applied behavior analysis (ABA) therapists and other professionals who work with students with autism and similar problems. This software is only accessible in the app store and is completely free. Telehealth mode for distance learning is one of the app's features. Designed specifically for DTT and intense instruction sessions, Create your flashcards or use the built-in activities. Use photographs from your gallery or look for pictures and gif animations on the internet. Data collecting and automatic grading Multiple student profiles are supported. Autism Read and Write [4]is a Google Play Store app that is primarily developed to enable ASD children in learning the fundamentals of reading and writing. To get started with the reading lessons, click on Start -> Reading lessons. Start -> Writing lessons if you want to learn how to write. The reading and writing classes are of varying degrees of difficulty. Change the levels by going to Settings - > Reading level or Settings - > Writing level.

Social story creator educators [82] is an iOS device software that can assist autistic children in better learning how to deal with various social events, as well as allow them

[3]https://apps.apple.com/us/app/aba-cards
[4]https://play.google.com/store/apps/details?id=com.benitez.autismrandwfree

to create tales about various occurrences. Autism is linked to reciprocal behaviors, making it difficult for persons with autism to react appropriately to a circumstance. As a result, this sort of tool can assist folks who are unaware of how to handle these circumstances. Rethink Ed [5] is a web-based platform that offers autistic children social-emotional learning, mental health awareness, and special education. It guarantees that autistic children have access to the most effective educators who can provide them with a high-quality education. Rethink Ed offers scalable professional development for students with autism, including video models, high-quality lesson plans, and a curriculum. Communication and social skills are emphasized to help individuals with ASD fully engage in their education. AsDetect [83] is a screening application for people with ASD. This software is a useful tool for detecting ASD early on. It is incredibly simple to use; simply sign up for the app, then enter the patient's information and complete an evaluation. After completing these procedures, one may check the results and determine whether or not a person has ASD. They created a demographic questionnaire by administering assessments such as BSID (Bayley Scales of Infant Development)- Third Edition. The goal of Language therapy for children with autism MITA [84] is to create unique, digital apps based on proven, evidence-based early-intervention therapies designed specifically for very young children with ASD. These apps have the potential to significantly narrow the gap between the quantity of therapy prescribed and the amount of therapy received by children with ASD while also improving care quality. MITA's activities use a methodical way of teaching the skill of responding to many cues. The exercises' greatest distinguishing characteristic is their ability to deliver instruction outside of the verbal realm, which is critical for children with ASD who are either nonverbal or just marginally spoken. While these youngsters may not be able to follow a spoken command (such as "pick up the red crayon beneath the table"), preliminary findings from our pilot research (see the "Initial Results" section, below) show that they can obey a command provided visually rather than vocally. MITA helps children develop their creativity and linguistic skills. The visual activities are organized in a methodical way to help your youngster learn to notice many characteristics of an item. MITA begins with easy tasks that educate a kid to focus on only one aspect of a situation, such as size or color. The tasks become increasingly challenging with time, requiring your youngster to focus on two things at once, such as color and size. After the kid has practiced paying attention to multiple puzzles and games. Finally, complex problems require the child to pay attention to an ever-increasing number of qualities.

[5] https://www.rethinked.com/edu

The Jade [6] Autism app aids in the development of cognitive skills by increasing information and improving development. It can act as a mediator in the process of acquiring skills and abilities such as attention and logical reasoning, in addition to assisting in problem-solving, strategic thinking, and decision-making. Activities are categorized into degrees of difficulty within each category, and each phase is only unlocked based on the child's performance, following the natural flow of learning and respecting each child's pace and personality.

2.6 Research Gap

The requirement for cost-effective ASD evaluation, combined with the global increase in ASD cases, necessitated the deployment of rapid and effective evaluations provided by data science and artificial intelligent, including DL and neural network algorithms. Although numerous efforts to evaluate ASD using ML techniques such as genetic data, eye tracking, and functional magnetic resonance imaging (fMRI), has been taken. These types strategy needs further investigations to prove worthy to be used in real life scenario. Maintaining accuracy, ensuring data balance, utilizing a benchmark dataset, and reducing diagnosis duration are all critical considerations in ASD classification. Despite this, a critical flaw in CI-inspired detection and management continues to be a lack of sufficient dataset size. Because ASD patients' symptoms vary, the number of patients willing to participate in data collection methods remains low. Because the majority of studies use data from no more than 50 to 300 ASD patients, the publicly available datasets [45, 46, 48, 47, 49, 50] are significantly smaller in size than other healthcare-related datasets. Additionally, data augmentation is frequently used in a variety of ways to increase the number of data instances. Again, the majority of datasets are based on questionnaires, which may be insufficient to capture ASD patients' subconscious behavior and social deficits. DL techniques have a limited number of implementations due to the scarcity of data. Additionally, time and computational complexity are reduced through the use of feature reduction and selection algorithms. Numerous modalities have been used to detect ASDs, including questionnaires, EEG, ECG, MRI, eye tracking, eye scan path, simulation and task completion, facial imaging, neuroimaging, behavior analysis, body movement analysis, video and audio analysis, and genetics. Though numerous attempts have been made to classify individuals using the aforementioned modalities, the majority of the study focuses on classification using only one or two modalities. Again, due to the

[6] https://jadeautism.com/en/home/

scarcity of data, the majority of studies employ only shallower ML and DL models.

CHAPTER III

Methodology

3.1 Overview

In this chapter, a questionniare based on all possible ASD screening methods is combined into one. The questionniare is used to conduct a survey. The data is analyzed to determine the most prevalent symptoms. Then, based on these symptoms, the next stage of data collection includes audio, behavioral, and video. Following that, a detailed description of the feature extraction process from audio and video data will be provided. Following that, several multimodal architectures for ASD classification were discussed.

3.2 Basic Workflow

The proposed methodology encompasses multiple facets of autism screening and type identification. Historically, autism screening methods have been more dependent on health professionals and checklists; thus, the proposed method will assist the professional rather than replace them. The majority of medical research is devoted to determining or correlating symptoms with ASD or other diseases. In technical terms, the majority of recent ML research is focused on a single aspect, such as a questionnaire, checklist, facial image, eye tracking, eye scan path, genetics, image data, brain imaging, MRI, f-MRI, motor movement, video analysis, audio analysis, task compilation, and simulation. Though in each study, one or two modality combinations with one or more ML methods are tested for autism screening. The purpose of this article is to propose combining the two disciplines into a single multimodal screening method. Thus, we began by compiling a list of all possible biomarkers for ASD and its subtypes. We conducted a small survey to determine the most distressing symptoms in ASD. This process included the creation of a scenario-based questionnaire. Then

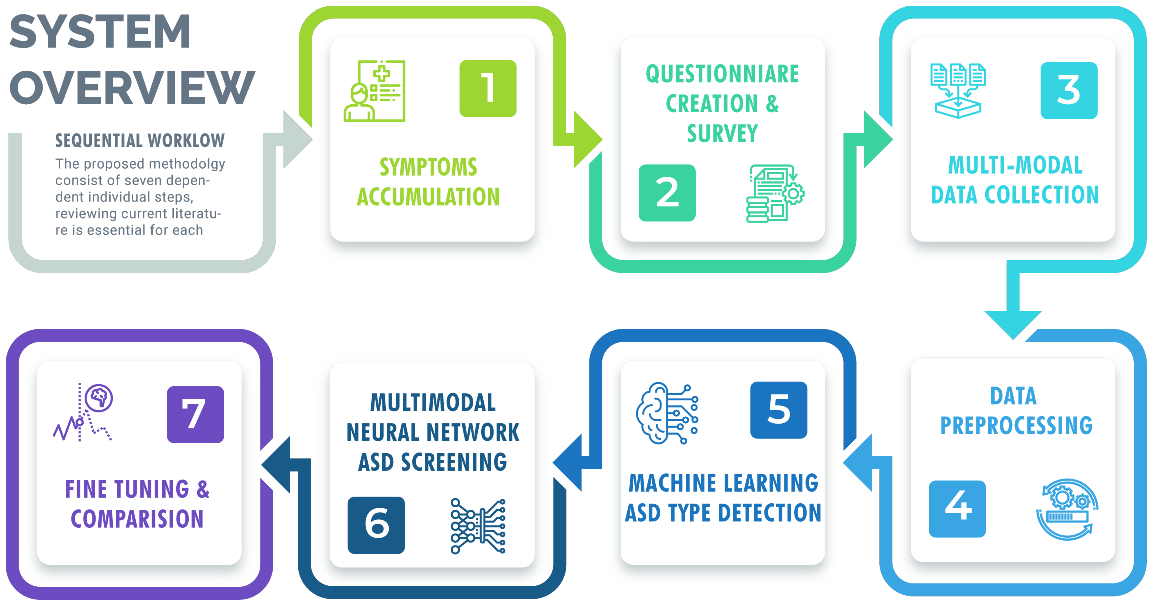

Figure 3.1: System pipeline of proposed methodology.

we expanded this questionnaire into an appropriate simulation and task, allowing us to elicit the subconscious response of the human mind. As an illustration, a scenario is created with the necessary objects and the reaction of a patient is collected for further investigation. This allows for the application of cutting-edge ML models and sensing techniques to a specific patient. We concatenated the predictions of the individual ML models to create a multimodal ML model for predicting overall screening and type detection results. Additionally, we have integrated a smartwatch (Fitbit) for monitoring and detection. This smartwatch can monitor various aspects of human health, including sleep patterns, sudden movement, acceleration, heart rate, abnormal breathing patterns, and sound volume. Daily patient reports can be generated from this data and used by health professionals and caregivers to take further action. Implementing all ML models for each component is beyond the scope of this study. Preprocessing, feature extraction, CNNs, and related neural networks can all be used to implement expressive imaging classification. Additionally, these algorithms can be used to classify eye tracking scan paths. When using video data for traditional eye tracking, eye area, head movement, and behavioral features can be extracted and shallow DL classifiers used. For ASD patient voice data, time series classifiers such as GRU-RNN, LSTM, and Transformer are most frequently used. By concatenating these neural networks, a diverse range of data types can be condensed into a single prediction for ASD Screening.

3.3 Numerical Detection Using Questionnaire:

Previous similar checklists such as Mchat[5] are primarily focused on specific age groups. On the other hand, though DSM-5 [2] gave an overview of symptoms in ASD, the direct questionnaire has not been provided. Again, a straightforward question by asking whether any of the symptoms are present or not in individuals may carry a certain level of human error. The severity of these issues may remain unclear. As a result, a scenario-based severity scaled question was also required. Thus the creation of a question set for determining ASD and its types was necessary. A questionnaire derived from the symptoms mentioned above has been listed in detail in Figure 2.1 with relevant ASD types, which allows measuring across the whole spectrum of autism. The questionnaire consists of 24 questions, with 82 fields representing options for these questions. These questions enable discrimination of the three major components of autism. Each of these options was then graded on a five-point scale. Age, gender, ASD types, and other miscellaneous questions also have been added to the survey.

3.3.1 Participation and Procedure:

The collection of data from various people of various ages with clinically diagnosed ASD has been the survey's main focus. The scenario of each question has been portrayed in such a way that it can be relatable with all kinds of individuals: toddlers,

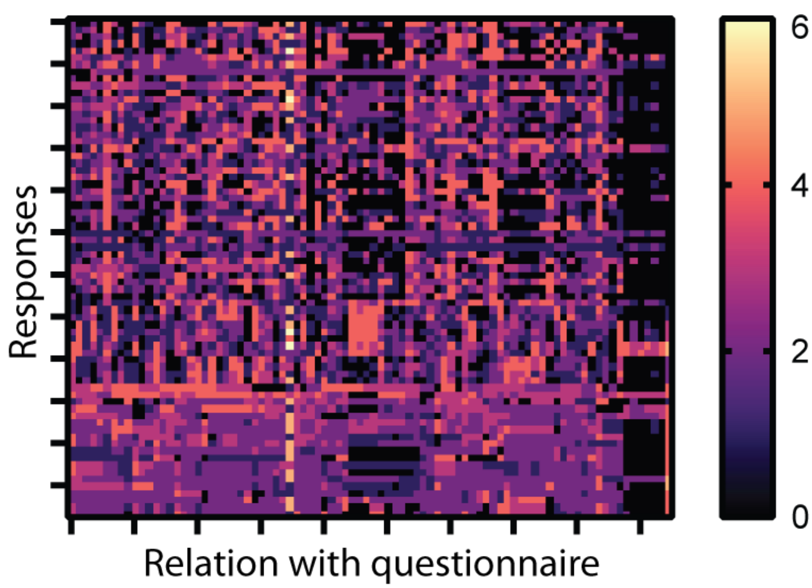

Figure 3.2: Values of responses corresponds to questionnaire

children, adolescents and adults. A google form has been prepared with multiplechoice options from the questionnaire. The form was sent to an autism specialized school, doctors, and students for completion. Filled results have been checked respectively to find out the anomaly. A separate form with the same questionnaire has been sent to ordinary educational institutions. Only form responses correspond to participants who had no prior disorders and were subsequently labeled as neurotypical.

3.3.2 Dataset Details and Data Distribution:

There are 71 data instances in the collection; all acquired from the same number of people [1]. The participants were split into two groups: 42 men and 29 women. The ages of the participants ranged from four to twenty-seven, with an average of 18.8 years. The participants filled out 38 forms, family members filled out 32 forms, and a health professional filled out one. Thirty-nine participants were neurotypical, while 32 were clinically diagnosed with ASD, including 16 AD, 4 AS, 4 CDD, 4 RS, and 4 PDD-NOS patients. Figure 3.2 represents the relationship of responses with the particular question, whereas color represents the value of each field. The last few questions in the proposed questionnaire delineate physical impairment, which is nonexistent for most ASD cases except for Rett syndrome. Hence those fields have been occupied with lower values. For the rest of the questionnaire, the values were evenly distributed.

3.4 ASD Detection from Questionnaire Using ML

Four ML techniques, namely support vector machine(SVM), k-nearest neighbors (KNN), random forest(RF), and artificial neural network(ANN), have been utilized for the classification of ASD and its types. SVM assumes data points as support vectors and uses hyperplanes to separate data into classes. One vs. one has been selected as a decision function shape in SVM, which calculates a hyperplane for two classes at a time. Radial Basis Function (RBF) has been utilized as the kernel. On the other hand, KNN groups together data points based on similarities or distance. The number of neighbors for ASD classification is selected as 20. RF is an ensemble classifier consisting of multiple decision trees, where each tree predicts the output, and the final prediction is given on the majority vote. In the experiment, the number of estimators is set as 20 with two random states and a max depth of 15. SVM, KNN, and SVM have been implemented using the Scikit-learn library. The proposed ANN consists of one input layer, three fully connected hidden layers, two batch normalization layers, two dropout layers, and an output layer. The number of neurons in hidden layers is

32, 256, and 64, respectively. The first two hidden layers utilize rectified linear unit (ReLU) as the activation function, whereas a sigmoid is used in the last hidden layer and the output layer. For loss function, categorical cross-entropy has been used with adam optimizer. All of the ML classifiers have been executed for 100 epochs.

3.5 Feature Extraction Algorithm

Multiple modalities of data were collected for datasets in this investigation. For the wide differences in data forms and the sheer volume of data, training and testing any neural network on a home computer is implausible. Thus, both video and audio data have been subjected to extensive preprocessing in order to extract required features.

3.5.1 Video Feature Extraction

The video feature extraction algorithm takes video sequentially from a directory path. The video files were combined into a single folder and then read loaded via a separate function. The eye iris center, the convex polygonal area of the eyes, the displacement of the head position, the convex polygonal area of the outer mouth, the eye scanpath, and the facial image are all extracted in this manner. For 60 seconds, these features are combined into a single data point, which will be discussed in greater detail later in this section. Mediapipe, Google's open-source framework, is used to detect faces and predict facial landmarks within the bounding box. With a minimum detection confidence of 60%, only one face with 478 points has been taken. To begin, all 478 indexes are listed. The OpenCV library is used to parse each video file. Only valid frames with detected faces are passed on to the next processing stage. Due to the fact that Face Mesh, one of the ML solutions in Mediapipe, only outputs the normalized value of 478 points, the frame's height and width are multiplied to obtain the actual coordinate value of any feature. In Face Mesh, the Iris position consists of four points. To determine the iris's center, a minimum enclosing circle with these four points is calculated using the center and radius. The center indicated by both the x and y axis corresponds to the actual iris center, which is saved in a separate list. Lips points are composed of multiple coordinates. The area of the lips is used to determine how much the necessary target opens, allowing us to determine whether participants are shouting or remaining silent at any given time. To begin, a convex hull is calculated to determine the convex polygon with maximum area. Then, using the summation value of the triangles in the polygons, the area can be calculated.

Similarly, the left and right eye areas are calculated using the points. The eye area is used to determine the participants' level of awareness and blinking. The movement is also determined using the coordinates of the head's center position. The second layer of preprocessing involves further data point reduction. Due to the fact that the majority of the video is 30 frames per second and the window is about 60 seconds long, a total of 1800 data points are listed for one feature. If we take two coordinate values for each eye and iris, the amount of data becomes unmanageable. Thus, two eye areas are averaged into a single eye area, just as two iris coordinates are averaged into a single iris coordinate. The iris coordinates were then plotted as a scatter plot to represent the path of the eyescan. The current position is subtracted from the head position using the two-point distance formula to determine the pixel distance. Only if the distance is greater than the average value summation with standard deviation is the head movement counted. Blinking and awareness in the eye area, as well as shouting and quietness in the mouth area, were also counted. Although these features reduce the model's capacity, the previously mentioned 1201 features are used in the final datasets.

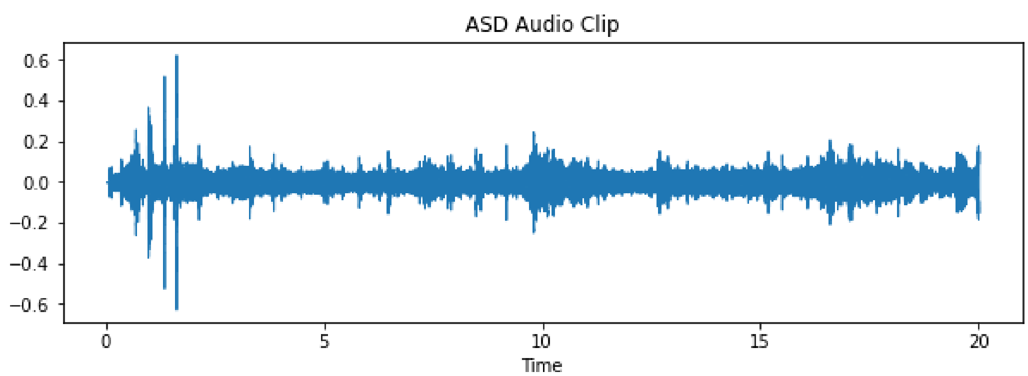

Figure 3.3: Spectrogram of imported audio data

3.5.2 Audio Feature Extraction

ASD patients lack social and behavioral communication abilities. Thus audio from the survey is further analyzed, and preprocessed to be trained on a neural network. At first every sound file is split with a time window of 60 seconds, and saved as a Waveform Audio file (wav). Since the available preprocessing python library utilize .wav file only, this conversion was necessary. For each of the cropped wav file zero crossing value,spectral centroids, spectral rolloff, Mel-frequency cepstral coefficients, chroma features from short-time Fourier transform (STFT), spectral bandwidth is calculated. The number of audio feature in the dataset is 73.

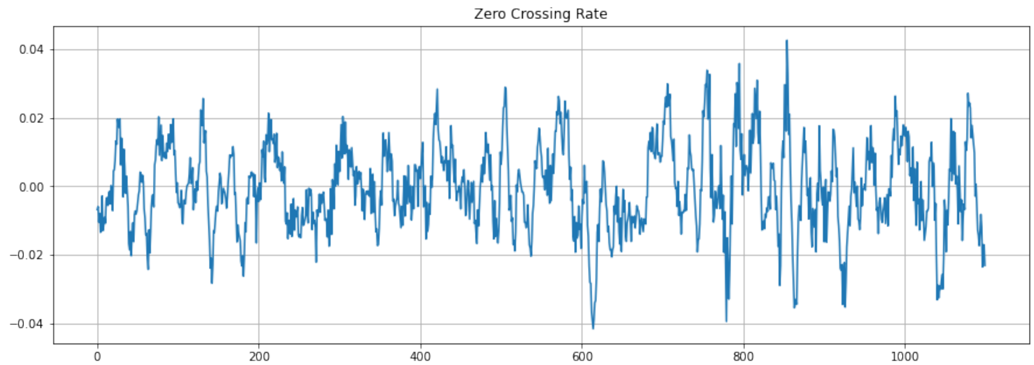

Figure 3.4: Depiction of wave crossing zero point

Zero Crossing Rate The Zero-Crossing Rate (ZCR)3.4 of an audio frame is defined as the rate at which the signal's sign changes across the frame. For example, the number of times the signal changes value, from positive to negative and vice versa, is divided by the duration of the frame to calculate a signal's frequency. The ZCR value can be regarded as a measure of the noise level of a signal in some situations.

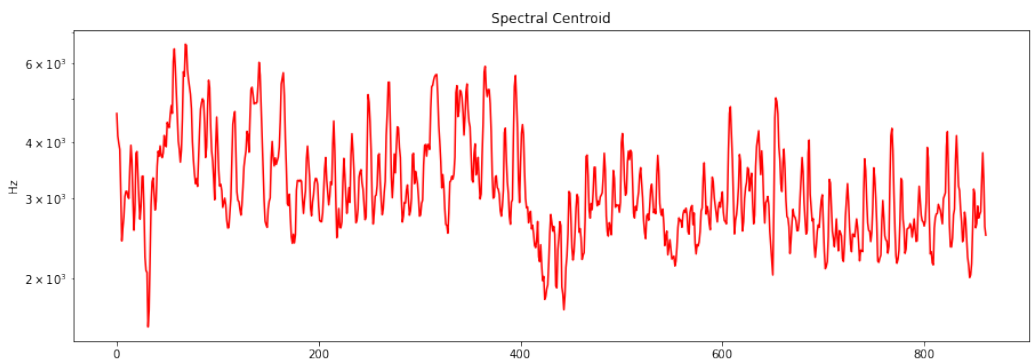

Figure 3.5: Centroids of The Spectrum of The Imported Audio Data

Spectral Centroids It is the centroid of the spectrum, and it is a measure used in digital signal processing to characterize it. It indicates where the electromagnetic spectrum's center of mass is located. It has a strong link with the impression of brightness caused by sound, according to perception.

Spectral rolloff It is the operation of a certain sort of filter that is meant to roll off frequencies that are outside of a certain range. The term "roll-off" refers to the

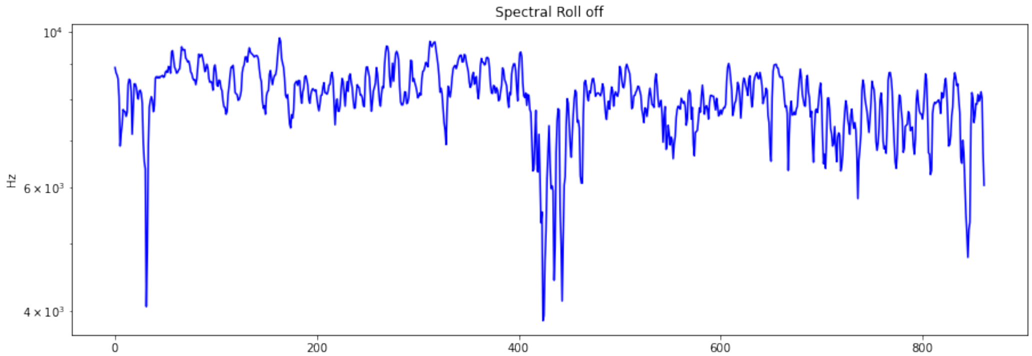

Figure 3.6: Rolloff of The Spectrum of The Imported Audio Data

progressive nature of the practice. Hi-pass and low-pass filters are two types of filters that can roll off the frequency from a signal that goes outside of their range. The spectral roll-off point in the power spectrum is the fraction of bins in which 85 percent of the power is at lower frequencies. Setting the roll percent to a value near 1 and 0 can be used to calculate the maximum and minimum.

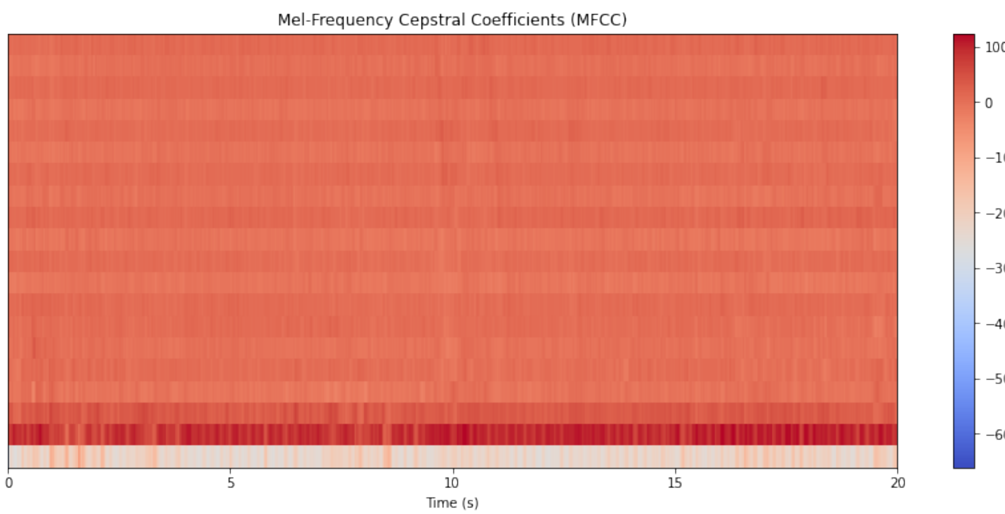

Figure 3.7: Mel-frequency cepstral coefficients of Imported Audio Data

Mel-Frequency Cepstral Coefficients The initial stage in any automatic speech recognition system is to extract features, that is, to identify the components of the audio signal that are useful for detecting the linguistic content while ignoring everything else, such as background noise, emotion, and so on. The critical aspect to grasp speech is that the sounds produced by humans are filtered through the form of the vocal tract, which includes the tongue and teeth. This shape dictates the type of sound that is produced. If we can precisely establish the form, we should be able to obtain an exact representation of the phoneme being produced. The form of the vocal tract is represented by the envelope of the short-time power spectrum, and MFCCs are responsible for appropriately representing this envelope.

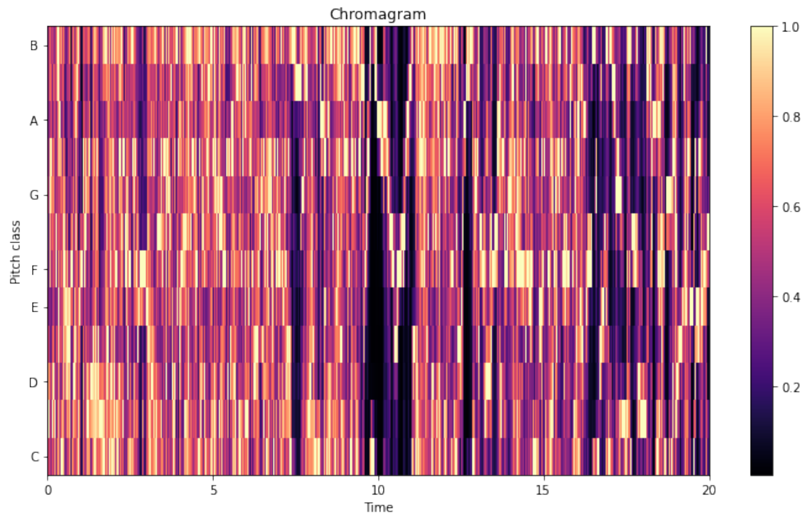

Figure 3.8: Chromagram of The Imported Audio Data

Chromagram Because the term "chromatography" relates to the process of separating distinct components from their combination, we may comprehend the term "chromagram" in the context of audio files. When it comes to audio file analysis, an audio file can contain up to 12 distinct pitch classes. These pitch class profiles are extremely valuable for audio file analysis. The term chromagram refers to a representation of the pitches contained within an audio file in one location, in order to facilitate the classification of the pitches contained within the audio files. Pitches are a feature of any sound or signal that enables file organization on a frequency-related scale. It is a measure of the sound's quality that aids in classifying the sound as high, low, or medium.

Root Mean Square The root-mean-square (RMS) amplitude or level analysis is used to determine the overall amplitude of a signal. It refers to the average signal amplitude conceptually. It is, however, not the same as calculating the arithmetic mean of a signal. Positive and negative amplitude values are possible for an audio signal. If we took the sine wave's arithmetic mean, the negative values would cancel out the positive values, yielding zero. This method provides no information about the average signal strength. This is when the RMS value comes in handy. It is based on the magnitude of a signal as a proxy for signal strength, regardless of its amplitude.

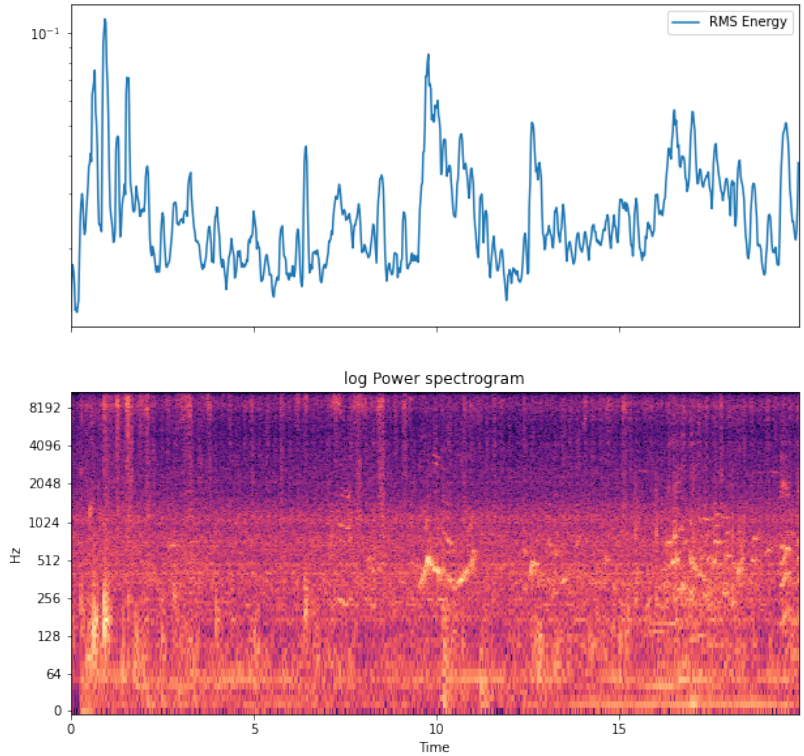

Figure 3.9: Root Mean Square of Signal to Noise Ratio

The magnitude is computed by first squaring each sample value (to ensure that they are all positive), then calculating the signal average, and then performing the square root operation. More precisely, the RMS level is "the square root of the signal's arithmetic mean squared.

3.6 Multimodal Dataset Details

The multimodal dataset consist of nine different modalities grouped into five sections. From the previously mentioned questionnaire studied, the question and task with highest correlation values are asked to the participants. Though Asperger's, PDD-NoS, AD, and Rhett syndrome patients were grouped in to the ASD class. The first one of the modalities, is the facial images of both ASD and neurotypical persons. Eye scan path is created from the video data. Then behavioral information is collected from the eye area, and mouth area. Motor movement is calculated from the head position and also from gyro accelerometer sensor of Fitbit watch. This watch also used to collect information like, heart rate, activities and calories consumed. Lastly the audio data is also taken. The final dataset consist of one audio input with 73 features, facial image, eye scan path, 1201 video extracted data and 8 data instance from the fitbit watch the total amount of data collected is 1960. It is necessary to mention that several instance is collected from one participates.

3.7 Neural Network Layers in Multimodal Classification

There are several layers for implementing each of of the neural networks necessary for multimodal neural networks.

3.7.1 Dense

In a neural architecture, dense layers are highly interconnected with one other and with the layers above them. This is the most prevalent layer in neural networks. Neurons in the dense layer receive input from neurons in the preceding layer during matrix-vector multiplication. In matrix-vector multiplication, the previous layers' row and column vectors are alike. The row vector and column vector must both have the same number of columns. The matrix and previously learned parameters may both be updated via backpropagation. Backpropagation can be used to train feedforward neural networks. Backpropagation is used to calculate the gradient of the loss function for single input or output in a neural network. Based on the available data, we may deduce that the dense layer's output is an N-dimensional vector. The size of the vectors is shrinking. Each neuron in a dense layer effectively changes the vector size.

3.7.2 Dropout

In order to simulate the simultaneous training of several different neural networks with various topologies, the dropout regularization approach is used. Certain layer outputs are "left out" at random during training. Thus, it appears and behaves as though it were a whole new layer, one with a distinct set of nodes and connections from the one it follows. Each "view" of the current layer is used to modify a layer during training. Dropout adds noise to the training process by requiring nodes inside a layer to take on a stochastic distribution of responsibility for the inputs. Dropout, in this model, separates instances when network echelons co-adapt to address flaws introduced by preceding levels, hence increasing the architecture's durability. Dropout emulates sparse initiation from a particular layer, which has the unintended consequence of encouraging the model to acquire a sparse representation. As a result, it may be used in place of activity regularization to promote sparse representations in autoencoder models. Due to the random subsampling of the outcomes of a layer with dropout, the capability of the network is reduced during training. As a result, when dropout is used, a larger network, i.e. additional nodes, may be required.

3.7.3 Batch Normalization

Normalization is a technique for translating quantitative data to a standard scale without altering the data's shape. We frequently balance the sizes of the numbers when entering data into a ML or DL system. To verify that our model generalizes correctly, normalization is required. Batch normalization is a method for expanding the quantity of layers in a deep neural network to improve its efficiency and reliability. On the input of the previous layer, the next layer conducts standardization and normalization processes. Batch data, which is a collection of input data, is frequently used to train a neural network. Batch normalization, on the other hand, works in batches rather than as a single input.

3.7.4 Convolution

When two functions are combined in an ordered fashion, convolution is the process of transforming one function into the other. Using convolutions in image processing has been around for a long time, but they can also be utilized for a variety of purposes. To emboss and enhance the edges, for example, CNN's ensure that neurons in nearby layers communicate with one other via the same connections. Image properties, such as edges, can be detected by CNN using the filters (also referred to as kernels) utilized by the algorithm. The first layer of a Convolutional Neural Network is often a Convolutional Layer, as the name suggests. The output of one layer is sent to the next via a series of convolutional layers. Receptive area pixels are all turned to one single value using convolution. It is possible to minimize the overall image size while still retaining all relevant information by using a convolutional algorithm on the image. The final output of the convolutional layer is a vector. It is possible to use different forms of convolutions for different problems and learning objectives.

3.7.5 The 2D Convolution Layer

It is the most often used type of convolution, and it is denoted by the abbreviation conv2D (for two-dimensional). A filter or kernel in a conv2D layer conducts component-wise multiplication by "sliding" through the 2D input data in a conv2D layer. The result will be an output pixel consisting of the total of all of the results obtained thus far. If the kernel slides over a point in a 2D matrix of features, it will convert that 2D matrix of attributes into another 2D matrix of features.

3.7.6 Maxpooling 1D

There are layers called max pooling after convolutional layers that let the inner convolutional layers get input from more of the original vector. This lets them work with a bigger part of the original vector than before. The maximum number of layers can be pooled at one time. If we think of convolutional layers as detectors of a unique feature, max-pooling only keeps the "strongest" value of a feature that is in the pooling rectangle. For each of the channels, there is a different way to go about it (and subsequently, each characteristic). Each window of data is swept over by an infinitely large pool (window) with a predetermined stride. The maximum is then found for each window of data.

3.7.7 Maxpooling 2D

When given a tensor of size (input width) by (input height) by (input channels), the 2D Global max pooling block computes the maximum value of all values throughout the whole (input width) x (input height) matrix for every one of the input channels in the input tensor (input channels). As a result, the output is a 1-d tensor of size.

3.7.8 Attention

The attention mechanism enables a neural network to store extended input sequences and is commonly employed in neural machine translation (NMT). As stated previously, the encoder compresses the sequential input to a context vector. We may employ attention to construct a link between all input and the context vector, with different weights for each output. Because input and context vector is linked, the context vector may access the full input, reducing the risk of missing extended sequences.

3.7.9 Self Attention

Intrinsic attention, also known as self-attention, occurs when an attention method is applied to a network in such a way that it may relate to various points of a same sequence while simultaneously calculating a representation of the same sequence.

3.8 ASD Detection with Multimodal Neural Network

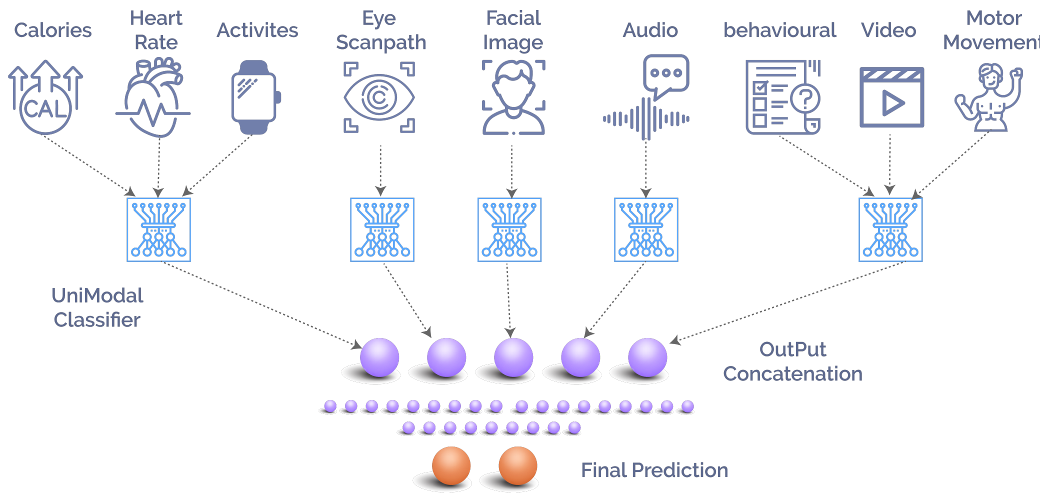

Figure 3.10: Proposed Multimodal network with Uni model concatenation

Multimodal neural networks takes several types of inputs and output the targeted label in the supervise learning. Although multiple instance or modalities increases the effectiveness and over all accuracy of any detection systems; there are still no golden rules for the model creation. Furthermore the necessity of all types of data some makes it harder for any data instances of less modalities. In this study, three types of multimodal neural networks has been proposed and implemented to test the effectiveness of each approach. The first approaches is to built neural networks for each of the modalities, output the predictions then concatenate those production further in a dense layer and train the whole architecture as a single one 3.10. For this study five such networks has been created. For image classification, convolutional neural networks with two dimensional convolution , max pooling , dropout has been utilized each of convolution part consist 3 by 3 kernel filter followed 5 by 5 kernels. After applying these operation several times, dense layer has been added for the classification purposes, the last layer before concatenation consist of one neuron with sigmoid activation. Two such network has been used for eye scan path, and facial images. Audio data has been classified with RNN. LSTM layer has been utilized with dropout to capture the temporal dimensionalities. After that sequential dense, dropout and batch normalization has been utilized. The last dense layer has been added for the classification purposes. The input from Fitbit data is fed into fully connected layer with dropout. Similar to previous network the last out layer contains one neuron with sigmoid activation. Behavioral data from video data is large in dimension. To reduce this, one dimensional convolution and one dimensional max pooling is utilized with dropout. Subsequently the inputs are reduced from 1201 to 256. Dense layer then added for the classification purposes. The last neurons from all of these layers then concatenated into one dense layer. Two hidden layer then applied before the final classification.

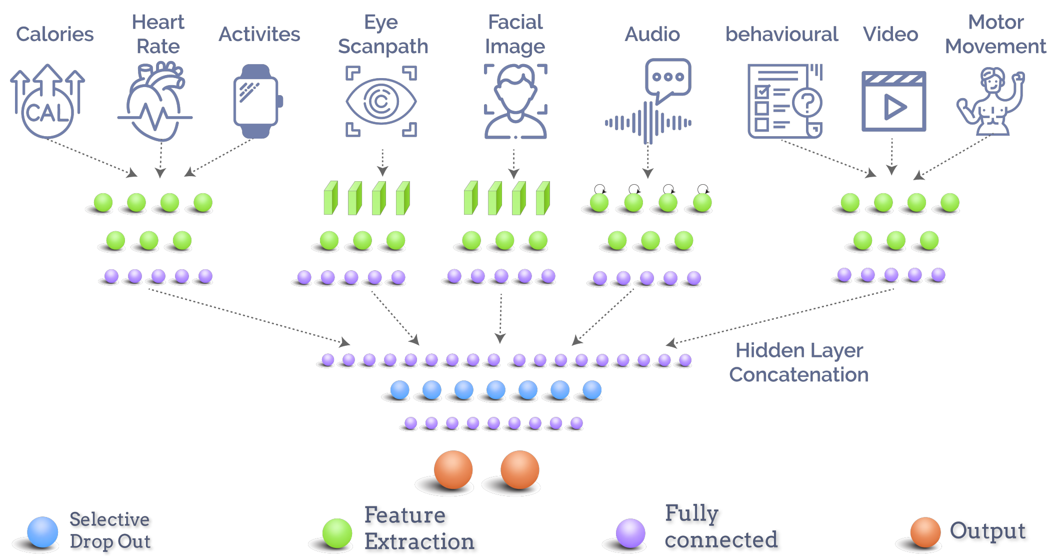

Figure 3.11: Proposed Multimodal network with hidden layer concatenation with selective dropout

The second model3.11, concatenation happens in the hidden fully connected layers of each models. This allowed shared hidden layer for each of the neural networks subsets. Similar to the previous approaches, two 2D-CNN.one 1D-CNN, one RNN with dense layer in the last , is concatenated for the classification. Except, in this approach the dense layer disjoint to each other consist of 64 to 256 units with ReLU activation. In some version of The analysis , attention layers also added after concatenation. This followed by a sequence of shared dense layer, batch normalization and dropouts. ReLU activation is applied in all of these dense layers. The final output layer consist of only one neurons, because of binary classification. The most prominent model, however, comes from the concatenation of features layers3.12. Then this concatenation is fed into fully connected layers shared by the all data modalities. Initially each of the data modalities is given into the feature extraction layers of neural networks like 2D convolution with excitation blocks, that is a smaller convolution followed by a larger kernel size and filter size then again reduced into smaller filter size again. These also reduces the computational complexity of large operation, at the cost of model complexity. Rest of the neural networks are similar to previous two models excepts there is no disjoint dense layer added to them. The shared dense layer consists of

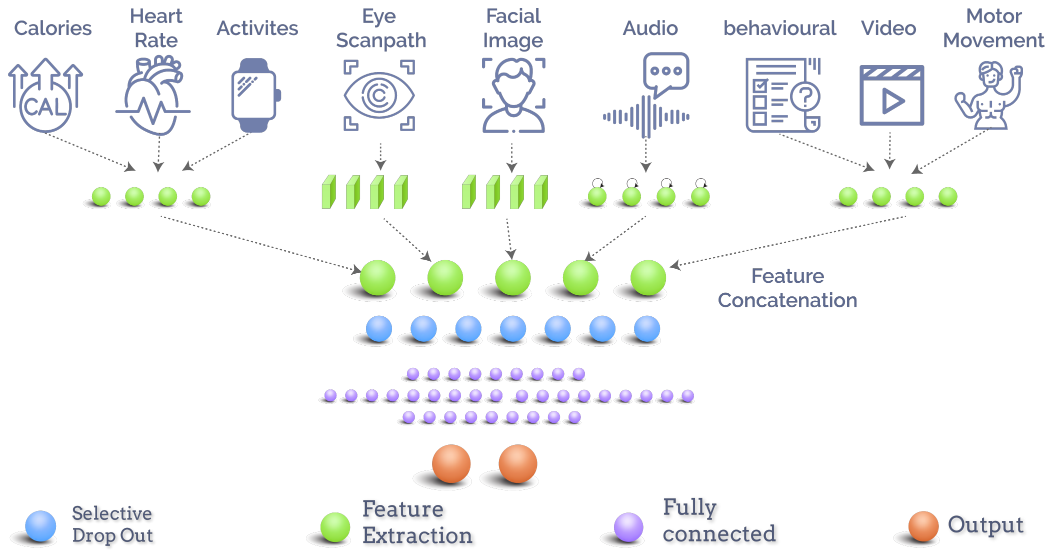

Figure 3.12: Proposed Multimodal network with feature layer concatenation with selective dropout

relatively smaller number of neurons then 5 model calculating separately. However there is two problems in the above mention models. The first one is to oscillation of loss that lead to the fluctuation of results also. Another problem, in the case of any input data consist of less modalities then expected, the out can be predicted.

3.9 Proposed Selective Dropout Layer

As mentioned previously, the models required to handle one or several missing data modalities. However learning using traditional dropout layer randomly dropout any of the given percentage of the neurons, does not imply training capacities for all type modalities. Again an weighted average prediction from single modality classifier can perform better than the multilabel in case of missing data, it just eliminate the model that has no data from the model stack. The idea behind the selective dropout is to drop one modalities randomly during each batch training to ensure that the model is capable to learn from all of the modalities. And it will be most effective to use after the concatenation layer. From previous layer input C, let each of modalities weight contains $C_{i-1} to C_i$. For all know i if we select one randomly and multiply with zero, the weights are dropped for i_{th} modalities. Thus enables better training and testing facilities. Algorithm 1 represents the idea of implementing the selective dropout layer.

Algorithm 1 Calculate Dropouts for tf.Layers

Require: $selectionArray$
Require: $previousLayer$
Ensure: $output = dropout$
  $len \leftarrow size(selectionArray)$
  $rand \leftarrow randominteger\%len$
  $start \leftarrow selectionArray[rand - 1]$
  $end \leftarrow selectionArray[rand - 1]$
  while $start \leq end$ **do**
    $previousLayer[start] \leftarrow 0$
    $start \leftarrow start + 1$
  end while

3.10 Proposed Multimodal Architecture

The proposed multimodal neural networks, has separate feature extraction layer for each of the modalities, outputting the smaller feature set, then concatenating those outputs in a dense layer and training the entire architecture as a single neural networks3.13. Binary cross entropy loss function and adam optimization has been utlized in this study. Five such networks were developed for this investigation. Convolutional neural networks using a efficient-net style block has been utlised for feature extraction. Each convolution phase consists of 3 by 3 kernel with N filters followed by 5 by 5 kernels with 2*N filters then again a 3 by 3 kernel with N filters. Two dimensional maxpooling, batch normalization along column axis and dropout layer is also added after the convolution in the block. After performing these block operations four times, the output from the last block is sent for the concatenation. The output size is 256. For eye scan paths and facial images, two such networks have been deployed. RNN was used to classify audio data, while two LSTM layer with dropout was used to capture temporal dimensionalities. Following that,a sequential dense, dropout, and batch normalization were used. Fitbit data is delivered into a fully connected layer that has a dropout. The last out layer, like the preceding network, has been kept seperate for concatenation. Behavioral data with dropout, one-dimensional convolution and one-dimensional max pooling is processed for feature extraction. All of these five network output then concatenated into a dense layer with a shape of 1024. Then the proposed selective dropout is applied. Which will reduce the dependencies on a specific modality. Though the dropout layer can randomly drops any connection from previous layer, limiting an specific modality in batch train results with better training. Then three fully connected layer with 256,256,64 neuron respectively is applied before the output.

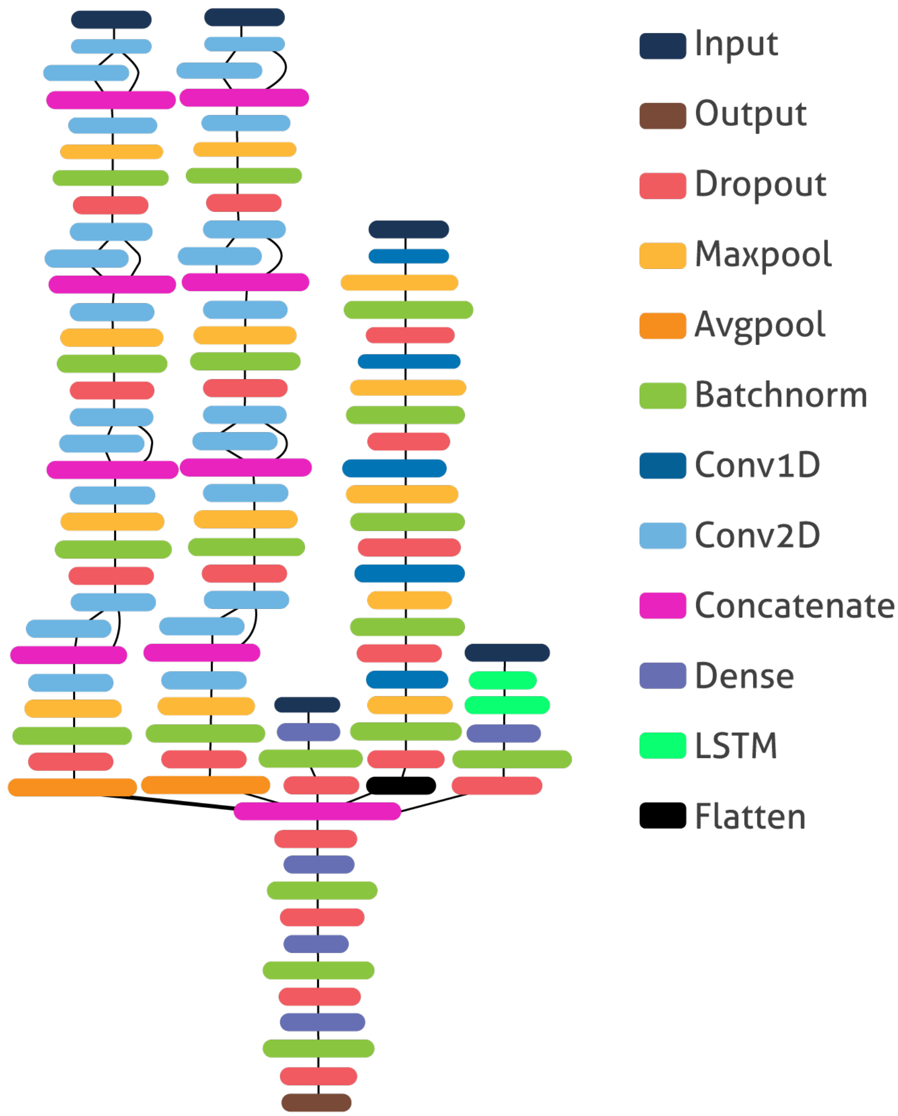

Figure 3.13: Proposed Multimodal network with selective dropout after Concatenation

CHAPTER IV

Experimental Analysis

4.1 Overview

In this section, first statistical metrics for comparisons are given followed by data analysis from two survey. Then ML based type detection and multimodal screening for different methods are compared.

4.2 Statistical Metrics

Accuracy This statistic reflects how well the method fits across all classes and is used to evaluate its accuracy. It is beneficial when all of the classes are of similar significance to the student. Heuristics are used to compute this as the ratio of correct guesses to the grand total of hypotheses.

Positive Predictive Value or Precision As the number of correctly categorized Positive samples divided by the total number of specimens classified as Positive, the precision is determined . As the name implies, precision assesses how accurate the model is in identifying positive samples. This leads to an increase in the numerator and a decrease in precision when the model generates many wrong Positive classifications or few right Positive classifications.

Recall The recall is determined by dividing the total amount of Positive samples by the quantity of Positive samples that were incorrectly labeled as Positive. The model's capability to identify Positive samples is measured by the recall. More positive samples are found when the recall is larger. Only the classification of positive samples is of interest to the recall. Regardless of how negative samples are categorised, for example, for accuracy, this is how they are analyzed: If the model mistakenly identifies all of the negative instances as Positive, the recall will still be

100% if the model labels all of the samples tested as Positive. Let's take a look at a few samples to get you started.

F1-score The harmonic mean of the accuracy and recall of a classifier is used to get the F1-score. You may use this to see how well two classifiers compare. Consider the following scenario: Classifier A has a greater recall, whereas Classifier B has a better precision. F1-scores may be used to compare the performance of the two classifiers in this situation.

AUC For model assessment, AUC (Area under the Curve) is a popular statistic. In most cases, it's employed to solve difficulties involving the binary categorization. The total two-dimensional area under the ROC curve is measured by AUC. As a rule of thumb, the AUC of a classifier is equivalent to the percentage that a randomly selected positive example ranks higher than a picked at random negative example. The ROC curve may be summarized using the Area Under the Curve (AUC), which allows a classifier to discriminate between different classes. The greater the AUC, the better the model's ability to differentiate between positive and negative classes is considered to be.

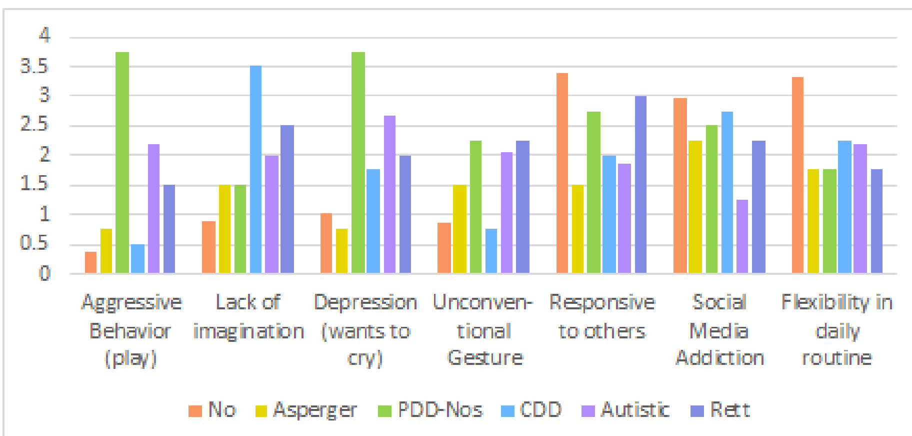

Figure 4.1: Mean value of symptoms relevant to ASD types and neuro-typical

Binary cross entropy loss In the case of binary cross entropy, each of the expected values is compared to the actual class output, which can either be 0 or 1. Afterwards, it generates the score, which disallows the possibilities based on how far they are from the predicted outcome. That is, how near or how distant the estimated value is from the real value.

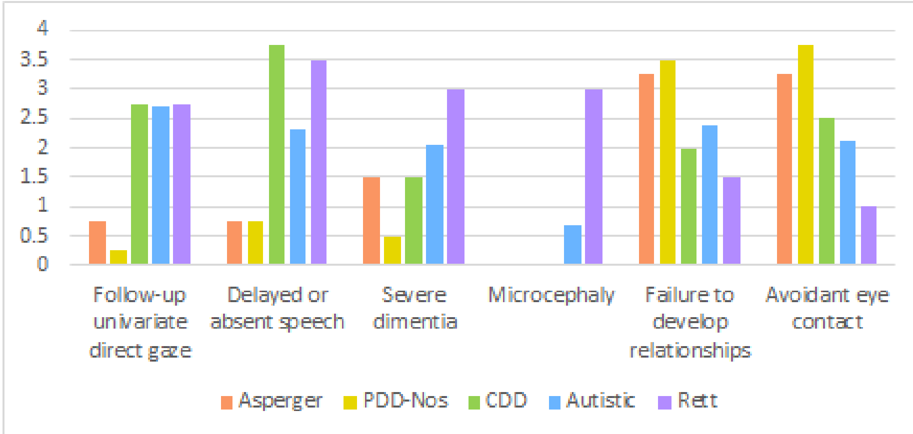

Figure 4.2: Mean value of symptoms relevant to ASD types

4.3 Questionnaire Analysis

Correlation analysis assesses the extent and orientation of the relationship between input and output variables; in this case values of each question and ASD categories. Figure 4.3 shows the top 18 symptoms with the highest correlation value, where blue represents negative correlation and red represents positive correlation. The sever-

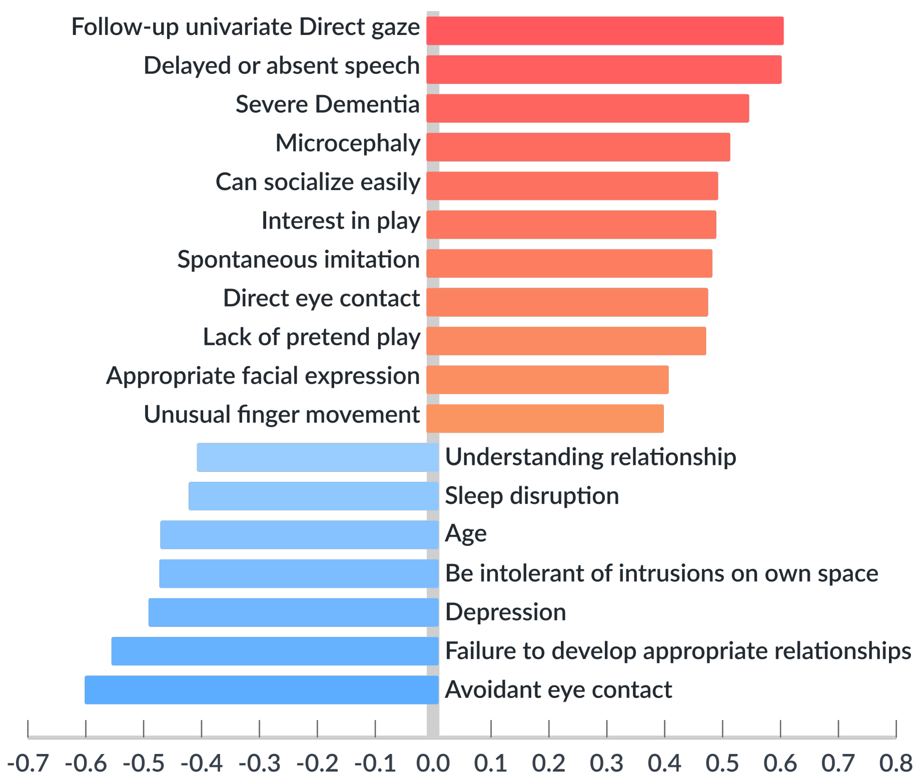

Figure 4.3: ASD symptoms with most correlation in Dataset

ity of the dataset has been transformed into numerical values ranging from 0 to 4.

The occurrence of any specific value was then determined using the mean value of

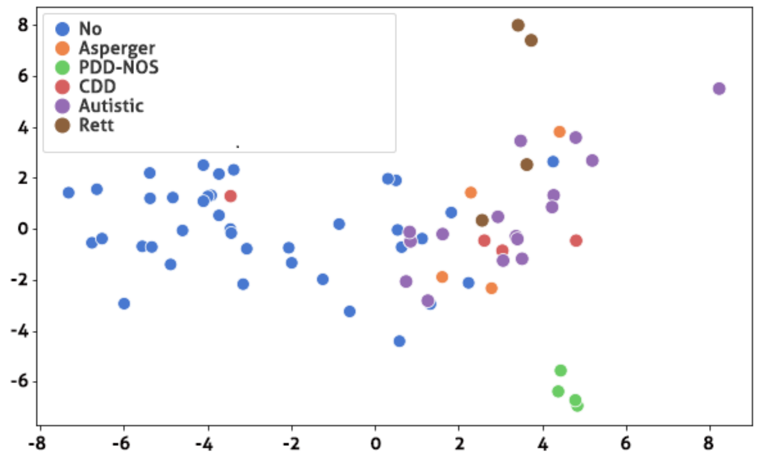

Figure 4.4: PCA analysis of dataset

the symptoms in ASD and neuro-typical individuals. Figure 4.1 depicts the seven symptoms with the highest association between ASD types and neuro-typical traits. To determine the most common symptoms among ASD categories, the correlation between ASD types and questionnaires was evaluated individually. In figure 4.2, the symptoms with the highest correlation have been depicted with a mean value. As

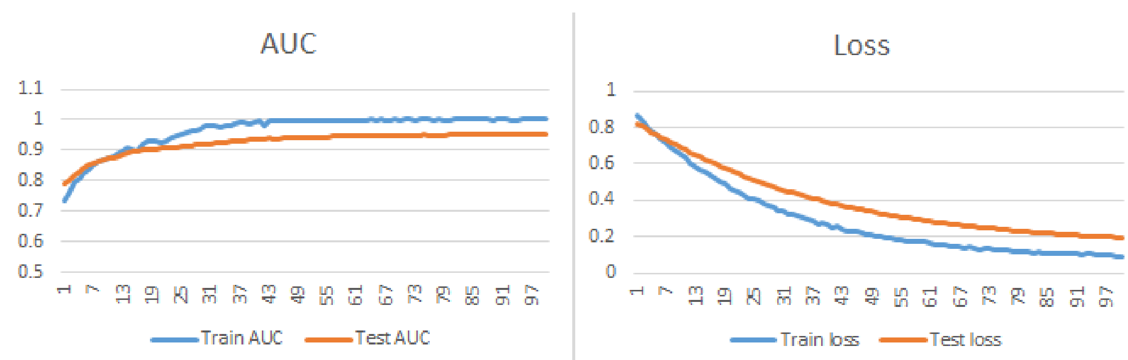

Figure 4.5: Epochwise Area Under Curve (AUC) and Categorical Cross-entropy loss of ANN

a result of principal component analysis (PCA), datasets become more interpretable while avoiding performance degradation. It accomplishes this by generating new negatively correlated parameters that sequentially optimize variance—principal component analysis of two components in the ASD dataset depicted in figure 4.4.

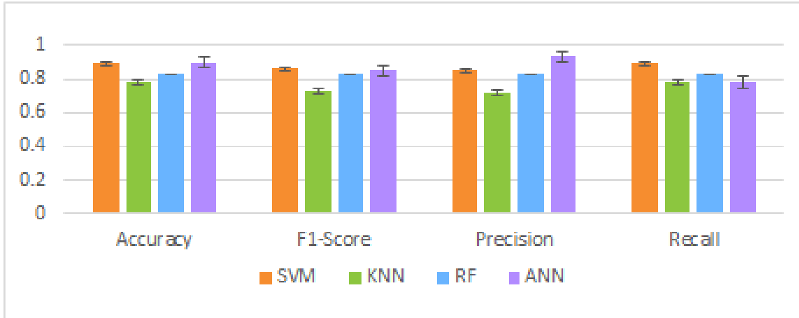

Figure 4.6: Comparison of performance metrics among classifiers

4.4 ML Detection from Questionnaire

The dataset has been split into 20% data for testing and 80% data for training. Four ML models(SVM, KNN, RF, ANN) have been trained and tested on the accumulated dataset. Accuracy and F1-Score have been calculated for model perfor-

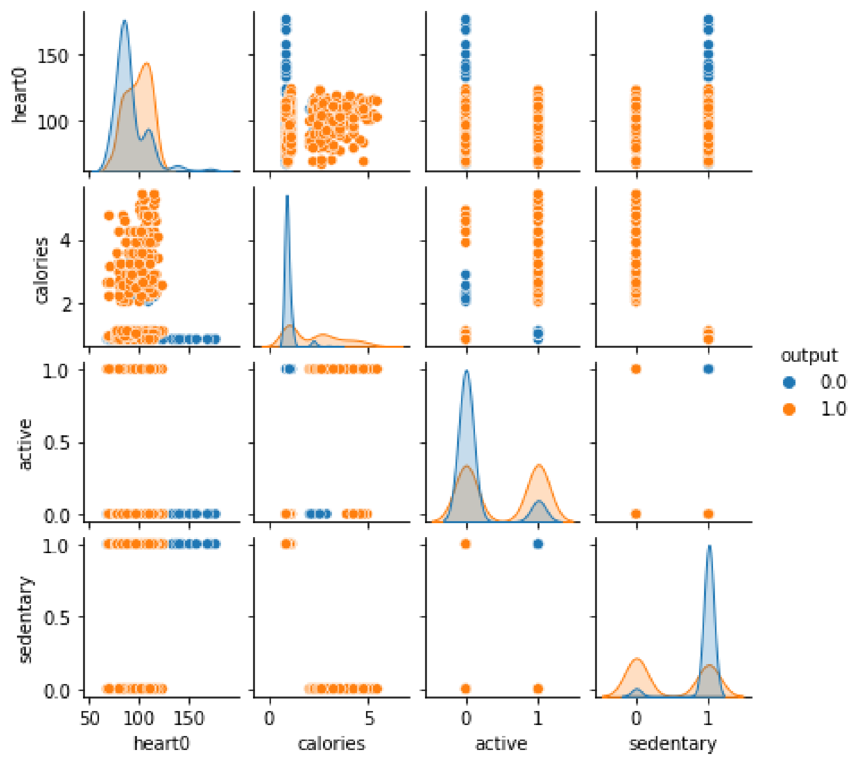

Figure 4.7: Fitbit Data Pairplot

mance and comparison. The percentage of correctly predicted classes, both positive and negative, is referred to as accuracy. F1-score is the weighted average of accurate classification. Among total positive predictions and valid classification among correct positive and false negative predictions. Testing accuracy for SVM, KNN, RF, and

ANN was 89%, 78%, 83%, and 89.8%, respectively, with training accuracy near 100% for all classifiers. The achieved F1-Score of SVM, KNN, RF, and ANN in testing data is 86%, 73%, 83%, and 85% subsequently. Figure 4.6 depicts the comparison of accuracy, recall, precision and F1-Score among ML classifiers. The evaluation metrics show that SVM and ANN perform significantly better than KNN and RF for ASD classification. The epoch-wise test and train AUC and loss of ANN is depicted at figure 4.5.

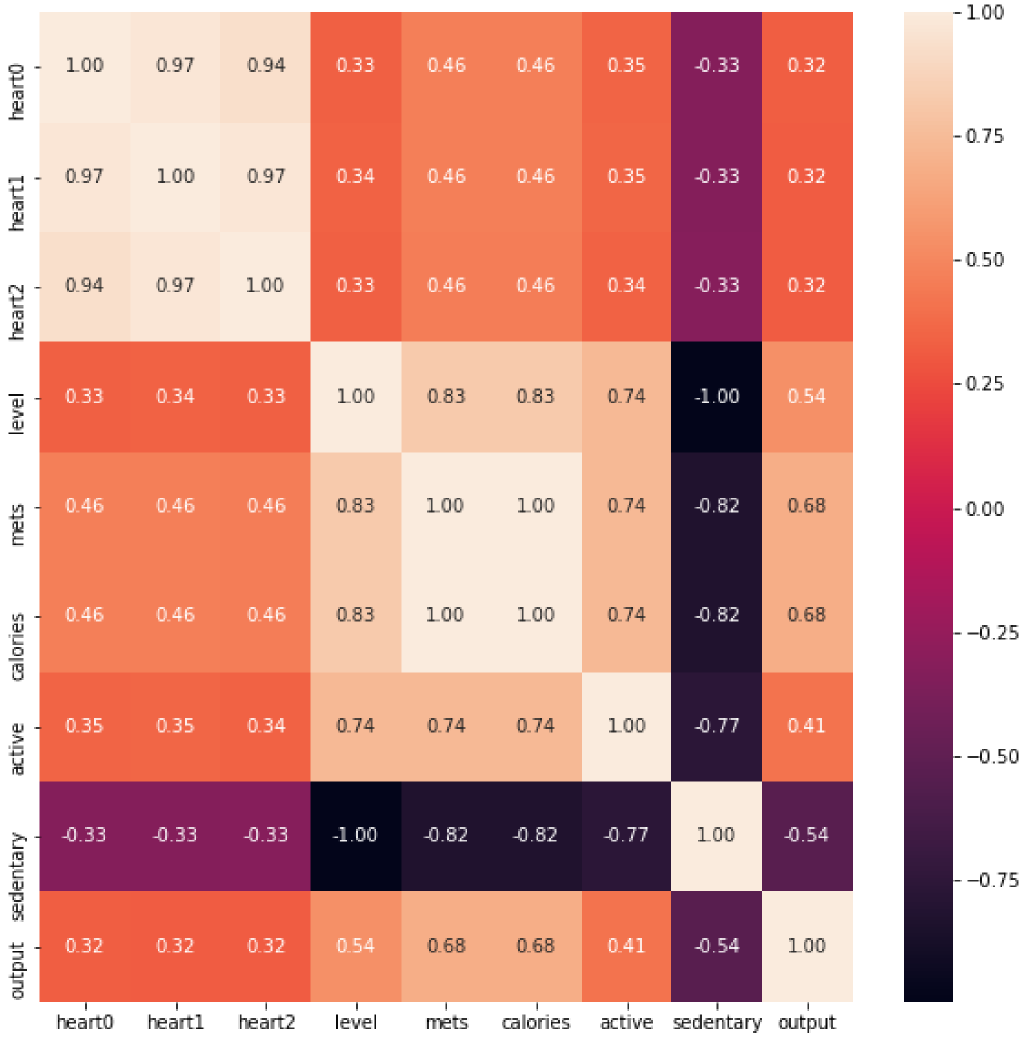

Figure 4.8: Fitbit Data Correlation Heatmap

4.5 Management Data from Smart watch

Eight features has been selected for the experimental analysis from the fitbit smart watch. This watch is widely available and have open API that can integrated into any application design. Which was also used for collecting data from cloud. Fitbit servers also track every data points for almost a year. Heart rate, mets, calories, activity and sedentary has been utilised. Figure 4.8 represents the correlation between each

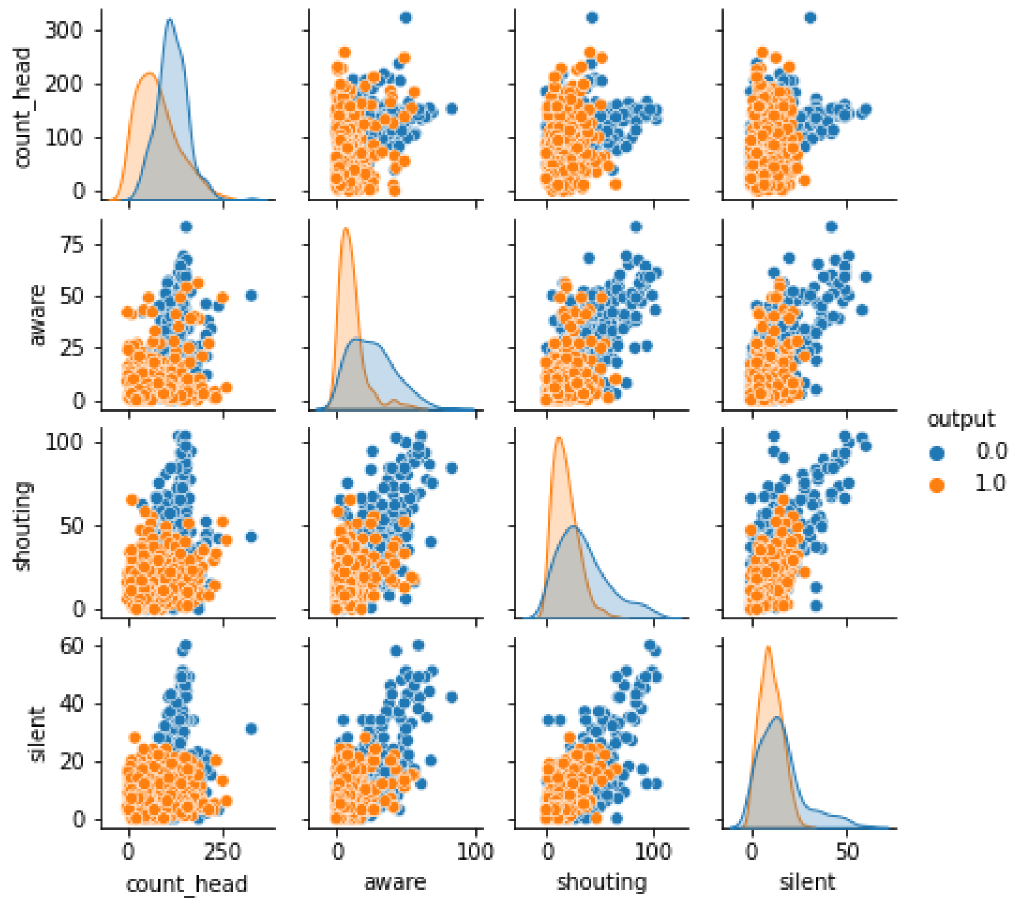

Figure 4.9: Behavior Data Pairplot

values, here heart0 represent the first heart rate. Among them calories has the highest spearman correlation value of 0.68. Heart rate has a significant 0.32 correlation with the output. The highest correlation value another pair plot has been generated in 4.7. Here the color represents 0 for neurotypical and 1 for ASD patient. This presentation visually distinguished the difference between the input feature with outputs.From the calories vs heart rate plot it is shown that the out put can be differentiated. On the other hands, behavioral data were further compressed into five features, although the multimodal models works with the uncompressed data, the visualization shows that it has a good amount of correlation with the outputs. Again similar pair plots at 4.9 shows the differences for the neurotypical persons and ASD in terms of different features. This also visualize the effect of each features for the classification of ASD and neurotypical.

4.6 Multimodal Comparisons

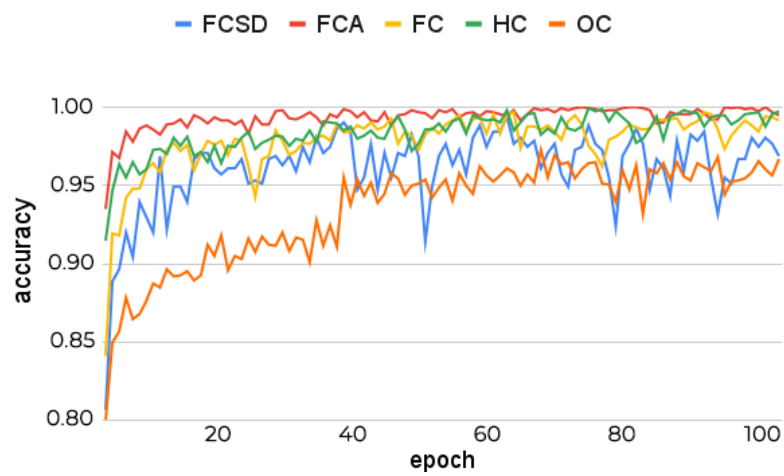

Figure 4.10: Train Accuracy for Different Multimodal Concatenation Technique

To compare the deference of several proposed multimodal architectures, a base model was first created. All of the neural networks shared characteristics such as RelU activation in the hidden dense layer, a dropout rate of 0.25, Adam optimizer, and binary cross entropy loss. The multimodal dataset is divided into two parts: 80% of each modality is used for training, and the remaining 20% is utilized for testing. All models were trained and tested for a total of 100 epochs.

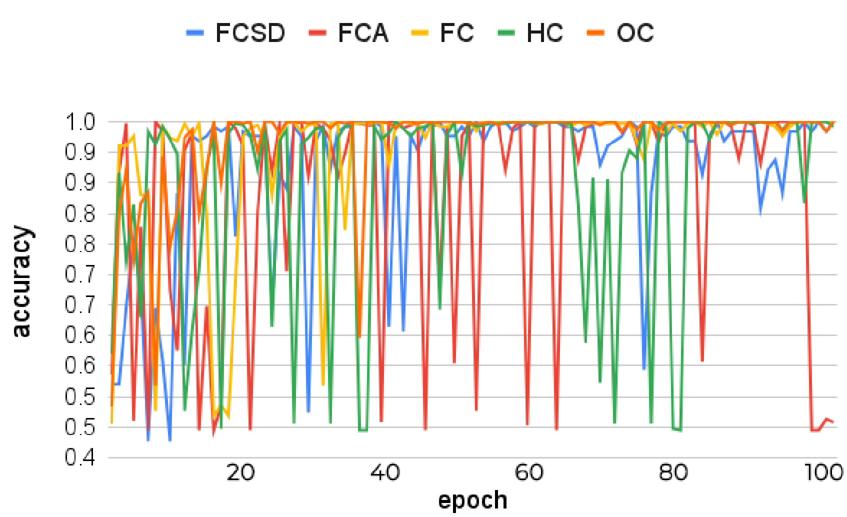

Figure 4.11: Test Accuracy for Different Multimodal Concatenation Technique

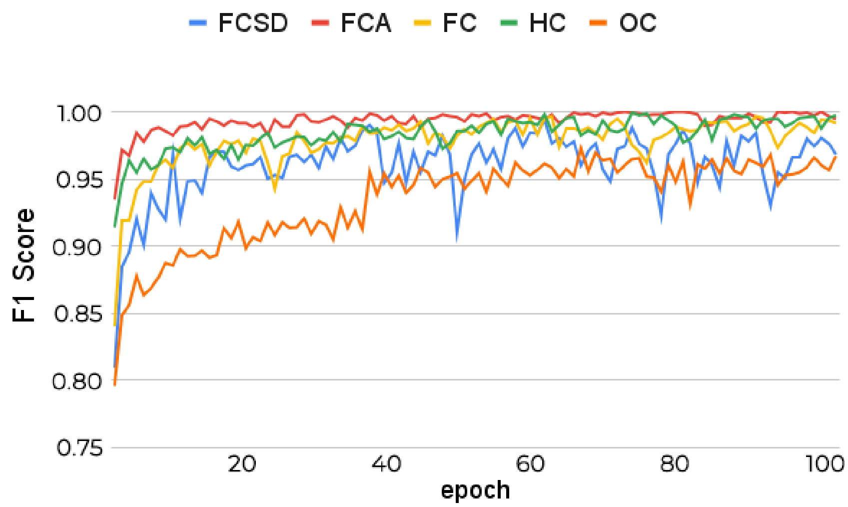

Figure 4.12: Training F-1 Score for Different Multimodal Concatenation Technique

Output Concatenation(OC) In output concatenation architectures only the output sigmoid neurons were concatenated. Each of the modalities were passed into relevant neural networks for feature extraction and classification. Thus each of the network before concatenation is exactly similar to those found in classification from single modalities or uni-modal. Each of the sigmoid output from five modalities then passed into fully conneceted layers with 256, 64 units were used in this architecture. Each of the dense layer were followed by batch normalization and dropout layers with

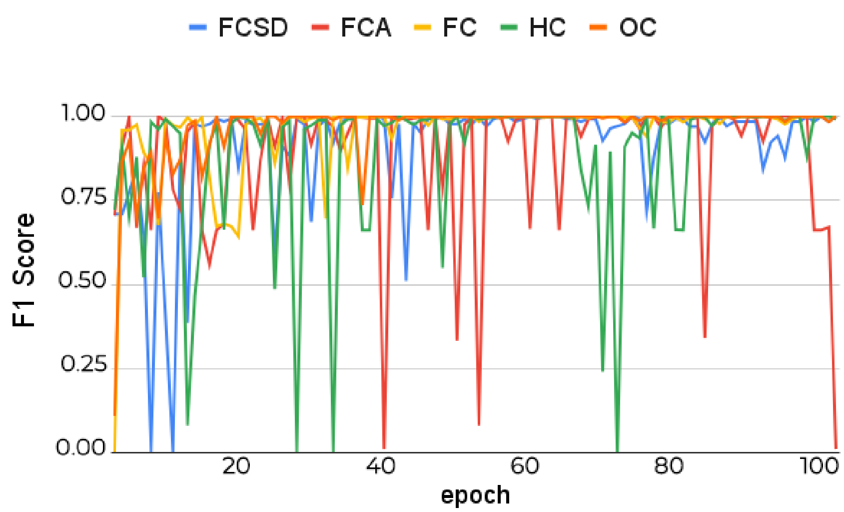

Figure 4.13: Testing F-1 Score for Different Multimodal Concatenation Technique

a dropout value of 0.25. The dropout value were similar in all of the models. The output layer had *sigmoid* activation, *binary cross-entropy* is set as the loss function and Adam is set as optimiser. After 100 epoch of training and testing, the model achieved maximum testing ,accuracy of 96.74%, , and AUC of 0.997 . The average

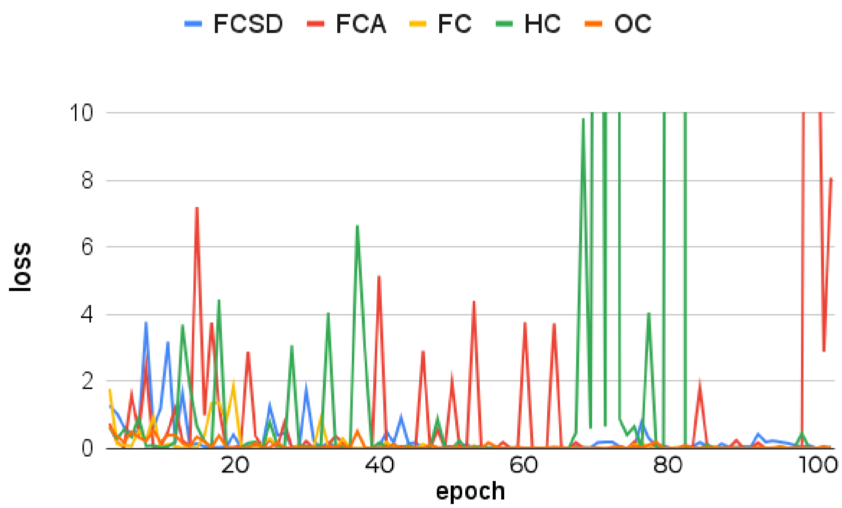

Figure 4.14: Testing Binary Cross Entropy Loss

training loss is 0.1660738825 and average testing loss is 0.08770136029 (fig: 4.14 4.15). It has a average training accuracy of 0.9342578125 whereas the average testing accuracy is 0.9705989558 for 100 epochs (fig: 4.10 4.11). The value show that the model suffer from over-fitting from problem since the test accuracy is larger than the training ones.

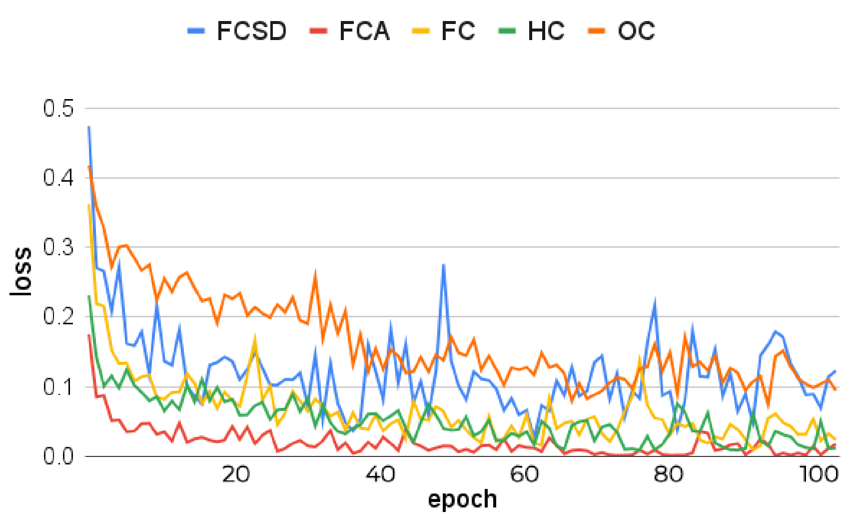

Figure 4.15: Training Binary Cross Entropy Loss

Hidden Layer Concatenation(HC) The output of dense layer neurons was concatenated in hidden layer concatenation architectures. That is, after extracting features from each modality, these features were passed to a distinct dense layer. However, unlike OC architectures, HC architectures concatenate intermediate dense layers of each sub network, with each layer containing more than one neuron. These dense layers have been concatenated to form a single layer. Then passed into dense layers with 256,256, 64 units. Each of the dense layer were followed by batch normalization and dropout layers with a dropout value of 0.25.. The output layer had *sigmoid* activation, *binary cross-entropy* is set as the loss function and Adam is set as optimiser. After training, it achieved a maximum validation accuracy of 97.31%, F1-Score of 96.99% (fig: 4.124.13), and AUC of 99%. The average training accuracy is 98.35% and testing accuracy is 90% for 100 epoch. However this models suffers from oscillation and has the lowest average testing recall of 89.59%.

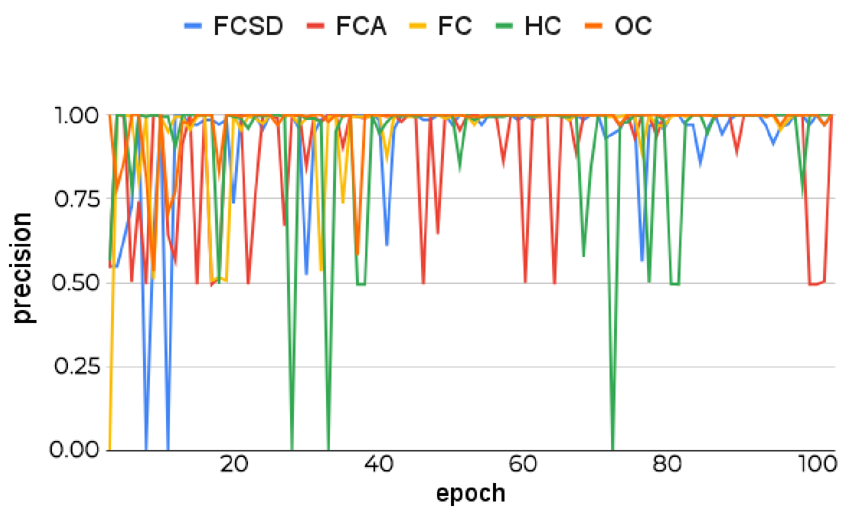

Figure 4.16: Testing Precision of for Different Multimodal Concatenation Technique

Feature Concatenation (FC)

When concatenating features, the features are first extracted from the corresponding input modalities. Except for numerical inputs, the feature extraction layers are distinct for each input modality. Two-dimensional convolutions with varying convolution sizes are used in this experiment to extract image features. For comparison purposes, the feature layers are identical for each of the modalities. The LSTM is used to extract audio features, while the Fitbit's numeric features are passed into a fully connected layer. These layers were then concatenated to create a single layer with 1024 neurons. Prior to the final output layer, a dense layer with 256,256,64 neurons with ReLU activation is used. The output layer is identical to the previous one; the activation function is a sigmoid. From the epochwise training, maximum testing accuracy of 98.7% and F1-Score of 98.93% is achieved. The average training

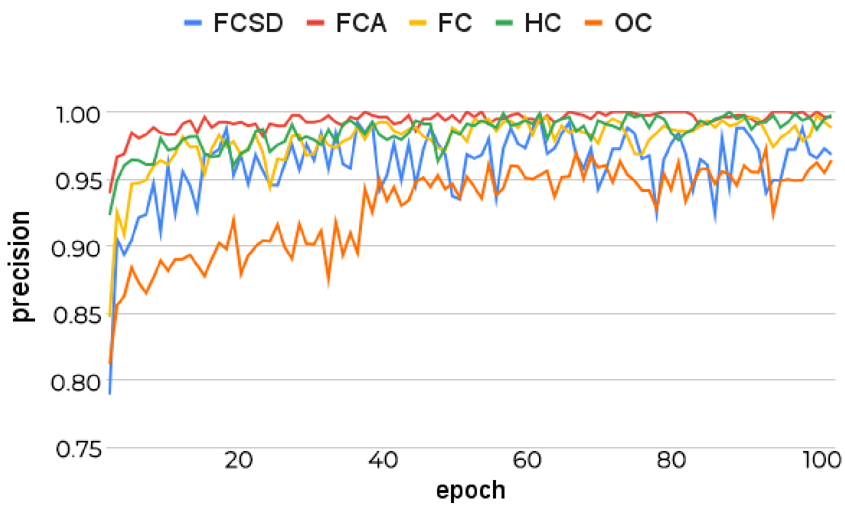

Figure 4.17: Training Precision of for Different Multimodal Concatenation Technique

and testing loss for this methods are 0.12 and 0.27 (fig: 4.15).

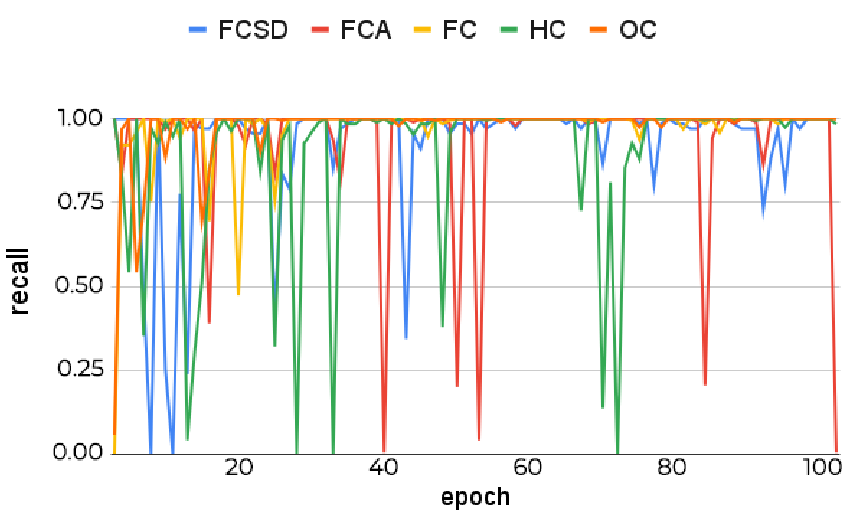

Figure 4.18: Testing Recall of for Different Multimodal Concatenation Technique

Feature Concatenation with Attention(FCA) Similar Multimodal networks as (HC) feature extraction with only one extra attention layers is also applied to see its effect on the Multimodal systems. With 512,256,64 three shared hidden dense layer, dropout of 0.25% and batch normalization,4.16 this model achieved an testing accuracy 98.7% with F1-Score of 98.63%. The average testing accuracy for this model were only 88%, lowest in all of experimented multimodal architectures.

Feature Concatenation with Selective Dropout(FCSD) Same neural network structure as FC has been utilized for model with selective dropout. This network

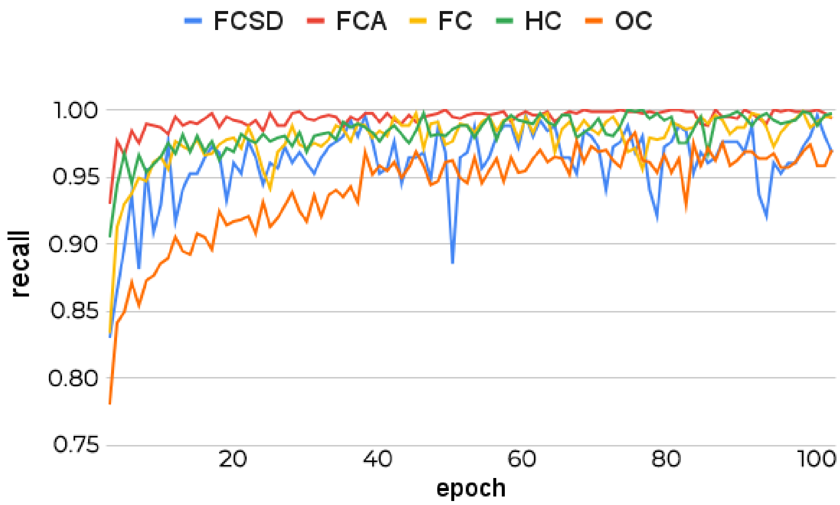

Figure 4.19: Training Recall of for Different Multimodal Concatenation Technique

performs better than all the other models. The feature were concatenated from respected model first. Then the proposed selective dropout is applied. Which will reduce the dependencies on a specific modality. Though the dropout layer can randomly drops any connection from previous layer, limiting an specific modality in batch train results with better training. With four shared hidden dense layer4.18, this model achieved an testing accuracy 99.7% with F1-Score of 99.56%.

4.7 Discussion

From each of the models stated above most them has a very high accuracy, F1-Score for bot testing and training aspects. From a single value they are not very far apart. But in terms of performances output concatenation, feature concatenation with attention models suffers severely from fluctuating results. That can be due to the fact that the batch size for all these run is quite small as 16 and random dropout or without drop out layers make this models to be trained with out one specific modalities. Thus whenever that type of modality data is significant for the classification , the model testing result can drop nearly to zero.Which is a cause of such accuracy,recall,precision and F1 Score fluctuations. Another fact is that the model is relatively large compared to 1900 data points taken for this experiments.With a larger dataset this problem can be removed, although data from all these modalities are not wide available. Selective drop out layer in the results of 4.11 reduces some of these problems, it also suffer from the same type oscillation initially. This dataset also trained using single classifiers such CNN for eye scan path and facial data and LSTM

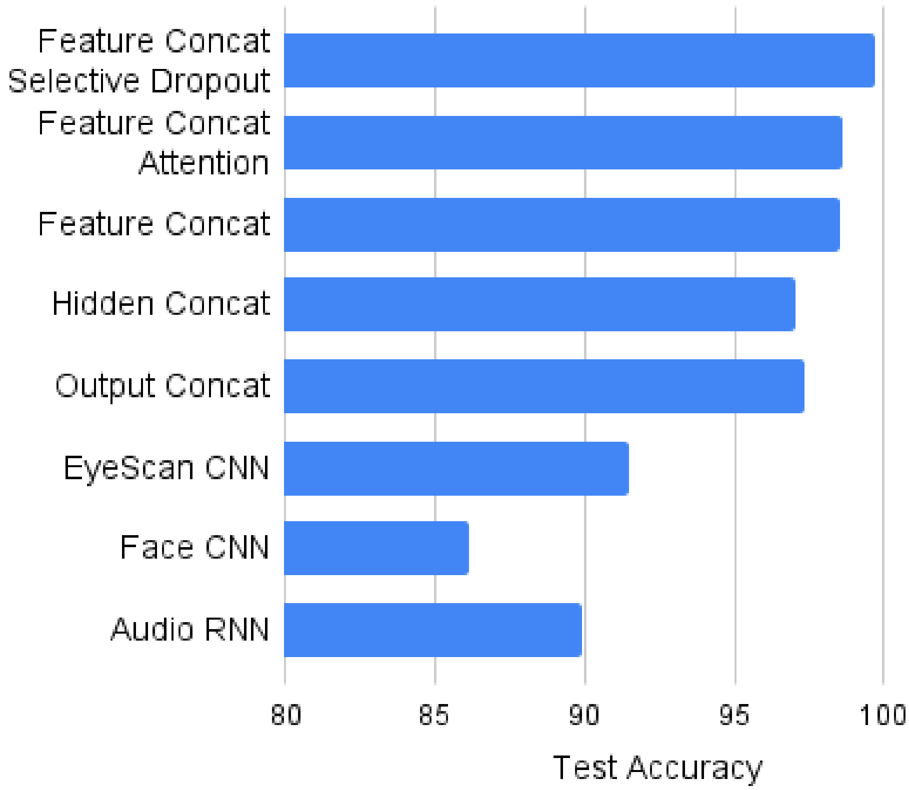

Figure 4.20: Comparison of performance metrics among Multimodal classifiers

network for the audio data. Figure 4.20 depicts a bar chart comparing the testing accuracy for multimodal classifiers with single label classifiers.Though it should be mentioned that all of these classifiers can be run extensively for fine tuning.

CHAPTER V

Conclusion and Future Work

5.1 Conclusion

The early and accurate detection of ASD enables early intervention and medical treatment, significantly lowering the risk. ASD is a broad term that encompasses a variety of psychological deficits that vary by individual. As a result, detecting autism has become more difficult by taking into account all possible physical and psychological problems. We compiled and analyzed a large number of ASD symptoms in this article and then converted them into a scenario-based questionnaire. A survey was conducted using a questionnaire to collect data. We collected physiological and video data from ASD patients and recorded their videos in order to extract additional modalities such as eye tracking, facial imaging, audio features, and behavioral practice. Fitbit, a smart watch, is used to track medical data such as heart rate and calories burned and can also be used to manage ASD pateints. By extracting these into these in formations and implementing multimodal models, we were able to obtain a satisfactory result that demonstrates how this type of data concatenation can result in improved ASD classification.

5.2 Limitations

Diverse types of ASD patients are in short supply for data collection. Though the initial questionnaire-based dataset contains five distinct types of ASD participants, the final dataset contains only 71. Due to the fact that ML models require a large amount of data, more precisely, the small dataset size is a significant limitation of this study. Again, when implementing a multimodal strategy, all modalities are initially required for each data instance. Due to the fact that the dataset may contain fewer modalities based on participants, this is another limitation of the dataset. Multimodal classification suffers from the same issue as the previously mentioned questionnare, namely a small dataset size. Additional research is required to ascertain the robustness of multimodal ASD classification.

5.3 Future Work

The primary future goal of this study is to expand the data sets and conduct additional comparisons with existing research. Because there are numerous types of neural networks, fine tuning can be done indefinitely in the future to determine the optimal network combinations. Due to the fact that the proposed method is based on multimodal data, it can be applied to other domains that require multiple input factors for proper recognition.

Publications

[Pub1] S. Ahmed, M. F. Hossain, S. B. Nur, M. S. Kaiser, and M. Mahmud, "Toward machine learning-based psychological assessment of autism spectrum disorders in school and community," *Proceedings of Trends in Electronics and Health Informatics: TEHI 2021*, p. 139.

References

[1] "Data & statistics on autism spectrum disorder," Sep 2020.

[2] A. P. Association *et al.*, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[3] C. Lord, T. S. Brugha, T. Charman, J. Cusack, G. Dumas, T. Frazier, E. J. Jones, R. M. Jones, A. Pickles, M. W. State, *et al.*, "Autism spectrum disorder," *Nature Reviews Disease Primers*, vol. 6, no. 1, pp. 1–23, 2020.

[4] T. Vos, S. S. Lim, C. Abbafati, K. M. Abbas, M. Abbasi, M. Abbasifard, M. Abbasi-Kangevari, H. Abbastabar, F. Abd-Allah, A. Abdelalim, *et al.*, "Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019," *The Lancet*, vol. 396, no. 10258, pp. 1204–1222, 2020.

[5] H. Faras, N. Al Ateeqi, and L. Tidmarsh, "Autism spectrum disorders," *Annals of Saudi medicine*, vol. 30, no. 4.

[6] G. Dawson, S. Rogers, J. Munson, M. Smith, J. Winter, J. Greenson, A. Donaldson, and J. Varley, "Randomized, controlled trial of an intervention for toddlers with autism: the early start denver model," *Pediatrics*, vol. 125, no. 1, pp. e17–e23, 2010.

[7] Z. Cidav, J. Munson, A. Estes, G. Dawson, S. Rogers, and D. Mandell, "Cost offset associated with early start denver model for children with autism," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 56, no. 9, pp. 777–783, 2017.

[8] L. J. Berlin, J. Brooks-Gunn, C. McCarton, and M. C. McCormick, "The effectiveness of early intervention: examining risk factors and pathways to enhanced development," *Preventive medicine*, vol. 27, no. 2, pp. 238–245, 1998.

[9] W. H. Organization *et al.*, "World health statistics: a snapshot of global health," in *World health statistics: A snapshot of global health*, 2012.

[10] C. on Children With Disabilities, S. on Developmental Behavioral Pediatrics, B. F. S. Committee, and M. H. I. for Children With Special Needs Project Advisory Committee, "Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening," *Pediatrics*, vol. 118, no. 1, pp. 405–420, 2006.

[11] S. Hamilton, "Screening for developmental delay: reliable, easy-to-use tools: win-win solutions for children at risk and busy practitioners," *Journal of Family Practice*, vol. 55, no. 5, pp. 415–423, 2006.

[12] M. L. Barton, T. Dumont-Mathieu, and D. Fein, "Screening young children for autism spectrum disorders in primary practice," *Journal of autism and developmental disorders*, vol. 42, no. 6, pp. 1165–1174, 2012.

[13] S. B. Mukherjee, S. Aneja, V. Krishnamurthy, and R. Srinivasan, "Incorporating developmental screening and surveillance of young children in office practice," *Indian pediatrics*, vol. 51, no. 8, pp. 627–635, 2014.

[14] C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter, "The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[15] M. Hosseinzadeh, J. Koohpayehzadeh, A. O. Bali, F. A. Rad, A. Souri, A. Mazaherinezhad, A. Rezapour, and M. Bohlouli, "A review on diagnostic autism spectrum disorder approaches based on the internet of things and machine learning," *The Journal of Supercomputing*, vol. 77, no. 3, pp. 2590–2608, 2021.

[16] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, "Applications of supervised machine learning in autism spectrum disorder research: a review," *Review Journal of Autism and Developmental Disorders*, vol. 6, no. 2, pp. 128–146, 2019.

[17] D. L. Robins, K. Casagrande, M. Barton, C.-M. A. Chen, T. Dumont-Mathieu, and D. Fein, "Validation of the modified checklist for autism in toddlers, revised with follow-up (m-chat-r/f)," *Pediatrics*, vol. 133, no. 1, pp. 37–45, 2014.

[18] S. K. Berument, M. Rutter, C. Lord, A. Pickles, and A. Bailey, "Autism screening questionnaire: diagnostic validity," *The British Journal of Psychiatry*, vol. 175, no. 5, pp. 444–451, 1999.

[19] C. Lord, M. Rutter, and A. Le Couteur, "Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders," *Journal of autism and developmental disorders*, vol. 24, no. 5, pp. 659–685, 1994.

[20] E. Schopler, R. J. Reichler, R. F. DeVellis, and K. Daly, "Toward objective classification of childhood autism: Childhood autism rating scale (cars).," *Journal of autism and developmental disorders*, 1980.

[21] Z. K. Zheng, J. E. Staubitz, A. S. Weitlauf, J. Staubitz, M. Pollack, L. Shibley, M. Hopton, W. Martin, A. Swanson, P. Juárez, *et al.*, "A predictive multimodal framework to alert caregivers of problem behaviors for children with asd (premac)," *Sensors*, vol. 21, no. 2, p. 370, 2021.

[22] M. Taj-Eldin, C. Ryan, B. O'Flynn, and P. Galvin, "A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers," *Sensors*, vol. 18, no. 12, p. 4271, 2018.

[23] L. Zwaigenbaum and M. Penner, "Autism spectrum disorder: advances in diagnosis and evaluation," *Bmj*, vol. 361, 2018.

[24] M. M. Seltzer, M. W. Krauss, P. T. Shattuck, G. Orsmond, A. Swe, and C. Lord, "The symptoms of autism spectrum disorders in adolescence and adulthood," *Journal of autism and developmental disorders*, vol. 33, no. 6, pp. 565–581, 2003.

[25] G. Dawson, A. N. Meltzoff, J. Osterling, and J. Rinaldi, "Neuropsychological correlates of early symptoms of autism," *Child development*, vol. 69, no. 5, pp. 1276–1285, 1998.

[26] M. A. Taylor, K. A. Schreck, and J. A. Mulick, "Sleep disruption as a correlate to cognitive and adaptive behavior problems in autism spectrum disorders," *Research in developmental disabilities*, vol. 33, no. 5, pp. 1408–1417, 2012.

[27] M. Solomon, M. Miller, S. L. Taylor, S. P. Hinshaw, and C. S. Carter, "Autism symptoms and internalizing psychopathology in girls and boys with autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 42, no. 1, pp. 48–59, 2012.

[28] F. R. Volkmar, D. V. Cicchetti, E. Dykens, S. S. Sparrow, J. F. Leckman, and D. J. Cohen, "An evaluation of the autism behavior checklist," *Journal of Autism and Developmental Disorders*, vol. 18, no. 1, pp. 81–97, 1988.

[29] J. H. Baskin, M. Sperber, and B. H. Price, "Asperger syndrome revisited," *Rev Neurol Dis*, vol. 3, no. 1, pp. 1–7, 2006.

[30] B. Mirkovic and P. Gérardin, "Asperger's syndrome: What to consider?," *L'encephale*, vol. 45, no. 2, pp. 169–174, 2019.

[31] R. Eisenmajer, M. Prior, S. Leekam, L. Wing, J. Gould, M. Welham, and B. Ong, "Comparison of clinical symptoms in autism and asperger's disorder," *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 35, no. 11, pp. 1523–1531, 1996.

[32] K. E. Macintosh and C. Dissanayake, "Annotation: the similarities and differences between autistic disorder and asperger's disorder: a review of the empirical evidence," *Journal of Child Psychology and Psychiatry*, vol. 45, no. 3, pp. 421–434, 2004.

[33] K. Karabekiroglu, "Pervasive developmental disorder-not otherwise specified: Specifying and differentiating," *Autism Spectrum Disorders: The Role of Genetics in Diagnosis and Treatment*, p. 17, 2011.

[34] A. V. Snow and L. Lecavalier, "Comparing autism, pdd-nos, and other developmental disabilities on parent-reported behavior problems: little evidence for asd subtype validity," *Journal of Autism and Developmental Disorders*, vol. 41, no. 3, pp. 302–310, 2011.

[35] V. K. Jensen, J. A. Larrieu, and K. K. Mack, "Differential diagnosis between attention-deficit/hyperactivity disorder and pervasive developmental disorder—not otherwise specified," *Clinical pediatrics*, vol. 36, no. 10, pp. 555–561, 1997.

[36] J. K. Buitelaar, R. Van der Gaag, A. Klin, and F. Volkmar, "Exploring the boundaries of pervasive developmental disorder not otherwise specified: Analyses of data from the dsm-iv autistic disorder field trial," *Journal of autism and developmental disorders*, vol. 29, no. 1, pp. 33–43, 1999.

[37] C. Mehra, A. Sil, T. Hedderly, M. Kyriakopoulos, M. Lim, J. Turnbull, F. Happe, G. Baird, and M. Absoud, "Childhood disintegrative disorder and autism spectrum disorder: a systematic review," *Developmental Medicine & Child Neurology*, vol. 61, no. 5, pp. 523–534, 2019.

[38] M. Elia, R. Ferri, S. A. Musumeci, S. Del Gracco, M. Bottitta, C. Scuderi, G. Miano, S. Panerai, T. Bertrand, and J.-C. Grubar, "Sleep in subjects with autistic disorder: a neurophysiological and psychological study," *Brain and Development*, vol. 22, no. 2, pp. 88–92, 2000.

[39] R. Militerni, C. Bravaccio, C. Falco, C. Fico, and M. T. Palermo, "Repetitive behaviors in autistic disorder," *European Child & Adolescent Psychiatry*, vol. 11, no. 5, pp. 210–218, 2002.

[40] B. Hagberg, F. Goutières, F. Hanefeld, A. Rett, and J. Wilson, "Rett syndrome: criteria for inclusion and exclusion.," *Brain & development*, vol. 7, no. 3, pp. 372–373, 1985.

[41] S. M. Kyle, N. Vashi, and M. J. Justice, "Rett syndrome: a neurological disorder with metabolic components," *Open biology*, vol. 8, no. 2, p. 170216, 2018.

[42] G. L. WENK, "Rett syndrome: neurobiological changes underlying specific symptoms," *Progress in neurobiology*, vol. 51, no. 4, pp. 383–391, 1997.

[43] J. Wulffaert, I. A. Van Berckelaer-Onnes, and E. M. Scholte, "Autistic disorder symptoms in rett syndrome," *Autism*, vol. 13, no. 6, pp. 567–581, 2009.

[44] Y. Nomura and M. Segawa, "Natural history of rett syndrome," *Journal of child neurology*, vol. 20, no. 9, pp. 764–768, 2005.

[45] F. F. Thabtah, "Autistic spectrum disorder screening data for children data set," *UCI machine learning repository*, 2017.

[46] S. Eraslan, Y. Yesilada, V. Yaneva, and S. Harper, "Autism Detection Based on Eye Movement Sequences on the Web: A Scanpath Trend Analysis Approach," Feb. 2020.

[47] R. Carette, M. Elbattah, G. Dequen, J.-L. Guérin, and F. Cilia, "Visualization of eye-tracking patterns in autism spectrum disorder: method and dataset," in *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*, pp. 248–253, IEEE, 2018.

[48] F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and dsm-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and health Informatics 2017*, pp. 1–6, 2017.

[49] H. Duan, G. Zhai, X. Min, Z. Che, Y. Fang, X. Yang, J. Gutiérrez, and P. Le Callet, "A Dataset of Eye Movements for the Children with Autism Spectrum Disorder," Apr. 2019.

[50] A. Zunino, P. Morerio, A. Cavallo, C. Ansuini, J. Podda, F. Battaglia, E. Veneselli, C. Becchio, and V. Murino, "Video gesture analysis for autism spectrum disorder detection," in *International Conference on Pattern Recognition (ICPR)*, 2018.

[51] N. Goel, B. Grover, D. Gupta, A. Khanna, M. Sharma, *et al.*, "Modified grasshopper optimization algorithm for detection of autism spectrum disorder," *Physical Communication*, vol. 41, p. 101115, 2020.

[52] F. Thabtah and D. Peebles, "A new machine learning model based on induction of rules for autism detection," *Health informatics journal*, vol. 26, no. 1, pp. 264–286, 2020.

[53] C. Küpper, S. Stroth, N. Wolff, F. Hauck, N. Kliewer, T. Schad-Hansjosten, I. Kamp-Becker, L. Poustka, V. Roessner, K. Schultebraucks, *et al.*, "Identifying predictive features of autism spectrum disorders in a clinical sample of adolescents and adults using machine learning," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[54] S. Levy, M. Duda, N. Haber, and D. P. Wall, "Sparsifying machine learning models identify stable subsets of predictive features for behavioral detection of autism," *Molecular autism*, vol. 8, no. 1, pp. 1–17, 2017.

[55] T. G. Pratama, R. Hartanto, and N. A. Setiawan, "Machine learning algorithm for improving performance on 3 aq-screening classification," *Communications in Science and Technology*, vol. 4, no. 2, pp. 44–49, 2019.

[56] R. Carette, F. Cilia, G. Dequen, J. Bosche, J.-L. Guerin, and L. Vandromme, "Automatic autism spectrum disorder detection thanks to eye-tracking and neural network-based approach," in *International conference on IoT technologies for healthcare*, pp. 75–81, Springer, 2017.

[57] R. Carette, M. Elbattah, F. Cilia, G. Dequen, J.-L. Guérin, and J. Bosche, "Learning to predict autism spectrum disorder based on the visual patterns of eye-tracking scanpaths.," in *HEALTHINF*, pp. 103–112, 2019.

[58] M. Elbattah, R. Carette, G. Dequen, J.-L. Guérin, and F. Cilia, "Learning clusters in autism spectrum disorder: Image-based clustering of eye-tracking scanpaths with deep autoencoder," in *2019 41st Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 1417–1420, IEEE, 2019.

[59] Y. Tao and M.-L. Shyu, "Sp-asdnet: Cnn-lstm based asd classification model using observer scanpaths," in *2019 IEEE International conference on multimedia & expo workshops (ICMEW)*, pp. 641–646, IEEE, 2019.

[60] M. Chita-Tegmark, "Social attention in asd: A review and meta-analysis of eye-tracking studies," *Research in developmental disabilities*, vol. 48, pp. 79–93, 2016.

[61] S. L. Oh, V. Jahmunah, N. Arunkumar, E. W. Abdulhay, R. Gururajan, N. Adib, E. J. Ciaccio, K. H. Cheong, and U. R. Acharya, "A novel automated autism spectrum disorder detection system," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2399–2413, 2021.

[62] M. Baygin, S. Dogan, T. Tuncer, P. D. Barua, O. Faust, N. Arunkumar, E. W. Abdulhay, E. E. Palmer, and U. R. Acharya, "Automated asd detection using hybrid deep lightweight features extracted from eeg signals," *Computers in Biology and Medicine*, vol. 134, p. 104548, 2021.

[63] J. J. Esqueda-Elizondo, R. Juárez-Ramírez, O. R. López-Bonilla, E. E. García-Guerrero, G. M. Galindo-Aldana, L. Jiménez-Beristáin, A. Serrano-Trujillo, E. Tlelo-Cuautle, and E. Inzunza-González, "Attention measurement of an autism spectrum disorder user using eeg signals: A case study," *Mathematical and Computational Applications*, vol. 27, no. 2, p. 21, 2022.

[64] M. N. A. Tawhid, S. Siuly, H. Wang, F. Whittaker, K. Wang, and Y. Zhang, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from eeg," *Plos one*, vol. 16, no. 6, p. e0253094, 2021.

[65] A. Lu and M. Perkowski, "Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application," *Brain Sciences*, vol. 11, no. 11, p. 1446, 2021.

[66] T. Eslami and F. Saeed, "Auto-asd-network: a technique based on deep learning and support vector machines for diagnosing autism spectrum disorder using fmri data," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pp. 646–651, 2019.

[67] 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*, vol. 17, pp. 16–23, 2018.

[68] U. Großekathöfer, N. V. Manyakov, V. Mihajlović, G. Pandina, A. Skalkin, S. Ness, A. Bangerter, and M. S. Goodwin, "Automated detection of stereotypical motor movements in autism spectrum disorder using recurrence quantification analysis," *Frontiers in neuroinformatics*, vol. 11, p. 9, 2017.

[69] L. Sadouk, T. Gadi, and E. H. Essoufi, "A novel deep learning approach for recognizing stereotypical motor movements within and across subjects on the autism spectrum disorder," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[70] E. Leblanc, P. Washington, M. Varma, K. Dunlap, Y. Penev, A. Kline, and D. P. Wall, "Feature replacement methods enable reliable home video analysis for machine learning detection of autism," *Scientific reports*, vol. 10, no. 1, pp. 1–11, 2020.

[71] J. Li, Y. Zhong, J. Han, G. Ouyang, X. Li, and H. Liu, "Classifying asd children with lstm based on raw videos," *Neurocomputing*, vol. 390, pp. 226–238, 2020.

[72] M. Alcaniz Raya, J. Marín-Morales, M. E. Minissi, G. Teruel Garcia, L. Abad, and I. A. Chicchi Giglioli, "Machine learning and virtual reality on body movements' behaviors to classify children with autism spectrum disorder," *Journal of clinical medicine*, vol. 9, no. 5, p. 1260, 2020.

[73] E. T. Sadek, N. A. AbdElSabour Seada, and S. Ghoniemy, "Computer vision techniques for autism symptoms detection and recognition: A survey.," *International Journal of Intelligent Computing and Information Sciences*, vol. 20, no. 2, pp. 89–111, 2020.

[74] F. J. Van Steensel, S. M. Bögels, and S. Perrin, "Anxiety disorders in children and adolescents with autistic spectrum disorders: a meta-analysis," *Clinical child and family psychology review*, vol. 14, no. 3, pp. 302–317, 2011.

[75] L. Jansen, C. C. Gispen-de Wied, V. M. Wiegant, H. G. Westenberg, B. E. Lahuis, and H. Van Engeland, “Autonomic and neuroendocrine responses to a psychosocial stressor in adults with autistic spectrum disorder,” *Journal of autism and developmental disorders*, vol. 36, no. 7, pp. 891–899, 2006.

[76] C. H. Vinkers, R. Penning, J. Hellhammer, J. C. Verster, J. H. Klaessens, B. Olivier, and C. J. Kalkman, “The effect of stress on core and peripheral body temperature in humans,” *Stress*, vol. 16, no. 5, pp. 520–530, 2013.

[77] M. Viqueira Villarejo and B. García Zapirain, “Méndez zorrilla, 2012,” *A.,“A Stress Sensor Based on Galvanic Skin Response (GSR) Controlled by ZigBee”, Sensors (Basel)*, vol. 12, no. 5, pp. 6075–6101, 2012.

[78] J.-J. Cabibihan, H. Javed, M. Aldosari, T. W. Frazier, and H. Elbashir, “Sensing technologies for autism spectrum disorder screening and intervention,” *Sensors*, vol. 17, no. 1, p. 46, 2016.

[79] T. Y. Tang, “Helping neuro-typical individuals to" read" the emotion of children with autism spectrum disorder: an internet-of-things approach,” in *Proceedings of the The 15th International Conference on Interaction Design and Children*, pp. 666–671, 2016.

[80] T. Notenboom, “Using technology to recognise emotions in autistic people,” B.S. thesis, University of Twente, 2017.

[81] C. M. Northrup, J. Lantz, and T. Hamlin, “Wearable stress sensors for children with autism spectrum disorder with in situ alerts to caregivers via a mobile phone,” *Iproceedings*, vol. 2, no. 1, p. e6119, 2016.

[82] T. Autism, “Social story creator educators,” Jun 2015.

[83] B. Mozolic-Staunton, M. Donelly, J. Yoxall, and J. Barbaro, “Early detection for better outcomes: Universal developmental surveillance for autism across health and early childhood education settings,” *Research in Autism Spectrum Disorders*, vol. 71, p. 101496, 2020.

[84] “Language therapy for children with autism (mita) - apps on google play.”