Empowering Autism Diagnosis: An Artificial Intelligence-Assisted Assessment in the Indian Context

A Dissertation
Presented to
The Academic Faculty

By

Ashwini B
Roll No. PhD18010

In partial fulfilment
of the requirements for the
Degree of Doctor of Philosophy

Under the Supervision of

Dr. Jainendra Shukla

INDRAPRASTHA INSTITUTE *of* INFORMATION TECHNOLOGY **DELHI**

Department of Computer Science and Engineering
Indraprastha Institute of Information Technology Delhi

March 2024

Empowering Autism Diagnosis: An Artificial Intelligence-Assisted Assessment in the Indian Context

A Dissertation
Presented to
The Academic Faculty

By

Ashwini B
Roll No. PhD18010

In partial fulfilment
of the requirements for the
Degree of Doctor of Philosophy

Under the Supervision of

Dr. Jainendra Shukla

Department of Computer Science and Engineering
Indraprastha Institute of Information Technology Delhi

March 2024

CERTIFICATE

This is to certify that the thesis titled, “Empowering Autism Diagnosis: An Artificial Intelligence-Assisted Assessment in the Indian Context”, being submitted by Ashwini B to the Department of Computer Science and Engineering, Indraprastha Institute of Information Technology Delhi, in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science is an authentic proof of work carried out by her under my supervision. In my opinion, the thesis has reached the standards of fulfilling the requirements of the regulations relating to the degree.

The work in this thesis has not been submitted in any form for another degree or diploma at any university or another institute.

July 1, 2024

Jainendra Shukla
Assistant Professor, Department of CSE
IIIT-Delhi

Date of Signature

DECLARATION

I hereby declare that the work presented in this thesis titled "Empowering Autism Diagnosis: An Artificial Intelligence-Assisted Assessment in the Indian Context", submitted to the Department of Computer Science, Indraprastha Institute of Information Technology Delhi, in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Computer Science and Engineering, is an authentic proof of my own work carried out from June 2018 to the present date under the supervision of Dr. Jainendra Shukla.

The work in this thesis has not been submitted in any form for another degree or diploma at any university or another institute. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given.

	01.07.2024
Ashwini B	Date of Signature

karmaṇy-evādhikāras te mā phaleṣhu kadāchana

mā karma-phala-hetur bhūr mā te saṅgo 'stvakarmaṇi

- Bhagavad Gita (2:47)

DEDICATION

This thesis is dedicated to my little one, Vibha Viswanathan.

ACKNOWLEDGEMENTS

First and foremost, I am truly indebted to Dr. Jainendra Shukla for his unwavering support and mentorship throughout this journey. Dr. Shukla's role extends far beyond that of a supervisor; he is an amazing teacher, mentor, and inspiring individual. Starting from the preliminaries of research to writing manuscripts and proposals, he has played a pivotal role in my development as a researcher. Beyond academia, he helped me develop as a better professional by instilling the value of teamwork and fostering respectful professional relationships. His receptiveness to ideas, willingness to engage in meaningful discussions, and open-minded approach have immensely enriched my academic journey. I am deeply thankful for his encouragement and constant motivation, particularly during moments of doubt or uncertainty.

I am grateful to my monitoring committee members, Dr. Vikram Goyal and Dr. Rajiv Ratn Shah, for their rigorous evaluation of my research progress and invaluable feedback, which has significantly enhanced the quality of my work. They have provided constructive criticism where necessary and equally motivated me to strive for better. I am grateful to Dr. Venkata Ratnadeep Suri for his invaluable support and insightful suggestions in our collaborative projects. I extend my gratitude to the anonymous members of the thesis evaluation committee - Prof. Uttama Lahiri, Prof. Albert Ali Salah and Dr. Puneet Gupta for their positive comments and their valuable feedback on the thesis, which have been instrumental in enhancing this thesis. Coincidentally, Dr. Puneet Gupta also served as my comprehensive exam evaluator, providing valuable feedback on my thesis work at an early stage, which contributed to the development of the research that culminated in this thesis. Additionally, his generosity in agreeing to be my referee and supporting my applications for postdoctoral positions has been invaluable. I sincerely thank him for his support.

Now, I express my gratitude to those without whose support I wouldn't have even embarked on this research journey, fulfilling a dream I once thought unattainable - my family. I am deeply grateful to my husband, Dr. Viswanathan Puthan Veedu, my companion, my biggest critic, and my greatest support. He pushed me through the initial phases of my PhD, motivated me, and stood by me like a pillar, helping me sail through the tough times and continues to do so. As a young mother struggling to juggle the roles of motherhood and researcher, yearning to be with my baby, he was the one who ensured that our daughter stayed happy and safe at home, giving me the peace of mind needed to focus on my research. Being a perfectionist and one who has experienced this journey, he has always made sure I put my best effort into my research. I am blessed with my little angel Vibha, who is the most understanding, adjusting, and encouraging child. Without any complaints, she has accepted her mom, who at times has little time to play with her. She even keeps tabs on whether I have completed my work hours when I work from home. Her understanding and support have been invaluable.

I thank my grandparents, Late Shri. Raghavan Nair and Smt. Gomathi

Thankachi for the love and blessings they have showered on me. I'am thankful to my parents, Mr. Velutheertha and Mrs. Bhamini Thankachi, for being the wonderful parents they are, for working hard to get me the privilege of a good education and life, never putting the pressure of their expectations in my academics and being their with me whenever I needed their support. I would also like to extend my sincerest gratitude to my mother-in-law, Smt. Indira Madhavan, for understanding and accepting my goals, never doubting me, and extending her unreserved support on this journey. My sister, Roshni, has been an incredible support. Her unconventional approach, filled with wit and taunts, has lightened stressful times and made them more bearable. Thanks to her and Avinash for their encouragement. Additionally, I would like to extend my gratitude to my friends Kirithika and Prof. Sivananthan for their constant encouragement and support throughout this journey. Further, I am also grateful to my extended family for their backing. Their blessings and support have been instrumental in helping me navigate this long journey with perseverance.

I am immensely grateful to our collaborator, Dr. Sheffali Gulati from AIIMS, for her invaluable support in executing the research studies and providing valuable insights throughout the research. I also express my gratitude to her for her willingness to be my referee despite her extremely demanding schedule. This thesis owes much to the invaluable contributions of other collaborators, such as Dr. Krishnaveni Achary and Ms. Tanushree Roy from Deepalaya NGO. Dr. Krishnaveni introduced me to the intricate world of autism spectrum disorder, providing invaluable insights that enriched the project with a humane perspective. Ms. Tanushree and her team at Deepalaya extended wholehearted support and welcomed our team with warmth, making the data collection process an enriching experience. Our interactions with the children at Deepalaya were always eagerly anticipated, thanks to their hospitality, love and cooperation.

Heartfelt thanks to Atmadeep Ghoshal for his wholehearted support throughout the final years of my PhD journey. His collaboration not only introduced a researcher with a computing background, like me, to the nuances of sociology but also helped me navigate through diverse research methodologies of qualitative study. This journey was both entertaining and cherishing. I extend my gratitude to my lab mates Saumya Yadav and Adarsh Pandey for making my time spent at the lab enjoyable. I cherish the moments we spent celebrating successes, both big and small and supporting each other through setbacks, our casual conversations during breaks and sharing their lunch. Thanks to my friends, Ramneek Kaur and Jithin, for wholeheartedly supporting me through tough times. Your presence in my PhD journey has been a blessing. I would also like to thank my friends Binu Ann Joseph, Yasmeena Akhter, Divyanshu Kumar Singh, Abhijeet Mishra, Asra Sakeen Wani, and Jasmeet Kaur for their help and support.

I am grateful to all the student collaborators who have contributed to this thesis: Vrinda Narayan, Saksham Bhupal, Deeptanshu, Sanmay Sood, Pratyush Jain, and Pruthivi Raj Behera. Special recognition goes to Arka

Sarkar, a collaborator-turned-friend, for his sincere support in translating our ideas into tangible outcomes.

I would like to thank Indraprastha Institute of Information Technology Delhi for giving me an opportunity to be part of the IIITD family as a researcher and for supporting this research with the necessary infrastructure and facilities. Special mention to Priti Patel, Priti Patwal (Assistant Manager (CSE)), Raju Biswas (Junior Manager (Academics)), Bhawani Shah (Assistant Manager, System Admin & Networking), Kapil Mishra Executive Assistant (CDNM) and Rajendra Singh (Sr. Manager, Library & Information Service) for their timely support.

I would like to extend my thanks to everyone who has directly or indirectly supported me throughout this journey. I am also grateful to those who have challenged, doubted, and criticized me, as their scepticism fueled my determination to prove myself.

Finally, I would like to express my gratitude to the Almighty for bestowing upon me kindness and blessings beyond measure.

ABSTRACT

The diagnostic process for Autism Spectrum Disorder (ASD) is complex, requiring extensive expertise to integrate information from diverse sources such as parental reports and clinical observations. However, limited access to diagnostic facilities, especially in suburban areas, poses a significant challenge, often necessitating multiple expert consultations before an accurate diagnosis can be made. Given the shortened neuroplasticity period experienced by children with ASD (CwA), early intervention is crucial for enhancing their social and communication skills. These challenges have prompted the research community to explore technology-based solutions to meet the needs of individuals with autism, caregivers, and professionals, with a focus on achieving reliable and objective diagnoses.

Artificial Intelligence (AI)-based behaviour analysis has emerged as a promising approach for identifying characteristic behaviours associated with ASD diagnosis. With this in mind, the thesis aims to develop an AI-assisted assessment system capable of identifying these traits in children with autism. Moreover, recognizing the preference of children with ASD for technology-assisted tools in therapeutic interventions, robot assistants were employed to interact with children during diagnostic procedures, aiming to create engaging and comfortable interactions for the children undergoing assessment. The overarching goal is to develop an assessment system that is context-aware,

robust in adhering to diagnostic standards, and transparent in its decision-making processes, thereby addressing current gaps and challenges in existing intervention approaches.

To accomplish these objectives, we initially assessed the feasibility of deploying the system in resource-constrained environments, particularly in the Global South, with a specific focus on India. Our investigation began by examining the response behaviour of children of Indian ethnicity, both with and without autism spectrum disorder, towards robot-assisted interventions for diagnosis. Through a study targeting children aged 3 to 6 years, we observed that the children effectively followed the robot's instructions during directive tasks and successfully completed them. Our analysis revealed the acceptance and benefits of employing robotic assistants as facilitators in various domains, including education, cognitive therapies, and healthcare.

Subsequently, we conducted a study to explore the perceptions of special educators regarding the use of social robots in robot-assisted interventions for diagnosis. Employing a mixed-methods approach involving interviews, workshops, and a panel discussion with 25 educators in India, we uncovered both challenges and opportunities associated with integrating social robots into autism interventions. While special educators expressed concerns about their functional capacity and apprehensions regarding potential redundancy resulting from the substitution of human efforts by social robots, they also acknowledged the importance of technological innovation in reshaping and enhancing their roles in autism therapy. Despite initial scepticism, professionals identified various strategies for effectively incorporating social robots into intervention programs. We further elucidate the implications of these findings for the development of context-aware solutions and policy-level initiatives essential for resource-constrained settings.

Having established the feasibility of robot-mediated interventions for children with autism, our focus shifted to developing interpretable machine-learning frameworks capable of identifying speech and facial expression behaviours essential for ASD assessment and diagnosis. Our initial endeavour involved the creation of a multi-source transfer learning approach to capture facial emotional attributes in children of Indian ethnicity, thereby aiding diagnostic decisions. Through optimization techniques aimed at enhancing multivariate correlation among source tasks, we achieved notable improvements in facial emotion recognition accuracy compared to existing methodologies.

Subsequently, we turned our attention to identifying speech behavioural characteristics critical for ASD diagnosis, emphasizing language impairments as key markers. Through automated methods for speech behaviour extraction and comprehensive speech evaluations, we characterized linguistic traits in children with autism speaking Hindi. Our extensive analysis encompassed a diverse set of acoustic and linguistic speech attributes, including lexical, syntactic, semantic, and pragmatic elements. Notably, our investigation also addressed the influence of linguistic diversity on speech-based ASD assessment, examining speech behaviours in both English and Hindi-speaking children. The results of our study, involving data from 76 English-speaking children and 33 Hindi-speaking children, demonstrated the reliability of automatically extracted acoustic and linguistic features as predictors of ASD, achieving an impressive macro F1-score of approximately 91.30%.

Concluding our thesis, we validated the AI-assisted system with a robot mediator for the diagnostic assessment of children with ASD through a pilot study involving 21 participants. This involved 12 typically developing (TD) children and 9 CwA. The results affirmed the feasibility and efficacy of the robot-mediated AI system in diagnosing ASD in Indian children. In sum, our thesis provides an initial exploration into the potential of utilizing robots in autism care in India, supported by AI-assisted multimodal behavioural analysis, paving the way for further advancements in this field.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI Artificial Intelligence

ASD Autism Spectrum Disorder

CDC Centers for Disease Control and Prevention

CwA Children with Autism

ADOS Autism Diagnostic Observation Schedule

ICD International Classification of Diseases

ISAA Indian Scale for Assessment of Autism

INDT-ASD INCLEN Diagnostic tool for ASD

CARS Childhood Autism Rating Scale

INCLEN International Clinical Epidemiology Network

DALY Disability Adjusted Life Years

DSM Diagnostic and Statistical Manual of Mental Disorders

JA Joint Attention

RN Response to Name

ML Machine Learning

AIIMS All India Institute of Medical Sciences

ABC Autism Behaviour Checklist

CBCL Child Behaviour Checklist

MISIC Malin's Intelligence Scale for Indian Children

VSMS Vineland Social Maturity Scale

TD Typically Developing

HRI human-robot interaction

ADHD Attention-deficit/hyperactivity disorder

RET Robot Enhanced Therapy

RAT Robot-assisted Therapy

RAS Robot-assisted Systems

ABA Applied Behaviour Analysis

SAR Socially-assistive Robots

HCI Human-Computer Interaction

RCI Rehabilitation Council of India

FER facial emotion recognition

RAF-DB Real-world Affective Faces Database

JAFFE Japanese Female Facial Expression

CAFE Child Affective Facial Expression

CFED Child Facial Expression Dataset

MCW Maximal Correlation Weighting

MSDA multi-source domain adaptation

ACE Alternating Conditional Expectation

MSTL-MCA Multi-source Transfer Learning using Multivariate Correlation Analysis

DECISION Data frEe multi-sourCe unsupervISed domain adaptatiON

NMC Network Maximal Correlation

PRL Prerequisite Learning

NLP Natural Language Processing

SHAP SHapley Additive exPlanations

LOO-CV Leave one out cross-validation

SALT Systematic Analysis of Language Transcripts

CHAPTER 1
INTRODUCTION

This thesis focuses on the development of an Artificial Intelligence (AI)-assisted multimodal behaviour analysis for the assessment of ASD in children with a robot assistant as the intervention mediator. ASD falls within the broader spectrum of pervasive developmental disorders and is characterised by challenges in social interaction, communication skills, and the presence of repetitive and restricted behaviours [1]. According to the Centers for Disease Control and Prevention (CDC) report from 2018, the prevalence of autism is estimated at 1 in 54 individuals. In contrast, the International Clinical Epidemiology Network (INCLEN) study suggests a prevalence of 1 in 89 among Indian children aged 2 to 9 years [2]. A recent study on the burden of mental disorders in India reveals that ASD contributed $3 \cdot 2\%$ ($2 \cdot 7\%$–$3 \cdot 8\%$) to the total mental disorders Disability Adjusted Life Years (DALY) in India, indicating the need for addressing the demand for experts and infrastructure to support the autism community [3]. India's vast population and diverse factors influencing autism place it in a unique global position, with an estimated 2 million children estimated to be living with the condition.

Children with Autism (CwA) experience a shortened period of neuroplasticity, emphasising the necessity for early intervention to enhance their social and communication skills [4]. Early diagnosis plays a pivotal role in maximising the benefits of subsequent interventions. For individuals diagnosed with ASD, early interventions, ideally administered between the ages of one to five, leverage the heightened learning potential of a young child's brain [5]. Despite the recognised importance of early intervention, the absence of absolute biological markers for autism necessitates a reliance on behavioural observations by expert clinicians, complemented by assessments based on parental responses regarding the child's behavioural history [6]. The diagnostic process for ASD involves a structured series of steps. Upon expressing concerns about a child's development by caregivers or parents, the standard practice involves conducting ASD screenings at 18 and 24 months of age. If these screenings reveal potential ASD symptoms, a comprehensive diagnostic assessment ensues, guided by the criteria outlined in the International Classification of Diseases (ICD)-11 or the Diagnostic and Statistical Manual of Mental Disorders (DSM)-5. This multidisciplinary evaluation incorporates standardised methods to assess both core and co-morbid symptoms, encompassing a detailed review of caregiver concerns, behavioural descriptions, medical history, and the completion of relevant questionnaires. Despite parents often identifying developmental issues as early as 18 months, a formal diagnosis typically occurs two years after the initial expression of parental concern. Moreover, the lack of uniform administration of diagnostic tools globally adds to the complexity. While DSM-5 provides standard diagnostic guidelines for ASD, diverse adaptations exist. For instance, Autism Diagnostic Observation Schedule (ADOS) is considered the gold standard in the Global North, whereas, in India, prevalent diagnostic tools include the Indian Scale for Assessment of Autism (ISAA) [7], INCLEN Diagnostic tool for ASD (INDT-ASD) [8], and Childhood Autism Rating Scale (CARS) [9]. This lack of indigenous tools and variability in existing diagnostic tools, combined with the intricate nature of the diagnostic procedure, contributes to misjudgments. The expertise required for accurate ASD diagnosis demands extensive training and experience, often concentrated in metropolitan areas and beyond the reach of a significant portion of the affected population [10]. Owing to the limited medical infrastructure and unavailability of experts, individuals often have to approach multiple experts before receiving a correct diagnosis.

These challenges have inspired the research community to seek technology-based solutions to meet the needs of individuals with autism, their caregivers, and professionals, aiming for reliable and objective diagnoses. AI and Machine Learning (ML) based behaviour analysis has been extensively used to identify characteristic behaviours such as JA [11–14], eye contact [15–17], imitation [18–21], atypicalities in facial expression [22, 23], speech [24–27] in CwA. This research has widely focused on imparting education and intervention for CwA, but their application in diagnosis is less explored. One of the factors catering to this is the complexity of the procedure for the assessment of the individual, which involves providing efficient reasoning of the diagnosis and algorithms for perceiving the behavioural characteristics. Further, most of these studies have been conducted in the Global North, and the generalisability of these studies to the Indian context remains uncertain due to the lack of representation of Indian-specific data in the training datasets. Additionally, there is a growing recognition that a multimodal behaviour analysis is necessary for a more comprehensive characterisation of autism in children. To tackle these challenges, we delved into the development of an AI-assisted assessment system designed to identify characteristic traits in CwA, thereby aiding in their diagnosis.

Moreover, CwA often exhibit a preference for interacting with technological tools over human beings due to the predictability and simplicity of these interactions compared to complex human social interactions [28]. Furthermore, robot-assisted interventions have gained increasing acceptance as a supportive tool for therapy and education for CwA. This motivated our decision to utilise robot assistants to interact with children during diagnostic procedures, aiming to make the interactions more engaging and comfortable for the children under assessment. This approach also alleviates the burden on the expert, who is relieved from simultaneously administering diagnostic tasks and monitoring the child.

Addressing the above-mentioned factors, we aimed to develop an assessment system which is context-aware, robust to diagnostic standards, and explainable, addressing the current gaps and challenges in existing intervention approaches.

1.1 Related Works

In this section, we explore the opportunities created by the rapid advancements in AI, which hold immense promise for innovative and effective diagnostics and treatments for individuals with ASD [29]. Exploring the existing literature, we first examine various facets of ASD diagnosis and intervention methods, with a specific focus on behavioural observations, diagnostic criteria, and the field of socially assistive robots in therapy. By analysing current research findings, this section aims to shed light on the evolving landscape of ASD treatment, offering insights into novel approaches that leverage AI technologies to enhance the quality of care and support for individuals with ASD.

1.1.1 Autism Spectrum Disorder and Clinical Diagnosis

ASD is a neurodevelopmental condition characterised by impairments in social interaction and communication, along with the presence of restricted and repetitive patterns of behaviours, interests, or activities. The term "spectrum" reflects the diverse range of symptoms and severity levels observed in individuals with ASD. According to the World Health Organisation (WHO), it is estimated that approximately 1 in 100 children worldwide has autism [30], with an estimated 2 million children in the age of 2-9 years affected in India [31]. ASD is more prevalent in males than girls, though the reasons for this gender difference remain unclear [32]. The etiology of ASD remains unclear, with various factors implicated in increasing the likelihood of its occurrence, including environmental, biological, and genetic influences [30].

ASD can be detected as early as 18 months of age, with a diagnosis typically considered reliable by the age of 2 when conducted by an expert. Early intervention plays a crucial role in improving the communication and social interaction skills of individuals with ASD, leading to significant improvements in their quality of life. Diagnosis of ASD is, therefore, critical and hinges primarily on observing the child's behaviour and gathering developmental history from parents, given the absence of biological markers for diagnosis. In the academic community, ongoing research is focused on understanding the etiology, epidemiology, and effective interventions for ASD.

The American Psychiatric Association's DSM, Fifth Edition (DSM-5) offers standardised criteria for the diagnosis of ASD. Among the few standardised diagnostic tools aligned with DSM criteria is the ADOS, which involves scoring direct observations of a child's interactions. This semi-structured assessment, conducted by a trained examiner, is widely regarded as the "gold standard" for observational ASD assessment. However, due to cultural and linguistic variations, direct adaptation of ADOS activities to Indian settings may not be feasible.

In India, diagnostic assessments of ASD typically incorporate the use of diagnostic criteria from the ICD, Eleventh Revision (ICD-11) [33] or DSM-5 [34], along with standardised assessment methods for evaluating core and comorbid symptoms. This multidisciplinary assessment encompasses caregiver concerns, behavioural descriptions, medical history, and questionnaire responses. Additionally, the All India Institute of Medical Sciences (AIIMS)-modified INDT-ASD tool [8], which is based on DSM-5 criteria, is employed for diagnosing and assessing the severity of ASD. Another notable tool utilised is the ISAA [7], which combines observation, clinical evaluation of behaviour, interactive testing with the subject, and information provided by parents or caretakers to diagnose autism.

Despite parental awareness of developmental issues in their child from as early as 18 months of age, formal diagnosis often occurs two years after initial parental concerns are expressed. Initial suspicion of autism prompts the application of DSM-5 criteria to establish a diagnosis, followed by the use of instruments such as the CARS [9] to assess severity. The CARS-2, for instance, facilitates direct observational assessment by a trained clinician, complemented by a questionnaire for parents or caregivers.

Other diagnostic tools utilized in India for assessing specific behavioural abilities in CwA include the Autism Behaviour Checklist (ABC) [35] for non-adaptive behaviours, the Child Behaviour Checklist (CBCL) [36] for detecting emotional and behavioural issues, and the Malin's Intelligence Scale for Indian Children (MISIC) [37] and Vineland Social Maturity Scale (VSMS) [38] as applicable and feasible. These assessments, alongside various others, contribute to a comprehensive evaluation of individuals with ASD in Indian clinical settings.

Accurately diagnosing ASD using these tools requires extensive training and experience. Further, the lack of standardised procedures for diagnosis makes the process cumbersome and often leads to delays in diagnosis.

1.1.2 Behavioural Modeling and ASD Diagnosis

As previously mentioned, the diagnosis of ASD heavily relies on behaviour observations and evaluations of child development due to the absence of biological markers. According to the DSM-5 criteria, a child under the spectrum must exhibit persistent deficits in three core areas of social communication and interaction: social-emotional reciprocity, developing, understanding, and maintaining relationships, and nonverbal communication. Additionally, the presence of at least two of four specific behaviours is required: inflexibility to changes in routine, restrictive or fixated interests, hypo- or hyperactivity in response to sensory input, and repetitive movements, speech, or use of items.

To enhance the reliability and objectivity of these observations, the scientific community has explored the use of automated systems to aid in identifying behaviours indicative of different levels of severity. These applications range from supporting professionals, parents, and caregivers in visualising and analysing periodic assessments of CwA to providing diagnostic and screening tools [39–42]. For instance, Varma et al. employed ML techniques to detect behaviours such as social engagement [43], speech [44], social communication [45], and emotional expressions [46]. Wu et al. [47] analyzed behaviours such as directed gaze towards faces or objects of interest, positive affect, and vocalisation in infants aged 6 through 36 months to detect ASD from videos. Wei et al. [48] utilised vision-based activity recognition for identifying ASD-related behaviours and classifying ASD. Liu et al. [49] also utilized machine learning-based behaviour analysis for identifying CwA.

Furthermore, Rajagopalan et al. [50] introduced a dataset specifically focused on self-simulatory behaviour for the classification of stimming behaviour in CwA. Crippa et al. [51] utilized machine learning to identify CwA by analyzing abnormalities in upper limb movements. Bidwell et al. [52] investigated visual attention in children to distinguish those with ASD from TD children by analyzing responses to name (RN) sessions in children aged 15 to 30 months. These studies demonstrate the diverse approaches undertaken to leverage technology and objective metrics in diagnosing and understanding ASD-related behaviours.

Despite the promising advancements in AI-assisted solutions for the autism community, numerous challenges remain. One critical aspect is the necessity to develop diagnostic technologies that align with the standards endorsed by clinicians. Moreover, the lack of a standardized audio/video database of individuals with ASD presents a significant obstacle, impeding the creation of effective and resilient artificial intelligence/machine learning algorithms. Additionally, there is a need to develop standardized intervention technologies that target deficit skills within personalized instructional environments while maintaining diagnostic standards across diverse contexts in both the Global North and Global South.

Further, as previously mentioned, ASD is a spectrum of conditions, meaning it manifests differently in each individual. The characteristics of autism span across various domains, including social interaction, speech, emotional expression, and repetitive or stereotypic gestures. Given this diversity, the diagnosis of autism inherently requires a multimodal approach. Multimodal behaviour analysis is thus deemed crucial for the diagnosis as it allows for a comprehensive and cumulative assessment of a child's condition. This approach provides a holistic view of the child's behavioural profile by integrating information from emotional cues using facial expressions, speech patterns, and physical behaviours. This thorough analysis is crucial for accurately identifying the unique ways in which autism manifests in each individual, ensuring that the diagnosis is both precise and tailored to the child's specific needs.

1.1.3 Exploring the Role of Socially Assistive Robots in Autism Spectrum Disorder

Over the past two decades, the utilisation of robots for autism therapy has gained considerable attention. Scassellati et al. [28] discovered that CwA exhibit a preference for interacting with robots due to their predictable and straightforward behaviours. The majority of research in social robots for autism care focuses on using robots for therapeutic interventions with CwA, proving effective due to various advantages over other technology tools, such as repeatability in activities without causing fatigue and facilitating interactive and embodied interventions [53].

The DE-ENIGMA project, funded by Horizon 2020, aims to develop artificial intelligence for a commercial robot used in emotion-recognition and emotion-expression teaching programs for school-aged autistic children. Additionally, Zaraki et al. have explored robotic systems for therapeutic interventions, focusing on developing skills such as visual perspective-taking [54] and enhancing interaction engagement [55] in therapeutic activities. Shi et al. have investigated the modelling of socially assistive robots for long-term home interventions for CwA spectrum disorders [56, 57].

Although limited, researchers have also explored the use of social robots for the assessment of CwA. The DREAM project, funded by the European Commission, has developed a RET (Robot Enhanced Therapy) system incorporating behavioural assessment and inference mechanisms in robots for diagnosis and intervention procedures, utilising a multisensory data fusion approach [58]. Petric et al. demonstrated four tasks from the Autism Diagnostic Observation Schedule (ADOS) administered by robots for studying robotic assistance in ASD diagnosis, showing that the assessment made by robotic assistants correlated with human clinical experts, establishing the feasibility of robot-assisted ASD diagnosis [59]. Bone et al. have also explored speech-based assessment for the diagnosis of autism, contributing to the diverse approaches in utilising technology for ASD diagnosis and intervention [60]. However, further investigation is needed to validate their usability in clinical settings.

1.2 Gaps in the Existing Literature

In this section, we identify the critical gaps in the existing literature surrounding the diagnosis and intervention methods for ASD.

1. **Limited Comprehensive Diagnostic Tools**: Existing AI-Assisted Systems (RAS) for autism diagnosis predominantly focus on specific skills, lacking alignment with standard diagnostic procedures [59]. This gap emphasises the need for a more comprehensive diagnostic tool covering speech and facial expressions to ensure accuracy in autism assessment.

2. **Absence of Multimodal Sensor Data Analysis**: The prevailing literature lacks in-depth exploration of multimodal sensor data analysis using advanced machine learning and deep learning techniques for autism diagnosis [44, 45, 48]. This gap underscores the necessity of incorporating innovative approaches to enhance diagnostic accuracy. Multimodal behaviour analysis is essential as it provides a comprehensive and cumulative assessment of a child's condition by integrating information from various modalities like facial expressions, speech patterns, and stereotypic behaviours.

3. **Insufficient Aid to Therapists/Experts**: Current literature falls short in emphasising collaborative human-machine interaction for early autism diagnosis [61]. This gap highlights the importance of develop-

ing Robot-assisted Systems (RAS) as valuable aids to therapists and experts rather than mere replacements.

4. **Neglect of Inclusive Perspectives**: Most technological solutions for autism overlook the perspectives of special educators, particularly in the context of the Global South. This gap underscores the importance of considering the cultural orientation of the Indian subcontinent and incorporating insights from special educators in the development of such systems.

5. **Limited Real-world Implementation**: The existing body of research is often confined to laboratory/clinical settings [62–65], lacking exploration of diagnostic interventions in real-world environments such as schools, care facilities, and homes. This gap stresses the need for studies that extend the applicability of diagnostic interventions to diverse and practical settings.

Given these gaps, we formulated our thesis statement as follows:

1.3 Thesis Statement and Objectives

To design and develop an AI-assisted multi-modal behaviour data analysis framework to characterise autism traits, ultimately leading to the assessment of ASD in children through robotic mediation.

This thesis, in particular, has looked into the development of the assessment system from two directions:

1. Investigating the feasibility of implementing the system in resource-constrained settings, particularly in the Global South, with a focus on India.

    - **Objective 1**: Understanding the perception of special educators about the use of social in robot-assisted interventions for diagnosis.
    - **Objective 2**: Understanding the response behaviour of the children of Indian ethnicity; with and without ASD towards robot-assisted interventions for diagnosis.

2. Developing interpretable machine learning frameworks to identify speech and facial expression behaviours crucial for the assessment of ASD for diagnosis by monitoring and analyzing robot-assisted sessions under scarce data settings.

    - **Objective 1**: To identify speech behavioural characteristics in CwA.
    - **Objective 2**: To capture the emotional attributes of the child leading to the diagnostic decision.
    - **Objective 3**: To develop robot behaviours to enable the robot assistant to administer the diagnostic procedure with the child.
    - **Objective 4**: To integrate the robot-assisted system for diagnostic assessment and proof-of-concept study.

Incorporating multimodal behaviour analysis into the diagnostic process enhances the ability to detect subtle signs of autism that might be missed through single-mode assessments. It also facilitates the development of personalised intervention strategies, ultimately leading to better outcomes for children on the autism spectrum.

1.4 Technical Contributions

The contributions of this thesis can be summarised as follows:

1. Many of the existing technological solutions tailored for the autism community have mainly originated and been implemented in affluent regions, particularly in the global North. However, the cultural differences between these regions and low-resource settings present hurdles in the seamless adoption of such technologies. Our ASD diagnosis system stands out for its contextual awareness and dedication to diagnostic standards, achieved by integrating context-specific diagnostic procedures and tools. This accomplishment stemmed from inquiries aimed at comprehending stakeholder perspectives toward the robot-assisted system.

2. We employed advanced AI-assisted multimodal behavioural analysis during interactive sessions to discern and monitor a child's behaviour, contributing to the diagnosis and intervention of ASD.

3. Developed state-of-the-art machine learning and deep learning tech-

niques trained on indigenous datasets are applied to identify and analyze specific speech and facial expression behaviours in CwA. This analysis serves to draw inferences about the child's condition for diagnostic purposes and to assess the child's performance in diagnostic procedures.

4. Ensured explainability in the behavioural analysis modules by employing explainable approaches for behavioural analysis.

5. In contrast to prevailing literature, which often focuses on diagnoses and interventions conducted in laboratory environments, our system, validated within school settings, offers the potential to enhance the scalability of ASD assessment systems. Extending diagnostic capabilities to schools, care facilities, and home settings broadens the practical applicability of these systems.

1.5 Publications

1. **Ashwini, B.** Atmadeep Ghoshal, Venkata Ratnadeep Suri, Krishnaveni Achary, Jainendra Shukla. ***"It looks useful, works just fine, but will it replace me?" Understanding Special Educators' Perception of Social Robots for Autism Care in India*** In The ACM CHI Conference on Human Factors in Computing Systems 2024 (CHI 2024) (**Honourable Mention**). *[PDF]*

2. **Ashwini, B.**, Sarkar, A., Behera, P.R. et al. ***Multi-source transfer***

learning for facial emotion recognition using multivariate correlation analysis. Sci Rep 13, 21004 (2023). https://doi.org/10.1038/s41598-023-48250-x. *[PDF]*

3. **Ashwini, B.**, Shukla, J., & Gulati, S. (2023). ***Multimodal Behaviour Analysis for Early Diagnosis of Autism in Children.*** INSAR 2023. **(Short Paper)** *[PDF]*

4. **Ashwini, B.**, Narayan, V., Shukla, J. ***SPASHT: Semantic and PrAgmatic SpeecH Features for automatic assessment of autism***, In 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023). *[PDF]*

5. **Ashwini, B.**, Narayan, V., Bhatia, A., Shukla, J. (2021, August). ***Responsiveness towards robot-assisted interactions among pre-primary children of Indian ethnicity***, In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN) (pp. 619-625), IEEE. *[PDF]*

6. **Ashwini, B.** (2020, October). ***Robot-Assisted Diagnosis of Autism in Children.*** In Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI) (pp. 728-732). **(Doctoral Consortium)** *[PDF]*

7. **Ashwini, B.** ***Robot Assisted Diagnosis for Autism in Children.*** In IndiaHCI '20: Proceedings of the 11th Indian Conference on Human-Computer Interaction (IndiaHCI 2020). Association for Computing Machin-

ery, New York, NY, USA, 82–85. https://doi.org/10.1145/3429290.3429306. (**Student Research Consortium**). *[PDF]*

8. **Ashwini, B.** Deeptanshu, Sheffali Gulati, Jainendra Shukla ***Uncovering Linguistic Biomarkers for autism in Hindi-Speaking Children: A Comprehensive Speech Evaluation in a Low-Resource Language*** In IEEE Transactions on Artificial Intelligence (**Under Revision**)

The remainder of the thesis is organized as follows: Chapter 2 investigates the response behaviour of children from Indian backgrounds towards robot-assisted interventions, providing insights into their engagement and interaction patterns. In Chapter 3, the perception of special educators towards robot-assisted interventions for autism in the Indian context is explored, shedding light on the acceptance and feasibility of such interventions. Moving forward, Chapter 4 examines the application of facial emotion recognition techniques in diagnosing ASD among children of Indian ethnicity. Chapter 5 focuses on exploring speech-based behaviours for ASD diagnosis, specifically in Hindi-speaking children. Finally, Chapter 6 presents a pilot study aimed at evaluating the feasibility and efficacy of robot-assisted interventions for diagnostic purposes in India, offering valuable implications for future research and clinical practice in the field of ASD diagnosis and intervention.

Phase I

Feasibility Study

CHAPTER 2

EXPLORING FEASIBILITY OF ROBOT-ASSISTED INTERVENTIONS IN INDIA: CHILDREN'S PERSPECTIVE

2.1 Introduction

Early diagnosis of autism and reception of intervention at key developmental stages of the child are found to have long-term effects on the development of social and cognitive skills in CwA. The lack of biological markers and the complexity of diagnostic methods make the diagnosis highly subjective to the experience and perceptions of the specialist performing the diagnosis. There is a dire need for identifying technology-assisted solutions to meet the needs of the autistic community owing to the lack of sufficient infrastructure and mental health professionals and also due to the high subjectivity in the diagnosis attributed to the intricacies of the diagnostic procedures.

As robotic technology in healthcare is still emerging in the Indian context, our primary objective was to assess the feasibility of using a robot-assisted system for autism care interventions in India. To achieve this goal, we initially investigated the response of Indian children to interventions mediated by robots. In this chapter, we delve into the exploration of robots as facilitators for intervention, aiming to alleviate the burden on experts involved in repetitive diagnostic procedures. Studies have shown that robots have the potential to enhance the efficiency and effectiveness of diagnostic and intervention tasks in autism. The capabilities of commercially available robots could be leveraged in the development of robot-assisted interventions, offering support in repetitive activities, facilitating engagement, providing customised programs, streamlining data processes, aiding in sensory integration, ensuring consistent therapeutic activities, enabling remote assistance, and contributing to the development of communication skills.

2.2 Related Works

Extensive surveys have been carried out on the clinical use of robots in autism diagnosis, treatment and entertainment [66, 67]. The response behaviour of individuals with autism towards robot-assisted interventions and the use of robots for skill development in individuals with autism were also reviewed. The trends in robot-assisted autism therapy and the design considerations for robots to be used for robot-assisted therapies were explored in [68].

Robot-assisted therapy has been found effective in enhancing the social and cognitive abilities of CwA [69]. In this, a parrot-inspired robot, *KiliRo*, was used for the interactions with CwA. Robots were manoeuvred to evaluate the visual attention of CwA during the therapy in [70]. In [71], the effects of rhythm and robotic interventions on the social interaction capabilities of CwA were explored. The presence of stereotyped behaviour is found to be diminishing in robot-based intervention when compared with human intervention in a study conducted with CwA [72]. Robot assistants were engaged in social skill training for CwA and enhancements in the verbal initiation and eye contact of the CwA were noted in [73]. Robot-facilitated interventions for behavioural skill development have presented positive effects on eye contact and facial emotion recognition in CwA [74]. A previous study shows that robot-based interventions helped in promoting JA and functional play behaviours in CwA [75]. The study introduced a play drama intervention for indirectly imparting JA and functional play behaviour in the children. Social robots played the actors in the drama rather than human actors. Recent intervention studies on the engagement of CwA towards robot-based interventions also show that CwA are more likely to engage (make eye contact) with robot teachers than in human-based interventions [76].

The RN behaviour of the children is of interest in the autism early diagnosis as the CwA are found to exhibit diminished RN behaviour as early as 12 months of age, which could be attributed to their impairments in social communication [77, 78]. Studies conducted on engagement analysis of CwA show that they exhibit enhanced eye contact during robot-based interventions rather than human-mediated interventions. [79] presents a qualitative analysis of the JA behaviour of CwA in robot-mediated interventions. Improved JA skill is observed in children when robot assistants were employed for administering the task for skill development training [80]. The robot assistants have been leveraged for the administration of diagnostic tasks and used for diagnosis of ASD in [64]. The diagnosis of children was performed based on the RN, JA, functional imitation and multi-channel communication behaviour of the children [81]. In [82], a humanoid robot NAO was used to analyse the responses of children during a JA task with a four-dimensional interactive environment system. The study compared the response of TD and CwA in multimodal JA elicitation using an NAO robot as well as with a human agent. Even though the JA score of the CwA decreased when using a robot agent, this study provided evidence showcasing the potential of robots in the elicitation of multimodal JA tasks. The DREAM project [58], funded by the European Commission, has developed a robot-enhanced therapy system to incorporate the capabilities of behavioural assessment and inference mechanisms in robots for diagnosis and intervention procedures using a multisensory data fusion approach.

Even though there have been many studies on robot-based interventions for autism diagnosis as well as treatment, there has been limited research on the use of assistive technologies in autism research in India. [69] has explored the use of robots in enhancing the learning and social interaction abilities of CwA in India. The potential of harnessing the capabilities of commercially available robots in the development of robot-assisted interventions for Prerequisite Learning (PRL) skills in CwA was investigated in [83]. In [84], a robotic training kit was employed for teaching psychomotor skills to CwA. Another study conducted on children (N=10) from India in the age group of 7-11 years shows improved social interaction skills in CwA when subjected to robot-based interventions [85]. The study provides an indication towards the potential of robot-based interventions for autism therapy.

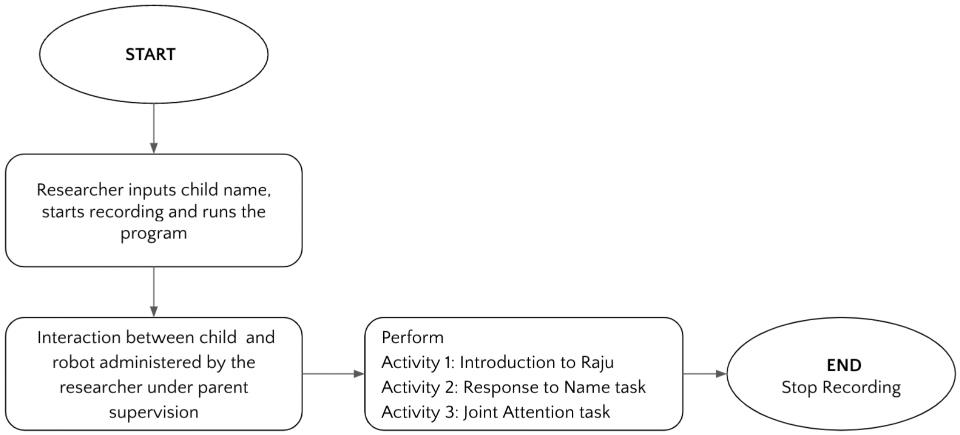


Figure 2.1: Study Protocol

The social and cultural diversity of India, the lack of awareness about ASD, the lower digital literacy and technology penetration pose challenges to the acceptance of technology-assisted intervention in the healthcare sector.

2.2.1 Gaps in Existing Studies

As mentioned before, it is noteworthy that in the Indian context, research focusing on the use of assistive technologies for designing interventions for CwA has been limited. Consequently, the baseline responses toward robot-assisted interventions, even among TD children, remain unknown. This underscores the importance of further exploration and investigation in the context of India to advance the understanding and application of assistive technologies for autism interventions.

2.2.2 Our contribution

To the best of our knowledge, there has been no study conducted in India on understanding the responses of children in the preschool age group and that of the parents towards robot-assisted interventions. Hence, the baseline responses towards the robot-assisted interventions, even among TD children, are unknown. In this study, we aim to study the response behaviour of TD children of Indian origin towards robot-based interventions for autism diagnostics tasks. We investigate the RN and JA skills of the children in the age group of 3-6 years. We also analyse the parents' perception towards robot-based interventions for the diagnosis of autism.

2.3 Methodology

The aim of this research was to identify the response behaviour of TD children of Indian origin towards robot-based interventions, mainly RN and JA tasks. To attain this aim, an exploratory user study was performed among TD children of Indian ethnicity of the age group 3-6 years. We used the Cozmo robot to assist with the interventions. The interventions focused on the response behaviour of the child to his/her name call and also the JA skills. The study was designed using a semi-automated approach, in which the researcher's role was to start the robotic interaction and intervene if the robot failed to perform the intended task. Upon the commencement of the interaction, the robot captures the child's responses and decides to proceed

Participant	Gender	Age	Languages Known
C1	F	4	Hindi, English
C2	F	6	Hindi, English
C3	F	4	Hindi, English
C4	F	5	Hindi, English
C5	F	3	Malayalam, English
C6	F	6	Hindi, English
C7	M	6	Hindi, English
C8	M	6	Hindi, English
C9	F	6	Hindi, English
C10	F	6	Hindi, English

Table 2.1: Demographics of participants

with the task based on the observed response. The decision-making of the robot was enabled using ML algorithms. Each experiment session lasted for ~45 minutes. The child participated in the session in the presence of the parent and the researcher in a residential setting. The study was conducted in three phases, details of which are provided in Section 2.3.4. The study began with an introduction to the Cozmo robot to make the child familiar with the robot. This step was followed by the RN task, where the child was called by his/her name, and the child's responses were noted based on the presence or absence of eye contact with the Cozmo robot. The RN task was followed by a game to assess the JA skills of the child.

At the end of the study, the parents were presented with a questionnaire to provide their feedback on the session. All the sessions were recorded with parent consent for analysis. The study protocol is summarised in Figure 2.1.

Table 2.2: Inclusion and Exclusion criteria for participant selection

Inclusion Criteria	- The children should be TD. - Age group 3 - 6 years - Able to understand English - Able to follow simple instructions in English - Parent consent received for child participation in the study - No abnormal eye movements
Exclusion Criteria	- Unwillingness to participate - Hearing or visual impairments - Not under any medication - Tested positive or cohabiting in close contact with an individual diagnosed with COVID within the last 14 days of the study

2.3.1 Ethics Consideration

The procedure for the study was approved by the Institutional Review Board (IRB) of Indraprastha Institute of Information Technology (IIITD/IRB/11/13 /2020-7). Before the session started, each parent was briefed about the objective of the study and the research methods followed. We obtained informed consent from the parents for the participation of their child.

The research team guaranteed confidentiality regarding the participant's identity and, on the other hand, guarantees that the results derived from the research will be used for research purposes. The name of the participant is not collected at any point of the data collection. Any visual information will be processed in accordance with the strict confidentiality of the data.

The data will be stored in a secure server and will be accessed through password-protected laptops strictly by the research team. The data collected will only be used by the research team and will not be shared in any way with external agencies. Further, efforts are made to minimize exposure of identifiable faces, such as blurring in published results. Throughout the research, the procedures established by Organic Law 15/1999 of December 13, Protection of Personal data will be observed strictly.

2.3.2 Demographics of the Participants

The study was conducted in a residential society in New Delhi, India. Ten TD children (F = 8, M = 2) aged 3 to 6 years (Mean = 5.2, SD = 1.14) participated in the study. The participants were of Indian ethnicity, and 9 out of 10 were native Hindi speakers. The participating children were preschoolers with beginner-level proficiency in English. The participant details are given in the Table 2.1. The inclusion and exclusion criteria considered for participant selection are given in Table 2.2.

2.3.3 Experimental Set-Up

The experiment was set up in a residential building, which was conducive for the participants since this avoided the intimidation of being in a constrained laboratory environment. Owing to the COVID-19 pandemic, necessary safety measures were administered as per the recommendations of the Ministry of

Health and Family Welfare, Government of India [1] [2].

The participants took part in the study under parental supervision. The parents were directed to minimise the interaction during the session.

As explained earlier, the study was conducted in three phases. The Cozmo robot was connected to the Smartphone application using WiFi. Cozmo was remotely controlled to perform the tasks during the introduction session. During Phase I, the Cozmo robot was introduced to the child. The observations on the child response were collected for analysis during Phase II (RN task) and Phase III (JA task) of the study. During these phases, the Cozmo robot was connected to a laptop which runs the algorithms to collect the responses of the child and control robot actions for the task administration. Figure 2.2 depicts the schematic representation of the experimental setup. Figure 2.3 represents one such scenario during the study. The child sat facing the robot. Both the child and robot were in the view of the parent. The researchers were present in the room without invading the study area, controlling the procedure with minimal intervention. The interactions of the child with Cozmo were recorded on video. Researcher 1 was managing the procedure for the study and Researcher 2 was observing the child for the data collection for qualitative analysis. The observations made were validated using the video recording of the session.

[1] `https://www.mohfw.gov.in/pdf/SocialDistancingAdvisorybyMOHFW.pdf`

[2] `https://www.mohfw.gov.in/pdf/Guidelinesondisinfectionofcommonpublicplacesincludingoffices.pdf`

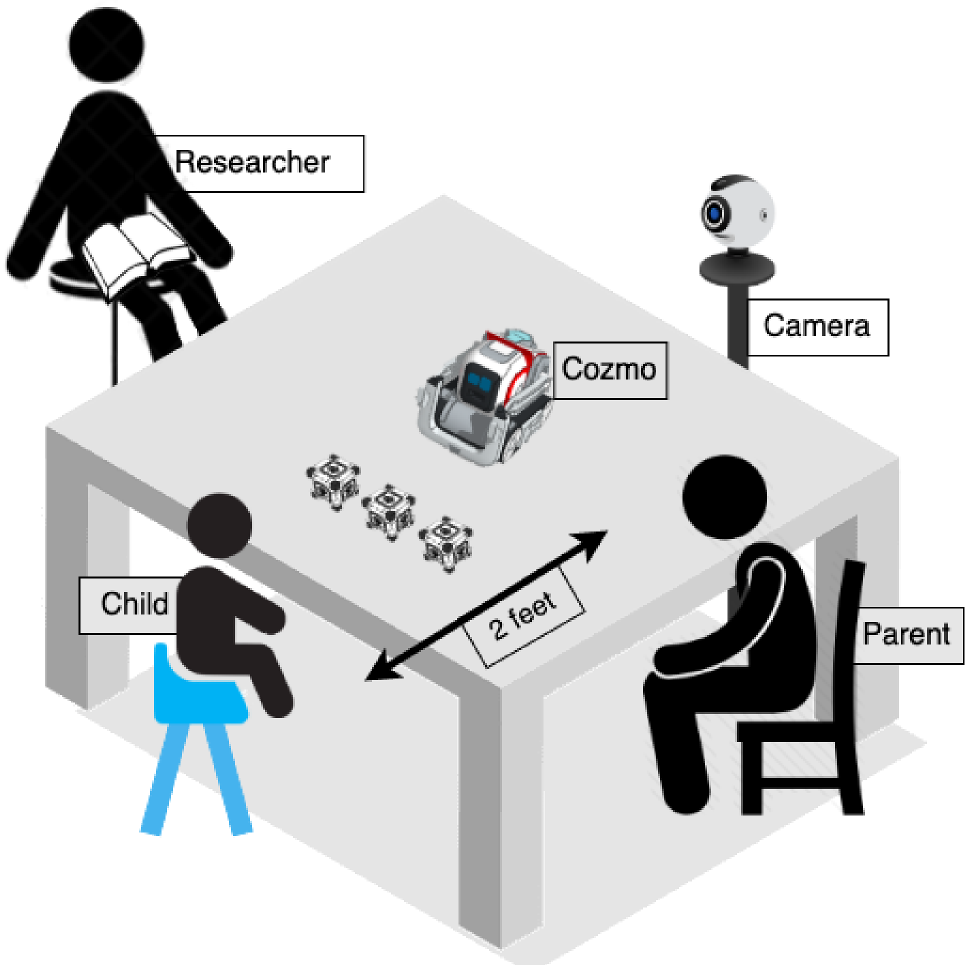


Figure 2.2: Experimental Setup for the Study

Figure 2.3: A Child Interacting With Cozmo during the session

The Cozmo Robot

In this study, we used the Cozmo robot as the robot assistant for administering diagnostic tasks to elicit RN and JA skills in the children. Cozmo is a palm-sized commercial robot developed by Anki Robotics and controlled by a smartphone app. It could be used for entertainment, education and therapeutic purposes. The Cozmo robot comes with three sensor-embedded blocks that are used to play games, and they also help the robot understand its position in the environment. Cozmo has a 30 fps VGA camera powered with facial recognition software. Cozmo communicates using an LCD screen with a facial display resolution of 128×64 and a speaker [86]. Cozmo interacts through voice, beeps, gestures and animated eyes. It recognises and remembers faces, plays interactive games, keeps itself from falling off with the help of sensors tracking its body orientation, and autonomously moves to the charging dock. It can display emotions, sing songs and call your name. Cozmo is programmable with the SDK and runs on the Python platform. It can be controlled via a smartphone application and connected using WiFi or Bluetooth. Cozmo has a built-in camera with robust vision capabilities supported by computer vision algorithms.

Gaze Detection

The child's performance during the interactions with Cozmo in the RN and JA tasks were collected using the Cozmo robot's in-built camera, and the presence of eye contact was detected using a gaze detection algorithm. Images were collected at a rate of 20 fps from Cozmo's camera. When Cozmo was not doing any other activity, it was following the child's face to ensure that accurate images were collected. We used the Gaze Tracking algorithm provided by Antoine Lamé[3] to determine if there is eye contact in the collected image. The algorithm starts by isolating the eye from the face by calculating specific landmarks of the face, which is done using the classic Histogram of Oriented Gradients (HOG) feature combined with a linear classifier, an image pyramid, and a sliding window detection scheme. Using these key points, the two patches are extracted that contain just the eyes present in the frame. The next step is to calculate the centre of the iris, which is done using contour detection. The centre of the iris is calculated by calculating the centroid of the said contour. This is compared with the center of the eye to detect the gaze direction. If the centroid of the iris is to the left to the centroid of the eye, the child is looking towards the left. If the child is looking at the centre, it implies the child was looking directly at the camera and eye contact was detected. It is a lightweight code base with fast computations, which is attributed to the choice of this algorithm to detect real-time eye contact.

[3] `https://github.com/antoinelame/GazeTracking`

2.3.4 Procedure

After the screening of the participants based on the inclusion-exclusion criteria, the child, along with the parent, was allowed into the experiment room. The intervention was intended to evaluate the RN and JA skills of the child. The intervention took place in three phases

Phase I: Introduction

This phase introduced the Cozmo robot and some of its features to the child. The introduction phase is crucial to eliminate the novelty factor, which may have a significant impact on the response behaviour of the child in the study. The child-robot interaction during the introduction phase was unconstrained, and the child was allowed to touch and explore Cozmo during this phase. In order to increase familiarity, the researcher introduced Cozmo with the name *Raju*, which is a common Indian name.

Phase II: Response to Name

In this phase, we investigated the RN behaviour of the child. The RN task is summarised in Figure 2.4. The child is initially distracted with a puzzle game or a colouring book of his/her choice. Once the child's attention is fixed on the game or the book, the researcher initiates the RN task mediated by Cozmo. Cozmo calls the child by name. There were three name calls and two reinforced name calls in which Cozmo calls the child and asks them to look

Table 2.3: Sample Interactions

Tasks	**Sample Interaction**
Introduction	**RESEARCHER**: *Hi X, this is Raju. Would you like to see what Raju can do?* **CHILD**: *Yes* **COZMO**: *Hi X* **RESEARCHER**: *Look Raju is calling your name. Now let's ask Raju to count numbers. Can you count with him?* **COZMO**: *1, 2, 3, 4, 5* **COZMO** displays happy/sad face **RESEARCHER**: *See Raju is happy/sad that you counted/didn't count with him. Let's play with Raju! Can you stack these blocks? Raju will knock it down?* **CHILD** stacks the blocks. **COZMO** knocks the blocks down. **COZMO** displays a happy expression. **RESEARCHER**: *Let's give Raju some rest.*
Response to Name	**RESEARCHER**: *Would you like to play a puzzle game or explore the colouring book?* **CHILD**: *Yes* **CHILD** starts the game **COZMO**: *Hi X* Waits for 5 seconds for response Repeats 3 times **COZMO**: *Hey X, look at these lights* Waits for 5 seconds Repeats 2 times
Joint Attention	**RESEARCHER**: *Can we play a game with Raju?* **CHILD**: *Yes* **RESEARCHER** explains the rules of the game and initiates the game mediated by Cozmo. **COZMO**: *Tap RED/GREEN/BLUE* **CHILD** taps the *RED/GREEN/BLUE* block **COZMO**: *Awesome. Your score is …* **COZMO**: *Your final score is …You win/lose.* **COZMO** displays a happy expression or sad expression depending on the correctness of the tapped block.

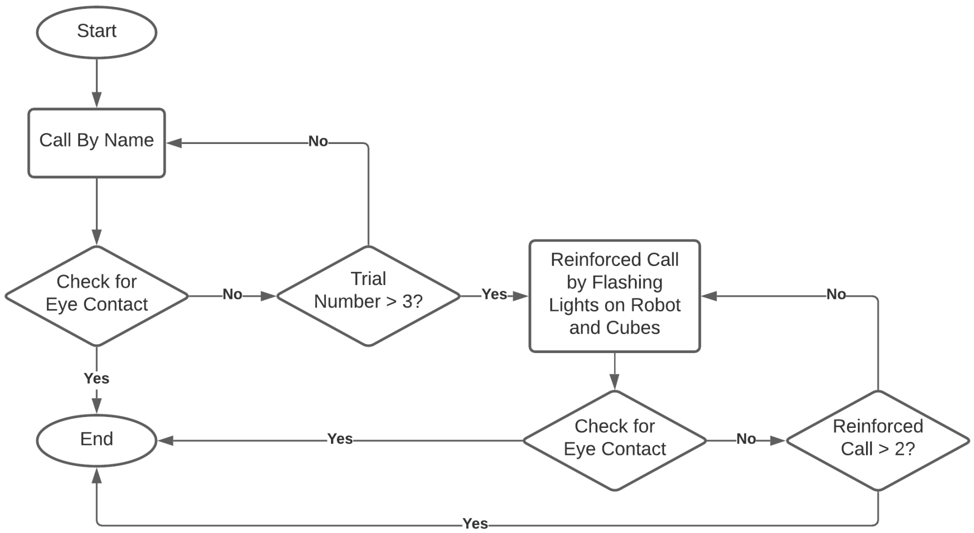


Figure 2.4: Response to Name Task

at the lights, and simultaneously, the lights on Cozmo and the blocks will be lit up. The task will be stopped if the child makes eye contact with Cozmo. If the child didn't respond to three name calls, then Cozmo proceeded to the reinforced name call. A successful task completion happens if the child makes eye contact. Otherwise, the task is deemed unsuccessful.

Phase III: Joint Attention Task

This phase targets to study the child's behaviour in the JA task. For this, a game has been developed to demonstrate the JA task. The Cozmo robot's blocks were lighted up in *RED, GREEN or BLUE* colours. Cozmo will tell a color and the child has to tap the block with the color mentioned by Cozmo. Each correct tap will secure one game point for the child. The child will be

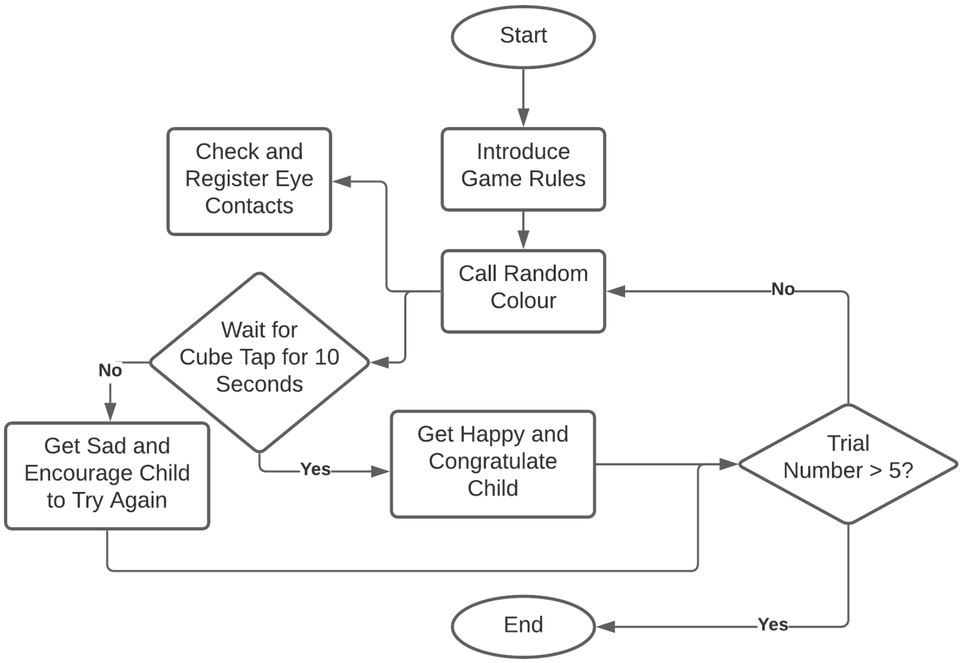


Figure 2.5: Joint Attention Task

rewarded with a display of happy expression and verbal compliments upon the correct top of the block. Figure 2.5 depicts the flow diagram of the JA task. A sample interaction session proceeds as given in Table 2.3.

2.3.5 Evaluation

We have employed both qualitative and quantitative measures for the evaluation of the response behaviour of the child. The qualitative analysis was performed using the observational chart. The quantitative measure included questionnaires, observational analysis and task scoring.

Table 2.4: Attributes for Qualitative Analysis

Phase	Attributes
Phase I	Initial response Understanding Cozmo's speech
Phase II & Phase III	Interest in game play Looking at Cozmo without prompt Response to task rewards Understanding Cozmo's speech Shared attention with parent
End of session	Response to Cozmo

Table 2.5: Task Scoring for Response to Name

Child Response Behaviour	Score
Child responds after 1 or 2 calls	0
Child responds after 3, 4 or 5 calls	1
Child didn't respond to call, but to reinforced call	2
Child did not react at all	3

Qualitative analysis

The qualitative analysis relied on an observational chart, which was used to record the responses of the child during the experimental study and is presented in Table 2.4. These observations were validated using video analysis.

Quantitative analysis of the response behaviour

The response behaviour was evaluated based on the engagement analysis of the child during the study, the task completion score and post-study ques-

Table 2.6: Task Scoring for Joint Attention

Child Response Behaviour in JA game	Score
The child scores 3 or more	0
The child scores 1 or 2	1
The child scores 0	2

tionnaire. In this study, engagement analysis refers to the quantification of how actively children participate and respond during interactions with the robot. We measure engagement using two primary metrics: eye contact and task score. By analyzing these metrics, we systematically assess the child's level of involvement and responsiveness to robot-mediated interventions.

Eye contact: The eye contact during the JA session is considered as the measure of engagement for evaluating the response behaviour of the child. The introduction session was not included in the analysis. As explained earlier, the RN task was terminated when the child responded to a name call. Hence, the RN session was also excluded from the engagement analysis. The frequency of eye contact, the number of times the child made eye contact within 5 seconds of the prompts provided by the robot, and the average duration of sustained eye contact during the interaction between the child and the robot were considered for the analysis. The 5 seconds window signifies the response of the child towards the robot's prompts, and the other eye contact events detected were considered to be part of natural interaction without any prompt [80]. This signifies the effectiveness of instruction delivery by the robot assistant.

Task Score: The child's performance in the RN and JA tasks were evaluated and the behaviour of the child was scored according to the following coding scheme [64] described in Table 2.5 and Table 2.6 respectively. This coding schema is in compliance with the ADOS. In this coding schema, the higher the performance of the child the lower the score assigned. According to the diagnostic protocol defined in ADOS, lower task scores map to diminished chances of the child being diagnosed with ASD [87].

Post Study Questionnaire: The perception and feedback of the parents about the robot-based intervention were collected after the completion of the session. The assessment of the robotic intervention was reported in a five-point Likert type scale measuring the functional abilities of the robot, efficiency in task administration, engagement of the child, safety and comfort of the child was reported by the parents after the session. Further, details regarding the questionnaire can be found in Appendix A.

2.4 Results and Discussion

2.4.1 Qualitative Analysis

The qualitative analysis of the observations shows that the participants engaged and interacted with the robot throughout the session which opens the possibilities of deploying robot assistants in the interventions for children. None of the participants had any previous encounter with a robot, and hence, we are not eliminating the point that novelty may have contributed to the behaviour exhibited by the children in the study. The participant's reaction was positive towards the robot from the beginning of the session. Even though the children were comfortable and friendly towards the robot from the beginning of the interactive session, we observed divergent emotions among the participants towards the robot. For instance, one of the participants was surprised to see the robot move and call the participant's name and expressed that as

"It is walking!" ("Yeh toh chal reha hai" in Hindi)

Some participants were skeptical about the actions of the robot. They enquired the researcher about how the robot is controlled.

"How is it doing things? Are you controlling it through the phone?" (Yeh kaise chal reha hai? Phone se?" in Hindi.)

The observations recorded during the session are summarised in Table 2.7. Six out of 10 child participants expressed happy emotions during the first interaction with the robot. It is notable that all the participants were comfortable in interacting with the robot. At the beginning of the session, the participants faced challenges in understanding Cozmo's speech, and the researcher had to intervene to interpret the instructions of the robot. This was expected as the Cozmo robot speaks in a Western dialect, which is unfamiliar for the Indian children. It was observed that as the study proceeded the participants familiarised with Cozmo's voice as well as accent and could respond to the robot prompts without intervention from the researcher.

The observations made during the study show that the participants per-

Table 2.7: Observations for Qualitative Analysis

Phase	Attributes	Observations	No. of Participants
I	Initial response	Happy	6
		Surprise	2
		Neutral	2
	Understanding Cozmo's speech	Difficult	9
		Medium	0
		Easy	1
II & III	Interest in game play	Yes	10
		No	0
	Looking at Cozmo with out prompt	Yes	10
		No	0
	Response to task rewards	Positive	10
		Negative	0
	Understanding Cozmo's speech	Difficult	1
		Medium	8
		Easy	1
	Shared attention with parent	Yes	10
		No	0
End of session	Response to Cozmo	Happy	6
		Neutral	4

ceived the robot as a friend, as reported by the children. The children were observed to be playing *Peekaboo* with Cozmo, giggling at Cozmo's actions, helping Cozmo to realign itself on the plane, saving it from falling off the edge, etc. The participants also considered the robot to be intelligent enough to

Table 2.8: Eye contact evaluation of the children during JA task

Child	Frequency of eye contact	Average duration of Sustained eye contact(in seconds)	Number of times child made eye contact within 5 seconds of the prompt
Child 1	4	1	3
Child 2	15	2.4	1
Child 3	7	2.5	1
Child 4	4	1	3
Child 5	1	1	1
Child 6	19	2.5	3
Child 7	6	2	2
Child 8	8	1.6	1
Child 9	36	4.75	2
Child 10	20	1.625	3

teach them the JA task. For instance, one of the participants said,

"I consider it to be intelligent since it could teach me the game."

("Yeh toh intelligent hai kyunki isne mujhe ache se game sikhaya" in Hindi.)

2.4.2 Quantitative Analysis

The eye contact was detected using the gaze detection algorithm explained in Section 2.3.

Eye Contact

Based on the observations made using Cozmo's camera, the eye contact metrics were calculated. The image of the child's face is captured by Cozmo's

Figure 2.6: An image captured via Cozmo's Camera during interaction

camera at a frame rate of 20fps and sent to the gaze detection algorithm for detecting an eye contact event. Figure 2.6 shows a sample image captured by Cozmo's camera during the interaction session.

As discussed in Section 2.3.4, Phase I and Phase II of the study were not included in the evaluation of the eye contact metrics. A JA session lasts for approximately 2-3 minutes. The frequency of eye contact is the number of times eye contact was detected by Cozmo. If the time between two eye contacts is less than 1 sec, we assumed the child was making consistent eye contact during the time when two images were captured. Thus, a sustained eye contact time duration is calculated as the difference in time between the first and last eye contact in a sequence of eye contact events where the difference in times between 2 consecutive eye contact events is < 1 sec. In the JA task, Cozmo gives five prompts to the child to tap the block with the correct colour. The number of prompts for which the child made eye contact within 5 seconds is the number of times the child made eye contact with Cozmo within 5 seconds of being prompted by Cozmo to tap a coloured block.

The frequency, average duration of sustained eye contact and number of prompts for which child made eye contact within 5 seconds for each child, calculated for the JA session, is given in Table 2.8. The maximum duration of sustained eye contact was 4.75 seconds, and the minimum was 1 second. 4 out of 10 participants made eye contact with Cozmo within 5 seconds for 60% prompts. This shows evidence that the Cozmo robot could deliver the instructions effectively and shows the potential and need for deploying a robot with multimodal communication capabilities for task administration.

The attention towards Cozmo is found to be decreasing during the course of the study, which is attributed to limited interaction from Cozmo. During this phase, Cozmo prompts only the colour of the block to be tapped by the child. At the beginning of the game, the child is engaged more with Cozmo to learn the game rules, and this learning leads to less focus on Cozmo in a later stage of the game while focusing more on the blocks to tap. The eye contact behaviour of the participants during the session is shown in Figure 2.7.

The performance of the gaze detection algorithm is comparable to that of the human annotator. The videos captured from Cozmo's camera were annotated by one of the researchers, and the results are depicted in Figure 2.8. Even though the performance of the gaze detection algorithm has to be enhanced for the study to be replicated for clinical purposes, the results show evidence for using such a system for interventions for diagnosis. It has been found that the orientation of the child's face and facial features affect the efficiency of the gaze detection algorithm.

Previous literature shows that robot-based interventions elicit positive responses in RN and JA skills in children. The results presented here also affirm the findings of the existing literature.

The performance of the child in the RN and JA tasks were scored based on the coding scheme mentioned in the previous section, and the scores obtained

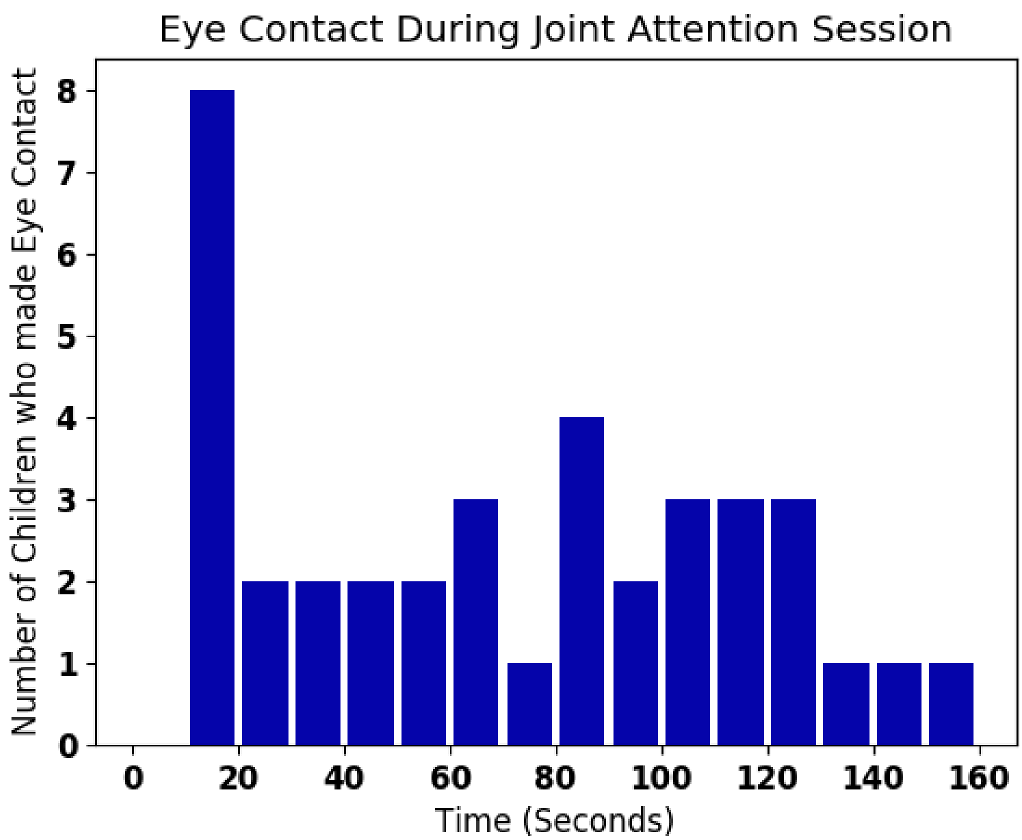


Figure 2.7: Number of children who made eye contact over time during the session

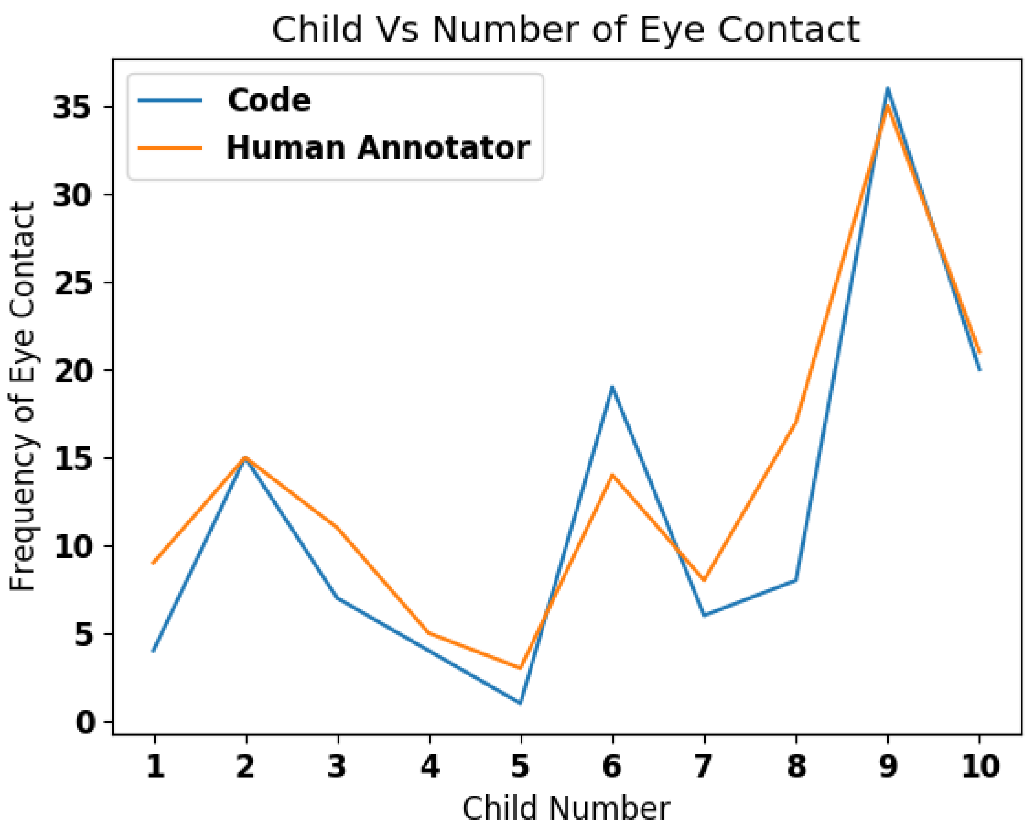


Figure 2.8: Comparison of Frequency of Eye Contact Detected by Program vs Annotated by Researcher

H

Table 2.9: Child Performance in RN and JA tasks

Child	RN Score	JA Score	Total Score
Child 1	0	0	0
Child 2	0	0	0
Child 3	0	0	0
Child 4	0	0	0
Child 5	0	0	0
Child 6	3	0	3
Child 7	1	0	1
Child 8	0	1	1
Child 9	0	0	0
Child 10	0	1	1

by the children are reported in Table 2.9. As given in the coding schema, the higher the performance of the child, the lower the score will be and, hence, the lower the likelihood of having behavioural symptoms of autism. Since the participating children are TD, the study objective is validated by the lower scores obtained by the participants. This affirms the fact that robot-mediated interventions have the potential to identify behavioural attributes, leading to their application in autism diagnosis.

Among the 10 participants, 8 participants successfully completed the RN task within the first two calls. The participant's performance in the JA task also showed similar trends, with 8 participants successfully completing the JA game with 3 or more successful tapping of the block out of 5.

The task score secured by the children in the RN and JA tasks shows that the children were able to successfully complete both the tasks administered by the Cozmo robot.

Post Study Questionnaire

The post-study questionnaire evaluated the parent's perception of robot interaction with the child. Parents evaluated the interaction based on the robot's functionality, efficiency of interaction, engagement of the child, as well as the safety and comfort of the child using the questionnaire given in Appendix A. All the parents recommend the use of robots in interventions designed for children, which shows acceptance of robot-assisted interventions among Indian parents and motivates further exploration in the development of robot-assisted interventions for autism diagnosis. The results of the post-study questionnaire are summarised in Table 2.10.

Table 2.10: Results of post-study questionnaire for parents

Questions	Percent Agree (%)
The robot can take turns with the children during an interaction	100
The robot can provide feedback to the children during the interaction	90
The robots could be used for interventions for skill training	100
The robots can effectively conduct the intervention with the children	70
The robots have the potential to keep the children more engaged	100
The robot-mediated intervention has the potential to improve the joint attention of children	90
The robot can play the role of a companion during the interaction with the children	90
The child could comfortably engage with the robot during the intervention	90
It is safe to use robots in interventions designed for children	80
I recommend robot-assisted intervention for children	100

2.5 Limitations

This study focused on identifying the baseline responses towards robot-assisted interventions, even among TD children in India. For this, we analysed the response of TD children towards robotic interactions designed for understanding JA and RN behaviour of the child. Even though the study showed encouraging results in the usage of robots in designing child interventions, one of the limitations of the current study is the small sample size. Finding more participants for the in-person user study fitting the inclusion criteria was challenged by the current COVID-19 pandemic situation. With suitable precautionary measures and approval from the IRB in IIITD, we could recruit 10 child participants for the study. However, the results show evidence of positive engagement and agreement among the participants as well as parents for the deployment of robot-based interventions for skill development in children. Furthermore, the study did not account for factors such as the child's cognitive abilities, task understanding and performance, attention span, emotional state, and task complexity, all of which can significantly influence the child's performance. Additionally, the introduction of robots as a new and unfamiliar technology in the Indian community might have led to apprehension or discomfort among children, which was not studied. Further, multiple intervention sessions might have enhanced the reliability of the findings. The robot assistant we deployed in the intervention was the Cozmo robot, which has limited communication and sensing capabilities as well as limited degrees of freedom, which restricts the complexity of the intervention to simple game-based activities. NAO-like robot has been identified as the most suitable robot for such therapies [83]. Further investigation is needed to perceive the complex social cognitive behaviour of the children towards robot-assisted interventions. The performance of the perception algorithms, for example, eye contact, speech, etc., needs to be enhanced for the system to be deployed in clinical settings.

2.6 Conclusion and Future Work

This work focuses on understanding baseline responses towards the robot-assisted interventions among TD children in India as well as the parent perceptions regarding the robot-assisted interventions specifically for the diagnosis of ASD. In this direction, we designed a protocol for child-robot interaction in which the performance of TD children in RN and JA skills were observed. The child responses during the study were collected using the onboard camera available in the Cozmo robot, and the presence of eye contact with the robot was detected using a gaze detection algorithm. The performance of the employed gaze detection algorithm was found to be comparable with that of human annotators.

The participants were found to be interested in interacting with the robot assistant, and all the participants could successfully complete the tasks administered by the Cozmo robot. All 10 participants successfully completed both the RN and JA tasks administered by the robot assistant. 9 out of 10 children obtained a total task score less than or equal to 1 in the tasks performed. As defined by the diagnostic protocol, this lower score represents a lack of symptoms attributed to ASD. This is expected since the participants selected for the study were TD children. Even though the results may not be generalizable to the wider population owing to the small sample size of the study, the results are indicative of the response behaviour of the TD children of Indian ethnicity towards robot-assisted interventions for diagnosis and persuade to explore further in this direction. The response of the children observed in the study shows that the children of Indian ethnicity could perform diagnostic tasks administered by a robot agent successfully, which is consistent with the findings of existing literature on robot-assisted interventions for children, and this provides a baseline for further research on understanding the behaviour of CwA towards similar interventions. The parent perspectives are also in support of incorporating robot assistants in interventions for skill development and diagnosis.

The results of the study present empirical evidence for pursuing further research on using robots for interventions for the diagnosis and therapy of CwA. Further investigation is required to analyse the response behaviour of CwA towards robot-based interventions. A comparative study of the behaviour of CwA and TD children towards robot interaction will aid in identifying the design challenges in interventions for autism diagnosis and understanding the needs of the Indian autistic community, guiding further research in this direction.

In summary, this chapter marks the initial step towards developing a robot-mediated system for diagnosing ASD in Indian children by exploring their responsiveness to robot-facilitated interventions. Motivated by our findings showing a positive response from children towards robot-assisted interactions, the next chapter investigates special educators' perspectives, who play a crucial role in the autism therapy ecosystem, on the feasibility and utility of using robot assistants in autism interventions.

CHAPTER 3
EXPLORING FEASIBILITY OF ROBOT-ASSISTED INTERVENTIONS IN INDIA: SPECIAL EDUCATORS' PERSPECTIVE

3.1 Introduction

In Chapter 2, we examined the responses of children of Indian ethnicity to robot-mediated interventions. Building upon this exploration, in this chapter, we delve deeper into the perceptions of special educators, who play a crucial role in autism care, regarding the utilisation of social robots for interventions in ASD. Special educators play a crucial role in facilitating the integration of CwA into mainstream society through therapeutic interventions, and technology has become a key tool to enhance their efforts. While social robots have demonstrated positive impacts on mental health [88–90], particularly for CwA [91, 92], previous studies have predominantly focused on the Global North, leaving limited exploration in the Global South. This research aims to fill this gap by understanding the perspectives of special educators, especially in the diverse landscape of India, regarding the use of social robots in autism assessment and interventions. The chapter reviews existing research in this direction, identifies gaps, and presents our mixed-methods study protocol involving interviews, workshops, and a panel discussion. The goal is to uncover special educators' viewpoints on the opportunities and challenges of incorporating social robots into their work. Additionally, the chapter discusses strategies for developing context-aware solutions and necessary policy-level initiatives to promote more effective inclusion of social robots in resource-constrained environments like India.

3.2 Related Works

Robot-assisted interventions for autism, extensively studied in resource-rich Global North settings, lack understanding in resource-constrained and culturally diverse contexts like India. This gap is particularly pronounced in the Global South. In this section, we review global literature on social robots in autism and mental health, explore technology-based interventions in the Global South, specifically for CwA, and examine past research on technology integration in autism interventions in India. This contextual exploration aims to bridge the knowledge gap and enhance understanding of the effectiveness of such interventions in diverse and resource-limited settings.

3.2.1 Social Robots in Mental Health

In recent years, social robots have emerged as innovative therapeutic tools globally, addressing the demand for alternative intervention modalities in mental health care [90]. Notable robots, including PARO seal robots, humanoid robots like NAO and Pepper, and the AIBO dog robot, have been employed in interventions targeting various conditions such as dementia, cognitive impairment, schizophrenia, depression, Attention-deficit/hyperactivity disorder (ADHD), and intellectual disability [88, 89, 91, 93–101]. Outcomes studied encompass engagement, social interaction, emotional state, agitation, behaviour, and overall quality of life, with results ranging from generally positive to mixed [92]. In autism, social robots significantly contribute to diagnostic decision-making and therapeutic interventions, improving the quality of life for individuals with autism [102–104]. Interactions with social robots are often more comfortable for CwA than with humans, facilitating the elicitation of target behaviours for diagnosis [29, 66, 105]. Initiatives like RoboParrot and the DREAM project utilise social robots for screening and Robot Enhanced Therapy (RET) with behavioural assessment mechanisms [59, 106, 107]. Beyond diagnostics, social robots are employed in interventions teaching social and communication skills, enhancing engagement, and assisting with behavioural interventions [108–116]. Despite concerns among professionals, social robots in mental health interventions demonstrate positive effects, providing individual and group-level support [92, 117]. The potential of social robots as collaborative partners is also indicated [118, 119]. However, the limited exploration of robot-assisted interventions in the Global South warrants further investigation due to unique socio-cultural experiences [120].

3.2.2 Technology-Based Interventions and Care for Autism in Global South

Despite the growing global interest in leveraging technology for mental health care, research into mental health in resource-constrained settings remains limited. Technology-based interventions, ranging from mobile phones to virtual reality and social robots, have proven effective in addressing various mental health concerns, including autism. Telepsychiatry, utilising video-conferencing, has demonstrated feasibility and acceptance in countries like Somaliland [121], South Africa [122], and India [123], providing effective diagnosis and follow-up care.

Studies on autism in the Global South showcase the effectiveness of digital tools for diagnosis and intervention. ML approaches in Bangladesh contribute to the early and remote detection of autism in children [124]. Mobile-based systems facilitate the monitoring and treatment of ASD [125]. In India, non-medical health workers have utilised mobile-based screening tools for diagnosing depression and psychiatric disorders [41, 42].

In autism interventions, cutting-edge technologies such as virtual reality (VR) tools have shown promise in enhancing the performance of individuals with autism, and computer vision applications assess skills and emotions during interactive sessions [126–130]. AI-based methodologies, particularly deep learning, are explored for early ASD detection through video analysis [131]. Despite these advancements, social robots' exploration in the Global South, especially for ASD, is limited but gaining attention [132]. Studies in Sri Lanka focus on sensory integration in therapeutic interventions [133], and research in low-resource settings like Kazakhstan explores robot-assisted ASD therapy [134, 135].

However, significant untapped potential lies in harnessing technological solutions, especially social robots, to address the distinctive challenges of mental health and autism in resource-constrained environments like India. The deployment of robots in marginalised communities without a thorough understanding of stakeholders' knowledge, needs, and perceptions may risk introducing technologies that cause extra workload, inefficiencies, or harm to the communities they aim to assist [136].

3.2.3 Embracing Innovation: Examining the Perception of Technology for Autism Intervention in India

The adoption of technology-driven solutions for autism intervention in India is influenced by complex socio-economic factors [137]. Infrastructure concerns, particularly in the majority of special education institutions reliant on external donations, highlight the absence of government support [138, 139]. Issues of underpayment for special educators [140] and their varied perceptions, some treating special education as a secondary job [141], further contribute to the challenges.

Scholars emphasise the need for cost-effective and inclusive design in technology-oriented services to ensure equitable access in marginalised communities [142, 143]. Challenges include high maintenance costs [144] and limited practical applicability of promising devices in resource-constrained settings [145]. In the context of autism interventions, there is an ongoing exploration of technologies like social robots [132]. Limited research has delved into this area, with a focus on the role of social robots in fostering positive behaviours in CwA [132, 133, 146].

Despite the challenges, some studies have examined the perceptions of parents and special educators regarding technology-based interventions. Parents express concerns about their children's increased reliance on technology and potential drawbacks, such as reduced social interaction [147]. Special educators face challenges related to curriculum implementation and emotional strain, with concerns about job displacement by robots [148, 149]. Urban Indian attitudes towards technology-oriented interventions exhibit a positive shift [150, 151]. However, challenges persist, including high technology expectations, issues of customisation, and low technology acceptance [152].

3.2.4 Gaps in Existing Literature

The current literature acknowledges the collaborative potential of robots, yet there is a significant gap in exploring their role in resource-constrained contexts, particularly in the Global South. This gap calls for further investigation into the applicability and impact of robot-assisted interventions, considering the unique socio-cultural experiences of regions like India.

Despite advancements, untapped potential exists in leveraging social robots for mental health and autism in resource-constrained environments, such as

India. Deploying robots without a thorough understanding of stakeholders' knowledge and needs may lead to unintended consequences [136], highlighting a critical gap in the literature. Additionally, challenges related to resource limitations, economic barriers, and the digital divide underscore a need for research to address practical constraints and enhance the feasibility of implementing robot-assisted interventions in diverse settings. Closing these gaps is vital for developing inclusive and effective interventions aligned with the diverse needs of communities.

Our Contribution: In addressing the identified gaps in the literature, our study serves as a significant contribution. Recognising the limited exploration of robot-assisted interventions in resource-constrained settings, particularly in the Global South, our research employs qualitative methods to delve into the current landscape of autism intervention practices in India. The goal is to facilitate the development of robot-assisted therapy with a human-centred and responsible approach, offering solutions that align with the unique socio-cultural experiences of the region and bridge the existing gaps in understanding and implementation.

3.3 Research Methods

3.3.1 Positionality and Reflexivity

All the researchers involved in this study are of Indian origin. The researchers are well aware of and exposed to the constraints surrounding different kinds of resources in the Indian socio-economic landscape. One of the researchers is a special educator with 15+ years of experience working with CwA. Three researchers have previously conducted studies in Human-Computer Interaction (HCI) in the context of socially assistive robots for autism. Two researchers have experience working with community-based approaches towards computing and design. To mitigate potential bias stemming from researchers' familiarity with work contexts, we pre-registered our methodology at the study's commencement, promoting transparency and reducing bias risks in analysis.

3.3.2 Ethical Considerations

Ethical approval for this study was obtained from the Institutional Review Board (IRB) of Indraprastha Institute of Information Technology (IIITD) (IIITD-IRB/FR/0-1/2022/04). Before the interviews, the participants were provided with an overview of the study and verbal and written consent was obtained. Participants were allowed to converse in Hindi, English, or Bengali based on their preference during the interviews.

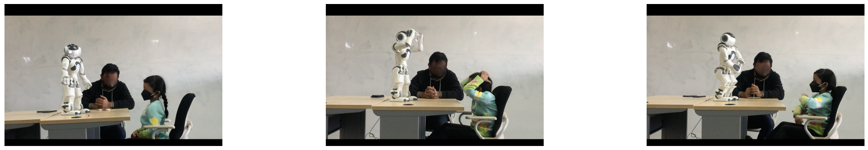
3.3.3 Participants and Recruitment

The study was conducted in urban settings within Delhi and Kolkata, two metropolitan cities in India, from February 2022 to June 2023. The study comprised 25 participants who were licensed special education practitioners (F=22, M =3, Age range (years): 24-58) in these cities, possessing at least five years of experience working with CwA registered with the Rehabilitation Council of India (RCI). RCI is a statutory body that regulates and monitors services given to persons with disability and maintains the centralised register of qualified professionals and personnel in the field of rehabilitation and special education. Informants were contacted and recruited through professional organisations, special schools, and rehabilitation centres using a combination of email solicitation, snowballing methods, and the researchers' contacts. The recruitment did not consider the participants' prior experience with or knowledge of technology or robotics-assisted interventions to mitigate potential bias in the study. However, many participants mentioned routinely incorporating some form of technology into their work. Each of our informants had extensive experience (Mean experience (years): 13.4) working with CwA. Appendix B presents the demographic information of our participants.

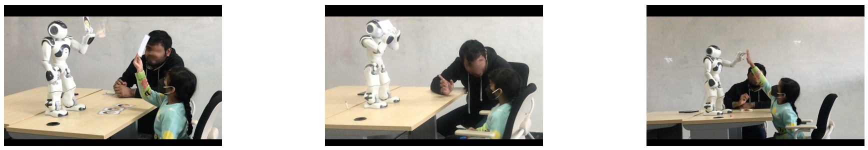
3.3.4 Methodology

Field Trips and Groundwork

Prior to data collection, the researchers conducted extensive field visits to 4 rehabilitation centres in Delhi. At each centre, three researchers conducted several field observations to understand the various activities and strategies that special educators working with CwA regularly use. Extensive notes and schematic sketches were developed to understand and document special educators' activities. Special attention was also paid to understanding how



(a) Joint Attention (b) Vocabulary (c) Imitation



(d) Emotions (e) Anticipation Play (f) Response to Name

Figure 3.1: Snapshots from the video probe depicting different intervention activities

different tools and technologies were used in these activities. Care was taken to ensure our field visits were not disruptive to the children's daily routine. No photographs or videos were taken to ensure the privacy of the children. As elaborated in subsequent portions of the methodology section, these field trips were invaluable for us to familiarise ourselves with the educational activities, rehabilitation strategies, work routines, and the challenges they face in their day-to-day professional lives. The data collected during these field visits was extremely useful for programming our NAO robot deployed to carry out the demo RAT activities, developing our semi-structured interview guide, and setting the workshop agenda.

Study Materials

Video for Elicitation Activity. Video provocation has been previously employed in literature [153, 154] to educate participants about the technology. Building on previous research, we employed video provocation to educate special educators about the capabilities and potential use cases of robots in interventions.

Based on our field observations, interactions with the special educators, and an extensive review of literature, we identified the most common set of activities performed with CwA across the globe and in India. The final set of exercises selected was as follows: RN, emotion identification using card matching, imitation, vocabulary building, and making eye contact using a peek-a-boo game. This set of exercises was widely used by special educators in India. These exercises were derived from a therapeutic approach based on Applied Behaviour Analysis (ABA), commonly used to improve social, communication, and learning skills in individuals with autism [155]. This approach utilises reinforcement techniques to encourage the development of desired behaviours and skills. ABA is an evidence-based and scientifically validated practice that has been demonstrated to be effective in enhancing the abilities of individuals with autism and other developmental disabilities. Taking these activities as a basis, we programmed an interactive social robot agent - the NAO robot, which is a widely used robot in autism research [83, 156, 157] to deliver intervention activities to CwA under the guidance of a special education expert. The robot was programmed to speak in Hindi (a native language widely spoken in North India, where the study took place).

Centred on the study's goal, the primary aim of the video was to emphasise the application of robots in intervention activities rather than highlighting the interactions between robots and the child. Consequently, we trained a seven-year-old non-ASD child to engage with the NAO robot and carry out the specified activities. Subsequently, a video with English subtitles was created of six RAT activities (RN, JA, imitation, vocabulary, emotions, anticipation play) outlined above to inform special educators about the potential use of robots in therapeutic practices. The video showcased both positive and negative use cases of using robots in such settings. One example of a positive use case demonstrated in the video was when the robot successfully completed a task without any errors. On the other hand, a negative use case was showcased where the robot failed to complete an activity due to technical glitches or hardware limitations. The exploratory video was of $\sim$10 minutes in duration.

3.3.5 Procedure

The study protocol incorporated a qualitative research approach, combining video provocation/elicitation with semi-structured interviews, workshops and panel discussions. The study unfolded in three phases.

Phase I: Video Presentation. Participants were contacted at their preferred time, either in person or through Zoom. In the beginning, one researcher briefly explained the scope of the study, shared the study information sheet and obtained the participants' consent. Then, the researcher presented an overview of NAO robot-based intervention activities and presented the exploratory video to them. Exploratory videos were used to elicit

(a)

(b)

(c)

Figure 3.2: Snapshots from workshop sessions

participants' responses and gain insights into their experiences and perspectives. Conti et al. utilised both video and oral presentations in their study to assess the readiness of psychologists to incorporate robots into their professional practice [158]. Previously, to examine the perceptions of Community Health Workers (CHWs) about integrating AI into their workflow and identifying the anticipated benefits and challenges, video provocation was utilised as an exploration artefact by Chinasa et al. [159]. Based on this previous work, in this study, we use video elicitation to inform special educators about the possibilities and challenges of incorporating social robots into their educational interventions for CwA. Snapshots of the video are presented in Figure 3.1.

Phase II: Semi-structured interviews. Once the video presentation had concluded, the informants participated in a semi-structured interview. The interview protocol was structured around four thematic categories: a) background of the educators, b) challenges and concerns faced by educators in facilitating routine autism interventions, c) exploring their perceptions regarding the usability of robots in interventions, and d) understanding perceived trust and meaningfulness of social robots in the Indian context. As recommended by Jacob and Furgerson [160], after each interview, we systematically reviewed and revised the questionnaire, incorporating additional prompts and follow-up questions. This iterative process continued until we reached saturation, ensuring the gathered information was comprehensive. The interview protocol is given in Appendix C. The interviews were conducted in Hindi, English, or Bengali according to each of the participant's language preferences and lasted 1 to 1.5 hours.

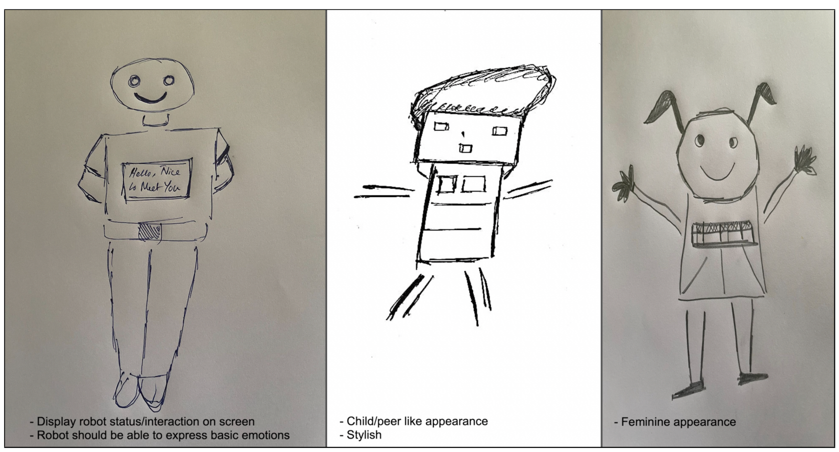


Figure 3.3: Participants' drawings of the social robot with their comments

Phase III: Expanding on the insights acquired during the interviews, the study's third phase included three one-day workshops for educators, providing them hands-on experience with the robot. This was followed by a single panel discussion, which brought together special educators (participants in our previous interviews) and technologists, including developers and researchers specialising in social robotics and AI.

Workshop. In each of these workshops, we facilitated the special educators to interact with the robot. The participants were afforded the opportunity to actively engage with the robot and experiment with the activities demonstrated in the video probe. Participants had the freedom to choose activities based on their preferences, and each participant performed at least four out of the six activities. Participants who did not complete all the activities mentioned reasons such as being technically unfamiliar, while a few found certain activities to be less challenging. During this time, we also answered any of the questions that they had in their mind. Each session had 8-9 participants and lasted between 1 to 1.5 hours. The screenshot from one of our workshop sessions is given in Figure 3.2.

Panel Discussion. Following the workshop, we extended invitations to these participants for a panel discussion to explore the future possibilities of integrating robots into autism intervention. Upon receiving confirmation of participation, the panel discussion was conducted one month after the workshops. It involved 14 special educators from the previous participant pool and four technology experts. The inclusion of panel discussions in the study protocol aimed to develop guidelines that are grounded in the specific needs of special educators and the technological feasibility of implementation. Throughout each session, participants were actively encouraged to share their ideas on how robots could be better built and designed to enhance integration, particularly in resource-constrained settings like India. Furthermore, they were prompted to articulate their perspectives on the role of robots in interventions and were encouraged to present and justify their views in an explainable manner. The sessions also saw many of our special educators provide us with pictorial depictions of social robots as they imagined themselves, as shown in Figure 3.3. It's crucial to acknowledge that design considerations were outside the scope of this study; nonetheless, their visualisations were taken into account to gain insights into their perceived role of robots in interventions. The technologists who formed a part of our panel discussion belonged mostly to research institutions across India and had an average experience of nine years in ideating, building, and maintaining assistive technologies.

As we took field notes from these conversations, every workshop session and panel discussion was audio-visually recorded with prior permission from all the respondents. The recordings from these sessions were transcribed for the purpose of further analysis. For taking part in these three sessions, participants were compensated with a 700 INR ($\sim$ USD 8.4) Amazon gift card.

3.3.6 Data Collection and Analysis

The data included 25 semi-structured interviews ($\sim$50 hours of audio) and $\sim$5 hours of video recordings of the three workshops and panels. Data also included detailed field notes made during the field visits, interviews, workshops, and a panel discussion. The audio recordings of interviews conducted in Hindi, English, and Bengali were translated and transcribed into English for analysis. For coding, the transcripts from interviews, workshops, and panel discussions were initially treated separately and first-level codes such as *"sharing of resources"*, *"robots are more animate than a tablet"*, etc. were extracted based on the emerging patterns in data. Each researcher independently analysed and open-coded the data using thematic analysis employing an inductive, constant comparison method [161]. Following this, we proceeded with the second iteration of open coding, identifying commonalities within the initial sets of the code list from different phases. To ensure accuracy and consistency, the researchers periodically met to compare the codes being generated, resolve any discrepancies, refine them and conceptualise themes to a higher level, such as *"Special educators talk about financial difficulties"*, *"Social robots can be partners in different activities", etc.* We repeated this process until all the interviews were coded, we reached data saturation, and all researchers reached a consensus on the identified themes. The final codebook with derived themes and the frequency of each code are presented in Appendix E. To protect the privacy of our participants, we used pseudonyms and anonymised the quotes in the manuscript.

3.4 Findings

This section is organised as follows. We begin by answering the first research question by highlighting the current socio-technical contexts of special education schools (Section 3.4.1) and the challenges that therapists face when using various technologies in their daily practice (Section 3.4.2). Next, we document Indian special educators' perspectives on the perceived benefits and challenges of using social robots as therapeutic interventions for CwA (Section 3.4.3) and the perceived transformation of their roles in therapy (Section 3.4.4) to answer the second research question. Lastly, to answer the third research question, we explore the initiatives that can be taken to integrate social robots into the Indian special education landscape (Section 3.4.5).

3.4.1 Current Socio-Technical Landscape of Special Education Schools

To understand and situate our work in such realities, we requested our interviewees to articulate the intricacies of their work environments. Our interviews revealed that nearly all the schools we investigated possessed fundamental infrastructural facilities, including essential tools for therapy. While some institutions managed to distribute these devices adequately, many centres had inadequate technological resources and a large number of faculty members or students who used them, thus creating a scarcity. For instance, respondents P12 and P13 highlighted that their school lacked devices such as laptops and tablets due to limited donor funds, which only covered locally manufactured desktops with outdated software, which limited their use.

" Our donors have made it explicitly clear that they do not want to invest all their money in hardware and software. Over time, we requested them for

another PC, but our request was denied. " (P12)

Distribution Hierarchies and Educator's Sentiments. Special education centres often employed creative strategies for addressing such deficits when resources were inadequate. For instance, respondents P1 and P2 highlighted that a rotating system was implemented every week in their school that ensured that nearly all teachers, regardless of their positions, could use technological equipment at least once every two weeks. These workarounds involving the sharing of devices have emerged as a common practice to overcome the lack of adequate resources in the Global South [162]. In addition, some participants also pointed out that factors like individual therapists' proximity to the centre's head and the attrition rate of children in their groups often influenced their access to and allocation of technological resources.

"Despite the policy for rotational equipment use, it often appears effective only on paper, influenced by proximity to the head and children's attrition affecting favour. High attrition is seen as the educator's inability, with no explanations permitted." (P1)

Educators, Conflicts and Emotional Impact. The challenges stemming from insufficient technological resources seemed to result in various conflicts among educators, resulting in a personal sense of dissatisfaction. For instance, respondents like P23, P24, and P25 conveyed their frustration around the scarcity of resources, noting that it occasionally led them to question their professional purpose and triggered feelings of being "unwanted" and "undesired." In many well-funded schools, we also observed that the special education classrooms were segregated into distinct areas within the regular institutions, amplifying these feelings. As highlighted by several respondents, this sense of othering was especially pronounced when they were physically segregated and denied access to basic amenities like internet connectivity, thus necessitating them to look for alternative arrangements.

"The setup appears odd to me. Our centre is part of the school, which caters to both neurotypical and neurodivergent children. I don't understand the need for a separate building outside the main premises. Teachers from there don't engage with us or share resources. We even have our own internet connection." (P23)

3.4.2 Understanding Educator's Needs and Challenges with Technology-Assisted Interventions for Autism

The therapists we interviewed employed a range of technological interventions for therapeutic purposes. These included Android applications such as *Avaz* and *Talk With Me*[1] to enhance speech, specifically curated YouTube videos to bolster social skills and vocabulary acquisition, and the integration of electronic toys and hand puppets to help with conversational interactions. Additionally, confident special educators utilised computer games installed on their centre's PCs to facilitate the learning of everyday activities. On the one hand, several therapists appreciated the value these technologies bring

[1]https://www.talkwithmeapp.com/

to their therapeutic practice in terms of enhanced learning outcomes. At the same time, they also expressed several apprehensions, as outlined below.

Expensive Machines and Educators' Frustrations. Several interviewees (n=13) revealed that using expensive devices like tablets and PCs was always risky and made them anxious. One special educator, P23, informed us that often, some centres and schools were "ruthless" and asked the educators to compensate for any damage done to the devices. P13, P18, and P23 acknowledged that such a policy fostered a perception that the institution lacked complete trust in their educators and prioritised the economic value of the devices over their expertise in delivering services using such devices.

"When I have to constantly worry about the chances of a portion of my earnings being deducted whenever I use a device, my frustration is directed more towards the device itself than the centre's trust in my responsibility. After all, it's all because the device is expensive, right? " (P13)

Challenges in Therapy Applications. Our interviewees had mixed opinions about using different therapy software. Some were concerned about the language options in popular apps like Avaz, which only offered seven Indian languages. Some interviewees felt this to be a lack of inclusiveness that showed a disregard for cultural and linguistic diversity across India. Some participants disagreed when discussing the feasibility of incorporating a large number of languages to be slightly difficult, suggesting that such apps were constructed without considering inclusiveness. At least three participants complained that such applications needed very powerful mobile devices with high storage capacities, and their inability to own such devices prevented them from using these Apps. Two other participants who worked independently providing door-to-door services found that subscribing to these applications was too expensive, and owning a mobile phone that would support such apps would add more cost to their practice.

"I have used the app called Avaz. I hear a lot of people using it these days. I have used other software as well. The problem with these technologies is that they lack in multilingual setups." (P7)

Technologies for Alleviating Fatigue in Autism Therapy. Reflecting on their professional challenges, our participants echoed narratives underscoring the pervasive issue of fatigue and emotional drain among educators. The repetitive nature of autism therapy tasks, coupled with extended periods of managing children with special needs, is noted as particularly demanding and emotionally taxing [148, 163]. Given these factors, our respondents expressed that they would find it helpful if technological interventions could be developed to share their work burden and alleviate their stress and fatigue.

"We do the same tasks every day, and it makes us exhausted. It would be good to have technologies that can share this exhaustion." (P4)

Needs, Requirements, and a Call for Collaboration. A common consensus that emerged among the experts we interviewed indicated that existing technology-mediated methods were falling short of meeting educators' needs across different contexts and circumstances. Particularly concerning software applications, they noted that these tools appeared to prioritise "the needs of the affected children" without adequately considering the operational capabilities and needs of the experts themselves. Nevertheless, despite facing several challenges, none of our participants were reluctant to use technological tools and applications for therapeutic interventions. Many special educators stressed the importance of collaborative tool development rather than merely critiquing technology for its inherent limitations. This sentiment was shared by experts across the spectrum, including both independent therapy service providers and those who work at institutions.

3.4.3 Educator's Perspectives on Benefits and Challenge of Robot-Assisted Therapy

In this section, we elaborate on the sentiments expressed by our participants about integrating social robots in their therapeutic practices and establish whether the latter would be a meaningful addition to the existing technological interventions.

Following the video demonstration, our respondents were prompted to share their opinions about the robot's engagement in diverse therapeutic activities. Many of our participants appreciated that social robots could replicate many therapeutic exercises they conducted regularly. Sixteen of the twenty-five therapists we interviewed confirmed that such technologies "defined the future of autism therapy." A few others remained "optimistic" about seeing if future robotic technological developments will address the specific needs of special educators, as this would increase the uptake and use of social robots in therapeutic practice.

“I am quite fascinated with how it works and understand everything. I think it is going to be really useful for the children as well as for myself. Unlike a tablet or a computer, this seems quite animate and should do well with the kids.” (P21)

Anthropomorphism and Perceived Benefits in Therapeutic Practice.

The anthropomorphic characteristics of the social robot captured the attention of almost all the participants. Some of them even projected their ideas using their drawings during our workshop sessions, as depicted in Figure 3.2. A few even pointed out that using social robots in therapy would be similar to their existing use of sock puppets that mimicked human-like features. Mirroring the participants’ views in Duffy’s study [164], several of our participants also agreed that incorporating human-like attributes would make working with robots easier than with other software tools or digital devices. Without exception, all the participants highlighted that the presence of humanoid traits made them perceive the robot as a "partner" capable of "walking and talking" and "interacting" like humans. Research also shows that the anthropomorphic attributes of social robots influence people to perceive them to be more sociable [165], leading to increased likability [166]. Some special educators, like P9 and P10, said that programming the robot to talk in Hindi and not English was a good step and an important one in countries like India, where many people don’t speak English. With the robots being able to speak Hindi, the educators felt that they would be "at ease" and "socially connected" while using the technology, which was crucial for the therapy process. Perceiving social robots to be similar to a helping hand, special educators also explained that participating in therapeutic activities with a robot by their side would make them more "confident" and feel more "empowered" when working with the children.

"Working with a Hindi-speaking robot would make me feel very proud and comfortable. I would know that I at least have an assistant with me who can perform some of my tasks if I am too fatigued. I would know I am not alone in this. Plus, I would always prefer a robot which looks more like a human." (P9)

Robots as the Non-Judgmental Aids. After learning that special educators contemplated utilising social robots as therapy partners and assistants, we proceeded to delve deeper into their rationale for envisioning social robots as work partners. We were especially interested in understanding if only their humanoid characteristics bolstered educators' confidence or if other factors influenced their decision. In response to this question, our participants pointed out that it did not matter if the robots were capable of actual human emotions or, for that matter, whether they had any sense of camaraderie towards their users. Instead, what mattered to them was a human-like entity that acted as a support system and could help in some of the exercises and activities consistently without fatigue and without exhibiting any prejudice or discriminatory behaviour towards the patients or causing physical harm to them.

"When called for assistance, our aides would sometimes think of us as incapable of working independently. I think a robot like this serves my purpose and, at the same time, will not judge me for using it as a helping hand." (P24)

Regulating Robot Engagement in Therapy. While acknowledging the importance of robot-assisted interventions for ASD patients, our interviewees also stressed the importance of finding the right balance between the level of human and robotic engagement in providing therapy. In this direction, several special educators pointed out that one way to achieve this was to deploy social robots customised to support specific repetitive tasks in their therapeutic routines under their supervision. Several participants saw this ability of social robots as potentially useful tools to reduce their workload and alleviate mental stress and fatigue, in concert with findings from previous studies [167].

"We get psychologically drained after our consultations. With this robot, I think I could escape from some of that. I could ask my assistants then to conduct some of the sessions for which the robot could be helpful." (P6)

Robots Could Lead to "Over-Professionalisation." Senior experts in our study expressed the need for educators in India to have a strong moral sense while working with social robots. This is in concert with previous studies in human-robot interaction (HRI) exploring ethical and accountability dimensions where researchers have highlighted the need for educators to be responsible towards the appropriate use of social robots [168]. These experts speculated that introducing social robots, like tablets and smartphones, might lead special educators to prioritise enhancing their professional identity and marketability over using the technology to benefit their patients. Therapists in our study also voiced concerns about the potential impact of robotic interventions on their capacity for empathy toward the children they work with. They were worried that over-reliance on social robots could lead to disregarding social and cultural issues faced by CwA such as stigma [169, 170].

"Computers initially gained popularity more for trendiness than therapy. Although not all educators were swayed, parental demand sometimes fueled market-driven approaches rather than thoughtful planning. Introducing robots could likely elevate professionalism for those who can afford them, potentially skewing therapy towards economic interests." (P9)

Concern for Damages and the Fear of Technical Complexity. In line with their anxiety about dealing with expensive digital devices, as elaborated in section 3.4.1, concerns among educators were also related to affordability. Explaining the rationale behind their apprehension towards social robots, participants like P1 and P2 articulated that, similar to other devices, their schools could ask them to pay for any damages inflicted on the robot. The educators were also concerned about the technical complexities of handling a social robot. They asked us if the therapy exercises could be controlled through an easy-to-use mobile application as per the needs of a child.

"I do not doubt the capabilities, just that using it would be risky if the kids throw it away or damage it. I might be asked to pay for the damages." (P2)

3.4.4 AI, Robots and Role Transformation: Fear of Replacement and Strategies for Collaboration

In this section, we discuss some of the anxieties and fears expressed by our participants to highlight the necessity of user perception in human-AI collaboration [171], with the hopes of contributing to developing AI-based technologies with a human-centred and responsible approach.

Fear of AI for Personal Reasons. On the same lines, concerns about potential job loss due to replacement by social robots emerged as a common theme in several of our interviews and panel discussions. In response, we explored the reasons behind these reservations to alleviate their fears and to draw insights and implications for the design community to address this issue from a practice perspective.

"As I am seeing technologies like ChatGPT become more and more a part of everyday life, I fear that, soon, robots like these could make us redundant. I was discussing this the earlier day with a colleague of mine, and even she felt the same way." (P10)

Our interviews revealed that much of the fears and anxieties of our participants were rooted in their in-direct experiences of seeing family members being laid off due to the introduction of computers, being exposed to speculative technology news reports, hearing stories of friends and distant relatives struggling to adapt to technological systems in industrial settings and personal experiences of losing occupational status with the introduction of technology-enabled therapeutic routines. Several respondents perceived robotic technologies as a force that only " devalued their work" or rendered their roles unnecessary.

"We suffered a lot as a family when both my parents lost their jobs. Their job was mostly administrative pen and paperwork. The problem was with the introduction of computers. My parents did not know anything about them, and the company office thought they were not needed anymore. That left me completely traumatised." (P11)

Taking a slightly different view, while admitting to her family member losing his job due to automation, participant P19 nevertheless advocated for using sophisticated technologies like robots in a controlled and collaborative manner.

"I like this robot; it's not like I don't like it. I am a bit sceptical. I have witnessed my father lose his job in the factory because of automation. (P16)

Collaboration for The Greater Good. Though ideas for collaborating with a social robot ranged across different tasks, almost all the therapists agreed that robots are highly beneficial in conducting therapeutic exercises that involve repetitive tasks designed to improve joint attention and imitation. They reasoned that it would make the process more enticing and appealing to CwA and reduce fatigue associated with repeated tasks for practitioners.

"I think for joint attention or imitation exercises, I can work out the robot to ensure better child engagement. Autistic children usually like these things and find them to be very attractive. I imagine in such cases, I and the robot can both do the exercise in unison and then the child can follow. I could maybe do a couple of jumps with my hand, and the robot can do the rest." (P7)

Further, many participants suggested that robots could perform a part of a particular task by themselves and let the educator complete the rest in a collaborative, iterative manner. These ideas we gathered from our participants were consistent with those documented in other related studies [172, 173]. Another critical area in which our participants found robots potentially useful is training special educators. Participants P9, P16, and P22, who owned their special education schools, proposed that social robots could be deployed to help young professionals practice therapy activities before they start working with CwA. They even enquired from us if the technology could be developed in such a way that it could assess an educator's capabilities before the individual is assigned to a working group.

"I think apart from doing the tasks, I could even make this work for teaching the educators themselves and letting them practise with it before they work with the children. Some educators have difficulties in working with children immediately after training, so this might help them in acclimatising with the

environment of therapy." (P5)

Two highly experienced experts discussed the possibility of using social robots as "positive behaviour influences" for CwA. Their emphasis was on robots being pivotal in cultivating socially acceptable behaviours, particularly when children resisted guidance from special educators. More importantly, these experts firmly believed that robots could foster a constructive experience for children and educators during therapy sessions. Lastly, drawing on their experiences, these senior practitioners gently dismissed the fears of facing obsolescence. They asserted that the role of special educators might transform in the years to come, fueled by the advent of robots. Rather than harbouring fears about being replaced, they advocated a proactive approach to harnessing technology and exploring innovative collaboration avenues to co-opt social robots in their practice.

"I believe the special education profession would have a different purpose and meaning with the rise of technology these days. I do not think it is going to wither away. But at the same time, we need to adapt to technology instead of simply postponing its use. Personally, I would advocate the use of social robot systems, but at the same time, our professional community has to work together to make it a prudent addition to their work." (P16)

3.4.5 On Trust, Belief and Reliability: Negotiating Meaningfulness of Social Robots in India

Previous research shows that understanding how users perceive technology in terms of trustworthiness and user-friendliness can be crucial to its successful implementation and use [174]. As social robots are relatively new in India compared to other assistive technologies, we agree with Kok and Soh's submission that it is crucial to investigate perceptions about their safety and reliability directly from their users. The section below documents our participants' conceptualisation of such crucial aspects to expand the scope of the broader HCI work revolving around user perception surrounding trustworthiness [175, 176] and making social robots meaningful [177] and appropriate for a professional setting [178].

Understanding Trust, Inequality, and Fragmentation. In our study, many special educators acknowledged the potential of robots as beneficial additions to their work. However, they also expressed concerns about their appropriateness, citing India's economic conditions and social inequalities. Such ambivalence in opinions has been reported in other studies as well [179, 180]. Some educators, like P12 and P18, expressed that robots might worsen existing differences, contributing to their hesitation despite recognising the technology's potential. Our findings align with previous work reporting similar attitudes that were due to factors such as anticipated issues of trust in robots [181], reluctance to embrace new technologies due to a negative bias [182] and a belief that technologies can be harmful to the society [183] emerging from personal first-hand experiences.

"There is already so much inequality in our country. I feel with this robot, those differences might get amplified." (P19)

Twenty out of the twenty-five interviewees highlighted two trust-related issues. First, they hesitated to accept social robots as trustworthy as they believed that it might cause them to be undervalued in a society where they were already marginalised despite their professional acumen. Secondly, recognising the differences within their fraternity, they believed that social robots could cause further divisions and may lead to financially well-off special educators looking down upon those who provided “door-to-door services” who were sometimes labelled as "not really professionals" or "pseudo-experts."

"The introduction of a robot could amplify existing inequalities in our profession. Those with access to robots could enhance efficiency, income, and prestige, while others struggle to earn a living. In India, our professional community is polarised, and I can’t fully endorse robots considering these dynamics." (P20)

Interestingly, some of the educators we spoke to linked the inequities ingrained in their profession to the broader social divisions in the country. They informed us that they preferred using "simple" tools without extravagant features since they believed it could potentially worsen societal and professional divides. One respondent, P11, particularly attuned to prevalent discrimination based on factors like caste and class, expressed reluctance to adopt technology that might accentuate these differences. Echoing these concerns, our participants suggested restrictions on the long-term use of robots in special education.

"I think, in a country like India where inequality is everywhere, I cannot knowingly let a machine disrupt the already fractured environment of my profession. If that happens, then the entire community will have to suffer, and the children will suffer more." (P22)

Support for Robots as The Future of Therapy. Educators supporting the assimilation of social robots into therapeutic interventions presented different reasons. In our study, participants P8, P11, and P13 underscored humanoid features and demonstrative abilities of social robots as crucial indicators of their positive outlook. They emphasised that the robot's impact on the child was pivotal in the Indian context, assessing the demonstrated features' appropriateness and utility. Participants like P10 and P16 highlighted that futuristic technologies like robotics were inherently trustworthy, seeing their success in other countries for RAT. They also embraced innovation despite poverty and inequality, foreseeing long-term benefits. They drew parallels to how computers and mobile phones, once expensive novelties, became everyday essentials, thus projecting a similar trajectory for robots.

"We are standing at such a time when India is growing as an economy, and of course, like every developing nation, we have our own problems. But that does not mean I will not support something for the children. As a pro-

fessional, I am not afraid to say I fully support social robots." (P8)

Training Special Educators for Using Social Robots. The two veteran special educators, P16 and P22, expressed that while robots might find utility and relevance in Indian settings, concerns about the use of such technologies arise with regard to their proper knowledge of handling and operation among professionals. Noting that technological interventions in autism therapy remain insufficiently accessible in rural and semi-urban areas due to a lack of proficient professionals, they emphasised the potential for mismanagement of these interventions in these contexts due to inadequate training and the risk of mishandling expensive devices. Hence, these experts called for adopting appropriate training programs advocating for government and civil society collaboration to ensure effective technological interventions among marginalised communities.

"I wholeheartedly support the use of social robots. From my experience, it would be my opinion that the government should work with NGOs and scientists to explore the options of making these cost-effective and widely available with sufficient training resources." (P16)

3.5 Discussion

Despite increased efforts to develop social robots in resource-constrained settings like India, limited progress has been made in understanding the needs and perceptions of end-users regarding robot-assisted interventions, with only a few studies addressing these aspects [184, 185]. Earlier implementations of technologies in low-resource clinical settings have demonstrated failures, resulting in added inefficiencies to clinical workflows and, at worst, the harm inflicted upon the communities they aim to benefit [186–188] necessitating proactive examination before deployment [153]. This is particularly crucial for addressing the vulnerability and marginalisation experienced by professionals in resource-constrained settings in the Global South, as observed in previous studies [189].

In our research, we aim to ensure social equity and justice in perception-oriented investigations of HRI, contributing to the glaring absence of scholarship in HRI literature. Organising our findings, the discussion section is structured into three segments: a) highlighting implications for designing social robotics in resource-constrained settings, b) providing guidelines for preparing educators in robot-assisted interventions, and c) suggesting institutional strategies for seamless integration of social robots in autism intervention.

3.5.1 Design and Development of Social Robotics for Resource Constrained Settings

In our endeavour to devise a robot-assisted therapeutic system tailored for resource-limited communities, the study underscores the need for a cautious approach in deploying robot-assisted therapeutic systems, emphasising educators' pivotal role in decision-making during therapy. The suggested strategies involve creating culturally suitable and lightweight activity modules, ensuring multilingual capabilities, endorsing open-source moderation, prioritising the improvement of existing applications rather than introducing entirely new systems for smooth integration into educators' routines and employing participatory design to ensure contextual appropriateness.

What should developers keep in mind while creating social robots?

Our findings underscore educators' preference for a clearly defined *strategic role* in robot-assisted interventions, emphasising the enhancement of their efforts in the therapy process. To align with this preference, developers, as suggested by Elbeleidy et al. [190], can focus on creating robots with reduced autonomy, fostering improved collaboration between humans and robots. To further address educators' insecurities, we propose involving them in the decision-making process during therapy, promoting confidence and mitigating potential complications in the event of technical failures. We also find it crucial to present the social robot as a peer or role model for children [185] as well as a non-judgemental aid, prioritising cost-effectiveness and robust build quality. To this end, we advocate for industry engagement in producing low-cost, durable devices that ensure longevity. Such an approach not only addresses concerns but also reduces fear, fostering the acceptance of advanced technologies, especially within marginalised communities, in a confident and approachable manner [191].

In light of our findings, prioritising cost reduction while preserving the social robot's functionality emerges as crucial. *Accessibility*, particularly for educators managing RAT, becomes a focal point [192]. Prior studies in the Global South have explored initiatives like using mobile phones to enhance online accessibility [193], reduce poverty [194], and provide free basic Internet connectivity [195]. Given the prevalent ubiquity of smartphones in marginalised communities, the limited digital literacy among educators, and storage challenges reported by our participants, a lightweight mobile application featuring an intuitive interface [196] stands out as a potential solution. Such an application could facilitate the curation of therapy exercises without necessitating direct robot reprogramming. Furthermore, we underscore the significance of integrating robotic intervention seamlessly into existing practices. This strategy avoids the imposition of an entirely new intervention approach tailored exclusively for robots. By adopting this approach, the transition from current tools to the incorporation of social robots into daily practice becomes smooth and natural.

Drawing from existing literature [197] and the input of our educators, we recognise the *diversity* in the needs of children on the spectrum. We thus advocate for flexibility in designing activity modules [198, 199], accommodating unique requirements and allowing for future improvements. Additionally, these applications should enable educators to customise and personalise therapeutic experiences for each child, aligning with fundamental autism therapy guidelines. Inspired by Barba, we also recommend making activity modules open-source by the companies developing social robots in order to promote transparency and reproducibility and allow for its iteration by educational institutions and non-profit research labs [200].

Educators also stressed the importance of ensuring *cultural appropriateness* in the robot's activity modules, e.g. such as performing positive reinforcement through clapping of hands and not by giving a flying kiss. Recognising the significance of aligning technology implementation with cultural understanding, as emphasised by Gross et al. [201], we follow Pal et al.'s approach to assistive technology in emerging regions and propose a collaborative design process involving professionals such as computer scientists, engineers, cultural theorists, interaction designers from various geographical contexts, and caregiving professionals in diverse regions [202].

Our educators also expressed the need for robots to communicate in vernacular languages beyond English. It's therefore recommended that these robots possess *multi-linguistic capabilities*, considering the diverse linguistic landscape in countries like India and the Global South, with the aim to include non-prominent vernacular languages as well in the future. Additionally, based on the perspective of some of our participants, we suggest that new applications for robot-assisted interventions should enhance existing therapy practices to facilitate seamless integration into educators' routines.

Participatory Design for Social Robotics: An Appropriate Answer for Contextual Appropriateness?

Based on the outlined recommendations, we advocate for a participatory design (PD) approach to responsibly develop and deploy social robots in resource-constrained environments. PD, as noted in prior research, offers a safe space for exploration and experimentation before technology deployment in marginalised communities, especially in the Global South [153]. Additionally, PD has demonstrated effectiveness in HRI, contributing to the creation of robots for depression management and mood stabilisation [203]. It also plays a significant role in ecosystem mapping for differently-abled populations, supporting democratic and respectful technology development [143, 204]. Considering the varied mechanisms of PD [205], our suggestion is anchored in fostering a dynamic power balance between users and designers, encouraging critical dialogue for user empowerment and collective decision-making [206]. In crafting social robots for environments like India, our proposed PD approach, with a focus on ethnography, must consider the socioeconomic backgrounds and everyday experiences of special educators. Emphasising the importance of acknowledging disparities in opportunities and resource awareness, inspired by Toyama [207], the design process should incorporate the voices of all educators, including those marginalised within the professional community. A more respectful and collaborative approach could nurture meaningful partnerships and contribute to deploying well-designed social robots for resource-constrained communities in the Global South.

3.5.2 Beyond Robotic Realities: Preparing the Educators for Robot-Assisted Interventions

All our participants stressed the need to prepare special educators for robot-assisted interventions. In light of the perceived technical complexity of social robots, we were given the understanding that their widespread adoption in marginalised communities could be a tough affair without proper training. Researchers in HCI have consistently recognised the crucial influence of *training and support initiatives* on the effectiveness of technology interventions [186, 188]. Typically, these training programs concentrate on instructing users on utilising new technology [188]. In our case, we turn to Irani et al.'s inspirational work on postcolonial computing [208] to put forward our recommendation about first identifying the level of digital literacy and understanding of social robots among special educators and then leveraging their willingness to use technological tools for the overarching benefit of the children. Following this, educators would need training to develop technical competence and proficiency in operating social robot systems. Simultaneously, they would need to engage in critical thinking processes to foster a balanced understanding of social robot systems, recognising both their strengths and weaknesses. This would be crucial towards making educators understand their significance in the therapeutic process despite using AI-enabled agents like social robots. It would also help in making them realise how AI would not be able to replace them entirely or not make their roles redundant [209].

Additionally, educators should be prepared for the risks and potential errors of using social robots in special educational settings. This preparedness could encourage adoption and boost confidence among them.

When addressing the education and training of special educators on assistive robots, it's crucial to recognise educators' *hesitancy towards adopting new technology*, stemming from their comfort with existing tools. Therefore, training efforts should prioritise educating educators about the potential benefits of social robots, providing insights into their strengths and weaknesses, and recognising and leveraging their strong commitment to the overarching goal of benefiting children. Additionally, during the training process, educators should be informed about ethical and privacy-preserving practices related to technologies like social robots, requiring programs to navigate diverse cultural perspectives and values within the context of privacy [153].

Additionally, drawing from the works of Uzorka et al. on professional development [210], we suggest that by actively involving educators and care workers as *participants in scientific projects* encompassing assistive technology, the role of robots as tools for empowerment rather than a burden could be strongly established in their minds. For the effective execution of training and exposure programs, it is imperative for academic institutions, research laboratories, and well-resourced researchers to proactively participate in outreach activities and capacity-building initiatives. This involvement includes collaborative efforts with various organisations such as special education schools, non-governmental organisations dedicated to differently-abled children, and publicly funded schools catering to children with special needs. This collaboration can take the form of workshops and joint seminars. In particular, there should be a concerted effort to encourage young researchers under supervision to take the lead in community-driven endeavours, fostering the dissemination of scientific knowledge across all echelons of society.

3.5.3 On the Future of RAT in India: Unpacking Strategies for the Adoption of Social Robots in Low-Resource Settings

Our final suggestion rests on the future of RAT in low-resource settings by envisioning a pathway for the adoption of Socially-assistive Robots (SAR) over a period of time. To that effect, we feel that there are three important institutions, namely government agencies, civil society organisations and special education schools themselves, that have to contribute and commit in different ways.

As discussed in existing literature, technology interventions from non-domestic sources for accessibility are often *costly* due to high procurement expenses, including taxes and import duties [138]. The United Nations Convention on the Rights of Persons with Disabilities [2] (UN-CRPD) forms a critical foundation for promoting low-cost Assistive Technology (AT) in developing nations by mandating signatory countries to ensure accessible conditions for their citizens. Based on this principle, we, therefore, agree with Pal et al. in suggesting the development of assistive technological interventions

[2]https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-Persons-with-disabilities.html

to materialise cost-effective solutions for resource-constrained communities [138, 202].

The Government of India has already implemented schemes such as *Make-in-India*[3] and *Digital India*[4], along with grants from the Science and Engineering Research Board[5] and the Department of Science and Technology[6], to promote domestically produced technology products. Under these projects, there have been initiatives that have materialised into the development of indigenous robots like *Vyommitra*, *Daksha*, and *Manav.* A significant development in the area of robotics in India has been the establishment of a *National Robotics Mission*[7] and the *Indian National Mission on Interdisciplinary Cyber-Physical Systems,* [8] specifically with centres for Robotics and Cobotics, emphasising indigenous technology creation and transfer. The proposal outlines a replicable model for other Global South countries, offering strategic support through institutional backing, tax exemptions, and facilitator grants to promote innovation and technology development.

Despite the program's emphasis on innovation and technology development, there remains a gap in addressing *public awareness* about technological advances and *educating individuals on technology building and innovations.* To address these gaps, we propose that the *National Education Policy*

[3] https://www.makeinindia.com/
[4] https://digitalindia.gov.in/
[5] https://serb.gov.in/
[6] https://dst.gov.in/
[7] https://static.mygov.in/innovateindia/2023/09/04/mygov-1000000000643330490.pdf
[8] https://nmicps.in/about-us

(NEP)[9] should prioritise Science, Technology, Engineering, and Mathematics (STEM) education, fostering a skilled workforce for technological advancements and creating an educational environment that encourages students to contribute to progress in diverse technological domains, especially in social robotics. Simultaneously, programs like India's *National Mental Health Programme* [10] that currently lacks avenues for technology inclusion require to be revamped to the integration of technology-enabled mental health interventions by incorporating technology education into their training and awareness programs. By providing comprehensive training programs, educators can overcome apprehensions and misconceptions, fostering a more informed and confident approach to incorporating these technologies into the learning environment.

Our findings further point towards the need to infuse *accountability* at the institutional level within special education schools and organisations. In this regard, we suggest special education institutions should take charge of the maintenance and preservation of social robots, assuming responsibility for their upkeep and, bearing the costs associated with accidental breakdowns, partnering with civil society organisations. Simultaneously, financially well-off schools can collaborate with civil society organisations to prioritise training and awareness for educators, fostering a collective understanding of social robot usage. We suggest that emphasising institutional accountability and

[9] `https://www.education.gov.in/national-education-policy`
[10] `https://nhm.gov.in/index1.php?lang=1&level=2&sublinkid=1043&lid=359`

responsible robotics development ensures a sustainable framework for adopting and utilising social robots, empowering special education professionals simultaneously.

We further recommend establishing *appropriate policies and ethical guidelines* to regulate robot-assisted interventions for diagnostic and therapeutic purposes. In contrast to the Global North, where guidelines exist for the safe and ethical use of such technologies [211], countries like India seem to lack such frameworks. While there have been initiatives in framing policies governing AI in healthcare [11], we strongly advocate developing similar frameworks to ensure the safe and responsible use of social robots in the Global South.

In summary, our multifaceted strategy envisions indigenous development, robust government support, targeted training, affordable pricing models, and collaborative efforts at societal and institutional levels. In materialising them, some of the challenges in the effective and responsible integration of social robots into the clinical settings of marginalised communities like India could be adequately addressed.

3.5.4 Limitation and Future Works

While our research has provided insights into technology adoption and HRI in the Global South, it is not without its limitations. Primarily influenced by perspectives from professionals in Indian metropolitan cities, our study lacks

[11] https://main.icmr.nic.in/sites/default/files/upload_documents/Ethical_Guidelines_\AI_Healthcare_2023.pdf

representation from care workers or special educators in rural and semi-urban regions. This limitation is particularly noteworthy given India's significant rural population. In future endeavours, we commit to addressing this gap by actively involving participants from non-urban areas.

Furthermore, our study provides insights from 25 special education professionals only, who were initiated for recruitment immediately after the pandemic. Initial hesitancy arose due to perceived time constraints, with educators expressing reservations influenced by past mistreatment by researchers. We anticipate that our set of recommendations will guide researchers to be more considerate of special educators and other respondents, fostering greater participation in future studies.

In addition, our current study centres on the viewpoints of special education professionals, recognising the necessity of parental consent for interventions like social robots in India. While valuable, we acknowledge the need to broaden our scope in future studies to include parents and other stakeholders. We encourage researchers in the Global South to adopt a comprehensive approach involving children, parents, and the larger community in their investigations. Such inclusivity is crucial, especially when the values of researchers, special educators, and communities may diverge at times, as seen in safeguarding personal health data privacy [153].

The integration of social robots in healthcare raises ethical concerns and is currently being discussed, particularly within the HRIcommunity. In low-resource settings like India, there's a notable absence of regulatory requirements for AI systems. As social robots address societal challenges, especially in healthcare in resource-constrained settings, there's an urgent need for regulatory frameworks. These should prioritise ethics, safety, and privacy, emphasising diverse values within a particular context. Our upcoming research aims to explore the ethical dimensions and tensions in real-world deployment, validating findings in healthcare facilities. A long-term study in clinics will further assess the effectiveness of recommendations, focusing on patient care and ethical considerations.

3.6 Conclusion and Future Work

In our qualitative exploration, we explored how Indian special educators perceive the integration of social robots in autism intervention. Employing a mixed-methods approach, including interviews, workshops, and a panel discussion with 25 educators, our goal is to uncover the challenges and opportunities associated with adopting social robots in autism intervention practices. The insights reveal significant concerns, particularly the urgency to democratise social robotics and AI in resource-constrained settings. Special educators, particularly in India, express reservations about the functional capabilities of these technologies, fearing potential redundancy. Despite initial scepticism, professionals suggest various ways to integrate social robots into their work, emphasising the importance of technological innovation in reshaping and enhancing their roles. Based on our findings, we discuss the implications of designing social robotics for resource-constrained settings, offering practical guidelines for preparing educators for robot-assisted interventions and proposing institutional strategies for seamless integration. While our insights may extend beyond India to similar low-resource settings dealing with autism, further research is needed to assess the generalisability of our findings. Additionally, understanding the perspectives of caregivers/parents, integral to the Indian intervention system, remains a crucial avenue for future exploration.

In this chapter, we explored the perceptions of special educators, who play a crucial role in autism care, regarding the utilization of social robots for interventions in ASD. Our findings highlight the positive attitudes of special educators towards integrating social robots as aids despite acknowledging the challenges posed by such expensive and new technologies. Chapters 2 and 3 laid the foundation and motivation for the development of an AI-assisted system for ASD diagnosis. Building upon this foundation, we proceeded to the next phase, which investigates AI-assisted multi-modal behaviour analysis, especially in the area of speech and facial expressions, to aid in capturing behavioural traits associated with ASD diagnosis.

Phase II

Behavioural Analysis

CHAPTER 4

FACIAL EXPRESSION-BASED EMOTION RECOGNITION IN CHILDREN WITH ASD

4.1 Introduction

Our investigations into the needs, challenges, and perceptions of special educators in autism care, as well as our understanding of how children of Indian ethnicity respond to robot-assisted interactions in previous chapters (Chapter 2 and Chapter 3), have laid the groundwork for the development of an AI-assisted system with a robot mediator for diagnostic interventions for autism in India. With this background, we proceeded to develop AI-assisted behaviour analysis modules aimed at detecting and analysing facial expressions and speech behaviours in CwA for assessment and diagnosis purposes. In this chapter, we focus on the identification of FER in children of Indian ethnicity, with subsequent exploration of FER in CwA.

Individuals with autism frequently encounter challenges in both understanding and expressing emotions, often manifested through distinct facial expressions. The social communication difficulties observed in this population are, to some extent, linked to deficits in emotion recognition, underscoring the challenge in comprehending and interpreting socio-emotional cues. Notably, impairments in emotion recognition and expression have been consistently proposed as fundamental criteria in autism diagnosis. Standard diagnostic tools such as the ADOS and INDT-ASD incorporate an Emotion Task designed to evaluate an individual's capacity to recognise and appropriately respond to emotions within a social context.

In pursuit of enhanced early detection and accurate diagnosis of autism, automated systems employing FER algorithms have been investigated. These systems aim to analyse facial expressions during diagnostic tasks, providing valuable insights into the emotional expressions of children undergoing assessment. Furthermore, the application of such technology extends beyond diagnosis, offering potential interventions. These systems can assist children in developing emotion recognition skills and contribute to establishing an affective loop in child–facilitator interactions. Leveraging FER technology holds promise in both diagnostic and therapeutic realms within the context of autism. In this chapter, we explore the existing literature related to FER, especially in children, identify gaps in existing research and present our proposed approach for recognising emotions from facial expressions, specifically addressing scenarios where data availability is limited.

4.2 Related Works

Deep learning techniques have proven to be effective in solving the FER problem. However, it demands a significant amount of supervision data, which is often unavailable due to privacy and ethical concerns. To overcome data scarcity challenges, recent research has employed transfer learning techniques, enabling the transfer of knowledge from one domain to another.

4.2.1 Transfer Learning in Facial Emotion Recognition

FER has witnessed a breakthrough with the advent of deep learning techniques, which eliminated the tedious pre-processing phase and provided end-to-end solutions from the input visual information to the emotion recognition. An end-to-end learning framework based on a deep region and multi-label was proposed for the detection of facial action units in [212]. Another approach shows that combining multiple networks shows better performance in automatic FER [213]. [214] introduced a method resilient to expression intensity variations through the learning of spatiotemporal feature representations in FER. Additionally, studies indicate that pre-processing images before inputting them into deep neural networks enhances classifier performance [215]. Convolutional neural network with attention mechanism (ACNN) has been shown efficient in perceiving the occlusion regions of the face and has been used to recognise facial emotions in the wild in the presence of occlusions [216].

To mitigate the need for extensive training data, transfer learning techniques have been proposed. For instance, [217] proposed an ensemble of industry-level face recognition networks pre-trained on large facial emotion databases, such as FER2013. Aly *et al.* [218] introduced a multi-stage Progressive Transfer Learning method, fine-tuning the AlexNet convolutional network and demonstrating FER performance on VT-KFER and 300W datasets.

Another approach by [219] utilises transfer learning with the SE-Resnet-50 model pre-trained on the VGG-Face2 database, incorporating a novel cluster loss function to transfer high-level features for FER. These approaches employ a single-source transfer learning approach, where source networks are trained on data from a single domain.

4.2.2 Multi-source Domain Adaptation

With the availability of a large number of datasets, even though with limited data samples, it is an intuitive step to take advantage of the diverse information comprehended by the different sources. Multi-source transfer learning has been explored widely in text classification [220], pattern recognition in EEG signals [221], speech recognition [222] etc. One of the approaches for multi-source transfer learning relies on the assumption that the target task can be represented as a weighted combination of the source tasks [223]. One prevalent approach to learning these combination weights in multi-source transfer learning is latent space transformation, which learns a common function across the different source tasks by optimising the overall loss function. Approaches like supervised multi-source domain adaptation [224] and the use of $\mathcal{H}$-divergence [225] or adversarial methods with Generative Adversarial Network (GAN) loss [226] aim to measure and minimise the distance between domain distributions. Meta-learning models (MAML) [227] and boosting approaches [228] address the divergence between source and target distributions, often using instance weighting or performance gaps as measures.

Many of the multi-source domain adaptation (MSDA) approaches train domain-specific classifiers and learn a weighted ensemble of these source classifiers for the target prediction [226, 229–231]. Some approaches, such as those by Guo *et al.* [224] and Yue *et al.* [232], employ point-set distance metrics or domain-invariant features with alignment loss for MSDA respectively. Ahmed *et al.* [233] tackle MSDA without accessing source data by using Information maximisation (IM) and a pseudo-labeling strategy, though this requires sufficient target data for training the ensemble source network.

Lee *et al.* [230] presents a multi-source transfer learning method in image classification that addresses data privacy concerns. By treating pre-trained source networks as black boxes, they employ bivariate maximal correlation analysis to train the ensemble of source networks. This method considers features independently, omitting group correlations within each source while combining networks. Recent studies indicate superior generalisation in multi-source transfer learning compared to single-source approaches in FER applications [234].

4.2.3 Limitations of Existing Literature

Current transfer learning methodologies presume access to source data for adapting source knowledge to the target domain. However, practical constraints, such as privacy and security concerns, often limit access to source data, allowing only the availability of trained source models. Recent research by Liang *et el.* [235] and the Data frEe multi-sourCe unsupervISed domain adaptatiON (DECISION) method [233] addresses this challenge by adapting single or multiple source models to the target domain without accessing the source data. This adaptation is done under the assumption of having sufficient target data. In domains like FER in CwA, obtaining ample training data is challenging due to the distinctive nature of the cohort. The existing literature indicates a scarcity of studies exploring multi-source transfer learning in automatic FER, with previous research mainly focusing on single-source transfer learning [236] or requiring access to source datasets for domain adaptation [218].

Our contributions: In our research, we delve into the realm of MSDA, wherein only multiple pre-trained source models are at our disposal to guide domain adaptation without access to the source datasets. Drawing inspiration from existing work, specifically Maximal Correlation Weighting (MCW) [230], we operate within a few-shot setting, assuming the availability of only a limited number of labelled target samples to supervise the adaptation process. We particularly look into the multivariate correlation [237] of the source features with the target domain, thereby capturing the complex association between the high-dimensional source features and the target. To sum up, the main contributions of this work are:

- We propose a multi-source domain adaptation approach for FER by leveraging the multivariate maximal correlation analysis using a few labelled target samples without access to source data.

- We evaluate our approach on the FER task by conducting extensive experiments on benchmark FER datasets. Experiments show that our approach consistently improves the results over the best single-source (best-SS) model. Further, our approach outperforms state-of-the-art FER - MCW [230] and DECISION [233] methods across multiple datasets.
- We also show the ability of the approach to generalise over domains outside FER by performing a general image classification task with the CIFAR-100 dataset.

4.3 Preliminaries

4.3.1 Multivariate Maximal Correlation Analysis

The maximal correlation was first introduced and developed by Hirschfeld [238], Gebelein [239], and Rényi [240] as a measure for the non-linear association between two random variables X_1 and X_2. It measures the strength of association among two random variables and characterises the non-linear transformations of the variables. We analyse the multivariate correlation of the features on the target classifiers and build an effective and computationally efficient approach for multi-source transfer learning.

Definition 1 (Maximal Correlation) *Given two jointly distributed random variables $X, Y \in \mathcal{X}$ with positive variance, the maximal correlation of*

(X, Y) is defined as:

$$\rho(X;Y) \triangleq (f^*, g^*) \triangleq \underset{\substack{f:\mathcal{X}\to\mathbb{R}, g:\mathcal{X}\to\mathbb{R} \\ \mathbb{E}[f(X)]=\mathbb{E}[g(Y)]=0 \\ \mathbb{E}[f^2(X)]=\mathbb{E}[g^2(Y)]=1}}{\arg\max} \mathbb{E}\left[f(X)^T g(Y)\right] \tag{4.1}$$

where expectations are with respect to joint distribution $P_{X,Y}$. (f^*, g^*) are referred as maximal correlation functions.

Maximal correlation is equal to the second largest singular value of a scaled joint probability distribution matrix and the singular vectors of the scaled probability distribution matrix could characterise the optimal transformations of the variables when they are discrete [237]. Given $f^* = \{f_1, f_2, ...\}$ and $g^* = \{g_1, g_2, ...\}$ with the associated singular values $\rho_1, \rho_2, ...$ the joint probability distribution $P_{X,Y}$ is given by [241] :

$$\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)} = \sum_{i=1}^{\infty} \rho_i f_i(x) g_i(y) \tag{4.2}$$

and

$$P_{Y|X}(y|x) = P_Y(y)\Big(1 + \sum_{i=1,2,...} \rho_i f_i(x) g_i(y)\Big) \tag{4.3}$$

In the case of the system of continuous random variables, most of the correlation measurements consider the pairwise relationship between the variables. In real-world datasets, data instances are represented as high dimensional multivariate random variables $(X_1, X_2, ...X_d)$. Extending definition 1 to multivariate random variables, maximal correlation among real-valued multivariate random variable $X = \{X_i\}_{i=1}^d$ can be given as

$$\rho^*(X_1, X_2, ...X_d) := \max_{f_1, f_2 \cdots f_d} \rho(f_1(X_1), f_2(X_2), ...f_d(X_d)) \tag{4.4}$$

Using bivariate measures to capture the multivariate relationships may not be efficient in capturing the association among the variables [242]. Methods like Maximal Information Coefficient (MIC) [243] and Canonical Correlation Analysis (CCA) [244] consider either two dimensions or linear correlations. In real-world scenarios, a feature may correlate weakly with the target class if considered individually, but when considered as a group, it can lead to a strong correlation [245]. Further, it is computationally expensive to evaluate all the pair-wise relations. Thus, the computation of maximal correlation in multivariate data eventually turns into an optimisation problem with complexity quadratic to the dimension, i.e. $O(n^2)$ where n is the feature dimension. By the above approach, for n dimensional data to find the correlation among the random elements, each X_i is paired with $n-1$ other elements, and solving the maximal correlation means optimising these $n(n-1)/2$ transformation functions. Multivariate maximal correlation analysis solves this by considering the group correlation among the features [242]. Maximal correlation eliminates the assumptions on data distribution and captures non-linear relations.

Based on Alternating Conditional Expectation (ACE), [237] proposed a computationally efficient method for addressing multivariate maximal correlation. It determines a single transformation function corresponding to each random variable, thereby reducing the computational complexity of computing multivariate maximal correlation. This approach maximises the aggregate inner products between transformed variables to optimise the correlation functions. Given a system of continuous random variables, this approach infers non-linear transformation functions assigned to each variable represented as vertices of a graph such that the aggregate pairwise correlations over the graph G are maximised. The ACE-based approach for computing multivariate maximal correlation is given in Algorithm 1

Definition 2 *Let* $G = (V, E)$ *be a graph with vertices* $V = \{1, 2, ...n\}$ *and edges* $E \subseteq \{(i, i') : i, i' \in V, i \neq i'\}$. *The multivariate maximal correlation of* $(X_1, X_2...X_n)$ *given* G *is*

$$\rho_G(X_1, X_2, ...X_n) := \sup_{(f_1, f_2, ...f_n)} \sum_{(i,i') \in E} \mathbb{E}[f_i(X_i), f_{i'}(X_{i'})] \tag{4.5}$$

such that $\mathbb{E}[f_i(X_i)] = 0$, *and* $\mathbb{E}[f_i(X_i)^2] = 1, \forall 1 \leq i \leq n$

4.4 Proposed Model for Multi-source Transfer Learning using Multivariate Correlation Analysis (MSTL-MCA)

Problem Setting: We formulate the FER with scarce data as a ACE problem, in which there are N labelled source domains and one target domain with few labelled samples. Let the input space be $\mathcal{X}$, and the classification is

Algorithm 1 ACE Algorithm to Compute Multivariate Maximal Correlation

Input: $G(V, E), X_1, X_2, ..., X_n$

Parameter: $\phi_1^{(0)}(X_1), ..., \phi_n^{(0)}(X_n)$ with mean zero and unit variance.

Output: Associated maximal correlations ρ_G and updated correlation functions $\phi_i^{(x)}(X_i)$

for k = 0 to S **do**

  for i = 1 to n **do**

$$\phi_i^{(k)}(X_i) = \mathbb{E}\sum_{j=1}^{i-1}\Big[\phi_j^{(k+1)}(X_j)|X_i\Big] + \mathbb{E}\sum_{j=i+1}^{n}\Big[\phi_j^{(k)}(X_j)|X_i\Big] \texttt{ for } j \in \mathcal{N}(i)$$

    update: $\phi_i^{(k+1)}(X_i) = \dfrac{\phi_i^{(k)}(X_i)}{\sqrt{\mathbb{E}[\phi_i^{(k)}(X_i)^2]}}$

  end for

$$\rho_G^{(k+1)} = \sum_{(i,j)\in E}\mathbb{E}\Big[\phi_i^{(k+1)}(X_i) \times \phi_j^{(k+1)}(X_j)\Big]$$

end for

return ρ_G, $\phi_n^{(x)}(X_n)$

among M categories. We represent the pre-trained source models as $\{\theta_S^j\}_{j=1}^N$ where the j^{th} model is represented as $\{\theta_S^j\} : \mathcal{X} \rightarrow \mathbb{R}^M$ is a classifier learnt from source dataset $D_S^j = \{x_{S_j}^i, y_{S_j}^i\}_{i=1}^{N_k}$, with N_k data points. $x_{S_j}^i$ denotes the i^{th} source image in source S_j and $y_{S_j}^i$ denotes the corresponding label. Given a target dataset $D_T = \{x_T^i, y_T^i\}_{i=1}^{N_T}$, with few samples, the problem we are addressing is to learn a classifier $\{\theta_T\} : \mathcal{X} \rightarrow \mathbb{R}^M$ using the ensemble of pre-trained source classifiers without access to source datasets. The data points are facial expression images represented by $(x_1, y_1), ...(x_n, y_n)$ where $(x, y) \in \mathcal{X} \times \{1, 2...M\}$, the feature $x \in \mathbb{R}^d$ is sampled from input space $\mathcal{X}$ and label $y \in \{1, 2, ..M\}$. In the absence of source training data, we lever-

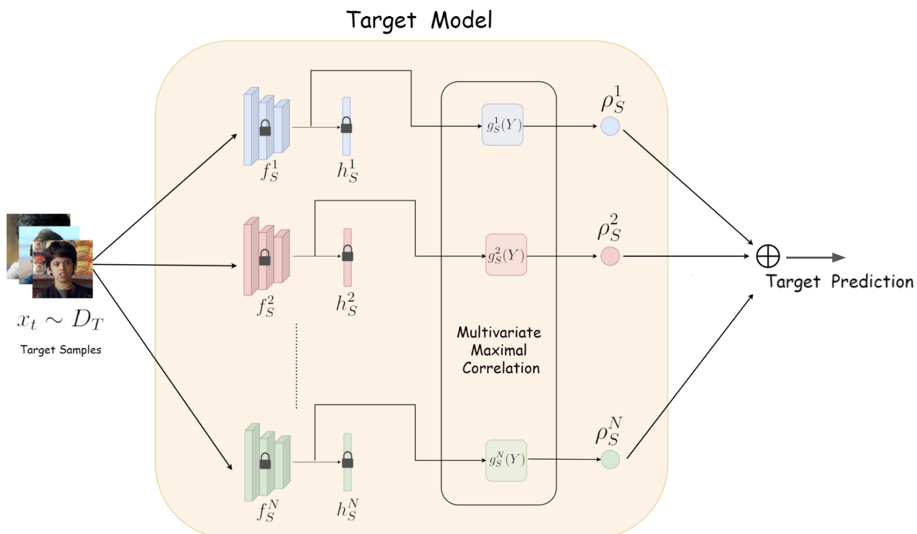


Figure 4.1: Proposed model architecture for MSTL-MCA

age the knowledge learned by N pre-trained networks trained on similar but different source datasets and learn the classifier θ_T, which has a low classification error on the target dataset. The high-level overview of the proposed architecture is given in Figure 4.1.

We represent each source model $\{\theta_S^i\}$ as the composition of two transformations :

(1) the mapping f_S^i transforming the input vector into feature vector of length d_i, $f_S^i : \mathcal{X} \to \mathbb{R}^{d_i}$ where d_i is the length of the feature vector of source i

(2) a classifier $h_S^i : \mathbb{R}^{d_i} \to \mathbb{R}^M$ from the feature vector into the output label, Y^{s_i}. This forms the hypothesis function.

Thus $\theta_S^i = (f_S^i \circ h_S^i)$

For the sake of better explainability, we have considered the feature-length to be the same for all the source tasks and $d_i = d_j, \forall i, j = 1, 2, ...N$. To build the target classifier, given N source tasks, with feature functions $\{f_S^i\}_{i=1}^N$, we optimise respective $\{g_S^i\}_{i=1}^N$ which is the hypothesis function such that the aggregate maximal correlation of functions f_S^i and g_S^i given by

$$\rho^* = \sum_{i=1,2,...N} \mathbb{E}_{\hat{P}^t_{X,Y}}[f_S^i(X)g_S^i(Y)] \tag{4.6}$$

where $\hat{P}^T_{X,Y}$ is the empirical joint distribution of the target data.

For each source, the optimal correlation function, g_S^i corresponding to feature function f_S^i and the corresponding correlation coefficient could be computed [230] as

$$g_S^i(Y) = \mathbb{E}_{\hat{P}^T_{X,Y}}[f_S^i(X)] \tag{4.7}$$

$$\rho(f, g) = \mathbb{E}_{\hat{P}^T_{X,Y}}[f_S^i(X)g_S^i(Y)] \tag{4.8}$$

While considering the high dimensional image data, it is interesting to analyse the group correlation of the multivariate data rather than the binary correlation among the individual features, $f_S^j(X)$ and $g_S^j(Y)$. Multivariate correlation analysis may reveal hidden complex interactions affecting the classification task [237]. Hence, we leverage the multivariate correlation among the group of features extracted by the feature extraction layer to compute the function $g_S^j(Y)$. In this direction, we apply network maximal correlation, an

ACE-based multivariate maximal correlation approach given in definition 2, which characterises the multivariate non-linear association between random variables.

We train the ensemble of the source classifiers on target samples to optimise g^i_S to maximise the aggregate maximal correlation given in equation 4.6 i.e.

$$g^i_S = \underset{\tilde{g}^i S}{\operatorname{argmax}} \; \rho* \tag{4.9}$$

Algorithm 2 Proposed MSTL-MCA approach

Input: source features $F = \{f^i_S\}_{i=1}^{N}$ from source task, target data $D_T = (x^i_T, y^i_T)_{i=1}^{N_T}$ where $x^i_T \in X$, $y^i_T \in Y$

for $i = 1$ `to` N **do** *// Iterate over source tasks*
    Randomly Sample m images $\{x_i, y_i\}_{i=1}^{m}$ from target dataset (D_T).
    Feed the target samples to the source feature extractor (f^i_S) to generate the source-specific feature representation.
    for $y \in Y$ **do** *// Iterate over target labels*
        Calculate the multivariate maximal correlation function.
        $g^i_S(y) \leftarrow \mathbb{E}_{\hat{P}^T_{X,Y}}[f^i_S(X)]$
        Calculate the correlation coefficient
        $\rho^i_S \leftarrow \mathbb{E}_{\hat{P}^t_{X,Y}}[f^i_S(x) g^i_S(y)]$
    end for
end for
return $\{\rho^i\}$, $\{g^i_S\}$

The correlation value for each pair of (f^i_S, g^i_S) gives the strength of association between the functions. Since we are considering the group correlation of features with the target, the ρ^i_S represents the combined weighted contribution of the feature functions of each source network to the ensemble classifier for the target domain.

$$\rho_S^i = \mathbb{E}_{\hat{P}^t_{X,Y}}[f_S^i(x) g_S^i(y)] \tag{4.10}$$

Finally, the prediction of the target label on the test data is given by

$$\hat{y} = \underset{y}{\mathrm{argmax}} \hat{P}_{Y|X}(y|x), \tag{4.11}$$

where

$$\underset{y}{\mathrm{argmax}} \hat{P}_{Y|X}(y|x) = \hat{P}_Y^t \Big(1 + \sum_{\substack{i=1,2,\ldots N \\ j=1,2,\ldots l_i}} \rho_S^i f_S^i(x) g_S^i(y)\Big) \tag{4.12}$$

The procedure for the Network Maximal Correlation (NMC)-based multi-source learning is given in Algorithm 2.

4.5 Experimental Setup

4.5.1 Task and Datasets

Facial Emotion Recognition To understand the performance of our approach, we designed a set of experiments on the FER task using four FER datasets: Facial Expression Recognition 2013, Real-world Affective Faces Database (RAF-DB), Japanese Female Facial Expression (JAFFE), and Child Affective Facial Expression (CAFE) under different source-target settings. Further, we investigated the efficiency of the approach on a novel FER dataset, the Child Facial Expression Dataset (CFED), curated by the authors. The dataset details are given below:

- FER2013 dataset [246] The 2013 Facial Expression Recognition dataset (FER2013) is a dataset provided by Kaggle, introduced at the International Conference on Machine Learning (ICML) in 2013 [247]. The dataset contains 35887 images, and each image has been categorised into 7 different types of emotion categories. The images in the dataset are registered; hence, the face appears in the centre of the image dataset.

- JAFFE dataset [248]: The JAFFE dataset consists of 213 images of different facial expressions from 10 different Japanese female subjects.

- RAF-DB dataset [249]: The RAF-DB dataset has 29672 real-world images labelled with 7 basic emotions and 12 compound emotions.

- CAFE dataset [250] The CAFE set features the facial expression data of a racially and ethnically diverse group of 2- to 8-year-old children posing for six emotional facial expressions and neutral emotion. The CAFE dataset consists of facial expression data of 90 female and 64 male children from varying ethnicities.

- CFED dataset: The CFED was collected, annotated, and prepared by our research group. There are limited annotated facial datasets for child facial emotion expression, especially in the global south, where active research in child emotion recognition is limited. The CFED dataset was collected by video search on child videos from YouTube under the Creative Common Licence, which allows the use of the videos for research. The manually retrieved video frames with expressed emotions were annotated by the research team. It consists of 606 images of children from Indian ethnicity representing 6 emotion classes - Anger, Fear, Happy, Neutral, Sadness, and Surprise.

For our experiments, we used the six emotional classes - Anger, Fear, Happy, Neutral, Sadness and Surprise from the FER datasets: FER 2013 (F), RAF-DB (R), JAFFE (J), and CAFE (C). Each domain has 600 labelled samples for training, i.e. 100 from each class label, and the testing set has 60 samples, i.e. 10 from each class label. Samples from each FER dataset are represented in Figure 4.2

Image classification We further considered the image classification to demonstrate the generalisability of the approach. For this, we conducted experiments on the benchmark image dataset CIFAR-100. We followed the specific experiment setting proposed by Lee *et al.* [230].

- CIFAR 100: The CIFAR-100 dataset has 100 classes containing 600 images each. There are 500 training images and 100 testing images per class. For our experiment, we have considered ten different source tasks, each consisting of 2 non-overlapping classes. All images were resized to 32x32, and the pixel values were normalised to zero mean and unit variance.

For our experiments, we randomly selected 10 non-overlapping class categories from the source task. For training, each source dataset had 500-labeled samples per class. Samples from the CIFAR-100 dataset are represented in Figure 4.3.

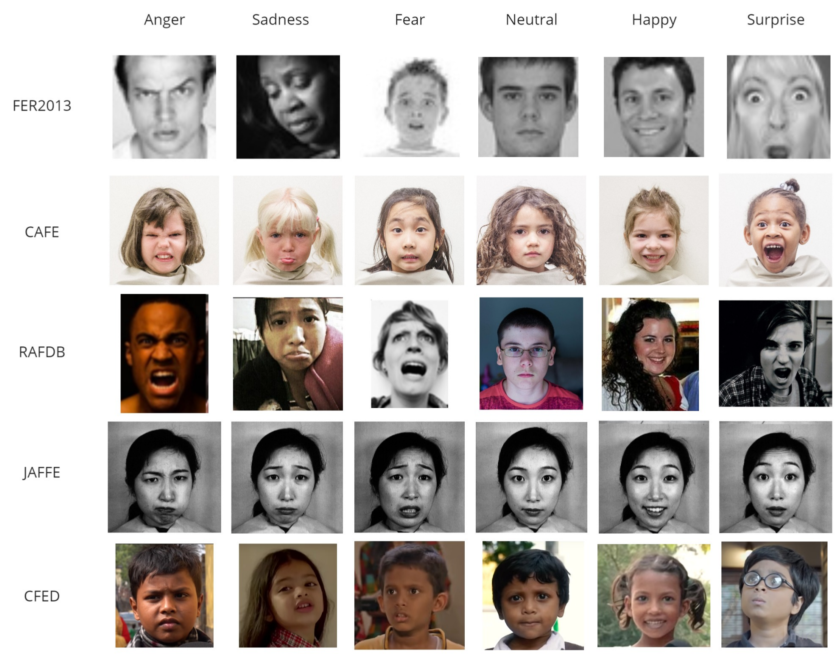


Figure 4.2: Sample images for FER datasets

4.5.2 Experiments

We investigated the performance of the Multi-source Transfer Learning using Multivariate Correlation Analysis (MSTL-MCA) approach on multiple FER datasets. In this, we compare our approach against two different baselines, which are commonly followed in the literature. The first one is the best single-source adaptation among the other sources (best-SS) [224], and the other is a

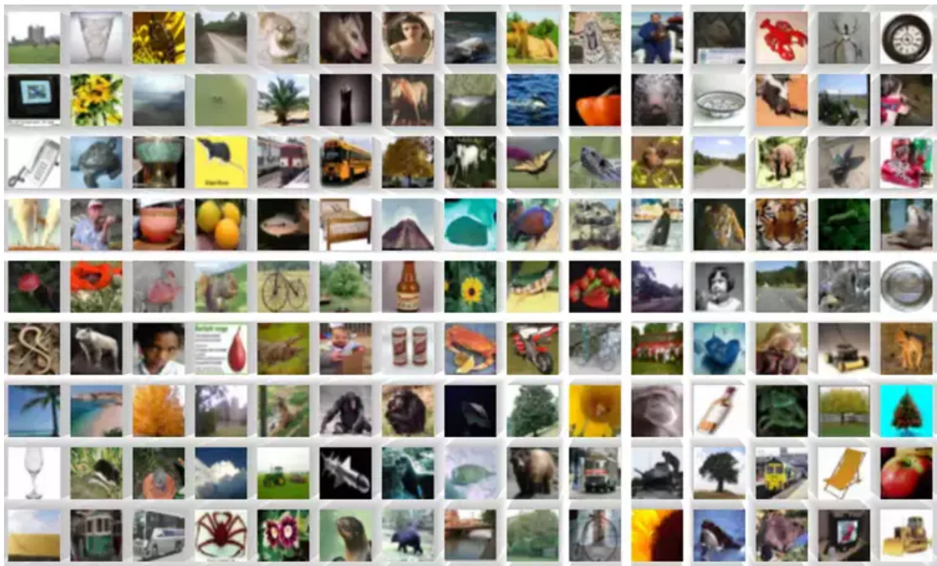


Figure 4.3: Sample images for the CIFAR-100

unified multi-source model (uni-MS) where all the source data are combined to form a single source and single-source domain adaptation is performed on the target data [251]. There is very limited work in multi-source domain adaptation without access to the source data. We compared our approach with Lee *et al.* [230], where they have considered feature independence while performing maximal correlation-based source weighting for multi-source domain adaptation. Further comparison was made with the DECISION [233] where unsupervised multi-source domain adaptation is being addressed using Information Maximisation loss and clustering-based pseudo labelling. We conduct experiments by selecting the target dataset in a round-robin fashion from among the set of FER datasets and keeping all the other datasets as the source datasets. We extended our experiments by evaluating the method on a novel CFED dataset as the target and the other standard datasets as the source task.

We further conducted experiments on image classification with the CIFAR-100 dataset. We followed the same experiment setting [230] for the comparison. For our experiment, we have considered ten different binary classification tasks as sources, each consisting of 2 non-overlapping classes. The source tasks were trained with 500 samples to generate the source network weights.

4.5.3 Implementation Details

Pre-training In our experiment on FER, we constructed 6-way (anger, sad, happy, surprise, neutral, fear) emotion classification on different FER datasets as the source tasks. The disgust class was discarded as it was not present in all the FER datasets considered. It is important to note that the source data samples were used only for pre-training the source tasks and not for training the target classifier. In other words, the source data samples were used to create a foundation or base knowledge for the source tasks but not for directly training the target classifier. This distinction is important because it highlights the importance of separating the pre-training and training stages and the potential benefits of using pre-trained networks for feature extraction. In real-world scenarios, the assumption is that these pre-trained source networks are available for feature extraction but are not trainable. Similarly, for the image classification task, we constructed binary classification tasks CIFAR-100 classes. We selected 10 non-overlapping pairs of classes from CIFAR-100 for classification in the source tasks.

All images were resized to 32x32, and the pixel values were normalised to zero mean and unit variance. We used ResNet18 architecture [252] similar to [233, 235] for pre-training the source tasks with parameters tuned for the specific dataset considered. We repeated all the experiments with LeNet architecture as well, which is a simple convolutional neural network architecture. The Cross-Entropy Loss was used as the loss function, and the Stochastic gradient descent (SGD) optimiser was used with the learning rate equal to 0.001, momentum set to 0.9, and the number of epochs to 100. These networks were considered as the black box pre-trained networks for the multivariate correlation analysis, where the features from the pre-trained networks will be extracted and further trained on target samples to compute correlation functions and coefficients for each set of features extracted from the pre-trained source networks, which will be used for the classification of the target test samples. Here, the black box implies that we do not have control over the training of the source networks but can only generate features pre-trained on these networks.

Training Once the pre-trained networks are available, the ensemble network is trained with 1, 5, 10, and 20 samples from the target task, during which the multivariate correlation functions and the correlation coefficient are computed. The training was done on Ubuntu Server 20.04 LTS, and the GPU used for training was Nvidia's RTX 3090. We used the PyTorch framework for all our implementations. To ensure reproducibility and to support open source, the code and the CFED dataset will be made available on request.

4.6 Results and Analysis

4.6.1 Facial Emotion Recognition

Our approach focuses on multi-source domain adaptation without the need for source data for domain adaptation while also addressing the challenge of limited target data, where only a small number of target samples are available for training. It should be emphasised that in this method, the source data is utilised solely to pre-train the source models. Most recent studies in multi-source domain adaptation, to the best of our knowledge, require labelled data from both source and target domains, as well as a mechanism for learning domain-invariant representations. For a fair evaluation, we compared our method with MCW [230], which is similar to our approach, which addresses source-free multi-source domain adaptation. Additionally, we compared our supervised approach with the DECISION [233] algorithm, which also tackles the problem of multi-source domain adaptation, even though it is an unsupervised approach.

We report our results on FER datasets in Table 4.1. We observe that our method consistently performs better across the different dataset settings and tasks. We observe a mean improvement of $\sim 12\%$ with respect to the best

Table 4.1: Experimental Results of MSTL-MCA on different FER datasets (RAF-DB (R), FER2013 (F), CAFE (C), JAFFE (J)) under different source (s) - target (t) settings. * indicates instances of *negative transfer*.

Setting	**s(R+C+J) - t(F)**	**s(F+C+R) - t(J)**	**s(F+C+J) - t(R)**	**s(F+J+R) - t(C)**	**Average**
Random	16.67%	16.67%	16.67%	16.67%	16.67%
uni-MS	23%	26%*	19%*	23%*	22.75%
best-SS	22%	34%	20%	27%	25.75%
DECISION (Ahmed et al., 2021)	26.43%	33%	22.14%	24.07%	26.4%
MCW (Lee et al., 2019)	33%	43%	31.67%	31.67%	34.83%
MSTL-MCA (LeNet)	**38%**	**43%**	**38.33%**	**33%**	**38.08%**
MSTL-MCA (ResNet-18)	**38.33%**	**43.33%**	**35%**	**35%**	**37.92%**

single source performance and compared to the uni-MS, our method gives $\sim$ 15% improvement in performance (Table 4.1 in *Average* column). Further, in cases of negative transfer, indicated as (*), our approach is performing better, indicating that it is robust to negative transfer. Negative transfer happens when transferring knowledge from a less related source, which may inversely affect the target performance. It is shown in cases where the best single-source model outperforms the unified multi-source model, indicating the adverse effect from unrelated sources. Compared to the MCW method, MSTL-MCA gives an improvement of 3.74% improvement. This signifies that group correlation among the features is capable of capturing the differentiating features in multi-source adaptation, and hence, the classification accuracy is higher. Even though an unsupervised algorithm, the DECISION approach addresses multi-source adaptation with similar settings. We compared our results with DECISION and obtained an improved performance of $\sim$ 11%.

Further, even with the newly curated CFED dataset, our proposed approach confirms its efficiency with similar trends in performance. The results for the CFED dataset are given in Table 4.2. The results reported are for 20 shots. With respect to the best-performing model, i.e. the MCW , it shows an improvement of 7% and $\sim$ 15% with DECISION . We have run the experiments for different shots, and the results are given in Table 4.3. The results show that the proposed method performs better in few-shot settings. This analysis illustrates that our algorithm's performance significantly improves

Table 4.2: MSTL-MCA results on CFED dataset

Model	**Accuracy**
uni-MS	19.00%
best-SS	17.00%
DECISION (Ahmed et al., 2021)	30.02%
MCW (Lee et al., 2019)	38.00%
MSTL-MCA (LeNet)	42.00 %
MSTL-MCA (ResNet-18)	**45.00**%

up to 20 shots, after which it gradually converges. At this point, the model with a joint training approach has received a sufficient number of samples to learn their parameters and the addition of more samples no longer yields significant knowledge gains. Finally, we evaluated the performance of the approach on a dataset with facial expressions of autistic children.

Table 4.3: MSTL-MCA elbow point analysis for CFED dataset with source as (F+R+C+J) and target as CFED

Dataset	1-shot	5-shot	**10-shot**	20-shot	25-shot	30-shot	60-shot
CFED	40%	40%	43%	**45**%	45.67%	45.01%	45.33%

Maximum correlation analysis To study the effect of multivariate maximal correlation in regulating the flow of knowledge from the source to the target task, we conducted the correlation analysis between source and target pairs. For this, we considered CAFE, FER-2013, RAF-DB, and JAFFE as the source datasets and CFED as the target dataset. We computed the correlation coefficient corresponding to each source task for 20 runs. We then compared it with the correlation weighting of the sources computed by MCW [230] under the same settings. The correlation coefficients for the different source tasks using MSTL-MCA and MCW are given in Figure 4.4. The results show that the correlation weighting of our approach for each source is clustered closely around the median when compared to the MCW method, where the weights learned are more variable to the input samples under consideration. This shows that our approach could produce more reliable and accountable results by consistently focusing on the relevant source knowledge over different runs. This accounts for the ability of the model to produce better results than state-of-the-art methods, as seen in Table 4.2.

For further analysis, we removed the source task with the highest correlation value given by our algorithm, i.e. JAFFE (J), and computed the accuracy for the adaptation task. We observed that average accuracy dropped to 41.99% with a relative drop of $\sim 7\%$. Likewise, removing the task with the lowest weightage given by our algorithm, CAFE (C), and keeping the other tasks dropped to 44.23% with a relative drop of $\sim 2\%$. With this, we can infer that removing the highly correlated sources leads to a significant drop in accuracy, showing that the source task with high correlation contributes higher to the target classifier learning. Similarly, we compared the effect of multi-variate correlation in the classification task. We compared the correlation strength of our proposed method with the MCW [230] approach, where binary correlation weighting has been used. The results in Table 4.2 show that multi-variate group correlation could capture the relevant source knowledge in a consistent and reliable way eventually leading to better performance.

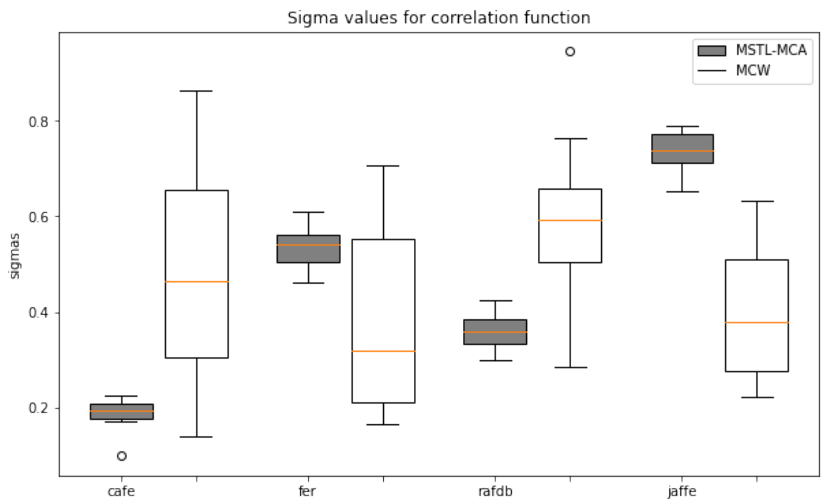


Figure 4.4: Maximal correlation analysis for CAFE, JAFFE, RAF-DB, FER-2013 as source and CFED as the target over 20 runs.

Statistical Analysis To further validate our results, we perform statistical analysis. For the null hypothesis, we assume that our proposed model works similarly to other algorithms and consider the average accuracy for all the algorithms. We tried 20 different samples for all classifiers on the CAFE dataset and then performed the Kruskal-Wallis H-test (also called one-way ANOVA test on ranks) and the Friedman test. We found the Kruskal-Wallis H statistic equal to 65.38, which shows significant statistical importance and outputs a very small $p = 9.33e-13$. Similarly, for the Friedman test, we got a statistical value of 72.37 and $p = 3.28e-14$. As the p-value is very small in both the tests and $p < 0.05$, we can safely reject the null hypothesis. Hence, we can infer that the performances of all algorithms are not equivalent.

Table 4.4: Average ranks for different methods in post-hoc tests.

Method	Average Rank
UNI-MS	5.75
BEST-SS	4.75
DECISION (Ahed et al., 2021)	4.5
MCW (Lee at al., 2019)	2.875
MSTL-MCA (LeNet)	**1.875**
MSTL-MCA (ResNet -18)	**1.25**

Considering that the null hypothesis was rejected, we have two scenarios for a post-hoc test [253]: (1) We perform the Nemenyi post-hoc to compare all algorithms with each other. (2) We perform the Bonferroni-Dunn post-hoc test to compare all the algorithms with a control algorithm (i.e., the proposed method). Both the posthoc tests are performed with alpha values 0.05 and 0.1 as suggested by [253].

To perform both the post-hoc tests, we calculated the average rank of each algorithm, as shown in Table 4.4. Average rank (or fractional rank) denotes the algorithm's performance, i.e. a lower-ranked algorithm performs much better than a higher-ranked algorithm. It is calculated by taking the mean of ordinal ranking, which is done by the simple ordering of the accuracies of respective algorithms. The results given in Table 4.4 show that our proposed method has a lower rank than other methods and hence outperforms others.

Then, we compute the critical differences (CD) as per Nemenyi and Bonferroni-Dunn tests plotted in Figure 4.5. In the CD diagram, closely performing algorithms are grouped into a single group. Figure 4.5 shows the graphical representation of the classification accuracies for our problem on the six different methods. In the CD diagram, the lowest (best) ranked algorithms are on the right side of the graph. Hence, the results reveal that UNI-MS, BEST-SS, and DECISION [233] perform significantly worse than MSTL-MCA (proposed method) and MCW [230]. Further, it can be observed that MSTL-MCA (for both LeNet and ResNet-18) have the lowest ranks among all. This implies that the MSTL-MCA outperforms the other approaches.

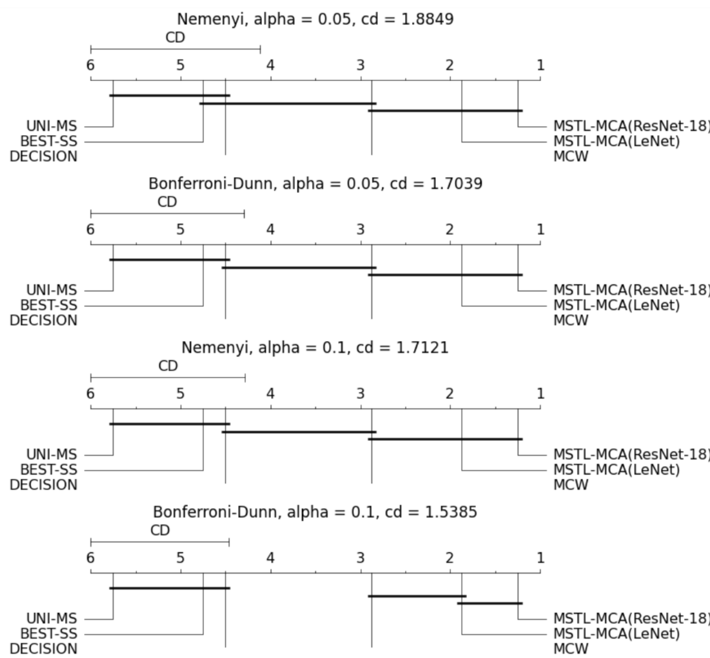


Figure 4.5: CD Diagram for Nemenyi and Bonferroni-Dunn test. The bold line represents the closely grouped algorithms together.

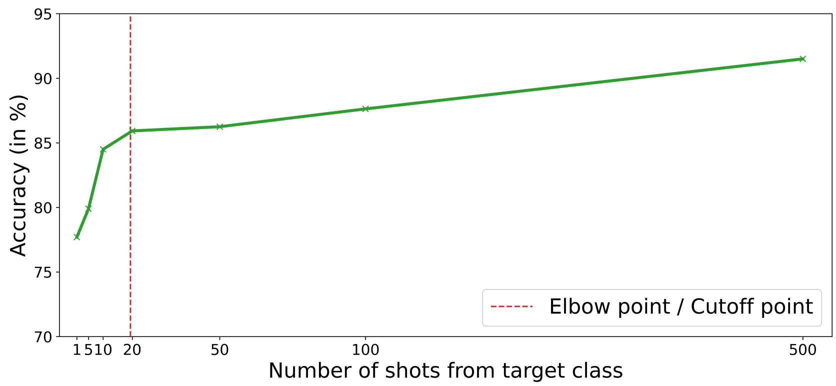


Figure 4.6: Plot for Accuracy v/s Number of shots for CIFAR-100. The orange line represents the Elbow point.

Table 4.5: MSTL-MCA results on CIFAR-100 for 10-shots.

Methods	**CIFAR- 100**
best-SS	60.00%
MCW [230]	78.10%
DECISION [233]	79.50%
MSTL-MCA (LeNet)	**83.50%**
MSTL-MCA (ResNet-18)	**84.53%**

4.6.2 Image Classification

The results of multi-source adaptation on image classification in the CIFAR-100 dataset are given in Table 4.5. We could see a similar performance of our method on the image classification task as in the FER task. Our method performs better with an improvement of $\sim 6\%$ in comparison with the state-of-the-art method MCW . It further shows comparable results with DECISION .

Elbow point analysis We performed the elbow point analysis on the

CIFAR 100 dataset to find the optimal k-value for the k-shot learning approach we used. We can observe from Figure 4.6 that in the CIFAR-100 dataset, after 20 shots, the rate of growth in the accuracy is significantly lower concerning shots. So, we can deduce that the elbow point or the knee of the curve is at 20 shots for the CIFAR dataset, and even with a smaller number of samples the algorithm is capable of training the classifier. This shows that our approach has utility in applications, including FER, where there is an unavailability of huge training datasets.

Maximum correlation analysis We conducted maximal correlation analysis on the CIFAR-100 dataset with the same settings given in Section 5.2.4. The weights for the source tasks for the CIFAR-100 dataset are given in Figure 4.7. Similar to the FER task, we can see that the correlation weighting of our approach is consistent across the different runs, as represented by the lower spread of the weights.

Following the validation of our approach's efficacy on FER datasets and Image Classification datasets, we assessed its feasibility on an ASD FER dataset. For this analysis, we utilised the Facial Expression Database provided by Alamgir et al. [254], comprising facial expression images collected from children aged 6 to 14 diagnosed with ASD. In total, the dataset comprises 113 images (F= 32, M= 81) representing four distinct facial expressions: Happy, Sorrow, Neutral, and Angry or Disgusted. The results of our approach on this dataset are presented in Table 4.6.

Our approach achieved an accuracy of 40.01%, significantly higher than

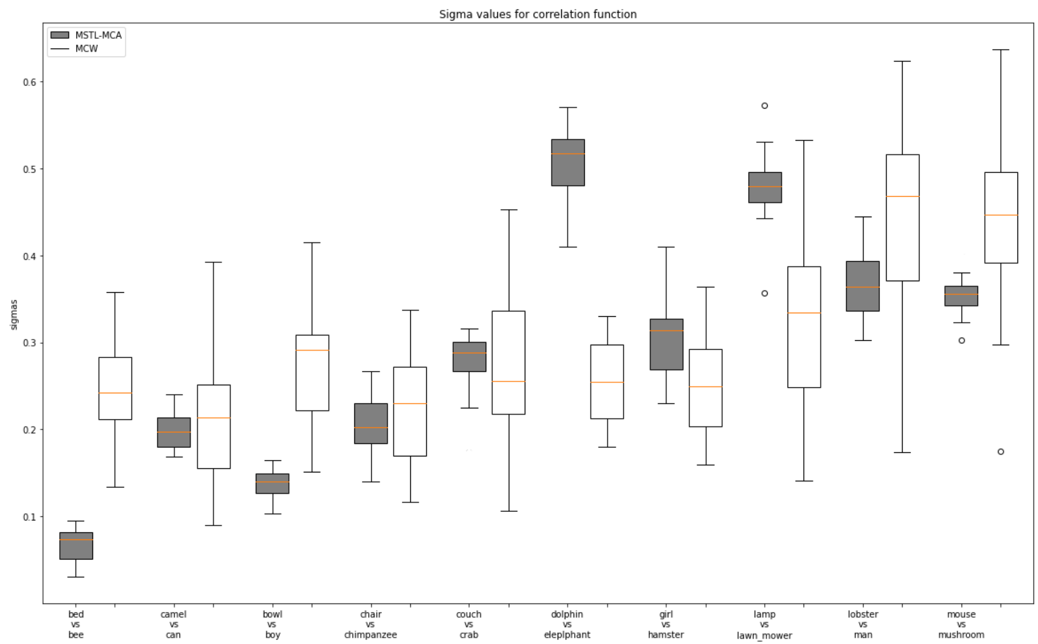


Figure 4.7: Maximal correlation analysis for CIFAR-100 dataset for 5-shots over 20 runs

Table 4.6: MSTL-MCA results on ASD FER dataset

Model	**Accuracy**
chance accuracy	25%
DECISION (Ahmed et al., 2021)	31.73%
MCW (Lee et al., 2019)	37.20%
MSTL-MCA (LeNet)	39.87 %
MSTL-MCA (ResNet-18)	**40.10**%

chance accuracy and outperforming unsupervised multi-source source-free transfer learning settings (DECISION - 31.73%). This demonstrates the efficacy of the approach in detecting facial expression information in CwA. However, it is important to note that while this method showcases the use of multi-source transfer learning for FER in ASD children, further investigation is warranted to enhance performance for clinical utility.

4.7 Conclusion

In this work, we proposed a multi-source transfer learning approach by leveraging the multi-variate maximal correlation of features extracted from an ensemble of source networks to build a target classifier. We measure the multivariate non-linear association among the features of the source networks using Network Maximal Correlation and optimise the aggregate multivariate maximal correlation over the source tasks to learn the target classifier. The results show that capturing the group correlation of the features with output, as proposed, significantly improves the learning of the target classifier.

We demonstrated the efficacy of our approach in FER using benchmark datasets. We verified and confirmed the performance on the novel CFED dataset with images from YouTube. We investigated the performance of the proposed method in the cross-cultural target classification task by considering the different FER datasets as the source dataset and our novel CFED dataset consisting of facial emotion data of children of Indian ethnicity and having limited samples. We then performed an image classification task using a standard image dataset, the CIFAR-100. We have also shown that the proposed method convincingly performs well even in smaller target datasets with our experiments of k-shot learning with k less than ten shots. Finally, we have demonstrated the feasibility of this approach in FER in ASD dataset as well.

The proposed method enables combining the knowledge from the multiple source networks in an effective and computationally efficient manner and can be leveraged where training data is limited. Further, since the knowledge gained by the source classifier is leveraged to build the target classifier without direct access to the input data in this approach, it ensures improved data privacy which is primal in facial emotion expression data. The proposed method can be generalised to other domains as well while applying transfer learning. The performance of the approach with heterogeneous source tasks with multimodal information can be explored in future work.

Moreover, the proposed automated FER systems play a pivotal role in the early detection and precise diagnosis of autism by analysing facial expressions during diagnostic assessments. This technology assists clinicians in discerning subtle emotional cues that may serve as indicative signs of ASD. Beyond diagnosis, FER technology extends its utility to behavioural intervention support, seamlessly integrating into intervention programs. By offering real-time feedback on a child's emotional expressions, facilitators can tailor interventions to address specific emotional challenges, fostering the development of crucial emotional regulation skills. Moreover, FER technology finds application in assistive communication devices for non-verbal individuals with autism. By interpreting facial expressions, these devices enhance communication, enabling individuals to convey emotions, needs, and preferences more effectively. Overall, the integration of FER technology contributes significantly to both diagnostic and therapeutic aspects in the realm of autism.

In this chapter, we explore a multi-source transfer learning approach to identify facial expressions in children, including a specific case study on FER in CwA. Our approach is generalizable and can be applied in domains where access to large annotated datasets is limited. After examining the facial expression characteristics in CwA, we proceed to investigate speech behaviours in CwA, which can aid in diagnosis. The details of this investigation are presented in the next chapter.

CHAPTER 5

SPEECH-FEATURES AS INDICATORS FOR AUTISM SPECTRUM DISORDER

In this chapter, we focus on the automated detection of speech behaviours, aiming to address the communication aspects of children for the diagnostic assessment of autism. Language impairments are prevalent and fundamental characteristics of ASD, significantly impacting communication and social interactions in affected individuals. The diagnosis of ASD heavily depends on the presence of impaired social interaction due to the inherent difficulties individuals with ASD face in communication and social interactions. In the literature, researchers have leveraged computational methods to develop linguistic biomarkers capturing specific language measures, aiming to provide an objective and reliable assessment of ASD. Most research has mainly concentrated on finding ASD biomarkers in English, but their usefulness is limited in resource-constrained settings of the Hindi language. To address this, we explored the manifestations of speech behaviours in both Hindi and English-speaking CwA. By utilising automated techniques, we extracted various acoustic and text-based speech features, including semantic-pragmatic features, to provide an objective and reliable assessment of ASD in diverse linguistic and cultural contexts. In this chapter, we discuss the current state of speech-based diagnosis for ASD, their limitations and analysis to identify the significance of semantic and pragmatic features in ASD diagnosis. Following this, we provide a cross-linguistic analysis to identify the manifestation of speech behaviours in CwA. Finally, provides the results on the use of a comprehensive speech-based feature set for the assessment of autism in children in the resource-constrained language setting of Hindi.

5.1 Related Works

5.1.1 Computational Methods for Extracting Speech Features in ASD

Due to the complexity of linguistic profiling, researchers often explore the automatic extraction of language measures of expressive language to determine the developmental state of the child [255, 256]. [257] presented a digital tool for the automatic extraction of social interaction measurements involving facial expression and voice characteristics. [258] extracted the acoustic features like intonation and rhythm of speech to analyse the prosodic nature of ASD manifestation in different cultural groups. They further analysed speech articulation in terms of prosodic characteristics and pragmatic ability and its correspondence with ASD [259]. Seven automated language markers were identified, and their contribution toward autism diagnosis was established by [260]. Natural Language Processing (NLP) techniques were leveraged to extract language measures featuring the linguistic characteristics of autism, and their feasibility in diagnosis is established in these studies [261]. All these studies provided evidence of the potential of technology-based solutions such as ML and deep learning techniques to improve the efficiency and effectiveness of applied behaviour analysis procedures [262]. Even though ML models were shown to be feasible in identifying vocal biomarkers, their generalisability over languages and different cultural groups is limited. This demands identifying explainable and meaningful features and developing context-specific ML models [263].

5.1.2 Cultural Context and Linguistic Behaviours

Studies exploring language atypicality across different cultural and linguistic contexts have revealed variations in the manifestation and severity of language impairments. Diagnosing children from culturally and linguistically diverse backgrounds poses additional complexities due to the potential for misinterpretation due to the cultural specificities in language profiles [264, 265]. This underscores the need for research that addresses specific linguistic populations, such as Hindi-speaking children. [266]. While there are universal aspects of meaning and language use, such as the ability to convey information and make inferences, the specific ways in which meaning is encoded and interpreted can vary across languages. In addition to the apparent variations in lexicon and syntax across languages, it is also important to note that the semantics and pragmatics of conversation are language-specific [267–269]. Semantics, which deals with the study of meaning, is closely tied to the structure and organisation of a particular language [270]. Different languages may have distinct lexical categories, grammatical structures, and semantic relationships between words and phrases. For example, how tense is expressed or the availability of grammatical gender can affect how meaning is conveyed in a language. Pragmatics, on the other hand, focuses on the use of language in context and how meaning is influenced by factors such as speaker intentions, social norms, and shared knowledge. Similarly, pragmatics is highly language-specific because it involves understanding the specific conventions, implicatures, and cultural nuances that shape communication within a particular linguistic community [271]. For instance, politeness strategies, speech acts, and the interpretation of indirect speech can vary significantly across languages and cultures. While there may be some cross-linguistic similarities and universal principles in semantics and pragmatics, such as the basic cooperative principle mentioned earlier, the specific ways in which meaning is expressed and understood are influenced by the linguistic and cultural context of a given language.

5.1.3 Limitations of Existing Literature

The majority of research on speech behaviour analysis in ASD has primarily focused on English-speaking populations. Existing literature primarily focuses on standardized measures and lacks comprehensive exploration of these specific language features, particularly in resource-limited language groups such as Hindi-speaking children. This gap limits our ability to accurately assess and diagnose ASD in these populations, hindering early identification and intervention strategies. In the context of Hindi-speaking children, there is a significant need for reliable predictors of ASD. Currently, there is a dearth of research on acoustic and linguistic features specific to this population. By identifying and establishing such predictors, an accurate diagnosis of ASD can be achieved, leading to targeted interventions and improved outcomes for Hindi-speaking CwA.

Our contributions: To address the existing gaps in the literature, in this work, we have made the following contributions:

1. Investigate the significance of specific semantic and pragmatic language features in English for diagnosing autism in children, focusing on aspects such as word use, discourse coherence, and pragmatic language abilities.

2. Develop automated techniques for extracting linguistic features from conversational data in the resource-limited Hindi language, enabling the automatic acquisition of relevant linguistic features, such as syntactic patterns and semantic cues, from the conversational data.

3. Identify and analyse the acoustic and linguistic features in Hindi that effectively differentiate CwA from TD children. This will involve employing rigorous methods, such as statistical analysis and ML, to identify the most discriminative features for ASD diagnosis.

Table 5.1: CHILDES data for semantic pragmatic speech analysis. All datasets are in English

Dataset	No. of participants	Age (years)
Eigsti [273]	ASD - 16	3 - 6
	TD - 16	
Flusberg [274]	ASD - 6	3 - 8
Nadig [275]	ASD - 13	3 - 7
	TD - 25	

5.2 Phase I: Significance of Semantic-Pragmatic Features in Autism Diagnosis

5.2.1 Datasets

We first conducted semantic and pragmatic speech feature analysis using transcripts obtained from the Child Language Data Exchange System (CHILDES) databank [272]. The transcripts were in English sourced from the Eigsti, Nadig, and Flusberg datasets, comprising interactions between CwA and TD children. In total, there were 76 children involved, with 35 children diagnosed with ASD and 41 TD children. The Eigsti dataset consisted of transcriptions from 30-minute free play sessions involving 48 children across ASD, developmental delay, and TD groups. The Nadig dataset, on the other hand, collected data during a 10-minute free-play task where the parent and child engaged in play. Lastly, the Flusberg dataset included language samples from natural interactions between the child and their mother, recorded when the child was between 12 and 26 months old. The dataset details are provided in Table 5.1.

5.2.2 Features

The first objective of this study was to investigate the discriminatory power of semantic-pragmatic speech features in child-partner conversations for the purpose of autism diagnosis. The features were extracted at the sample level and captured the interaction behaviours of the child with their partner. The choice of features aligned with the standard diagnostic tool for ASD - ADOS [276]. The features focused on language behaviours commonly observed in CwA, such as echolalia (repetitive speech) [276], difficulty using pronouns [277], and other semantic-pragmatic features, including discourse coherence [278], vocabulary size [279], back-and-forth conversational exchanges, question-asking, and building upon statements [276]. We extracted the four sets of linguistic features for this analysis: (i) Computerised Language Analysis (CLAN) based linguistic features, (ii) lexical features, (iii) syntactic complexity features, and (iv) semantic and pragmatic features. The details of the features are given below:

- CLAN-based linguistic features (baseline): We extracted the linguistic features from the conversational data between the child and their partner, which we obtained from the CHILDES databank. CLAN was used to extract morphosyntactic linguistic features like duration, percentage of word errors, number of repetitions, etc. [272]. CLAN provides a basic statistical analysis of linguistic behaviour; hence, these features are considered the baseline for our analysis.

- Lexical features: These features focused on the distribution and usage of different word categories.

- Features representing syntactic complexity: These features aimed to capture syntactic properties, such as clauses per sentence and mean sentence length.

- Semantic and pragmatic features capturing stereotypical and repetitive speech characteristics: These included measures such as *Mean Length of Utterances in words* (*MLUM*), *Number of Different Word Roots* (*NDWR*), *Initiative to Ask Questions*, *Repetition Proportion*, *Child-Child Discourse Coherence*, *Child-Partner Discourse Coherence*, *Echolalia*, *Unintelligible Proportion*, and *Unexpected Words*.

A total of 111 features were extracted and analysed in this study. Features that depend on the number of speech units, such as Echolalia, Unintelligible Proportion, Unexpected Words, and so on, were normalised based on the conversation length of each respective child. For a summary of the extracted linguistic features, refer to Table 5.2.

5.2.3 Feature extraction

- **Lexical Features**

  - POS tags: We extracted the POS tags using the POS tagger from Natural Language Toolkit (NLTK) [280] library. The features

Table 5.2: Linguistic features used for diagnosis. (The numbers in parenthesis show the no. of features in each category)

Feature Name	Description
MLUM (1)	The mean length of morphemes spoken by the child.
NDWR (1)	Number of distinct word roots spoken by a child.
Initiative to ask Questions (1)	Number of questions asked by a child.
Repetition Prop (1)	Phrases repeated in immediate sentences uttered by the child.
Child-Child Discourse coherence (1)	Cosine similarity between BERT embeddings of consecutive sentences of the child.
Child-Partner Discourse Coherence (1)	Cosine similarity between BERT embeddings of consecutive sentences of therapist and child.
Echolalia (1)	Number of c-units with immediate repetition of bigrams by the child (originally spoken by the therapist).
Unintell Prop (1)	Number of partially or fully illegible c units
Unexpected Words (1)	Number of unexpected words spoken by the child using TF-IDF Vectors.
POS Tag Frequencies (45)	The Parts-of-Speech tags and the corresponding frequency in the conversation.
Syntactic Complexity (23)	Measures of arrangement of words into a meaningful conversation, or the complexity of conversation .
CLAN-based features (34)	Morphosyntactic features.

involved the POS tags from Penn Treebank [281] and their frequencies.

- **Syntactic Features**
  - **Syntactic Complexity features** The syntactic features were extracted using the L2 Syntactic Complexity analyser (L2SCA) [282] software.
- **Semantic and Pragmatic Features**
  - *Initiative to ask Questions*: This is quantified using the number of questions asked by the child. Calculated using the number of times the question mark ('?') occurred in the utterances by the child.
  - *Echolalia*: The sentences are first tokenised using NLTK word tokeniser. Then all the bigrams from this tokenised texts of the therapist and child are compared. The count is updated if the

child uses the bigrams previously spoken by the therapist in their last sentence. The total count is returned.

- *Repetition Prop*: The sentences are first tokenised using NLTK word tokeniser. Then all the bigrams from this tokenised text of the immediate sentences spoken by the child are compared. The count is updated if the child uses the bigrams previously spoken by them in their last sentence. The total count is returned.

- *Metaphors*: To extract the metaphors, a BERT-based model that tagged words as metaphors was used and afterwards, the proportion of metaphors in the conversation was computed. A metaphor detection model and a tokeniser trained for Hinglish were used. The number of metaphors divided by the total number of words was returned.

- *Child-Child Discourse Coherence*: The embeddings of the consecutive utterances by the child are computed using a BERT-based model trained on Hinglish data. The cosine similarity between all such utterance pairs is calculated, and then the mean of it is returned.

- *Child-Partner Discourse Coherence*: The embeddings of the utterance by the therapist and then consecutive utterances by the child are computed using a Bert-based model trained on Hinglish data. The cosine similarity between all such utterance pairs is

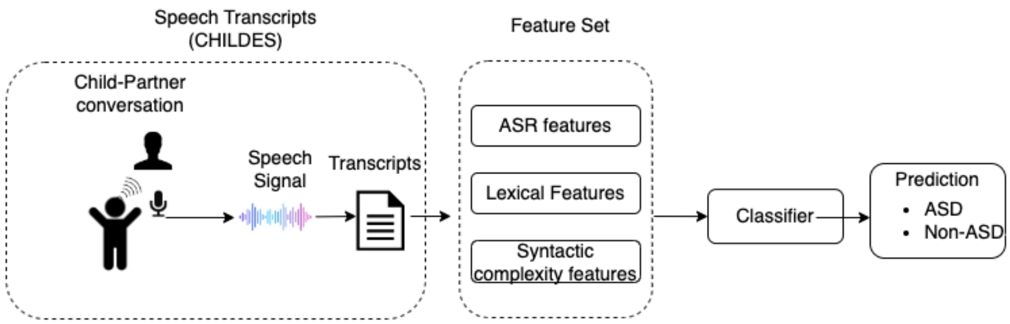


Figure 5.1: System architecture

calculated, and then the mean of it is returned.

5.2.4 Experiments

From the CHILDES transcripts, we separated the speaker information and cleaned the dataset for further analysis. After preprocessing the CHILDES transcripts and conducting feature selection, we obtained a final set of 63 features out of the initial 77 without considering the baseline. To guide our feature elimination process, we performed SHapley Additive exPlanations (SHAP) analysis on the initial feature set. This analysis reveals the contribution of features to the model's generated predictions. We excluded features such as um-propagation (a measure of uh and um usage in conversation) and certain POS tags (e.g., SYM-symbol, MD-model auxiliary) that did not significantly contribute to the dataset. Further, the CLAN-based features were used only in the baseline settings. Next, we focused on a binary classification problem: determining whether a child has ASD or not. The clinical status labels in all three datasets were binary, facilitating this classification task. To benchmark our results, we compared them with the majority classifier, which assigns every sample to the majority class in the training set. With a relatively small dataset of 76 children, maximizing the amount of training data is critical and hence, for the classification task, we split the data into training and testing sets using a 90:10 ratio. The split was performed at the subject level. The system architecture, as illustrated in Figure 5.1, showcases the overall flow of our approach for classifying ASD based on the CHILDES transcripts.

5.2.5 Results and Discussion: Do Semantic and Pragmatic Features Matter?

In Phase I analysis, our objective was to investigate the diagnostic potential of semantic and pragmatic speech features for ASD. The ML classifiers commonly identified from the literature for speech analysis were considered for the evaluation, and hyper-parameter tuning was done to identify the best model for the data. To determine the optimal hyperparameters, we employed grid search. Table 5.3 presents the performance of different classifiers on all the extracted linguistic features. Among the different classifiers evaluated, the Support Vector Machine (SVM) with a radial basis function (RBF) kernel achieved the highest classification accuracy of 94.1% for recognising ASD from speech signals. Hence, we selected the SVM classifier for further analysis.

In the subsequent step, we conducted an ablation study to evaluate the impact of previously explained four different feature sets on classification de-

Table 5.3: Results with different classifiers

Model Name	**Accuracy(%)**	**Macro-Precision (%)**	**Macro-Recall (%)**	**Macro-F1 Score (%)**
Majority Classifier	56.4	24	50	32
K-Nearest Neighbours	64.7	69	66	67.47
Logistic Regression	76.5	83	78	80.42
Random Forest	88.2	90	89	89.5
Gradient Boost	76.5	76	76	76
Support Vector Machine	**94.1%**	**95**	**94**	**94**

cisions. The results of the diagnosis with different feature sets are provided in Table 5.4. Using the SVM classifier selected in the previous experiments, we assessed the performance of various linguistic feature sets. We considered the CLAN-based linguistic features as the baseline feature set. We then compared the performance of other feature sets considered individually as well as in combination. We observed that each feature set outperformed the majority classifier significantly. Moreover, all feature sets, except for the syntactic complexity features, demonstrated superior performance compared to the baseline features. The syntactic complexity features exhibited comparable performance to the baseline features. Among the evaluated settings involving different feature sets, it is evident that integrating all the features together resulted in the highest performance. This comprehensive feature set showcased a significant improvement in terms of accuracy, achieving approximately a 37% relative enhancement over the majority classifier and a 23.5% relative improvement compared to using baseline features. It can be observed from Table 5.4 that similar observations were found to be true for other evaluation metrics as well. The findings suggest that including semantic and pragmatic features provides a more effective representation of

Table 5.4: Feature sets and their contribution to diagnosis

Model Name	Accuracy(%)	Macro-Precision (%)	Macro-Recall (%)	Macro-F1 Score (%)
Majority classifier	56.4	24	50	32
CLAN-based (baseline) features	70.59	70	71	70
Syntactic complexity only (I)	70	84	58	55
POS tag features only (II)	82	83	86	82
Semantic and Pragmatic features only (III)	82.3	83	82	81
POS and Semantic- Pragmatic (II + III)	88.2	90	89	89
All features(I+II+III)	**94.1**	**95**	**94**	**94**

communication behaviour than relying solely on POS tag features or syntactic complexity features. By incorporating semantic and pragmatic aspects, we gain a deeper understanding of the nuanced aspects of communication exhibited in ASD.

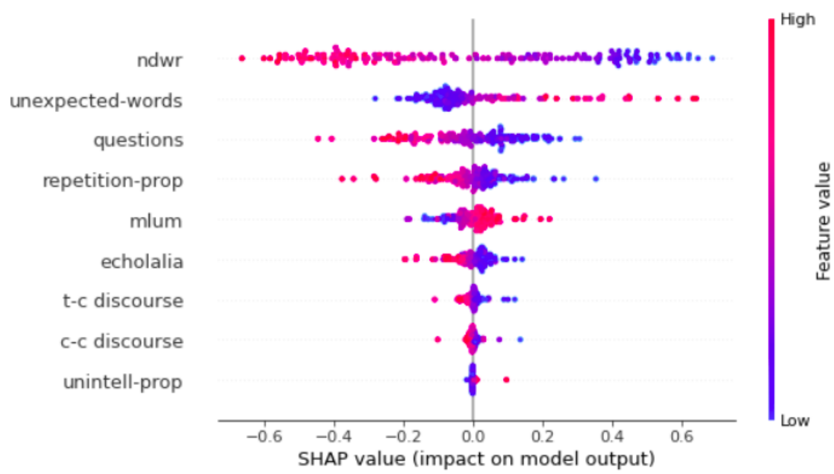


Figure 5.2: SHAP Analysis of linguistic features

To gain insights into the interpretability of the model predictions, we conducted a SHAP analysis. Figure 5.2 displays the SHAP values of the semantic-pragmatic features. Among these features, the number of distinct word roots emerges as the most influential in determining the child's diagnosis status. This finding suggests that vocabulary richness plays a crucial role in distinguishing CwA from TD children. This aligns with previous literature that has explored linguistic behaviour in CwA [283]. Studies have shown that CwA may struggle to maintain meaningful conversations and exhibit a lack of global coherence in their speech [284]. Our SHAP analysis aligns with these findings, revealing that features such as *unexpected-words*, *trainer-child (t-c) discourse*, and *child-child (c-c) discourse* prominently contribute to the predicted output. Their presence among the top 10 contributors underscores the importance of semantic coherence in conversations and its potential to guide the classification of ASD. In summary, the SHAP analysis provides compelling evidence for leveraging semantic-pragmatic speech features in ASD diagnosis. These features demonstrate a discriminative nature and offer improved interpretability by explaining the diagnostic process more effectively. The details are depicted in Figure 5.2.

5.3 Phase II: Acoustic and Linguistic Speech Biomarkers in Hindi for ASD Diagnosis

Phase II of our study aimed to comprehensively investigate speech features for ASD diagnosis, incorporating acoustic and linguistic aspects with a particular emphasis on the Hindi language context. The linguistic features included lexical, syntactic, semantic, and pragmatic dimensions of speech. Our underlying hypothesis for this analysis was that linguistic behaviours associated with ASD might display variations within different language contexts. To address the lack of available speech data for Hindi-speaking CwA to guide the analysis, we conducted data collection specifically for this study. Our objective was to gather a suitable dataset that would allow us to explore and analyse the linguistic behaviours associated with ASD in the Hindi-speaking population. The subsequent sections provide detailed information about the data collection process and experimental settings employed for this investigation.

5.3.1 Ethical Considerations

The procedure for the study was approved by the Institutional Review Board (IRB) of both the affiliating institutions of the researchers (IIITD/IRB/11/13-/2020-4, IEC-644/03.09.2021, RP-01/2021). Before the interviews, the teachers/parents of the participants were provided with an overview of the study, and written consent was obtained from the parents to participate and use data for scientific research. Participating children were informed about the study procedure, and verbal consent was obtained before the commencement of data collection.

5.3.2 Participants

Participants aged 4 to 14 were recruited for the collection of conversational speech data from the All India Institute of Medical Sciences (AIIMS), New Delhi, and Deepalaya Learning Centre, New Delhi. The data collection was conducted from May 2022 to February 2023. Potential ASD participants were identified by the expert team from AIIMS after the screening. During their visit, informed written consent was obtained from the parents of the participant children, and assent was obtained from the children to take part in the procedure. The children were of Indian ethnicity and were native Hindi speakers. All children in the ASD group were assessed by experienced child psychiatrists and clinical psychologists with a minimum of 3 years of experience as three years who confirmed their diagnosis based on DSM-V criteria. ASD was ruled out in the TD group by the clinical experts. Out of the 33 ASD participants, 15 participants were selected for the analysis. The rest of the participants were excluded due to failure to complete the sessions, poor data quality, and non-alignment with the inclusion criteria. Similarly, out of the 25 TD candidates who were recruited, 18 participants were selected for the final study. The age range of ASD candidates was between 4 to 14, and that of TD candidates was 4-10. The age range of the ASD group was kept slightly higher than the TD, considering the diminished verbal ability in the ASD group.

Table 5.5: Participant details for speech analysis

Category	**ASD**	**TD**
No. of participants	15	18
Age group (years)	4-14	4-10
Male-Female count	12-3	10-8

5.3.3 Data Collection Procedure

To keep the conversations close-ended and for a fair comparison, a diagnostic context was set for the data collection procedure. For the data collection, we designed a session with eight diagnostic activities: response to name-call, conversation, joint attention, functional imitation, the anticipation of routine with objects, turn-taking, facial expressions, and free play, which are some of the activities administered for the diagnosis of ASD. These activities were selected from the AIIMS-modified INDT-ASD [8]. INDT-ASD is a culturally coherent diagnostic tool used for the ASD diagnosis developed to meet the specifics of the Indian community and is validated against items from the DSM-5, which is considered the gold standard for the diagnosis of ASD. We strategically utilised the INDT-ASD tool for data collection in the Indian context. This decision was made after considering the potential limitations of ADOS-like tools that could arise due to their lack of cultural sensitivity, as recommended by the literature surrounding autism diagnosis. Our approach allowed us to select tasks that were more appropriate and relevant for the Indian population, leading to more accurate and meaningful data collection. It further helped to identify culturally dependent speech features like asking questions, metaphors, etc., for autism diagnosis in Hindi-speaking children. The activities were planned as simple games, and the adult conversation partner (therapist/caregiver) provided frequent prompts during these tasks to evoke verbal responses. Each interaction session has an average duration of ∼10 minutes. The parent/caregiver of the child was also present during the session, but they were instructed to minimise the prompts or interventions during the child-partner interaction unless there is any critical situation requiring their intervention arises. The speech data of the conversation between the child and the conversation partner was recorded using a camera and microphones. The conversation was carried out in the Hindi language. The recorded videos were carefully transcribed manually following Systematic Analysis of Language Transcripts (SALT) guidelines [285]). Two annotators were involved in the processing of each video. One junior annotator transcribed and annotated the video, and the senior annotator reviewed the transcriptions and annotations for corrections and consistency. Statistical analysis was deemed infeasible due to the complexity of the annotation task with multiple complex speech behaviours and longer conversation data. Hence, motivated by the literature [286, 287], we employed qualitative methods to evaluate inter-rater agreement, specifically consensus meetings. Any areas of disagreement between the two annotators were discussed and resolved. The final dataset contained a total of 320 minutes from 33 children (ASD - 15, TD - 18). The details of the participants are given in Table 5.5.

5.3.4 Features

For the speech analysis in the Hindi language, along with the semantic and pragmatic features mentioned in Phase I, we extracted the linguistic and acoustic speech features from dyadic Hindi conversations between children (with and without ASD) with a conversation partner. The details of the extracted features are in the Feature Extraction section.

5.3.5 Feature Extraction

- **Baseline Features** Similar to the previous study [288], we use N-gram features as the baseline features for ASD classification. N-grams give a preliminary impression of the vocabulary of the child, indicating the language capabilities and understanding of the child, while they lack to represent the nuances of communicative behaviour. For this study, we were interested in the language capabilities of children and hence extracted unigrams and bigrams from the conversational data. We preprocessed the transcripts to remove punctuations and stop words, and further case normalisation is done before the extraction of N-grams using NLTK [280] library. Due to the diversity in conversation, the dimensions of n-gram features were high, and hence we selected the n-grams with a count above the minimum threshold frequency. Finally, there were ten such n-gram features selected for analysis.

- **Linguistic Features** We extracted the lexical, syntactic, semantic, and pragmatic linguistic features from the child's dyadic conversation for the ASD classification. The number of different word roots (NDWR) and unintelligible utterance proportions were extracted from child utterances to represent the lexical diversity in the language used by ASD and TD children. Features such as part-of-speech (POS) tags, the mean length of utterances (MLUM), and unexpected words were extracted. We further quantified semantic-pragmatic features like echolalia, repetitive use of words, use of metaphors, discourse coherence among child's utterances as well as child-partner utterances, and ability to ask questions. These features were extracted using tools and libraries supporting Hindi/Hinglish language like iNLTK [289] and Stanford NLP [290] and modules developed by the authors. For the POS tagging, we used the universal POS (UPOS) tags by Stanza tool [291]. Adding these features helps in better interpretability of predictions on conditions of autism. Overall, the linguistic features extracted were 24, excluding the N-grams. The details of the feature extraction are as follows.

- **Lexical Features**
  - *Unintell Prop:* During the transcription process, unintelligible words were transcribed as per the SALT guidelines. Then, the proportion of these words was calculated by dividing the number of unintelligible words by the total number of words.
  - *NDWR:* First, the stemming of the words was performed. The stemmer used was the same as the one used for extracting n-grams. After the stemming, the number of distinct words spoken by the child was returned.
- **Syntactic Features**
  - *MLUM:* Using the morfessor and polyglot library, we were

able to extract all the morphemes from the Hinglish data. Then the mean length for these morphemes was returned.

* *POS tags:* Using the Stanford NLP group's pipeline from Stanza Library, we extracted the UPOS tags for our Hinglish data. Because of the lack of resources for Hinglish, we had to first transliterate it to Hindi so that it could be used with the existing library.

* *Unexpected Words:* We used the *TfidfVectorizer* from the *sklearn.features_extraction* library. A list of Hinglish stop words was prepared using a repository from GitHub. The vectoriser was fitted over all the utterances by the child. Then, using the vectoriser, each word was converted into a vector. The vectoriser automatically removes all the stop words and then tokenises and vectorises the remaining words. A TF-IDF vector value of over 0.9 was counted as unexpected. The final value returned was the number of unexpected words divided by the total number of words.

– **Semantic and Pragmatic**

The extraction of semantic and pragmatic features is similar to phase I, but to incorporate the specific characteristics of the Hindi language, we developed separate modules for extracting features in Hindi. The code used for the feature extraction will be made publicly available on request.

- **Acoustic Features** We segmented the conversation into child segments, partner segments, overlapping segments, and non-speech segments. Since we were interested in the child's speech characteristics for the ASD diagnosis, the child's conversation was considered for the acoustic feature extraction. The segments where the child interacted with their parents or the parent prompting is involved are removed for the analysis. The acoustic features included in the extended Geneva Acoustic Parameter Set (eGeMAPS) were extracted using openSMILE using the eGeMAPSv02 [292] configuration file with a sampling rate of 44.1 kHz. The acoustic characteristics that comprise the extended Geneva Acoustic Parameter Set (eGeMAPS) were obtained by utilising openSMILE [293] software using the eGeMAPSv02 [15] configuration file, with a sampling rate of 44.1 kHz.

  The eGeMAPS contains spectral parameters Mel-frequency cepstral coefficients (MFCC) 1–4, Spectral flux and Formant 2–3 Bandwidth, Equivalent Sound Level, Voiced and unvoiced region inclusions in addition to the minimalistic parameter set, which includes 18 Low-level descriptors (LLD) grouped into parameters associated with frequency, amplitude/energy and spectral information. Further, temporal features of voiced and unvoiced regions are added along with functionals (arithmetic mean and coefficient of variation) on all these LLDs. the

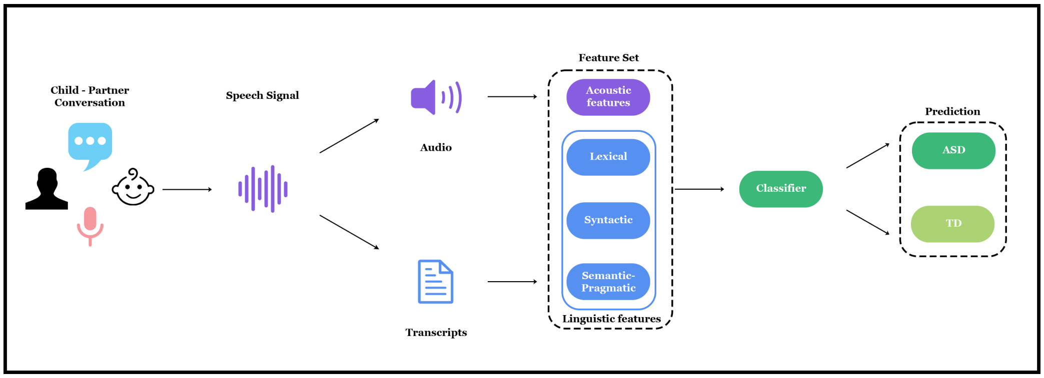


Figure 5.3: System architecture for audio-linguistic feature extraction and classification for ASD.

features in eGeMAPS add up to 88 features. The details of the features are given in Table 5.6.

5.3.6 Experiments

To analyse the linguistic behaviours in English and Hindi and their language-specific manifestations, we considered the transcripts of conversation data collected from 76 English-speaking children from the CHILDES dataset and 33 Hindi-speaking children. The system architecture is depicted in Figure 5.3.

We analysed all the extracted linguistic features as mentioned in Table 5.2 using different models. Similar to the experiment setting in Phase I, we evaluated the classifiers that have been commonly found in the literature for the speech-based diagnosis of ASD with all the extracted linguistic features for the analysis. As a result of the analysis, the Random Forest classifier with 30 *estimators* and *gini* criterion was chosen as it was well-suited to handle the

Table 5.6: Acoustic features used for diagnosis. (The numbers in parenthesis show the no. of features in each category)

Feature Name	Description
eGEMAPS (88)	Extended Geneva Minimalistic Acoustic Parameter Set. It is a set of acoustic parameters developed to provide a compact yet comprehensive set of features that capture various aspects of the human voice [292].

non-linear association between the features and the target variable, thereby giving the best performance among the classifiers. The common practices in ML were used for the training. With a relatively larger dataset of 109 children, we followed a standard train-test split of 80:20 at subject level, further ensuring that there is no overlap of users in the split, and to give an estimate of true error, 5-fold cross-validation is performed.

Following this, to identify the speech biomarkers in Hindi, we performed a comprehensive analysis of the speech features considering the acoustic and linguistic features. With 33 children as the participants, our sample size is limited, and using all the extracted linguistic and acoustic $\mathbf{(24+88 = 112)}$ features may overfit the data. Hence, we performed exploratory data analysis followed by Spearman's correlation analysis to select the most representative features for the diagnostic task by finding the monotonic relation between different features. The details of this analysis can be found in Figure 5.4 and Figure 5.5.

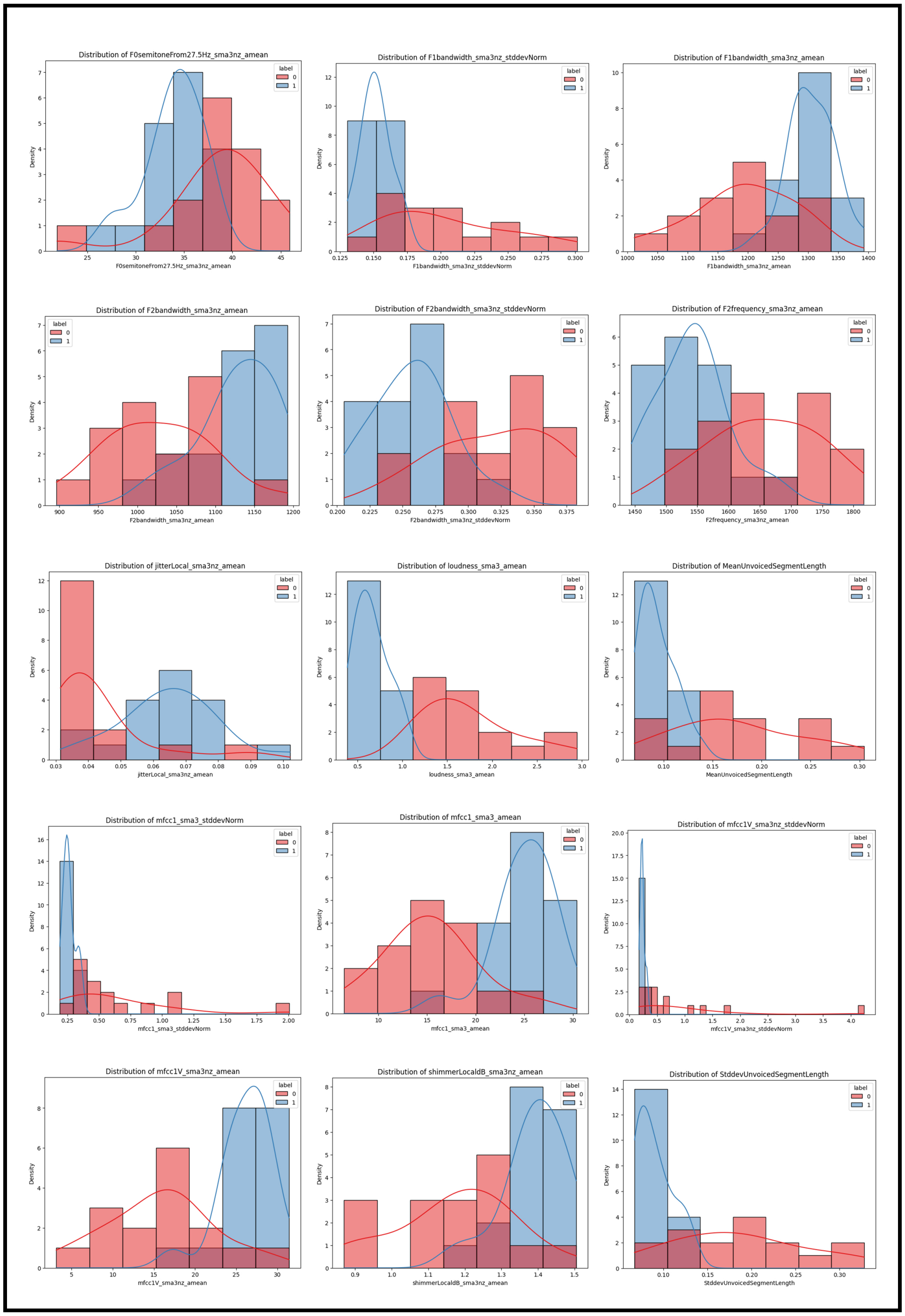


Figure 5.4: Analysis on Acoustic features (top-15)

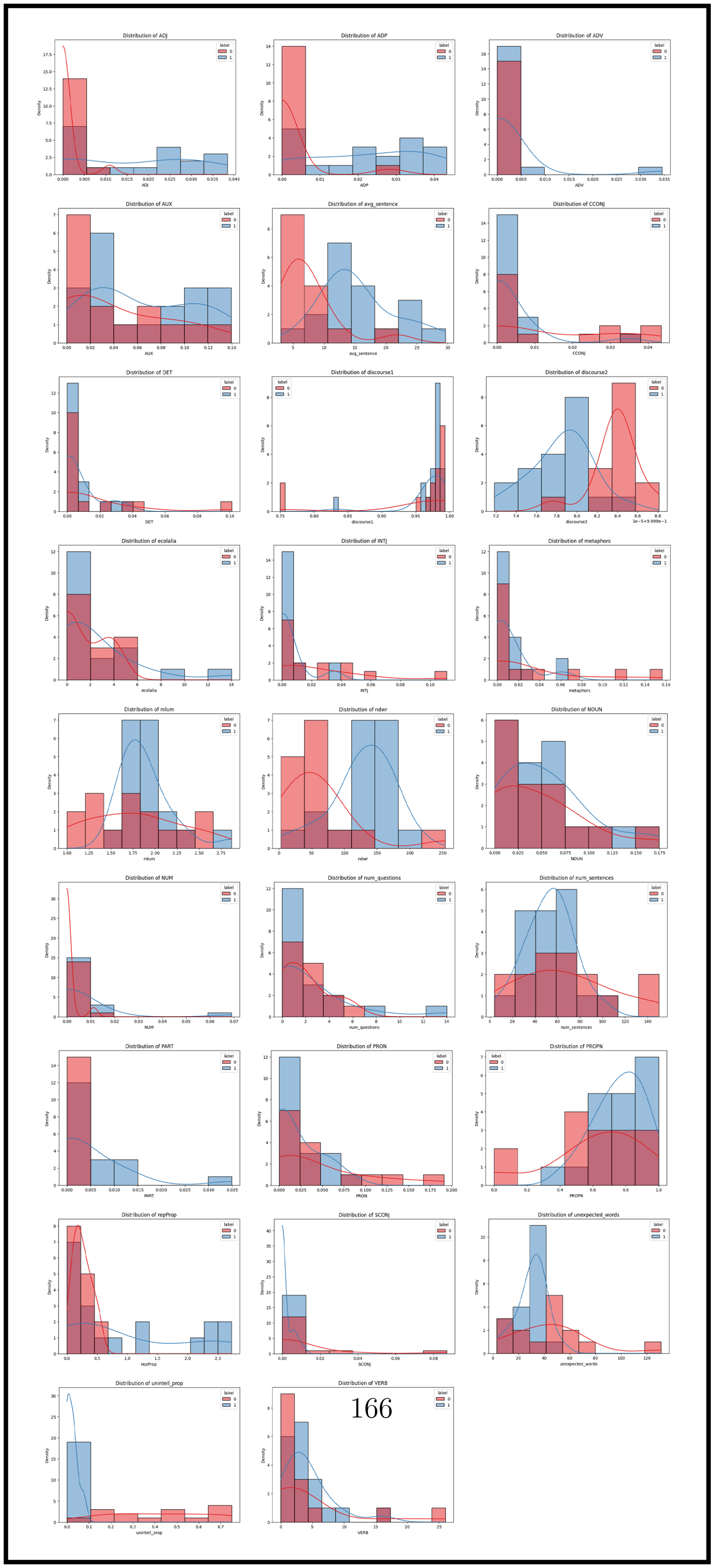


Figure 5.5: Analysis on Linguistic features

Hence we performed Spearman's correlation analysis to select the most representative features for the diagnostic task by finding the monotonic relation between different features. We then applied correlation analysis separately on linguistic and acoustic features to better select features for ASD classification based on their relative contribution towards classification. In this process, POS tags such as SCONJ (subordinating conjunctions) representing the subordinating conjunction, DET (determiner) representing the determiners like *'yeh'*, *'vah'*, metaphors, etc. were some of the features eliminated after the correlation analysis. Further, the acoustic features were reduced to a set of 15 features, which, along with a reduced set of 6 linguistic features, formed the final feature set of 21 features.

In conducting the analysis, we utilised settings that were similar to those employed in the previous experiments. The ML classifiers which were commonly used for the speech analysis, namely SVM (87.3%), ANN (55.38%), Random Forest (91.30%), KNN (82.35%), and Linear Regression (86.91%) were evaluated, and hyper-parameter tuning is done to identify the best model for the data. After evaluating their performances, the random forest classifier with 30 *estimators* and *gini* criterion was selected as it was performing best capturing the non-linear characteristics of the features, with an F1 score of 91.3 ($\pm$6.7) and 95% confidence. We used a train-test split of 80:20, further ensuring that there is no overlap of users, and to give an estimate of true error, 5-fold cross-validation is performed. Additionally, for the feature selection, we used training data so that the test data remained unseen by the classifier. This approach is commonly used to prevent data leakage and to ensure that the classifier generalises well. The individual and cumulative contributions and the complementary nature of features were analysed using different permutations of features for the classification. The entire workflow of the experiments and analysis is given in Figure 5.6.

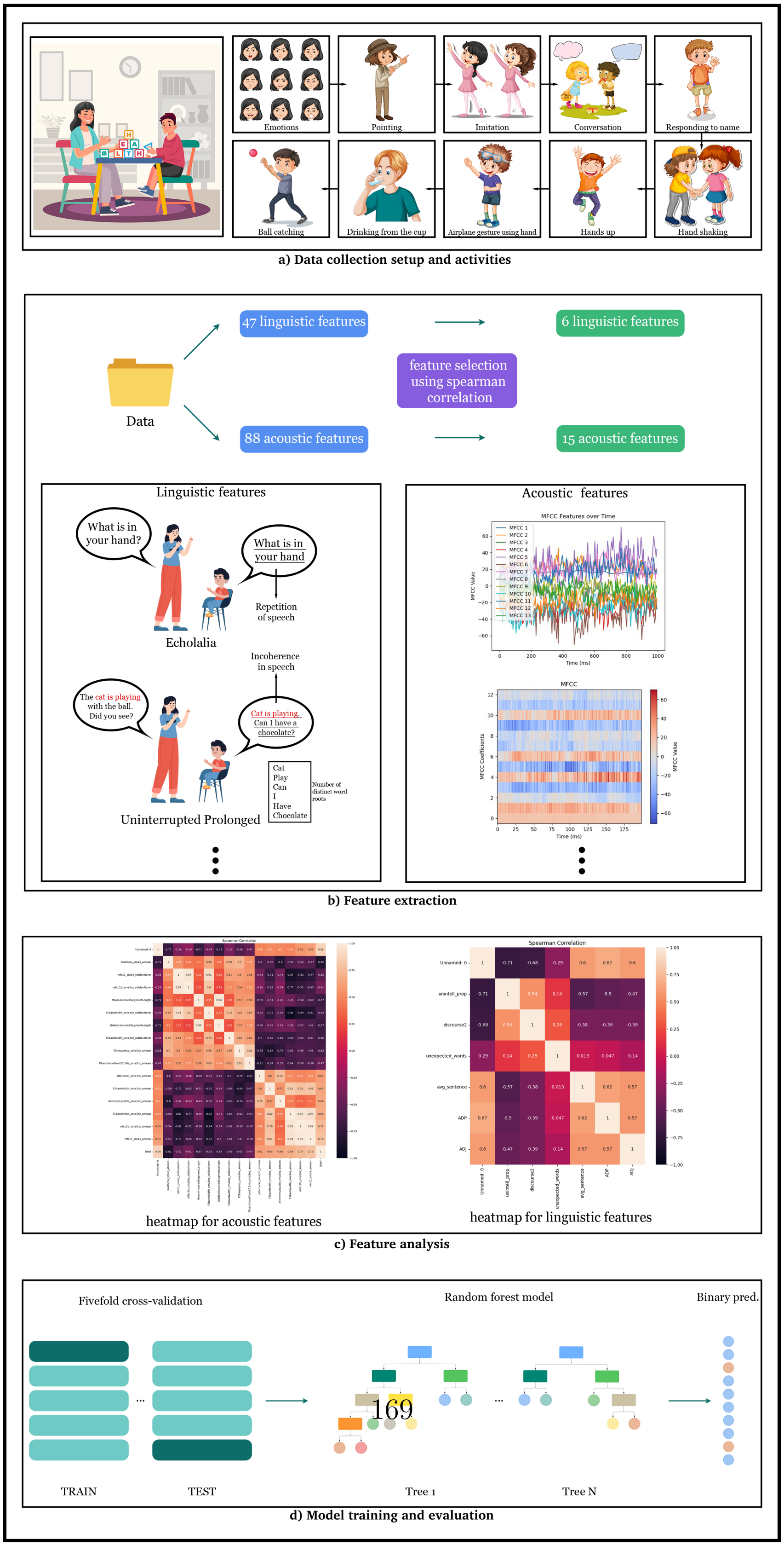
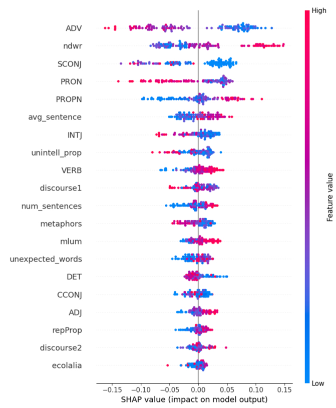


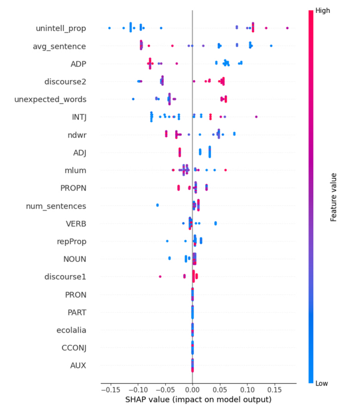
Figure 5.6: Analysis of features for classification of ASD vs TD

5.3.7 Results and Discussions: Cross-linguistic Speech Analysis for Hindi-Speaking Children

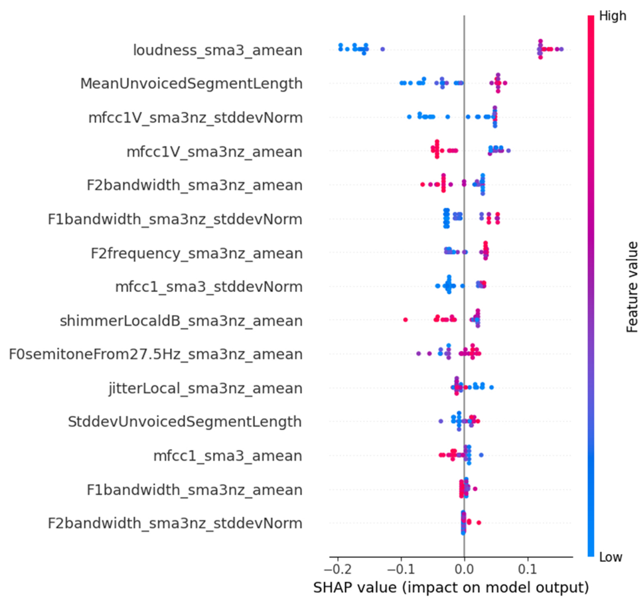
In Phase II of our investigation, our primary focus was to explore speech biomarkers for diagnosing ASD, specifically in Hindi-speaking individuals. As mentioned before, prior to delving into this objective, we analysed linguistic behaviours and their manifestations in both Hindi-speaking and English-speaking populations. To this end, we analysed the predictive contribution of linguistic features to the diagnostic state of the children. We observed an accuracy of approximately $\sim 87\%$ for the prediction of ASD in Hindi-speaking children with linguistic features as predictive factors, and for the English-speaking group, the accuracy of prediction was $\sim 75\%$. These findings suggest initial evidence of differences in the manifestation of linguistic atypicalities in ASD. To gain further insights, we performed an analysis of the individual feature contributions to the classification outcome using SHAP analysis. The results of the SHAP analysis, depicting feature importance in identifying ASD within the Hindi and English groups, are presented in Figure 5.7 (a1) and Figure 5.7 (a2), respectively.



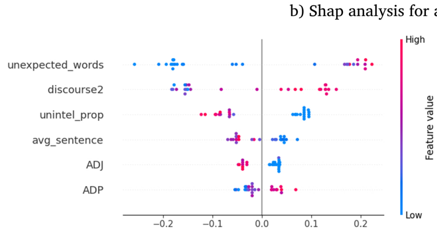
a1) Cross linguistic feautres analysis for ASD diagnosis for English Group



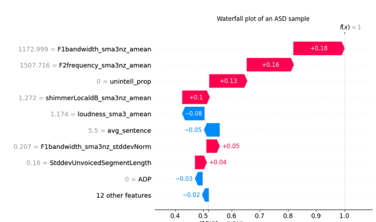
a2) Cross linguistic feautres analysis for ASD diagnosis for Hindi Group



b) Shap analysis for acoustic features



c) Shap analysis for linguistic features



d) Waterfall plot for an ASD participant

Figure 5.7: Analysis of features for classification of ASD vs TD

As shown in Figure 5.7, specific linguistic features were of particular significance in the diagnostic classification of ASD for Hindi-speaking children in comparison to English-speaking children. Specifically, features such as *unintell_prop*, the average length of utterances, POS tag representing the use of adpositions (*ADP*), and *discourse coherence* were found to contribute to the classification in Hindi-speaking children. On the other hand, features representing the use of adverbs (*ADV*), extend of vocabulary (*ndwr*), POS tags like subordinate conjunctions (*SCONJ*), and pronouns (*PRON*) exhibited greater importance in the classification of ASD for English-speaking children. These variations in performance could potentially be attributed to the differences observed in the grammatical structures and linguistic characteristics between Hindi and English. English places a greater emphasis on word order and word choice, whereas Hindi exhibits a more flexible word order and relies on inflections and post positions [294]. Hindi's more flexible word order and sentence structure allow for greater variability in expression and communication styles. For example, Hindi adpositions can change their form to agree with either the noun or the verb in the sentence ("mere papa" and "meri maa"), enabling more nuanced and varied expressions compared to English adpositions, which remain fixed ("my father" and "my mother"). This flexibility poses challenges for individuals with verbal impairments, such as those with ASD, who must grasp and apply grammatical rules related to gender, number, and case agreement.

This flexibility can contribute to a higher accuracy in predicting ASD because the model can capture a wider range of linguistic features that are indicative of the disorder, which is also found in our results. Conversely, English has stricter grammatical rules and less variability in sentence construction compared to Hindi. This can sometimes limit the model's ability to detect subtle linguistic cues or variations in communication patterns that are relevant for ASD prediction.

In our analysis, we leveraged these linguistic differences to tailor our model's features and training methods to each language group accordingly. We ensured that the model was sensitive to the specific linguistic characteristics of Hindi and English, which likely contributed to the observed differences in prediction accuracy. Moving forward, we acknowledge the importance of further exploring how linguistic variability impacts prediction accuracy across different languages and dialects. Further, it is crucial to consider the inherent variability in age and verbal ability among the children when interpreting these results. By validating our hypothesis that linguistic behaviours in ASD vary with language, our study underscores the importance of leveraging language-specific methods for diagnostic classification. Despite the presence of multiple parameters influencing these analyses, our study offers preliminary evidence to explore cross-language analysis for ASD classification.

Guided by our results in understanding the variations of linguistic behaviours in Hindi, we extended our investigation to explore speech biomarkers for ASD diagnosis in Hindi. We extracted and analysed the acoustic features along with the linguistic features for a comprehensive characterisation of speech behaviour. The classification performance evaluation using the different sets of speech features identified is presented in Table 5.7.

Table 5.7: F1-score of Random Forest classifier with different feature sets

Model Name	**Random Forest (%)**
Majority classifier	57.00
N-grams (baseline) features	60.87
Linguistic	86.95
Acoustic	78.26
Linguistic + Acoustic	**91.30**
N-grams + Linguistic + Acoustic	78.26

We assessed the classifiers based on the macro F1-score, considering the N-grams (baseline), linguistic, and acoustic features separately. Additionally, we examined the performance when these features were combined together. Furthermore, we evaluated the classifier's performance when all features, including the baseline features, were considered collectively. Even with individual feature sets, the classifier performed better than the majority classifier. We can see that the linguistic and acoustic features together gave an F1-score of $\sim 91.3\%$ in classifying the children with or without ASD. The linguistic and acoustic features together are found to be better predictors than considered individually, with an average improvement of $\sim 8.95\%$. With the baseline N-gram features added to the overall feature set, it is seen that the accuracy reduced slightly, which is expected due to the curse of dimensionality as the feature dimension increases considerably.

Further, the results of the individual contribution of linguistic features were analysed using SHAP analysis and are shown in Figure 5.7(b). Speech features like *unexpected_words*, *unintelligible utterances*, *discourse coherence*, *adpositions*, *adjectives*, and *average sentence length* were identified as most representative of speech characteristics in individuals with autism. These findings underline the impairments seen in individuals with autism. Individuals with autism exhibit impairments in building semantically coherent conversations and expressing themselves clearly [295]. Our analysis aligns with these observations, as speech features like *unintelligible utterances* and *discourse coherence* emerged as top contributors in the classification decision.

Similarly, CwA faces language and communication difficulties, creating challenges in the use of different components of language, which is evident in the significant role played by *adpositions* and *adverbs* in the diagnosis [296]. Further, the difficulty faced by individuals with autism in organising and structuring their thoughts coherently is reflected in our analysis through the presence of the *average length of sentences* feature [297]. Moreover, the analysis of acoustic features, as shown in 5.7(b)and 5.7(d), emphasises the importance of speech rate and loudness features in differentiating CwA from TD. This finding is consistent with prior research indicating impairments in speech rate entrainment among individuals with autism [298]. In summary, the above analysis effectively showcases the model's capability to accurately capture and utilise relevant speech features for the classification of ASD.

5.4 Conclusion

The primary objective of our study was to improve the interpretability of automated diagnostic predictions for CwA through the analysis and integration of semantic and pragmatic language features. Leveraging sophisticated AI techniques, we quantified and validated the significance of semantic and pragmatic speech features in automated ASD diagnosis. These features outperform the commonly used speech features such as POS, syntactic complexity, and morphosyntactic features when considered individually. The cumulative integration of these speech features further positions them as potential biomarkers for the accurate automated diagnosis of autism in children. Their inherent relevancy makes them accessible to experts and non-experts, fostering a deeper interpretability of the automated diagnosis and its implications. The results obtained from the speech analysis conducted in the English language highlight the importance of incorporating semantic and pragmatic features in addition to other linguistic features for ASD diagnosis. To further evaluate the relevance of the findings from English-speaking children in the context of ASD diagnosis in the Hindi-speaking population, we conducted additional investigations. The main objective was to benchmark the performance of the automated diagnostic model for Hindi-speaking CwA. As Hindi differs from English and other languages in various aspects, including word order, grammatical gender, vocabulary, etc., we investigated and identified the variations in speech behaviours in Hindi from English. In order to achieve this, we developed appropriate methods to effectively extract linguistic features, including semantic and pragmatic features, from Hindi speech as we discovered that extracting Hindi-specific linguistic features, such as discourse coherence, echolalia, etc., presents challenges due to its unique semantics and pragmatics. It is also found that feature extraction in Hindi is not natively supported by commonly used speech analysis tools like CLAN and Linguistic Inquiry and Word Count (LIWC) [299]. Further, our comprehensive analysis of the speech data demonstrated that the acoustic and linguistic features, including lexical, syntactic, semantic, and pragmatic speech features, are reliable predictors of ASD with a performance score of $\sim 91.3\%$ F1-score. In the absence of previous work in the Hindi language, we performed an ablation study with different sets of features to further validate our hypothesis. We observed that acoustic and linguistic features, including semantic and pragmatic features, together performed significantly better with an average improvement of $\sim 9\%$ compared to using only acoustic or linguistic features for the diagnosis. Furthermore, the identified semantic pragmatic features align with established diagnostic procedures, such as the INDT-ASD tool, thereby providing further validation of their potential in the Indian context. Despite being the most widely spoken language among over 325 million native speakers in India and ranking as the 5th largest spoken language worldwide, speech analysis in Hindi remains relatively unexplored in the context of ASD research. Consequently, our study holds significance as it is the first to conduct speech analysis for ASD diagnosis in the low-resource Hindi language.

In summary, our study contributes to the advancement of automated ASD diagnosis by emphasising the significance of semantic and pragmatic speech features. It further enhances our understanding of speech-based indicators for the automatic diagnosis of autism in the Hindi-speaking population. Further research could investigate the factors that may influence the varied manifestation of linguistic atypicalities in ASD across different languages and populations. By leveraging these speech-based language biomarkers, our findings have the potential to improve early identification and intervention strategies for individuals with ASD, ultimately leading to more effective outcomes, particularly for those from linguistically diverse backgrounds.

In Chapters 2 and 3, we explored the feasibility of using social robots in autism interventions from the perspectives of both children and special educators. In Chapters 4 and 5, we examined the facial expression and speech-based behaviours of CwA. Building on these behavioural insights, we now move forward to evaluate the AI-assisted system for assessing children for ASD diagnosis. The next chapter provides a detailed discussion of the pilot study conducted to evaluate the system, both quantitatively and qualitatively.

CHAPTER 6

ROBOT-MEDIATED ARTIFICIAL INTELLIGENCE-BASED SYSTEM FOR OBJECTIVE MULTIMODAL ASSESSMENT OF AUTISM: A PILOT STUDY

Despite a substantial body of research investigating the use of social robots and artificial intelligence-based systems in autism care, the majority of these studies are conducted in the Global North. Limited research in the Global South, including India, has explored the use of artificial intelligence-based solutions or the utilisation of social robots in this context. Notable studies in this direction include Prakash et al.'s exploration of computer vision-based video analysis for assessing and diagnosing ASD in children [300, 301], Rane et al.'s investigation into movement planning in CwA using virtual reality technology [302], and their further exploration of tablet-based interventions to enhance joint attention skills in CwA [14]. Additionally, Raj et al. have delved into machine learning techniques for classifying individuals with ASD from TD individuals using patient screening data [303], while Berlin et al. have employed the RGBPose-SlowFast deep network for identifying stimming behaviour in CwA [304]. However, most of these approaches rely on end-to-end machine learning or deep learning architectures, leading to a lack of interpretability in the decision-making process. Furthermore, in the context of India, there has been limited exploration of the use of social robots in autism care, with only a few modest attempts documented [83–85].

To address these challenges, we embarked on the development of a robot-mediated AI-assisted system designed to conduct multimodal behaviour analysis for identifying speech and facial expression behaviours in CwA. Within this framework, the robot assistant takes the role of administering assessment tasks, while an AI-based behaviour analysis system will be employed to detect pertinent behaviours throughout the assessment process. This integrated approach seeks to support experts in making more informed decisions during assessments for CwA. In the preceding chapters, we investigated the feasibility of a robot-assisted intervention procedure for autism diagnosis. This involved understanding the perceptions of special educators and examining the response behaviour of children of Indian ethnicity, who are key stakeholders in the system. Subsequently, we delved into the application of machine learning and deep learning technologies to analyse and detect speech and facial expression behaviours in CwA, facilitating diagnosis. In this chapter, we demonstrate the performance of the integrated system within the Indian context.

6.1 Methodology

6.1.1 Selection of Robot Assistant

We selected the NAO [1] humanoid robot developed by SoftBank Robotics as our interactive robot assistant, considering its widespread usage in autism care for children [305]. Standing approximately 58 centimetres tall, NAO boasts a humanoid appearance with two arms and legs, along with an array of sensors, cameras, and microphones. Programmable and customisable, NAO can execute a diverse range of tasks, encompassing interaction with humans, language capabilities, facial recognition, non-verbal communication to some extent, and voice recognition. Its interactive and engaging nature renders NAO particularly effective in autism care for children. Programmable to deliver structured and repetitive interventions, NAO caters to the needs of CwA who thrive in predictable environments. Additionally, its humanoid design fosters relatability and minimises intimidation, encouraging children's active engagement in therapy activities. To make the interactions more personal, we named the NAO robot "Chintu", which is a Hindi term of endearment often used for small children.

6.1.2 Selection of Diagnostic Tool

The first step in the development of a robot-mediated assessment system for diagnosis is to identify the tool for diagnosis of ASD in children. We

[1] `https://us.softbankrobotics.com/nao`

conducted interviews with diagnostic experts, psychologists, researchers and medical professionals. We identified DSM-5 based AIIMS (All India Institute of Medical Sciences) modified INDT- ASD (INCLEN Diagnostic Tool for Autism Spectrum Disorder) tool, which is used to diagnose and assess the severity of autism in India. Using this, the child's behaviour in the 6 domains- Social relationship and reciprocity, Emotional responsiveness, Speech-language and communication, Behaviour patterns, Sensory aspects, and Cognitive component will be assessed for aiding the diagnosis. For the diagnostic protocol, we derived the tasks adhering to the domains specified by INDT-ASD.

The tasks that have been identified for the diagnosis procedure follow a structured behavioural approach called discrete trial training (DTT). Working from the assumptions of this approach, the training tasks take place in a highly structured learning environment and are directed by the interaction partner.

6.1.3 Diagnostic Task Protocol

We included Response to name, Response to joint attention, Functional and Symbolic imitation, Turn-taking, Facial expression, Anticipation of a routine with an object, Pointing and Conversation. These activities were selected from the INDT-ASD [8] tool recommended by the experts. Since the participants were native Hindi speakers, the interactions were planned in Hindi language. The protocol (in English Version) is detailed below:

Conversation

1. NAO: *"Hi I'm Chintu ! I'm from the HMI lab, IIITD. Hi <name of the researcher>! How are you?"*

2. Researcher: *"I'm good Chintu, how are you?"*

3. NAO: Hi I'm good, thank you. Hi <name of the child>! How are you?

4. Wait for the child to respond.

5. If the child responds, NAO says *"Good to hear that"*

6. else NAO says, *"It's okay, You know I can do activities like dance, frog jump, teach you body parts and so on. Would you like me to show you some?"*

7. If the child responds, Chintu asks, "*Which activity do you prefer to try?*" and lists out the activities.

8. else says, "It's okay. It was nice meeting you. Let's meet another time."

Response to name

1. Distract the child with a puzzle game/toy

2. Call the child by his/her name/nickname /pet name/family name

3. If the child responds with eye contact/head movement or verbally, stop the task

4. Else repeat the call 3 times
5. If no response in 3 calls, call the child with reinforcement, (with mention of toys/characters, or music or lights)
6. Repeat reinforced call 2 times
7. End the task

Response to joint attention

1. Direct the attention of the child to NAO by calling the child by name/ with lights/sound/actions
2. Two objects are kept on the right and left side of the child
3. With head movement, pointing and verbal prompt (Repeat 2 times)
4. NAO turns and points towards the object kept on the right and says "Look! Is this a *book or eraser*"
5. NAO turns and points towards the object kept on the left and says "Look! Is this a *book or eraser*?"
6. End the task

Functional imitation

1. Direct the attention of the child to NAO by calling the child by name/ with lights/sound/actions

2. NAO instructs the child to imitate its actions

   Action 1

   (a) Ask for a fist bump

   (b) If the child responds and imitates, stop the task, else

   (c) Repeat the task 3 times

   Action 2

   (a) NAO moves the hand to depict a frog jump

   (b) If the child responds and imitates stop the task, else

   (c) Repeat the task 3 times

Anticipation of a routine with object

1. Direct the attention of the child to NAO by calling the child by name/ with lights/sound/actions

2. Asks the child to play “peek-a-boo”/”chuppachuppi”

3. NAO holds a card and covers face for 3 seconds

4. Lowers the card and says “BOOM”

5. Repeats this 3 times

6. Repeat the fourth time but the face is kept covered for 10 seconds

7. Monitor for child response (look for gesture/facial expression)

8. Repeat this 3 times

9. End the task

Vocabulary: learning body parts

1. Direct the attention of the child to NAO by calling the child by name/ with lights/sound/actions

2. NAO touches it's head

3. NAO asks: "Can you touch your head?"

4. Waits for 10 sec to respond

5. If the child doesn't respond, give repeated prompts "Let us try again, Can you touch your head?"

6. If the child doesn't respond after prompts Repeat 2 times

7. This is repeated for eyes, chest and hands as well.

8. End task

Emotions

1. Display stimuli (facial expressions on the card)

2. Ask the child to imitate the emotion displayed on the card

Figure 6.1: Experiments

3. Wait for 10 seconds for child response

4. Repeat this 5 times

5. End task

Imitation by action song

1. NAO performs the "Rolly-Polly" rhyme with actions and asks the child to imitate

2. Wait for 3 secs for the child's response

3. Repeat this 2 times

4. End task

The snapshot of the experimental setup is as shown in Figure 6.1

6.2 Ethical Considerations

The procedure for the study was approved by the Institutional Review Board (IRB) of both the affiliating institutions of the researchers (IIITD/IRB/11/13 /2020-4). Before the interactions, the teachers/parents of the participants were provided with an overview of the study, and written consent was obtained from the parents to participate and use data for scientific research. Participating children were informed about the study procedure, and verbal consent was obtained before the commencement of data collection.

6.3 Participants

We recruited a total of 21 participants (ASD - 9, TD - 12, Female:Male - 7:14) aged 4 to 14 for the interaction sessions with the NAO robot and subsequent behavioural data collection at the Deepalaya Learning Centre in New Delhi. Data collection took place between October 2023 and January 2024. Potential ASD participants were identified by the expert team from Deepalaya under the Sambhav project following screening procedures. Prior to participation, informed written consent was obtained from the parents of the participant children, with assent also obtained from the children themselves. All participants were of Indian ethnicity and native Hindi speakers. In the ASD group, diagnosis confirmation was conducted by experienced child psychiatrists and clinical psychologists with a minimum of three years of experience, based on DSM-V criteria. Further, ASD was ruled out in the

TD group by clinical experts.

6.4 Data Collection Procedure

To ensure a standardised and equitable comparison, a diagnostic framework was established for the data collection procedure. We devised a session comprising activities outlined in Section 6.1.3, structured as simple games. An adult conversation partner, such as a therapist or caregiver, prompted the child frequently during these tasks to elicit verbal responses. Each interaction session lasted approximately 10 minutes on average. While the child's parent or caregiver was present during the session, they were instructed to minimise prompts or interventions unless a critical situation necessitated their involvement. Each child engaged in the activities with both a human administrator (HA) and a robot assistant (RA). The participants were divided into two groups: one group (TD-6, ASD-5) where the child interacted with HA first and then RA, and the other group (TD-6, ASD-4) where the children interacted with the robot first, followed by the human. A gap of one week was maintained between the two interactions to mitigate any potential effects of familiarity with the activities during the second interaction. Behavioural data, particularly speech and facial expressions, were recorded using cameras and microphones. The conversations were conducted in Hindi. Subsequently, the recorded videos underwent meticulous manual speech transcription following SALT guidelines. Two annotators were engaged in processing each video: a junior annotator transcribed and annotated the video, while a senior annotator reviewed the transcriptions and annotations for accuracy and consistency. The final dataset encompassed approximately 200 minutes of recorded interactions. These videos were later used for qualitative analysis of the response behaviour of the children.

6.5 Data Analysis and Inference Generation

Following the collection of data from the diagnostic sessions, we conducted preprocessing to extract speech features and analysed the speech data to identify speech behaviours in the child, as outlined in Chapter 5. Subsequently, we extracted frames from the video capturing the child's facial expressions during interaction. Leveraging our MSTL-MCA module detailed in Chapter 4, we analysed the facial expression behaviours of the child. The system architecture is as given in Figure 6.2.

After the data analysis, a report was generated for each of the children, samples of which are given in Tables 6.1 and 6.2

6.6 Results and Discussion

After applying the FER and speech analysis modules, we extracted speech and facial emotion features of the child during the interactions. Subsequently, we applied ML classifiers commonly used for ASD classification, namely Support Vector Machines (SVM), Random Forest, and K-nearest Neighbours (KNN). These classifiers were evaluated using Leave one out cross-

Table 6.1: Sample report of a TD child session with the robot. {The report indicates coherence in discourse, further with the presence of 5 unexpected words. Additionally, there's a noted high average conversational unit length, along with the use of adjectives showcasing verbal communication proficiency. According to the AIIMS-modified INDT-ASD tool, these factors suggest TD behaviour.}

Text Features	
Unintell	0.0
Discourse	1.00
Unexpected Words	5.0
Avg C-unit Length	22.71
Adposition	0.0
Adjectives	0.01
Facial Emotion Recognition	
Percentage expression of emotions	Anger:22.21 % Fear: 2.67% Happy: 2.71% Neutral: 30.95% Sadness: 14.36% Surprise: 27.1%
True Label	TD
Predicted Label	TD

Table 6.2: Sample report of an ASD child session with the robot. {The report reveals a lack of coherence in discourse (value-0), along with a notably high presence of unexpected words (14). The average conversational unit length is 2, and there is an absence of adpositions or adjectives, indicating limited communication ability. Based on the AIIMS-modified INDT-ASD tool, these factors point towards characteristics consistent with ASD behaviour.}

Text Features	
Unintell	38
Discourse	0.0
Unexpected Words	14.0
Avg C-unit Length	2.0
Adposition	0.0
Adjectives	0.0
Facial Emotion Recognition	
Percentage expression of emotions	Anger: 29.05% Fear: 6.03% Happy: 3.04% Neutral: 32.54% Sadness: 8.58% Surprise: 20.77%
True Label	ASD
Predicted Label	ASD

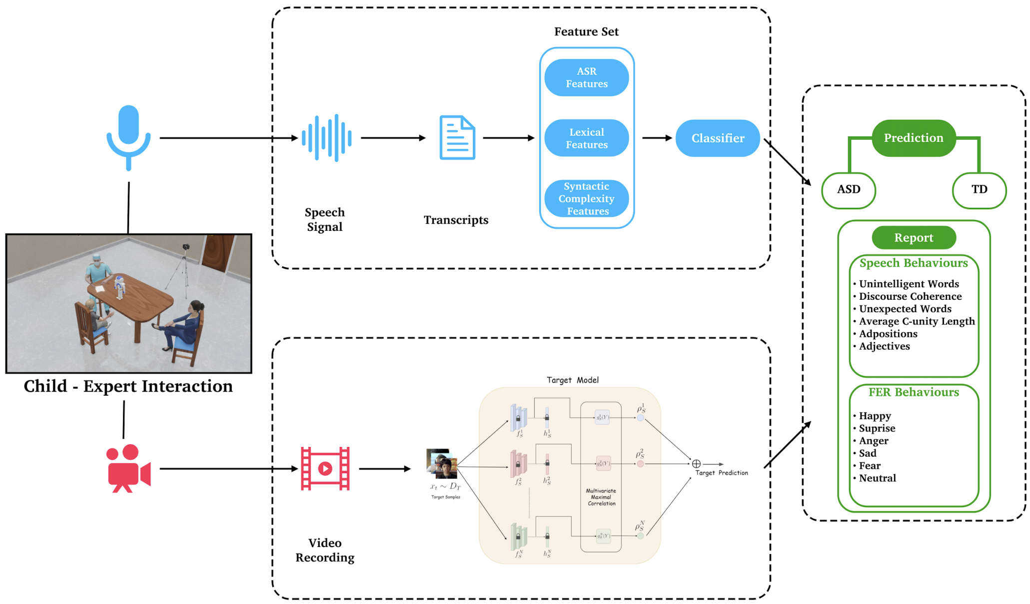


Figure 6.2: System architecture for AI-assisted ASD assessment

validation (LOO-CV). Among these classifiers, Random Forest achieved the highest performance, with a mean accuracy of 76.19%, which is significantly higher than the chance accuracy. The results reported a precision of 64.28%, recall of 100%, and an F1 score of 78.26%. Hyperparameter tuning was conducted to optimise the models, and Random Forest emerged as the best-performing classifier. The confusion matrix for the classification is given in Figure 6.3. LOO-CV, each sample is used as a test sample exactly once, with the remaining samples used for training. Therefore, the confusion matrix obtained in this context represents an aggregate of the predictions from all individual models generated during the cross-validation process. Therefore, the confusion matrix obtained in this context represents an aggregate

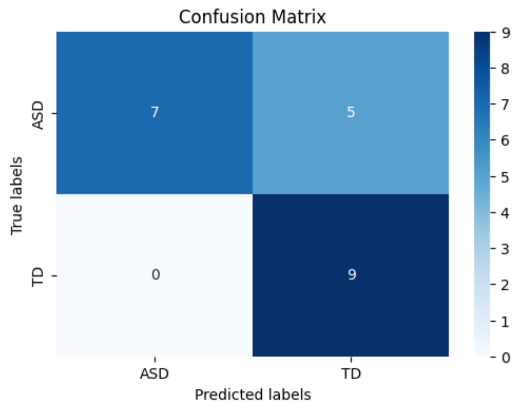


Figure 6.3: Confusion matrix for ASD vs TD classification.

of the predictions from all individual models generated during the cross-validation process. Since we are using LOO-CV, we have reported the cross-validation accuracy by averaging the accuracy over different iterations of cross-validation. Hence, mean accuracy is 76.19% (SD - 42.59). Further, we performed McNemar's test on the predicted and ground truth values to assess the similarity between the predicted and actual outcomes. The McNemar's test resulted in a p-value of 0.074 and a statistic of 3.2. This suggests that there is no statistically significant difference between the predicted and ground truth values at the conventional significance level ($p<0.05$).

During the diagnostic sessions with the robot, 10 out of 12 (TD) children successfully completed all the activities. One TD child completed 6 out of 8 tasks, while another completed 4 out of 8 tasks. On the other hand, among the 9 CwA, 3 completed all the tasks, while all others completed at least four tasks during the sessions. A similar trend was observed during interactions with humans. Specifically, 11 out of 12 TD children successfully completed all the activities, with one completing 6 out of 8 tasks. Among the nine children with ASD, 2 completed all the tasks, while all others completed at least four tasks during the sessions. The ASD children had minimal verbal ability with two non-verbal participants. While the response behaviour of TD children towards tasks administered by both the robot and human were comparable, significant differences were observed in the response behaviour of CwA. Children with ASD demonstrated better eye contact with the robot assistant and exhibited closer interaction with the robot, often engaging in tactile responses towards it compared to their interactions with the human facilitator of the tasks. Moreover, these children demonstrated an enhanced attention span during robot-assisted intervention.

6.7 Conclusion

In this chapter, we showcase the feasibility of employing social robots for interventions in assessing CwA. Furthermore, we illustrate the utility of multimodal behaviour analysis in making behavioural observations for diagnostic decision-making in ASD. CwA exhibits better interaction with a robot partner compared to their human counterpart, a trend consistent with existing literature. However, due to the limited sample size, we acknowledge the lack of generalisability of this finding. Further investigation with a larger sample size is warranted to validate and generalise this observation.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This thesis spans the multidisciplinary domain of human-robot interaction and behavioural modelling. Our efforts have revolved around the intersection of artificial intelligence, machine learning, and related fields to empower the autism community in environments with limited resources.

To this end, our work has focused on the development of a AI-assisted robot-mediated system for the assessment of ASD in children, leading to an objective and reliable diagnosis. In the absence of reliable biological markers, the diagnosis of ASD is often based on behavioural observation, which makes it a cumbersome process. It also makes the diagnostic procedure completely dependent on the subjective analyses of the experts, who are again only sporadically accessible to those affected in some of the world's emerging regions. As previously indicated in existing research, a reliable and objective diagnosis is essential to initiate timely interventions in children with autism, ensuring their efficacy in fostering the necessary cognitive development. Our efforts in this realm have involved modelling relevant behaviours that can ultimately assist experts in diagnosing ASD. This has resulted in the creation of interpretable frameworks for recognizing facial expressions of emotion in children and identifying speech-based behaviours indicative of ASD.

This thesis in particular, has looked into the development of the robot-assisted system from two broad directions:

1. Developing interpretable machine learning frameworks to identify behaviours crucial for the assessment of ASD by monitoring and analyzing robot-assisted diagnostic sessions under scarce data settings.

2. Investigating the feasibility of implementing such a system in resource-constrained settings, particularly in the Global South, with a focus on India.

In developing an AI-based framework for analyzing behaviour in autism, we encountered a significant challenge stemming from the scarcity of relevant datasets due to the sensitive nature of the target population. To bridge this gap, we initially tackled the issue of facial emotion recognition using limited data. This was formulated as a problem of multi-source domain adaptation, where multivariate-maximal correlation was harnessed to amalgamate knowledge from multiple sources for classifying facial emotions. This approach obviated the need for source datasets during training, thereby preserving privacy. Subsequently, we delved into the identification of language and communication impairments, recognized as core features of ASD. In this regard, we addressed the quantification challenge of language atypicalities in autism by extracting and analyzing semantic and pragmatic language features in children with autism, exploring their relevance in diagnosis. We extended these efforts to extract speech features indicative of ASD, particularly in resource-constrained languages like Hindi.

In response to a dearth of research on robotic interventions for children in the Global South, particularly in India, our study explored the response behaviour of 3-6-year-old Indian children during directive tasks with robots to address this knowledge gap. However, since technologies such as robots have often faced numerous issues in being adopted among marginalized communities, our next task was to understand the contextual appropriateness of RAD in resource-limited settings for which we engaged in participatory research work along with collaborators from different disciplinary backgrounds. This study catered to the understanding of how professionals who could use social robots for autism therapy perceived such kinds of technologies in the backdrop of their own inhibitions and the larger social inequities grappling India's socio-economic landscape.

7.1 Future Directions

Some future directions for this work are as follows:

i The current exploration could be extended to developing computational methods to include the identification of stereotypic gestures as indicative markers of ASD under data-scarce settings. Stereotypic gestures, characterized by semi-voluntary, repetitive movements, are significant indicators in the diagnosis of ASD. These gestures, observed in individuals with ASD, encompass behaviours like repetitive hand movements and body rocking. These stereotypic gestures provide valuable insights for clinicians conducting comprehensive evaluations of individuals with ASD. Due to the subtle and varied characteristics of stereotypic behaviours, researchers have turned to computational methods to offer a systematic and standardized means of quantitatively analyzing these intricate patterns. Deep learning has indeed achieved remarkable success in the domain, but its effectiveness is often constrained by a substantial need for labelled images, which can be both resource-intensive and challenging to obtain publicly in the case of autism research.

ii This research could be extended to explore the application of these algorithms for enhancing the effectiveness of interventions in the context of autism. The algorithms could identify behavioural indicators, acting as benchmarks for establishing measurable development goals, supporting continuous progress monitoring, and facilitating the adaptation of therapeutic techniques. Collectively, these contributions aim to improve the overall well-being and quality of life for CwA through more effective and personalized interventions.

iii ASD is a spectrum of conditions where the characteristics of affected individuals can vary significantly in terms of their symptoms and levels of impairment. Therefore, developing accurate methods and standardized assessment tools to identify and classify the severity levels of ASD is crucial. This is essential for improving diagnostic accuracy, tailoring interventions, advancing research, and ultimately enhancing the quality of life for individuals with ASD. Further, accurate differentiation between ASD and other neurodevelopmental disorders (NDDs) that share similar characteristics, such as developmental delays, Rett syndrome, intellectual disabilities, language disorders, and ADHD, is a possible future direction. Differential diagnosis ensures that individuals receive the appropriate diagnosis and interventions tailored to their specific needs, optimizing therapeutic outcomes.

Appendices

APPENDIX A

QUESTIONNAIRE FOR PARENT PERCEPTION TOWARDS ROBOT-ASSISTED INTERVENTION

The following questions were selected to evaluate the perception of parents towards robot-assisted interventions and their application in autism diagnosis. The parents have to score the robot interaction on a scale of 1-5, with 1 being strongly disagree and 5 being strongly agree.

1. The robot can take turns with the children during an interaction
2. The robot can provide feedback to the children during the interaction
3. The robot can play the role of a companion during the interaction with the children
4. The robots can effectively conduct the intervention with the children
5. The robots have the potential to keep the children more engaged
6. The robot mediated intervention has the potential to improve the joint attention of children.
7. The child could comfortably engage with the robot during the intervention.
8. The robots could be used for interventions for skill training

9. It is safe to use robots in interventions designed for children

10. I recommend robot-assisted intervention for children

APPENDIX B

PARTICIPANT DEMOGRAPHICS

Table B.1: Participant Demographics

Serial No.	Respondent	Experience (in years)	Category	Educational Qualification	Exposure to Technological Resources
1	P1	12	Employed at Institute	Bachelor of Education	Computer, Tablet, Mobile Applications, Electronic toys, feedback receiving pens
2	P2	15	Employed at institute	Bachelor of Education	Computer, Tablet, Mobile Applications,
3	P3	16	Door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications,
4	P4	8	Employed at Institute	Bachelor of Education	Computer, Tablet, Mobile Applications,
5	P5	9	Door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications,
6	P6	25	Door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications,
7	P7	11	Door-to-door	Master of Education with specialisation in Special Education,	Computer, Tablet, Mobile Applications
8	P8	14	Employee at institute	Master of Arts in Psychology, Bachelor of Education,	Computer, Tablet, Mobile Applications
9	P9	28	Own Institute	Bachelor of Arts in English, Master of Education with specialisation in Special Education,	Computer, Tablet, Mobile Applications,
10	P10	11	Door-to-door	Master of Arts in Psychology, Bachelor of Education	Computer, Tablet, Mobile Applications,
11	P11	11	Door-to-door	Master of Arts in Psychology	Computer, Tablet, Mobile Applications,
12	P12	13	Employed at institute	Bachelor of Education	Computer, Tablet, Mobile Applications,
13	P13	12	Employed at Institute	Bachelor of Education	Computer, Tablet, Mobile Applications, Social robots
14	P14	16	Door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications
15	P15	7	Door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications,
16	P16	18	Own institute	Bachelor of Education	Computer, Tablet, Mobile Applications
17	P17	15	door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications
18	P18	12	Employed at Institute	Bachelor of Education	Computer, Tablet, Mobile Applications
19	P19	9	Door-to-door	Bachelor of Education	Computer, Tablet, Mobile Applications
20	P20	8	Door-to-door	Master of Arts in Psychology, Bachelor of Education	Computer, Tablet, Mobile Applications
21	P21	7	Door-to-door	MPhil in Psychology, Diploma in Special Education	Computer, Tablet, Mobile Applications
22	P22	19	Own institute	Bachelor of Education	Computer, Tablet, Mobile Applications
23	P23	15	Employed at institute	Bachelor of Education	Computer, Tablet, Mobile Applications,
24	P24	13	Employed at institute	Master of Arts in Psychology, Bachelor of Education	Computer, Tablet, Mobile Applications
25	P25	11	Employed at institute	Bachelor of Education	Computer, Tablet, Mobile Applications

APPENDIX C

INTERVIEW PROTOCOL

- BACKGROUND QUESTIONS
  - Could you please introduce yourself?
  - Could you please tell me about your daily routine at your centre?
  - Could you please tell me about the specific programmes for CwA at your school/centre?
  - What role do you play as a special educator in order to facilitate these activities?
  - What are the challenges that you face in your work?
  - What kind of strategies do you use as a special educator?
  - What kind of tools/props do you use in your activities as a special educator? Do you make use of technological equipment? If yes, then what are they and are you satisfied using them?
  - Do you find the technology that you use or any kind of technology to be helpful in your work? What are your needs and requirements in general from a social robot that you would like to point out?
  - The tools and technologies that you use when it comes to CwA, do they fulfil your needs and requirements?

- What are these tools presently lacking?
- Why do you think they are lacking in your needs and how do you think that can be rectified?

- ANXIETY
  - Are you aware of the use of social robots as interventions by special educators in autism spectrum disorder?
  - Have you ever used one of them, or have seen it being used?
  - Do you think the kind of robot that we showed here would be able to help you in your work?
  - Given a situation in which we give you such a robot as it has been shown in this video, would you be inclined to use it? If yes, then why? If not, then why?
  - What would be confident in using if you were given to use such a robot, in case you have not used a robot before?
  - Would you be comfortable using such a robot in your work?
  - In case you find using such a robot to be an uncomfortable exercise, which part(s) exactly would you find to be uncomfortable?
  - What do you make of social robots, since you have seen in the videos that they can fail, too? Does that make you anxious?

- PERCEIVED USABILITY

- In what way do you think the robot will be helpful towards a) your work? and b) children?
- Do you think the use of social robots would make your work a) easy b) efficient?
- What are some of the activities that you think the robot can do all by itself, where it does not require your help or support? Give us some probable situations. What exactly would be the functionality of the robot in this case?
- What are some of the activities that you can think of that you can do with the robot, that is you and the robot can work together? Give us some probable situations. What exactly would be the functionality of the robot in this case?
- What kind of activities do you think you can let the robot do alone with the children, that is, activities where there would be minimum participation and supervision from your side?
- What do you think could be any other use of social robots apart from therapy?

• CONTEXTUAL MEANINGFULNESS

- What would you think would be the biggest challenges in a country like India if social robots were to be used in autism?
- Can you identify personal challenges on your part if you were to

use social robots?

  - What do you think could be the various ways in which these challenges could be addressed?
  - Personally, would you encourage/not encourage the use of social robots in autism in a country like India? If yes, then why and if no, then why?

- TRUST
  - Would you find a robot to be a trustworthy intervention in your work?
  - In case you find it to be trustworthy, then could you please indicate what aspect about it would make it seem untrustworthy to you?
  - How do you think a) parents, b) peers, and c) children perceive the use of social robots? Would you take into account their perception if/before you are to use it?
  - Let us say that you perceive the robot to be a trustworthy intervention, but others do not (parents/peers). Would you try to influence them to trust the use of social robots or do something that would make them trust it?
  - Do you find SARs to be safe as an intervention when it comes to CwA?

APPENDIX D

NUDGES FOR PANEL DISCUSSION

- Having viewed in the videos as well as in your interactions that a social robot might fail and it is not foolproof, do you still think it will replace you?
- When it comes to working with the robot, how do you view yourself as a specialist? Then, do you think your role changes, or does it have any new dimensions?
- What are your suggestions about making the use of social robots more contextualised in your practice, based on your perception?
- What would be your idea about making social robots more trustworthy as an intervention in your practice?
- What would be your suggestions on balancing the role of social robots in your practice?
- Did you feel that because this is a humanoid robot, it can replace you?
- Can you draw a robot that you might be looking for in case of therapy?
- Could you please explain why you drew such a robot?

- Some of you have highlighted that social robots can help not only in giving therapy but also in performing other activities, such as helping in building an inclusionary environment. How do you think your role as a special educator would transform should you use the robot in a non-therapeutic context such as this?
- How well do you find yourself prepared for handling a social robot in therapy?
- If no, what kind of support/assistance do you expect in making yourself prepared?
- Some of you have pointed out that CwAs can get attached to robots, which might turn out to be dangerous. Do you have some suggestions on how we can use social robots and, at the same time, not have children get attached to them? What do you think would be a healthy trade-off in that case?
- Do you have any other suggestions when it comes to social robots and their use for Autism therapy?

APPENDIX E

CODEBOOK AND THEMES FOR QUALITATIVE ANALYSIS

Theme / Code	Count
Infrastructural Issues & Resource Constraints	**214**
Special educators talk poor infrastructural facilities	12
Special educators talk about financial difficulties	25
Special educators provide instances of resource constraint	23
Special educators outline her concerns about how such infrastructural issues affects them	11
Special educators inform about institutional practices of ensuring responsibility by creating an atmosphere of fear	61
Special educators describe such an atmosphere of fear by signifying instructions from schools to repay the cost of damages if expensive devices are damaged.	22
Special educators point out fear of using expensive technologies in general	60
Present technologies, needs and challenges for ASD	**471**
Special educators use different kinds of computer software/mobile applications/YouTube videos to perform activities.	54
Special educators come across a wide variety of challenges in using technological interventions	63
Special educators are supportive of technological interventions	27
Special educators are not supportive of technological interventions	3
Special educators do not use technological intervention unless absolutely necessary	6
Special educators use technological intervention as and when required depending on their own experience and the requirement of children	28
Applications do not allow linguistic diversity	33
Some applications are not lightweight	19
Applications need high speed internet	11
Technological interventions are not culturally contextualised	29
Expensive software with costly annual subscription becomes difficult to maintain	15
Special educators feel their work is monotonous	51
Special educators feel judged by their aids when they need help	38
Special educators find it difficult to cope with the psychological burden of their work	60
Senior special educators have concern for YouTube videos	2
Special educators need to customise training as per needs and requirements of the child	17
Some mobile applications have poor UI	15
Trust, Reliability & Appropriateness of Social Robots in India	**336**
India has pressing issues of social inequality	37
Special educators are fragmented as a single professional class	26
Not all special educators command the same manner of societal reverence	45
Not all special educators are financially well-off	31
Social robots could fragment the professional fraternity of special educator	29
Social robots could widen the margin of difference between special educators	6
As therapeutic agent social robots are trustworthy	14
Social robots could not be trusted unless implemented on large scale and effects are seen	12
RAT has been successful in other nations and hence reliable inherently	2
Social robots could follow the trajectories of mobile phones in India and become inexpensive in the long run	11
Government programs should be implemented for building of low -cost robots	9
Government programs should be planned to make educators more aware about social robots	24
Government and NGOs should partner with each other for the training of special educators with social robots	37
Social robots could cause digital divide	50
Rural India should be focused upon while developing social robots	3

Theme / Code	Count
Special Educators' perception towards RAT	**410**
Special educators understand RAT	33
Special educators are aware about the features of RAT	28
Special educators feel RAT will help them have unbiased and non-judgmental aids	50
Special educators feel RAT is futuristic	41
Special educators feel RAT could help them battle psychological fatigue	60
Robots could do repetitive tasks	21
Robots could be good role models	11
Robots need to speak in different languages	7
Robots need to be less expensive	37
Erosion of parental touch and affection	9
Overprofessionalisation due to RAT	3
Humanoid robots are better for therapy	66
Children should not develop an attraction for robots	19
Provision for customisation of activities	11
Provision for robot control through mobile phone-based applications	14
Fear of AI and strategies for collaboration	**370**
Special educators have positive attitude toward AI	25
Special educators have negative attitudes towards AI	5
Special educators feel that technologies like AI and those machines using them will soon make them redundant	45
AI based systems need through cultural auditing before being deployed	21
Fear of AI on part of some educators is based on personal experiences	11
Fear of AI on part of some educators is based on professional experiences	3
Educators welcome the use AI in social robots as new age technology	54
Human-AI collaboration at all levels is necessary for betterment of society and humankind	6
Social robots can be partners in different activities	33
Social robots can serve as useful training agents for educators who are beginning their careers	29
Social robots can keep children engaged	30
Social robots would transform the role of the special educators	51
Social robots would augment the efforts of special educators and help in ensuring better therapy practices	41
AI use in countries like India should be regulated in all forms	2
AI could replace teachers and educators in schools for atypical children first	14

REFERENCES

[1] CDC, "About CDC", *Centers for Disease Control and Prevention*, Aug. 2022. [Online]. Available: `https://www.cdc.gov/about/index.html`.

[2] R. Tiwari, K. Purkayastha, and S. Gulati, "Public health dimensions of autism spectrum disorder in india: An overview", *Journal of Comprehensive Health*, vol. 9, no. 2, pp. 57–62, 2021.

[3] R. Sagar *et al.*, "The burden of mental disorders across the states of india: The global burden of disease study 1990–2017", *The Lancet Psychiatry*, vol. 7, no. 2, pp. 148–161, 2020.

[4] P. Desarkar, T. K. Rajji, S. H. Ameis, and Z. J. Daskalakis, "Assessing and stabilizing aberrant neuroplasticity in autism spectrum disorder: The potential role of transcranial magnetic stimulation", *Frontiers in psychiatry*, vol. 6, p. 124, 2015.

[5] F. R. Volkmar, S. J. Rogers, R. Paul, and K. A. Pelphrey, *Handbook of autism and pervasive developmental disorders, volume 1: diagnosis, development, and brain mechanisms*. John Wiley & Sons, 2014.

[6] P. T. Shattuck and S. D. Grosse, "Issues related to the diagnosis and treatment of autism spectrum disorders", *Mental retardation and developmental disabilities research reviews*, vol. 13, no. 2, pp. 129–135, 2007.

[7] S. Chakraborty, P. Thomas, T. Bhatia, V. L. Nimgaonkar, and S. N. Deshpande, "Assessment of severity of autism using the indian scale for assessment of autism", *Indian journal of psychological medicine*, vol. 37, no. 2, pp. 169–174, 2015.

[8] M. Juneja *et al.*, “Inclen diagnostic tool for autism spectrum disorder (indt-asd): Development and validation”, *Indian pediatrics*, vol. 51, pp. 359–365, 2014.

[9] E. Schopler, R. J. Reichler, and B. R. Renner, *The childhood autism rating scale (CARS)*. Western Psychological Services Los Angeles, CA, 2010.

[10] RCI, “35th annual report 2021-22”, *Rehabilitation Council of India*, 2022. [Online]. Available: `http://www.rehabcouncil.nic.in/writereaddata/AnnualReport21-22.pdf`.

[11] K. Chanyoung *et al.*, “Ai-assisted initiation to joint attention evaluation for autism spectrum disorder detection”, in *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, 2022, pp. 1–6.

[12] Z. Guo, V. Chheang, J. Li, K. E. Barner, A. Bhat, and R. L. Barmaki, “Social visual behavior analytics for autism therapy of children based on automated mutual gaze detection”, in *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies*, 2023, pp. 11–21.

[13] F. Alnajjar, M. Cappuccio, A. Renawi, O. Mubin, and C. K. Loo, “Personalized robot interventions for autistic children: An automated methodology for attention assessment”, *International Journal of Social Robotics*, vol. 13, pp. 67–82, 2021.

[14] V. Jyoti and U. Lahiri, “Portable joint attention skill training platform for children with autism”, *IEEE Transactions on Learning Technologies*, vol. 15, no. 2, pp. 290–300, 2022.

[15] H. Drimalla *et al.*, “Detecting autism by analyzing a simulated social interaction”, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, Springer, 2018, pp. 193–208.

[16] G. Wan *et al.*, “Applying eye tracking to identify autism spectrum disorder in children”, *Journal of autism and developmental disorders*, vol. 49, pp. 209–215, 2019.

[17] E. Chong *et al.*, “Detection of eye contact with deep neural networks is as accurate as human experts”, *Nature communications*, vol. 11, no. 1, p. 6386, 2020.

[18] B. Li, A. Sharma, J. Meng, S. Purushwalkam, and E. Gowen, “Applying machine learning to identify autistic adults using imitation: An exploratory study”, *PloS one*, vol. 12, no. 8, e0182652, 2017.

[19] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson, “Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis”, *Scientific reports*, vol. 10, no. 1, p. 8346, 2020.

[20] C. J. Zampella, E. Sariyanidi, A. G. Hutchinson, G. K. Bartley, R. T. Schultz, and B. Tunç, “Computational measurement of motor imitation and imitative learning differences in autism spectrum disorder”, in *Companion Publication of the 2021 International Conference on Multimodal Interaction*, 2021, pp. 362–370.

[21] A. Taheri, A. Meghdari, and M. H. Mahoor, “A close look at the imitation performance of children with autism and typically developing children using a robotic system”, *International Journal of Social Robotics*, vol. 13, pp. 1125–1147, 2021.

[22] T. Guha *et al.*, “On quantifying facial expression-related atypicality of children with autism spectrum disorder”, in *2015 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, 2015, pp. 803–807.

[23] J. Awatramani and N. Hasteer, “Facial expression recognition using deep learning for children with autism spectrum disorder”, in *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, 2020, pp. 35–39.

[24] D. Bone *et al.*, “Spontaneous-speech acoustic-prosodic features of children with autism and the interacting psychologist”, in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[25] J. Deng, N. Cummins, M. Schmitt, K. Qian, F. Ringeval, and B. Schuller, “Speech-based diagnosis of autism spectrum condition by generative adversarial network representations”, in *Proceedings of the 2017 international conference on digital health*, 2017, pp. 53–57.

[26] K. Han, D. Yu, and I. Tashev, “Speech emotion recognition using deep neural network and extreme learning machine”, in *Fifteenth annual conference of the international speech communication association*, 2014.

[27] S. Amiriparian *et al.*, “Recognition of Echolalic Autistic Child Vocalisations Utilising Convolutional Recurrent Neural Networks”, in *Proc. Interspeech 2018*, 2018, pp. 2334–2338.

[28] B. Scassellati, H. Admoni, and M. Matarić, “Robots for Use in Autism Research”, *Annual Review of Biomedical Engineering*, vol. 14, no. 1, pp. 275–294, 2012, _eprint: https://doi.org/10.1146/annurev-bioeng-071811-150036. [Online]. Available: `https://doi.org/10.1146/annurev-bioeng-071811-150036`.

[29] B. Scassellati, “How social robots will help us to diagnose, treat, and understand autism”, in *Robotics research: Results of the 12th international symposium ISRR*, Springer, 2007, pp. 552–563.

[30] “Autism”, *Autism Spectrum Disorder*, 2023. [Online]. Available: `https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders`.

[31] D. Chaudhary *et al.*, “Development of a cost of illness inventory questionnaire for children with autism spectrum disorder in south asia”, *BMC health services research*, vol. 22, no. 1, pp. 1–10, 2022.

[32] R. Loomes, L. Hull, and W. P. L. Mandy, "What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis", *Journal of the American Academy of Child & Adolescent Psychiatry*, vol. 56, no. 6, pp. 466–474, 2017.

[33] WHO, "International classification of diseases 11th revision", [Online]. Available: `https://icd.who.int/en`.

[34] D. A. Regier, E. A. Kuhl, and D. J. Kupfer, "The dsm-5: Classification and criteria changes", *World psychiatry*, vol. 12, no. 2, pp. 92–98, 2013.

[35] E Rellini, D Tortolani, S Trillo, S Carbone, and F Montecchi, "Childhood autism rating scale (cars) and autism behavior checklist (abc) correspondence and conflicts with dsm-iv criteria in diagnosis of autism", *Journal of autism and developmental disorders*, vol. 34, pp. 703–708, 2004.

[36] T. Nolan *et al.*, "Child behaviour checklist classification of behaviour disorder", *Journal of paediatrics and child health*, vol. 32, no. 5, pp. 405–411, 1996.

[37] A. Malin, "Malin's intelligence scale for indian children", *Nagpur (India): Child Guidance Center*, 1969.

[38] B. N. Roopesh, "Vineland social maturity scale: An update on administration and scoring", *Indian J Clin Psychol*, vol. 46, no. 2, pp. 91–102, 2019.

[39] E. Billing *et al.*, "The dream dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy", *PloS one*, vol. 15, no. 8, e0236939, 2020.

[40] A. A. Ramirez-Duque, A. Frizera-Neto, and T. F. Bastos, "Robot-assisted autism spectrum disorder diagnostic based on artificial reasoning", *Journal of Intelligent & Robotic Systems*, vol. 96, pp. 267–281, 2019.

[41] I. Dubey *et al.*, "Using mobile health technology to assess childhood autism in low-resource community settings in india: An innovation to address the detection gap", *Autism*, p. 13 623 613 231 182 801, 2021.

[42] P. Arun and B. S. Chavan, "Development of a screening instrument for autism spectrum disorder: Chandigarh autism screening instrument", *The Indian journal of medical research*, vol. 147, no. 4, p. 369, 2018.

[43] M. Varma *et al.*, "Identification of social engagement indicators associated with autism spectrum disorder using a game-based mobile app: Comparative study of gaze fixation and visual scanning methods", *Journal of Medical Internet Research*, vol. 24, no. 2, e31830, 2022.

[44] N. A. Chi *et al.*, "Classifying autism from crowdsourced semistructured speech recordings: Machine learning model comparison study", *JMIR Pediatrics and Parenting*, vol. 5, no. 2, e35406, 2022.

[45] Y. Penev *et al.*, "A mobile game platform for improving social communication in children with autism: A feasibility study", *Applied clinical informatics*, vol. 12, no. 05, pp. 1030–1040, 2021.

[46] H. Kalantarian *et al.*, "The performance of emotion classifiers for children with parent-reported autism: Quantitative feasibility study", *JMIR mental health*, vol. 7, no. 4, e13174, 2020.

[47] C. Wu *et al.*, "Machine learning based autism spectrum disorder detection from videos", in *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*, IEEE, 2021, pp. 1–6.

[48] P. Wei, D. Ahmedt-Aristizabal, H. Gammulle, S. Denman, and M. A. Armin, "Vision-based activity recognition in children with autism-related behaviors", *Heliyon*, 2023.

[49] W. Liu, "Identifying children autism spectrum disorder via machine learning based behavior analysis", Ph.D. dissertation, Carnegie Mellon University, 2022.

[50] S. Rajagopalan, A. Dhall, and R. Goecke, “Self-stimulatory behaviours in the wild for autism diagnosis”, in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 755–761.

[51] A. Crippa *et al.*, “Use of machine learning to identify children with autism and their motor abnormalities”, *Journal of autism and developmental disorders*, vol. 45, pp. 2146–2156, 2015.

[52] J. Bidwell, I. A. Essa, A. Rozga, and G. D. Abowd, “Measuring child visual attention using markerless head tracking from color and depth sensing cameras”, in *Proceedings of the 16th International Conference on Multimodal Interaction*, 2014, pp. 447–454.

[53] E. Deng, B. Mutlu, M. J. Mataric, *et al.*, “Embodiment in socially interactive robots”, *Foundations and Trends® in Robotics*, vol. 7, no. 4, pp. 251–356, 2019.

[54] A. Zaraki, L. Wood, B. Robins, and K. Dautenhahn, “Development of a semi-autonomous robotic system to assist children with autism in developing visual perspective taking skills”, in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE, 2018, pp. 969–976.

[55] T. Zorcec, B. Robins, and K. Dautenhahn, “Getting engaged: Assisted play with a humanoid robot kaspar for children with severe autism”, in *ICT Innovations 2018. Engineering and Life Sciences: 10th International Conference, ICT Innovations 2018, Ohrid, Macedonia, September 17–19, 2018, Proceedings 10*, Springer, 2018, pp. 198–207.

[56] Z. Shi, T. R. Groechel, S. Jain, K. Chima, O. Rudovic, and M. J. Matarić, “Toward personalized affect-aware socially assistive robot tutors for long-term interventions with children with autism”, *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 11, no. 4, pp. 1–28, 2022.

[57] S. Jain, B. Thiagarajan, Z. Shi, C. Clabaugh, and M. J. Matarić, “Modeling engagement in long-term, in-home socially assistive robot

interventions for children with autism spectrum disorders", *Science Robotics*, vol. 5, no. 39, eaaz3791, 2020.

[58] "Dream:development of robot-enhanced therapy for children with autism spectrum disorders", Accessed November 23, 2020. [Online]. Available: `https://dream2020.github.io/DREAM/`.

[59] F. Petric and Z. Kovačić, "Hierarchical pomdp framework for a robot-assisted asd diagnostic protocol", in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, 2019, pp. 286–293.

[60] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: Effectiveness, efficiency, and multi-instrument fusion", *Journal of Child Psychology and Psychiatry*, vol. 57, no. 8, pp. 927–937, 2016.

[61] G. Nie *et al.*, "Analytical modelling of affective data in human-machine interaction system designed for children with autism spectrum disorders", Ph.D. dissertation, 2021.

[62] J. Hashemi *et al.*, "Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants", *Autism research and treatment*, vol. 2014, 2014.

[63] Z. Chang *et al.*, "Computational methods to measure patterns of gaze in toddlers with autism spectrum disorder", *JAMA pediatrics*, vol. 175, no. 8, pp. 827–836, 2021.

[64] F. Petric *et al.*, "Four tasks of a robot-assisted autism spectrum disorder diagnostic protocol: First clinical tests", in *IEEE Global Humanitarian Technology Conference (GHTC 2014)*, 2014, pp. 510–517.

[65] A. Duquette, F. Michaud, and H. Mercier, "Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism", *Autonomous Robots*, vol. 24, no. 2, pp. 147–157, 2008.

[66] J. J. Diehl, L. M. Schmitt, M. Villano, and C. R. Crowell, "The clinical use of robots for individuals with autism spectrum disorders: A critical review", *Research in autism spectrum disorders*, vol. 6, no. 1, pp. 249–262, 2012.

[67] H. Kozima, M. P. Michalowski, and C. Nakagawa, "Keepon", *International Journal of Social Robotics*, vol. 1, no. 1, pp. 3–18, 2009.

[68] D. J. Ricks and M. B. Colton, "Trends and considerations in robot-assisted autism therapy", in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 4354–4359.

[69] J. Bharatharaj, L. Huang, R. E. Mohan, A. Al-Jumaily, and C. Krägeloh, "Robot-assisted therapy for learning and social interaction of children with autism spectrum disorder", *Robotics*, vol. 6, no. 1, p. 4, 2017.

[70] A. Di Nuovo, D. Conti, G. Trubia, S. Buono, and S. Di Nuovo, "Deep learning systems for estimating visual attention in robot-assisted therapy of children with autism and intellectual disability", *Robotics*, vol. 7, no. 2, p. 25, 2018.

[71] S. M. Srinivasan, I.-M. Eigsti, T. Gifford, and A. N. Bhat, "The effects of embodied rhythm and robotic interventions on the spontaneous and responsive verbal communication skills of children with autism spectrum disorder (asd): A further outcome of a pilot randomized controlled trial", *Research in Autism Spectrum Disorders*, vol. 27, pp. 73–87, 2016.

[72] L. I. Ismail, S. Shamsudin, H. Yussof, F. A. Hanapiah, and N. I. Zahari, "Robot-based intervention program for autistic children with humanoid robot nao: Initial response in stereotyped behavior", *Procedia Engineering*, vol. 41, pp. 1441–1447, 2012.

[73] E. Y.-h. Chung, "Robot-mediated social skill intervention programme for children with autism spectrum disorder: An aba time-series study", *International Journal of Social Robotics*, pp. 1–13, 2020.

[74] S.-S. Yun, J. Choi, S.-K. Park, G.-Y. Bong, and H. Yoo, “Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system”, *Autism Research*, vol. 10, no. 7, pp. 1306–1323, 2017.

[75] W.-C. So *et al.*, “A robot-based play-drama intervention may improve the joint attention and functional play behaviors of chinese-speaking preschoolers with autism spectrum disorder: A pilot study”, *Journal of Autism and Developmental Disorders*, vol. 50, no. 2, pp. 467–481, 2020.

[76] W.-C. So *et al.*, “Who is a better teacher for children with autism? comparison of learning outcomes between robot-based and human-based interventions in gestural production and recognition”, *Research in Developmental Disabilities*, vol. 86, pp. 62–75, 2019.

[77] J. A. Osterling, G. Dawson, and J. A. Munson, “Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation”, *Development and psychopathology*, vol. 14, no. 2, pp. 239–251, 2002.

[78] M. Miller, A.-M. Iosif, M. Hill, G. S. Young, A. Schwichtenberg, and S. Ozonoff, “Response to name in infants developing autism spectrum disorder: A prospective study”, *The Journal of pediatrics*, vol. 183, pp. 141–146, 2017.

[79] B. Robins, P. Dickerson, P. Stribling, and K. Dautenhahn, “Robot-mediated joint attention in children with autism: A case study in robot-human interaction”, *Interaction studies*, vol. 5, no. 2, pp. 161–198, 2004.

[80] Z. E. Warren *et al.*, “Can robotic interaction improve joint attention skills?”, *Journal of autism and developmental disorders*, vol. 45, no. 11, pp. 3726–3734, 2015.

[81] Z. Wang, J. Liu, K. He, Q. Xu, X. Xu, and H. Liu, “Screening early children with autism spectrum disorder via response-to-name pro-

tocol", *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 587–595, 2021.

[82] S. M. Anzalone *et al.*, "How children with autism spectrum disorder behave and explore the 4-dimensional (spatial 3d+ time) environment during a joint attention induction task with a robot", *Research in Autism Spectrum Disorders*, vol. 8, no. 7, pp. 814–826, 2014.

[83] J. Shukla, V. R. Suri, J. Garg, K. Verma, and P. Kansal, "Mapping robotic affordances with pre-requisite learning interventions for children with autism spectrum disorder", in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2019, pp. 1–8.

[84] R. S. Moorthy and S Pugazhenthi, "Teaching psychomotor skills to autistic children by employing a robotic training kit: A pilot study", *International Journal of Social Robotics*, vol. 9, no. 1, pp. 97–108, 2017.

[85] J. Bharatharaj, L. Huang, C. Krägeloh, M. R. Elara, and A. Al-Jumaily, "Social engagement of children with autism spectrum disorder in interaction with a parrot-inspired therapeutic robot", *Procedia computer science*, vol. 133, pp. 368–376, 2018.

[86] V. L. Pires Kusumota, R. Vidal Aroca, and F. N. Martins, "An open source framework for educational applications using cozmo mobile robot", in *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*, 2018, pp. 569–576.

[87] C. Lord, M. Rutter, P. C. Dilavore, and S. Risi, *ADOS: Autism diagnostic observation schedule.* Hogrefe Boston, MA, 2008.

[88] T. Shibata, "Integration of therapeutic robot, paro, into welfare systems", in *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, 2010, pp. 3–3.

[89] H. S. Kang, K. Makimoto, R. Konno, and I. S. Koh, "Review of outcome measures in paro robot intervention studies for dementia care", *Geriatric Nursing*, vol. 41, no. 3, pp. 207–214, 2020.

[90] A. A. Scoglio, E. D. Reilly, J. A. Gorman, and C. E. Drebing, "Use of social robots in mental health and well-being research: Systematic review", *Journal of medical Internet research*, vol. 21, no. 7, e13322, 2019.

[91] N. Rakhymbayeva, N. Seitkazina, D. Turabayev, A. Pak, and A. Sandygulova, "A long-term study of robot-assisted therapy for children with severe autism and adhd", in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 401–402.

[92] I. Guemghar, P. Pires de Oliveira Padilha, A. Abdel-Baki, D. Jutras-Aswad, J. Paquette, and M.-P. Pomey, "Social robot interventions in mental health care and their outcomes, barriers, and facilitators: Scoping review", *JMIR Mental Health*, vol. 9, no. 4, e36094, 2022.

[93] S. Shamsuddin *et al.*, "Humanoid robot nao: Review of control and motion exploration", in *2011 IEEE international conference on Control System, Computing and Engineering*, IEEE, 2011, pp. 511–516.

[94] A. K. Pandey and R. Gelin, "A mass-produced sociable humanoid robot: Pepper: The first machine of its kind", *IEEE Robotics & Automation Magazine*, vol. 25, no. 3, pp. 40–48, Sep. 2018. [Online]. Available: `https://ieeexplore.ieee.org/document/8409927/`.

[95] M. Fujita, "On activating human communications with pet-type robot aibo", *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1804–1813, 2004.

[96] N. Natarajan, S. Vaitheswaran, M. R. Lima, M. Wairagkar, and R. Vaidyanathan, "Acceptability of social robots and adaptation of hybrid-face robot for dementia care in india: A qualitative study", *The American Journal of Geriatric Psychiatry*, vol. 30, no. 2, pp. 240–245, Feb. 1, 2022. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S1064748121003183`.

[97] S. Whelan, K. Murphy, E. Barrett, C. Krusche, A. Santorelli, and D. Casey, “Factors affecting the acceptability of social robots by older adults including people with dementia or cognitive impairment: A literature review”, *International Journal of Social Robotics*, vol. 10, no. 5, pp. 643–668, 2018.

[98] L. Cohen *et al.*, “Influence of facial feedback during a cooperative human-robot task in schizophrenia”, *Scientific reports*, vol. 7, no. 1, p. 15 023, 2017.

[99] T. Paulsson, M. Zhong, I. García Velázquez, and G. Castellano, “Exploring mothers’ perspectives on socially assistive robots in peripartum depression screening”, in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 486–490.

[100] S. Shamsuddin, H. Yussof, L. I. Ismail, S. Mohamed, F. A. Hanapiah, and N. I. Zahari, “Initial response in hri-a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot nao”, *Procedia Engineering*, vol. 41, pp. 1448–1455, 2012.

[101] S. S. Balasuriya, L. Sitbon, M. Brereton, and S. Koplick, “How can social robots spark collaboration and engagement among people with intellectual disability?”, in *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, 2019, pp. 209–220.

[102] B. Scassellati *et al.*, “Improving social skills in children with asd using a long-term, in-home social robot”, *Science Robotics*, vol. 3, no. 21, eaat7544, 2018.

[103] B. Robins, K. Dautenhahn, R. T. Boekhorst, and A. Billard, “Robotic assistants in therapy and education of children with autism: Can a small humanoid robot help encourage social interaction skills?”, *Universal access in the information society*, vol. 4, pp. 105–120, 2005.

[104] L. Dickstein-Fischer, E. Alexander, X. Yan, H. Su, K. Harrington, and G. S. Fischer, “An affordable compact humanoid robot for autism spectrum disorder interventions in children”, in *2011 Annual Interna-*

tional Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 5319–5322.

[105] D. J. Ricks and M. B. Colton, "Trends and considerations in robot-assisted autism therapy", in *2010 IEEE international conference on robotics and automation*, IEEE, 2010, pp. 4354–4359.

[106] P. S. Dehkordi, H. Moradi, M. Mahmoudi, and H. R. Pouretemad, "The design, development, and deployment of roboparrot for screening autistic children", *International Journal of Social Robotics*, vol. 7, pp. 513–522, 2015.

[107] K. Richardson *et al.*, "Robot enhanced therapy for children with autism (dream): A social model of autism", *IEEE Technology and society magazine*, vol. 37, no. 1, pp. 30–39, 2018.

[108] B. Robins, N. Otero, E. Ferrari, and K. Dautenhahn, "Eliciting requirements for a robotic toy for children with autism-results from user panels", in *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, 2007, pp. 101–106.

[109] A. Raptopoulou, A. Komnidis, P. D. Bamidis, and A. Astaras, "Human–robot interaction for social skill development in children with asd: A literature review", *Healthcare Technology Letters*, vol. 8, no. 4, pp. 90–96, 2021.

[110] H. Kumazaki *et al.*, "The impact of robotic intervention on joint attention in children with autism spectrum disorders", *Molecular autism*, vol. 9, no. 1, pp. 1–10, 2018.

[111] D. Ghiglino, P. Chevalier, F. Floris, T. Priolo, and A. Wykowska, "Follow the white robot: Efficacy of robot-assistive training for children with autism spectrum disorder", *Research in Autism Spectrum Disorders*, vol. 86, p. 101 822, 2021.

[112] N. Rakhymbayeva, A. Amirova, and A. Sandygulova, "A long-term engagement with a social robot for autism therapy", *Frontiers in Robotics and AI*, vol. 8, p. 669 972, 2021.

[113] K. Dautenhahn and A. Billard, "Games children with autism can play with robota, a humanoid robotic doll", in *Universal access and assistive technology: Proceedings of the Cambridge workshop on UA and AT'02*, Springer, 2002, pp. 179–190.

[114] C. A. Huijnen, M. A. Lexis, R. Jansens, and L. P. de Witte, "How to implement robots in interventions for children with autism? a co-creation study involving people with autism, parents and professionals", *Journal of autism and developmental disorders*, vol. 47, pp. 3079–3096, 2017.

[115] G. Lakatos, L. J. Wood, D. S. Syrdal, B. Robins, A. Zaraki, and K. Dautenhahn, "Robot-mediated intervention can assist children with autism to develop visual perspective taking skills", *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 87–101, 2020.

[116] B. Vanderborght *et al.*, "Using the social robot probo as a social story telling agent for children with asd", *Interaction Studies*, vol. 13, no. 3, pp. 348–372, 2012.

[117] O. Mihoko, K. Setsuya, S. Y. OtaKen, and B. Daisuke, Eds., *New Frontiers in Artificial Intelligence*, Springer, 2017.

[118] J. K. Westlund *et al.*, "Lessons from teachers on performing HRI studies with young children in schools", in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Christchurch, New Zealand: IEEE, Mar. 2016, pp. 383–390, ISBN: 978-1-4673-8370-7. [Online]. Available: `http://ieeexplore.ieee.org/document/7451776/`.

[119] M. Fridin and M. Belokopytov, "Acceptance of socially assistive humanoid robot by preschool and elementary school teachers", *Computers in Human Behavior*, vol. 33, pp. 23–31, Apr. 2014. [Online].

Available: `https://linkinghub.elsevier.com/retrieve/pii/S0747563213004652`.

[120] S. R. Pendse *et al.*, "Mental health in the global south: Challenges and opportunities in hci for development", in *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, 2019, pp. 22–36.

[121] Y. A. Abdi and J. Y. Elmi, "Internet based telepsychiatry: A pilot case in somaliland", *Medicine, Conflict and Survival*, vol. 27, no. 3, pp. 145–150, 2011.

[122] J. Chipps, P Brysiewicz, and M Mars, "Effectiveness and feasibility of telepsychiatry in resource constrained environments? a systematic review of the evidence", *African journal of psychiatry*, vol. 15, no. 4, pp. 235–243, 2012.

[123] S. Naskar, R. Victor, H. Das, and K. Nath, "Telepsychiatry in india-where do we stand? a comparative review between global and indian telepsychiatry programs", *Indian journal of psychological medicine*, vol. 39, no. 3, pp. 223–242, 2017.

[124] Q. Tariq *et al.*, "Detecting developmental delay and autism through machine learning models using home videos of bangladeshi children: Development and validation study", *Journal of Medical Internet Research*, vol. 21, no. 4, e13822, Apr. 24, 2019.

[125] R. Aburukba, F. Aloul, A. Mahmoud, K. Kamili, and S. Ajmal, "Autiaid: A learning mobile application for autistic children", in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2017, pp. 1–6.

[126] P. R. Krishnappa Babu, P. Oza, and U. Lahiri, "Gaze-sensitive virtual reality based social communication platform for individuals with autism", *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 450–462, 2018.

[127] S. Kuriakose and U. Lahiri, “Understanding the psycho-physiological implications of interaction with a virtual reality-based system in adolescents with autism: A feasibility study”, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 4, pp. 665–675, 2015.

[128] S. Kuriakose, S. Kunche, B Narendranath, P. Jain, S. Sonker, and U. Lahiri, “A step towards virtual reality based social communication for children with autism”, in *2013 International Conference on Control, Automation, Robotics and Embedded Systems (CARE)*, 2013, pp. 1–6.

[129] V. Jyoti and U. Lahiri, “Virtual reality based joint attention task platform for children with autism”, *IEEE Transactions on Learning Technologies*, vol. 13, no. 1, pp. 198–210, 2020.

[130] V. G. Prakash *et al.*, “Computer vision-based assessment of autistic children: Analyzing interactions, emotions, human pose, and life skills”, *IEEE Access*, vol. 11, pp. 47 907–47 929, 2023.

[131] M. Kohli, A. K. Kar, V. G. Prakash, and A. Prathosh, “Deep learning-based human action recognition framework to assess children on the risk of autism or developmental delays”, in *International Conference on Neural Information Processing*, Springer, 2022, pp. 459–470.

[132] I. Dubey *et al.*, “Using mobile health technology to assess childhood autism in low-resource community settings in india: An innovation to address the detection gap”, *Autism*, p. 13 623 613 231 182 801, Jul. 17, 2023, Publisher: SAGE Publications Ltd. [Online]. Available: `https://doi.org/10.1177/13623613231182801`.

[133] A. I. Soysa and A. Al Mahmud, “Technology for children with autism spectrum disorder: What do sri lankan parents and practitioners want?”, *Interacting with Computers*, vol. 31, no. 3, pp. 282–302, Aug. 2019, _eprint: https://academic.oup.com/iwc/article-pdf/31/3/282/33525057/iwz020.pdf. [Online]. Available: `https://doi.org/10.1093/iwc/iwz020`.

[134] A. Sandygulova *et al.*, "Interaction design and methodology of robot-assisted therapy for children with severe asd and adhd", *Paladyn, Journal of Behavioral Robotics*, vol. 10, no. 1, pp. 330–345, 2019.

[135] A. Zhanatkyzy, Z. Telisheva, A. Amirova, N. Rakhymbayeva, and A. Sandygulova, "Multi-purposeful activities for robot-assisted autism therapy: What works best for children's social outcomes?", in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, 2023, pp. 34–43.

[136] E. Beede *et al.*, "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy", in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–12.

[137] B. Nardi *et al.*, "Computing within limits", *Commun. ACM*, vol. 61, no. 10, pp. 86–93, Sep. 2018. [Online]. Available: `https://dl.acm.org/doi/10.1145/3183582`.

[138] J. Pal *et al.*, "An Accessibility Infrastructure for the Global South", in *Proceedings of the Eighth International Conference on Information and Communication Technologies and Development*, Ann Arbor MI USA: ACM, Jun. 2016, pp. 1–11, ISBN: 978-1-4503-4306-0. [Online]. Available: `https://dl.acm.org/doi/10.1145/2909609.2909666`.

[139] N. Singal, "Inclusive Education in India: International concept, national interpretation", *International Journal of Disability, Development and Education*, vol. 53, no. 3, pp. 351–369, Sep. 2006, Publisher: Routledge _eprint: https://doi.org/10.1080/10349120600847797. [Online]. Available: `https://doi.org/10.1080/10349120600847797`.

[140] V. Thirumurthy and V. Thirumurthy, "Special Education in India at the Crossroads", *Childhood Education*, vol. 83, no. 6, pp. 380–384, Sep. 2007, Publisher: Routledge _eprint: https://doi.org/10.1080/00094056.2007.10522955. [Online]. Available: `https://doi.org/10.1080/00094056.2007.10522955`.

[141] M. Kalyanpur, “The Paradox of Majority Underrepresentation in Special Education in India: Constructions of Difference in a Developing Country”, *J Spec Educ*, vol. 42, no. 1, pp. 55–64, May 2008, Publisher: SAGE Publications Inc. [Online]. Available: `https://doi.org/10.1177/0022466907313610`.

[142] P. Chandra, S. I. Ahmed, and J. Pal, “Market Practices and the Bazaar: Technology Consumption in ICT Markets in the Global South”, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17, New York, NY, USA: Association for Computing Machinery, May 2017, pp. 4741–4752, ISBN: 978-1-4503-4655-9. [Online]. Available: `https://dl.acm.org/doi/10.1145/3025453.3025970`.

[143] L. Kirabo, “"Our perspective matters.": Using universal design goals to guide technology design in the global south”, *SIGACCESS Access. Comput.*, no. 128, 3:1–3:4, Dec. 2020. [Online]. Available: `https://dl.acm.org/doi/10.1145/3441497.3441500`.

[144] R. Heeks, “Digital inequality beyond the digital divide: Conceptualizing adverse digital incorporation in the global south”, *Information Technology for Development*, vol. 28, no. 4, pp. 688–704, 2022.

[145] N. Belorgey, “The Aadhaar battle: Why some players in the corporate world needed a biometric ID?”, *Global Policy*, vol. n/a, no. n/a, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1758-5899.13172. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-5899.13172`.

[146] J. Shukla, V. R. Suri, J. Garg, K. Verma, and P. Kansal, “Mapping robotic affordances with pre-requisite learning interventions for children with autism spectrum disorder”, in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, IEEE, 2019, pp. 1–8.

[147] P. Chauhan, J. Pujari, P. Yadav, and S. Guha, “Challenges and opportunities of technology related instruction for children with autism

spectrum disorder: Parents perspective", *MIER Journal of Educational Studies Trends and Practices*, pp. 320–337, 2022.

[148] N. C. Brunsting, M. A. Sreckovic, and K. L. Lane, "Special education teacher burnout: A synthesis of research from 1979 to 2013", *Education and Treatment of Children*, vol. 37, no. 4, pp. 681–711, 2014, Publisher: West Virginia University Press. [Online]. Available: https://muse.jhu.edu/pub/20/article/556432.

[149] E. Dahlin, "Are robots really stealing our jobs? perception versus experience", *Socius: Sociological Research for a Dynamic World*, vol. 8, p. 237 802 312 211 313, Jan. 2022. [Online]. Available: http://dx.doi.org/10.1177/23780231221131377.

[150] S. Sharma *et al.*, "Promoting joint attention with computer supported collaboration in children with autism", in *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, 2016, pp. 1560–1571.

[151] B Ashwini, V. Narayan, A. Bhatia, and J. Shukla, "Responsiveness towards robot-assisted interactions among pre-primary children of indian ethnicity", in *2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN)*, 2021, pp. 619–625.

[152] S. Sharma *et al.*, "Designing gesture-based applications for individuals with developmental disabilities: Guidelines from user studies in india", *ACM Transactions on Accessible Computing (TACCESS)*, vol. 11, no. 1, pp. 1–27, 2018.

[153] C. T. Okolo, S. Kamath, N. Dell, and A. Vashistha, ""it cannot do all of my work": Community health worker perceptions of AI-enabled mobile health applications in rural india", in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 6, 2021, pp. 1–20, ISBN: 978-1-4503-8096-6. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445420.

[154] I. Medhi and K. Toyama, "Full-context videos for first-time, non-literate pc users", in *2007 International Conference on Information and Communication Technologies and Development*, IEEE, 2007, pp. 1–9.

[155] J. O. Cooper, T. E. Heron, W. L. Heward, *et al.*, *Applied behavior analysis.* Pearson/Merrill-Prentice Hall Upper Saddle River, NJ, 2007.

[156] M. Spitale, S. Silleresi, F. Garzotto, and M. J. Matarić, "Using socially assistive robots in speech-language therapy for children with language impairments", *International Journal of Social Robotics*, pp. 1–18, 2023.

[157] B. Arsić, A. Gajić, S. Vidojković, D. Maćešić-Petrović, A. Bašić, and R. Z. Parezanović, "The use of nao robots in teaching children with autism", *European Journal of Alternative Education Studies*, vol. 7, no. 1, 2022.

[158] D. Conti, A. Cattani, S. Di Nuovo, and A. Di Nuovo, "Are future psychologists willing to accept and use a humanoid robot in their practice? italian and english students' perspective", *Frontiers in psychology*, vol. 10, p. 2138, 2019.

[159] C. T. Okolo, "Optimizing human-centered AI for healthcare in the Global South", *Patterns*, vol. 3, no. 2, p. 100 421, Feb. 2022. [Online]. Available: `https://linkinghub.elsevier.com/retrieve/pii/S2666389921003044`.

[160] S. A. Jacob and S. P. Furgerson, "Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research.", *Qualitative report*, vol. 17, p. 6, 2012.

[161] V. Braun and V. Clarke, *Thematic analysis.* American Psychological Association, 2012.

[162] N. Rangaswamy and N. Sambasivan, "Cutting chai, jugaad, and here pheri: Towards UbiComp for a global community", *Personal and Ubiq-*

uitous Computing, vol. 15, no. 6, pp. 553–564, Aug. 1, 2011. [Online]. Available: https://doi.org/10.1007/s00779-010-0349-x.

[163] K. L. Kilgore and C. C. Griffin, "Beginning special educators: Problems of practice and the influence of school context", *Teacher Education and Special Education*, vol. 21, no. 3, pp. 155–173, Jul. 1, 1998, Publisher: SAGE Publications Inc. [Online]. Available: https://doi.org/10.1177/088840649802100302.

[164] B. R. Duffy, "Anthropomorphism and the social robot", *Robotics and Autonomous Systems*, vol. 42, no. 3, pp. 177–190, Mar. 2003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0921889002003743.

[165] S. Kiesler, A. Powers, S. R. Fussell, and C. Torrey, "Anthropomorphic Interactions with a Robot and Robot–like Agent", *Social Cognition*, vol. 26, no. 2, pp. 169–181, Apr. 2008, Publisher: Guilford Publications Inc. [Online]. Available: https://guilfordjournals.com/doi/10.1521/soco.2008.26.2.169.

[166] Castro-González, H. Admoni, and B. Scassellati, "Effects of form and motion on judgments of social robots animacy, likability, trustworthiness and unpleasantness", *International Journal of Human-Computer Studies*, vol. 90, pp. 27–38, Jun. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S107158191600032X.

[167] J. Shukla, M. Barreda-Ángeles, J. Oliver, and D. Puig, "Effectiveness of socially assistive robotics during cognitive stimulation interventions: Impact on caregivers", in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, ISSN: 1944-9437, Aug. 2017, pp. 62–67.

[168] M. Smakman, P. Vogt, and E. A. Konijn, "Moral considerations on social robots in education: A multi-stakeholder perspective", *Computers & Education*, vol. 174, p. 104 317, Dec. 1, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360131521001949.

[169] S. Patra and B. Kumar Patro, “Affiliate stigma among parents of children with autism in eastern india”, *Asian Journal of Psychiatry*, vol. 44, 45–47, Aug. 2019. [Online]. Available: http://dx.doi.org/10.1016/j.ajp.2019.07.018.

[170] M. U. Desai, G. Divan, F. J. Wertz, and V. Patel, “The discovery of autism: Indian parents’ experiences of caring for their child with an autism spectrum disorder”, *Transcultural Psychiatry*, vol. 49, no. 3–4, 613–637, Jun. 2012. [Online]. Available: http://dx.doi.org/10.1177/1363461512447139.

[171] J. Rezwana and M. L. Maher, “Understanding user perceptions, collaborative experience and user engagement in different human-AI interaction designs for co-creative systems”, in *Proceedings of the 14th Conference on Creativity and Cognition*, ser. C&C ’22, New York, NY, USA: Association for Computing Machinery, Jun. 20, 2022, pp. 38–48, ISBN: 978-1-4503-9327-0. [Online]. Available: https://doi.org/10.1145/3527927.3532789.

[172] A. M. Alcorn *et al.*, “Educators’ Views on Using Humanoid Robots With Autistic Learners in Special Education Settings in England”, *Frontiers in Robotics and AI*, vol. 6, 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2019.00107.

[173] S. Lemaignan, N. Newbutt, L. Rice, and J. Daly, ““it’s important to think of pepper as a teaching aid or resource external to the classroom”: A social robot in a school for autistic children”, *International Journal of Social Robotics*, Oct. 31, 2022. [Online]. Available: https://doi.org/10.1007/s12369-022-00928-4.

[174] *Trust in Technology: A Socio-Technical Perspective — doi.org*, https://doi.org/10.1007/1-4020-4258-2, [Accessed 03-12-2023].

[175] X. Zhu, X. Zhang, Z. Chen, Z. Dong, Z. Gu, and D. Chang, “The trusted listener: The influence of anthropomorphic eye design of social robots on user’s perception of trustworthiness”, in *CHI Conference on Human Factors in Computing Systems*, New Orleans LA USA: ACM,

Apr. 29, 2022, pp. 1–13, ISBN: 978-1-4503-9157-3. [Online]. Available: https://dl.acm.org/doi/10.1145/3491102.3517670.

[176] T. Beelen, E. Velner, K. P. Truong, R. Ordelman, T. Huibers, and V. Evers, “Children’s trust in robots and the information they provide”, in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA ’23, New York, NY, USA: Association for Computing Machinery, Apr. 19, 2023, pp. 1–7, ISBN: 978-1-4503-9422-2. [Online]. Available: https://dl.acm.org/doi/10.1145/3544549.3585801.

[177] S. Park, P. G. T. Healey, and A. Kaniadakis, “Should robots blush?”, in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan: ACM, May 6, 2021, pp. 1–14, ISBN: 978-1-4503-8096-6. [Online]. Available: https://dl.acm.org/doi/10.1145/3411764.3445561.

[178] A. Sauppé and B. Mutlu, “The social impact of a robot co-worker in industrial settings”, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI ’15, New York, NY, USA: Association for Computing Machinery, Apr. 18, 2015, pp. 3613–3622, ISBN: 978-1-4503-3145-6. [Online]. Available: https://dl.acm.org/doi/10.1145/2702123.2702181.

[179] M. Smakman, P. Vogt, and E. A. Konijn, “Moral considerations on social robots in education: A multi-stakeholder perspective”, *Comput. Educ.*, vol. 174, no. 104317, p. 104 317, Dec. 2021.

[180] M. H. J. Smakman, E. A. Konijn, P. Vogt, and P. Pankowska, “Attitudes towards social robots in education: Enthusiast, practical, troubled, sceptic, and mindfully positive”, *Robotics*, vol. 10, no. 1, p. 24, Mar. 2021, Number: 1 Publisher: Multidisciplinary Digital Publishing Institute. [Online]. Available: https://www.mdpi.com/2218-6581/10/1/24.

[181] C. Esterwood and L. P. Robert, “Having the right attitude: How attitude impacts trust repair in human-robot interaction”, in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot*

Interaction, ser. HRI '22, Sapporo, Hokkaido, Japan: IEEE Press, Mar. 7, 2022, pp. 332–341.

[182] D.-A. Frank, P. Chrysochou, and P. Mitkidis, "The paradox of technology: Negativity bias in consumer adoption of innovative technologies", *Psychology & Marketing*, vol. 40, no. 3, pp. 554–566, 2023, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mar.21740. [Online]. Available: `https://onlinelibrary.wiley.com/doi/abs/10.1002/mar.21740`.

[183] R. Bellini, "Paying the price: When intimate partners use technology for financial harm", in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, Hamburg Germany: ACM, Apr. 19, 2023, pp. 1–17, ISBN: 978-1-4503-9421-5. [Online]. Available: `https://dl.acm.org/doi/10.1145/3544548.3581101`.

[184] B Ashwini, V. Narayan, A. Bhatia, and J. Shukla, "Responsiveness towards robot-assisted interactions among pre-primary children of indian ethnicity", in *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, Vancouver, BC, Canada: IEEE, Aug. 8, 2021, pp. 619–625, ISBN: 978-1-66540-492-1. [Online]. Available: `https://ieeexplore.ieee.org/document/9515520/`.

[185] D. Kumar Singh, S. Sharma, J. Shukla, and G. Eden, "Toy, tutor, peer, or pet? preliminary findings from child-robot interactions in a community school", in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20, New York, NY, USA: Association for Computing Machinery, Apr. 1, 2020, pp. 325–327, ISBN: 978-1-4503-7057-8. [Online]. Available: `https://doi.org/10.1145/3371382.3378315`.

[186] Y. Anokwa, "Stories from the field: Reflections on HCI4d experiences", *Information Technologies and International Development*, vol. 5, no. 4, 2009.

[187] E. Beede *et al.*, "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy",

in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, ser. CHI '20, Honolulu, HI, USA: Association for Computing Machinery, 2020, 1–12, ISBN: 9781450367080. [Online]. Available: https://doi.org/10.1145/3313831.3376718.

[188] N. Dell and N. Kumar, "The ins and outs of hci for development", in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, San Jose, California, USA: Association for Computing Machinery, 2016, 2220–2232, ISBN: 9781450333627. [Online]. Available: https://doi.org/10.1145/2858036.2858081.

[189] S. Serholt *et al.*, "The case of classroom robots: Teachers' deliberations on the ethical tensions", *AI & Soc*, vol. 32, no. 4, pp. 613–631, Nov. 2017. [Online]. Available: https://doi.org/10.1007/s00146-016-0667-2.

[190] S. Elbeleidy, T. Mott, and T. Williams, "Practical, ethical, and overlooked: Teleoperated socially assistive robots in the quest for autonomy", in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Sapporo, Japan: IEEE, Mar. 7, 2022, pp. 577–587, ISBN: 978-1-66540-731-1. [Online]. Available: https://ieeexplore.ieee.org/document/9889573/.

[191] K. Shinohara and J. O. Wobbrock, "Self-conscious or self-confident? a diary study conceptualizing the social accessibility of assistive technology", *ACM Transactions on Accessible Computing*, vol. 8, no. 2, 5:1–5:31, Jan. 13, 2016. [Online]. Available: https://dl.acm.org/doi/10.1145/2827857.

[192] S. Shahid, O. Mubin, A. Al Mahmud, Z. Iftikhar, and R. Arshad, "Child-computer interaction in the global south: Designing for children on the margins", in *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*, ser. IDC '21, Athens, Greece: Association for Computing Machinery, 2021, 655–657, ISBN: 9781450384520. [Online]. Available: https://doi.org/10.1145/3459990.3460518.

[193] S. Qureshi, "What is the role of mobile phones in bringing about growth?", *Information Technology for Development*, vol. 19, no. 1,

1–4, Jan. 2013. [Online]. Available: `http://dx.doi.org/10.1080/02681102.2013.764597`.

[194] A. S. Sife, E. Kiondo, and J. G. Lyimo-Macha, "Contribution of mobile phones to rural livelihoods and poverty reduction in morogoro region, tanzania", *THE ELECTRONIC JOURNAL OF INFORMATION SYSTEMS IN DEVELOPING COUNTRIES*, vol. 42, no. 1, 1–15, Jul. 2010. [Online]. Available: `http://dx.doi.org/10.1002/j.1681-4835.2010.tb00299.x`.

[195] V. Chari, "Internet. org and the rhetoric of connectivity", *Communication and Critical/Cultural Studies*, vol. 19, no. 1, pp. 54–73, 2022.

[196] P. Ravi, A. Ismail, and N. Kumar, "The pandemic shift to remote learning under resource constraints", *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, 314:1–314:28, CSCW2 Oct. 18, 2021. [Online]. Available: `https://dl.acm.org/doi/10.1145/3476055`.

[197] J. L. R. Rubenstein and S. O. Moldin, Eds., *Understanding Autism*, 1st edition, Boca Raton: CRC Press Inc, Apr. 25, 2006, 552 pp., ISBN: 978-0-8493-2732-2.

[198] G. A. Boy, "A framework for flexibility analysis in sociotechnical systems", in *Design for Flexibility: A Human Systems Integration Approach*, ser. Human–Computer Interaction Series, G. A. Boy, Ed., Cham: Springer International Publishing, 2021, pp. 7–19, ISBN: 978-3-030-76391-6. [Online]. Available: `https://doi.org/10.1007/978-3-030-76391-6_2`.

[199] G. Allanwood and P. Beare, *User Experience Design: A Practical Introduction*, 2nd edition. London New York Oxford New Delhi Sydney: Bloomsbury Visual Arts, May 30, 2019, 176 pp., ISBN: 978-1-350-02170-9.

[200] L. A. Barba, "Defining the role of open source software in research reproducibility", *Computer*, vol. 55, no. 8, pp. 40–48, Aug. 2022, Conference Name: Computer. [Online]. Available: `https://ieeexplore.ieee.org/document/9847328`.

[201] S. Gitau and G. Marsden, “Fair partnerships – working with NGOs”, in *Human-Computer Interaction – INTERACT 2009*, ser. Lecture notes in computer science, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 704–707.

[202] J. Pal, M. Pradhan, M. Shah, and R. Babu, “Assistive technology for vision-impairments: Anagenda for the ICTD community”, in *Proceedings of the 20th international conference companion on World wide web*, Hyderabad India: ACM, Mar. 28, 2011, pp. 513–522, ISBN: 978-1-4503-0637-9. [Online]. Available: https://dl.acm.org/doi/10.1145/1963192.1963365.

[203] N. Gasteiger *et al.*, “Participatory design, development, and testing of assistive health robots with older adults: An international four-year project”, *ACM Transactions on Human-Robot Interaction*, vol. 11, no. 4, 45:1–45:19, Sep. 8, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3533726.

[204] L. Hudson *et al.*, “Co-designing digital tools to enhance speech and language therapy training in ghana”, in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA: ACM, Apr. 21, 2020, pp. 1–13, ISBN: 978-1-4503-6708-0. [Online]. Available: https://dl.acm.org/doi/10.1145/3313831.3376474.

[205] M. Axelsson, R. Oliveira, M. Racca, and V. Kyrki, “Social robot co-design canvases: A participatory design framework”, *ACM Transactions on Human-Robot Interaction*, vol. 11, no. 1, pp. 1–39, Mar. 31, 2022. [Online]. Available: https://dl.acm.org/doi/10.1145/3472225.

[206] J. Simonsen and T. Robertson, Eds., *Routledge international handbook of participatory design*, Routledge international handbooks, New York London: Routledge Taylor & Francis, 2013, 295 pp., ISBN: 978-0-415-72021-2 978-0-415-69440-7.

[207] K. Toyama, “Designing for aspirations”, *Interactions*, vol. 27, no. 3, pp. 61–63, Apr. 17, 2020. [Online]. Available: `https://dl.acm.org/doi/10.1145/3394059`.

[208] L. Irani, J. Vertesi, P. Dourish, K. Philip, and R. E. Grinter, “Postcolonial computing: A lens on design and development”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’10, New York, NY, USA: Association for Computing Machinery, Apr. 10, 2010, pp. 1311–1320, ISBN: 978-1-60558-929-9. [Online]. Available: `https://dl.acm.org/doi/10.1145/1753326.1753522`.

[209] P. Mikalef, K. Conboy, J. E. Lundström, and A. Popovič, “Thinking responsibly about responsible ai and ‘the dark side’ of ai”, *European Journal of Information Systems*, vol. 31, no. 3, 257–268, Feb. 2022. [Online]. Available: `http://dx.doi.org/10.1080/0960085X.2022.2026621`.

[210] A. Uzorka, S. Namara, and A. O. Olaniyan, “Modern technology adoption and professional development of lecturers”, *Education and Information Technologies*, vol. 28, no. 11, pp. 14 693–14 719, Nov. 1, 2023. [Online]. Available: `https://doi.org/10.1007/s10639-023-11790-w`.

[211] S. Langman, N. Capicotto, Y. Maddahi, and K. Zareinia, “Roboethics principles and policies in europe and north america”, *SN Applied Sciences*, vol. 3, no. 12, Nov. 2021. [Online]. Available: `http://dx.doi.org/10.1007/s42452-021-04853-5`.

[212] K. Zhao, W.-S. Chu, and H. Zhang, “Deep region and multi-label learning for facial action unit detection”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3391–3399.

[213] Y. Fan, X. Lu, D. Li, and Y. Liu, “Video-based emotion recognition using cnn-rnn and c3d hybrid networks”, in *Proceedings of the 18th ACM international conference on multimodal interaction*, 2016, pp. 445–450.

[214] D. H. Kim, W. J. Baddar, J. Jang, and Y. M. Ro, “Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition”, *IEEE Transactions on Affective Computing*, vol. 10, no. 2, pp. 223–236, 2017.

[215] A. T. Lopes, E. De Aguiar, A. F. De Souza, and T. Oliveira-Santos, “Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order”, *Pattern recognition*, vol. 61, pp. 610–628, 2017.

[216] Y. Li, J. Zeng, S. Shan, and X. Chen, “Occlusion aware facial expression recognition using cnn with attention mechanism”, *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439–2450, 2018.

[217] B. Knyazev, R. Shvetsov, N. Efremova, and A. Kuharenko, “Convolutional neural networks pretrained on large face recognition datasets for emotion classification from video”, *arXiv preprint arXiv:1711.04598*, 2017.

[218] S. F. Aly and A. L. Abbott, “Facial emotion recognition with varying poses and/or partial occlusion using multi-stage progressive transfer learning”, in *Scandinavian Conference on Image Analysis*, Springer, 2019, pp. 101–112.

[219] T. Q. Ngo and S. Yoon, “Facial expression recognition on static images”, in *International Conference on Future Data and Security Engineering*, Springer, 2019, pp. 640–647.

[220] R. Gupta and L.-A. Ratinov, “Text categorization with knowledge transfer from heterogeneous data sources.”, in *AAAI*, 2008, pp. 842–847.

[221] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, “Multisource transfer learning with convolutional neural networks for lung pattern analysis”, *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 76–84, 2017.

[222] J. Meyer, "Multi-task and transfer learning in low-resource speech recognition", Ph.D. dissertation, The University of Arizona, 2019.

[223] Y. Mansour, M. Mohri, and A. Rostamizadeh, "Domain adaptation with multiple sources", *Advances in neural information processing systems*, vol. 21, 2008.

[224] H. Guo, R. Pasunuru, and M. Bansal, "Multi-source domain adaptation for text classification via distancenet-bandits", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 7830–7838.

[225] H. Zhao, S. Zhang, G. Wu, J. M. Moura, J. P. Costeira, and G. J. Gordon, "Adversarial multiple source domain adaptation", *Advances in neural information processing systems*, vol. 31, pp. 8559–8570, 2018.

[226] R. Xu, Z. Chen, W. Zuo, J. Yan, and L. Lin, "Deep cocktail network: Multi-source unsupervised domain adaptation with category shift", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3964–3973.

[227] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks", in *International Conference on Machine Learning*, PMLR, 2017, pp. 1126–1135.

[228] B. Wang, J. Mendez, M. Cai, and E. Eaton, "Transfer learning via minimizing the performance gap between domains", in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019.

[229] D. J. Shah, "Multi-source domain adaptation with mixture of experts", Ph.D. dissertation, Massachusetts Institute of Technology, 2019.

[230] J. Lee, P. Sattigeri, and G. Wornell, "Learning new tricks from old dogs: Multi-source transfer learning from pre-trained networks", *Advances in neural information processing systems*, vol. 32, pp. 4370–4380, 2019.

[231] X. Qin, J. Wang, Y. Chen, W. Lu, and X. Jiang, "Domain generalization for activity recognition via adaptive feature fusion", *ACM Trans. Intell. Syst. Technol.*, 2022, Just Accepted. [Online]. Available: https://doi.org/10.1145/3552434.

[232] X. Yue, Z. Zheng, H. P. Das, K. Keutzer, and A. S. Vincentelli, "Multi-source few-shot domain adaptation", *arXiv preprint arXiv:2109.12391*, 2021.

[233] S. M. Ahmed, D. S. Raychaudhuri, S. Paul, S. Oymak, and A. K. Roy-Chowdhury, "Unsupervised multi-source domain adaptation without access to source data", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 103–10 112.

[234] Y. S. Kong, V. Suresh, J. Soh, and D. C. Ong, "A systematic evaluation of domain adaptation in facial expression recognition", *arXiv preprint arXiv:2106.15453*, 2021.

[235] J. Liang, D. Hu, and J. Feng, "Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation", in *International Conference on Machine Learning*, PMLR, 2020, pp. 6028–6039.

[236] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, "Impact of deep learning approaches on facial expression recognition in healthcare industries", *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5619–5627, 2022.

[237] S. Feizi, A. Makhdoumi, K. Duffy, M. Kellis, and M. Medard, "Network maximal correlation", *IEEE Transactions on Network Science and Engineering*, vol. 4, no. 4, pp. 229–247, 2017.

[238] H. O. Hirschfeld, "A connection between correlation and contingency", in *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press, vol. 31, 1935, pp. 520–524.

[239] H. Gebelein, "Das statistische problem der korrelation als variations- und eigenwertproblem und sein zusammenhang mit der ausgleich-

srechnung”, *ZAMM-Journal of Applied Mathematics and Mechanics/ Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 21, no. 6, pp. 364–379, 1941.

[240] A. Rényi, “On measures of dependence”, *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 10, no. 3-4, pp. 441–451, 1959.

[241] S.-L. Huang, A. Makur, L. Zheng, and G. W. Wornell, “An information-theoretic approach to universal feature selection in high-dimensional inference”, in *2017 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2017, pp. 1336–1340.

[242] H. V. Nguyen, E. Müller, J. Vreeken, P. Efros, and K. Böhm, “Multivariate maximal correlation analysis”, in *International Conference on Machine Learning*, PMLR, 2014, pp. 775–783.

[243] D. N. Reshef *et al.*, “Detecting novel associations in large data sets”, *science*, vol. 334, no. 6062, pp. 1518–1524, 2011.

[244] H. Hotelling, “Relations between two sets of variates”, in *Breakthroughs in statistics*, Springer, 1992, pp. 162–190.

[245] X. Zhang, F. Pan, W. Wang, and A. Nobel, “Mining non-redundant high order correlations in binary data”, in *Proceedings of the VLDB Endowment International Conference on Very Large Data Bases*, NIH Public Access, vol. 1, 2008, p. 1178.

[246] I. J. Goodfellow *et al.*, “Challenges in representation learning: A report on three machine learning contests”, in *International conference on neural information processing*, Springer, 2013, pp. 117–124.

[247] R Pascanu, T Mikolov, and Y Bengio, “Proceedings of the 30th international conference on international conference on machine learning-volume 28, icml’13”, 2013.

[248] M. J. Lyons, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets (ivc special issue)”, *arXiv preprint arXiv:2009.05938*, 2020.

[249] S. Li, W. Deng, and J. Du, “Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild”, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017, pp. 2584–2593.

[250] V. LoBue and C. Thrasher, “The child affective facial expression (cafe) set: Validity and reliability from untrained adults”, *Frontiers in psychology*, vol. 5, p. 1532, 2015.

[251] H. Zhao, S. Zhang, G. Wu, J. P. Costeira, J. M. F. Moura, and G. J. Gordon, “Multiple source domain adaptation with adversarial training of neural networks”, 2017. arXiv: `1705.09684 [cs.LG]`.

[252] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[253] J. Demšar, “Statistical comparisons of classifiers over multiple data sets”, *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[254] F. M. Alamgir, S. M. H. Saif, S. M. Hossain, A. A. Hadi, and M. S. Alam, “Facial expression database of autism spectrum disorder children”, vol. 12, no. 04, 21109–21120, 2023.

[255] A. C. Salem *et al.*, “Evaluating atypical language in autism using automated language measures”, *Scientific reports*, vol. 11, no. 1, pp. 1–12, 2021.

[256] G. Gagliardi and F. Tamburini, “The automatic extraction of linguistic biomarkers as a viable solution for the early diagnosis of mental disorders”, in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 5234–5242.

[257] H. Drimalla *et al.*, “Towards the automatic detection of social biomarkers in autism spectrum disorder: Introducing the simulated interaction task (sit)”, *NPJ digital medicine*, vol. 3, no. 1, p. 25, 2020.

[258] J. C. Lau *et al.*, "Cross-linguistic patterns of speech prosodic differences in autism: A machine learning study", *Plos one*, vol. 17, no. 6, e0269637, 2022.

[259] J. C. Lau, M. Losh, and M. Speights, "Differences in speech articulatory timing and associations with pragmatic language ability in autism", *Research in Autism Spectrum Disorders*, vol. 102, p. 102 118, 2023.

[260] H. MacFarlane, A. C. Salem, L. Chen, M. Asgari, and E. Fombonne, "Combining voice and language features improves automated autism detection", *Autism Research*, vol. 15, no. 7, pp. 1288–1300, 2022.

[261] H. MacFarlane *et al.*, "Consistency and reliability of automated language measures across expressive language samples in autism", *Autism Research*, 2023.

[262] L. Neely *et al.*, "The case for integrated advanced technology in applied behavior analysis", *Advances in Neurodevelopmental Disorders*, pp. 1–11, 2022.

[263] A. Rybner *et al.*, "Vocal markers of autism: Assessing the generalizability of machine learning models", *Autism Research*, vol. 15, no. 6, pp. 1018–1030, 2022.

[264] R. L. Rhodes, S. H. Ochoa, and S. O. Ortiz, *Assessing culturally and linguistically diverse students: A practical guide*. Guilford press, 2005.

[265] P. Perepa, "Cultural basis of social 'deficits' in autism spectrum disorders", *European Journal of Special Needs Education*, vol. 29, no. 3, pp. 313–326, 2014. eprint: `https://doi.org/10.1080/08856257.2014.908024`. [Online]. Available: `https://doi.org/10.1080/08856257.2014.908024`.

[266] T. Muskett, "Examining language and communication in autism spectrum disorder–in context", *Re-thinking autism: Diagnosis, identity and equality*, pp. 300–316, 2016.

[267] J. Dunn, “Global syntactic variation in seven languages: Toward a computational dialectology”, *Frontiers in Artificial Intelligence*, vol. 2, 2019.

[268] R. Bayley, K. Greer, and C. Holland, “Lexical frequency and syntactic variation: A test of a linguistic hypothesis”, *University of Pennsylvania Working Papers in Linguistics*, vol. 19, no. 2, p. 4, 2013.

[269] M. Gasser, “How language works”, *Indiana University (electronic resource*, 2006.

[270] J. I. Saeed, *Semantics*. John Wiley & Sons, 2015.

[271] Y. Huang, *Pragmatics*. Oxford University Press, USA, 2014.

[272] B. MacWhinney, *The CHILDES project: The database*. Psychology Press, 2000, vol. 2.

[273] I.-M. Eigsti, L. Bennetto, and M. B. Dadlani, “Beyond pragmatics: Morphosyntactic development in autism”, *Journal of autism and developmental disorders*, vol. 37, no. 6, pp. 1007–1023, 2007.

[274] H. Tager-Flusberg, S. Calkins, T. Nolin, T. Baumberger, M. Anderson, and A. Chadwick-Dias, “A longitudinal study of language acquisition in autistic and down syndrome children”, *Journal of autism and developmental disorders*, vol. 20, no. 1, pp. 1–21, 1990.

[275] J. Bang and A. Nadig, “Learning language in autism: Maternal linguistic input contributes to later vocabulary”, *Autism Research*, vol. 8, no. 2, pp. 214–223, 2015.

[276] C. Lord, M. Rutter, P. C. DiLavore, S. Risi, K. Gotham, S. L. Bishop, *et al.*, “Ados”, *Autism diagnostic observation schedule. Manual. Los Angeles: WPS*, 1999.

[277] R. R. Jordan, “An experimental comparison of the understanding and use of speaker-addressee personal pronouns in autistic children”, *In-*

ternational Journal of Language & Communication Disorders, vol. 24, no. 2, pp. 169–179, 1989.

[278] K. Stenning and M. Van Lambalgen, “Reasoning and discourse coherence in autism spectrum disorder”, in *Thinking, reasoning, and decision making in autism*, Routledge, 2019, pp. 135–155.

[279] C. M. Hart and S. Curtin, “Trajectories of vocabulary development in children with autism spectrum disorder across multiple measures”, *Journal of Autism and Developmental Disorders*, pp. 1–16, 2021.

[280] E. Loper and S. Bird, “Nltk: The natural language toolkit”, *arXiv preprint cs/0205028*, 2002.

[281] M. Marcus, B. Santorini, and M. A. Marcinkiewicz, “Building a large annotated corpus of english: The penn treebank”, *Computational linguistics*, vol. 19, no. 2, pp. 313–330, 1993.

[282] X. Lu, “Automated measurement of syntactic complexity in corpus-based l2 writing research and implications for writing assessment”, *Language Testing*, vol. 34, no. 4, pp. 493–511, 2017.

[283] E. Leinonen and D. Kerbel, “Relevance theory and pragmatic impairment”, *International Journal of Language & Communication Disorders*, vol. 34, no. 4, pp. 367–390, 1999.

[284] T. Jolliffe and S. Baron-Cohen, “Linguistic processing in high-functioning adults with autism or asperger’s syndrome. is global coherence impaired?”, *Psychological medicine*, vol. 30, no. 5, pp. 1169–1187, 2000.

[285] J. Heilmann, J. Miller, and A Nockerts, “Using language sample databases”, vol. 41, no. 1, 84–95, 2010.

[286] S. Cho, M. Liberman, N. Ryant, M. Cola, R. T. Schultz, and J. Parish-Morris, “Automatic detection of autism spectrum disorder in children using acoustic and text features from brief natural conversations.”, in *Interspeech*, 2019, pp. 2513–2517.

[287] Y.-K. Kim *et al.*, "Analyzing short term dynamic speech features for understanding behavioral traits of children with autism spectrum disorder.", in *Interspeech*, 2021, pp. 2916–2920.

[288] M. Kumar, R. Gupta, D. Bone, N. Malandrakis, S. Bishop, and S. S. Narayanan, "Objective language feature analysis in children with neurodevelopmental disorders during autism assessment.", in *Interspeech*, 2016, pp. 2721–2725.

[289] G. Arora, "Inltk: Natural language toolkit for indic languages", *arXiv preprint arXiv:2009.12534*, 2020.

[290] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit", in *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, 2014, pp. 55–60.

[291] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A python natural language processing toolkit for many human languages", *arXiv preprint arXiv:2003.07082*, 2020.

[292] F. Eyben *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing", *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

[293] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: The munich versatile and fast open-source audio feature extractor", in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.

[294] V. Gambhir, *Syntactic restrictions and discourse functions of word order in standard Hindi*. University of Pennsylvania, 1981.

[295] L. B. Silverman, "Verbal semantic coherence", *Encyclopedia of Autism Spectrum Disorders*, pp. 5039–5039, 2021.

[296] K. Wittke, A. M. Mastergeorge, S. Ozonoff, S. J. Rogers, and L. R. Naigles, "Grammatical language impairment in autism spectrum dis-

order: Exploring language phenotypes beyond standardized testing", *Frontiers in psychology*, vol. 8, p. 532, 2017.

[297] K. M. Bigelow, *Reliability of the Mean Length of Utterance Measure in Samples of Children's Language*. Brigham Young University, 2012.

[298] C. J. Wynn, S. A. Borrie, and T. P. Sellers, "Speech rate entrainment in children and adults with and without autism spectrum disorder", *American Journal of Speech-Language Pathology*, vol. 27, no. 3, pp. 965–974, 2018.

[299] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: Liwc and computerized text analysis methods", *Journal of language and social psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[300] V. G. Prakash *et al.*, "Computer vision-based assessment of autistic children: Analyzing interactions, emotions, human pose, and life skills", *IEEE Access*, 2023.

[301] V. G. Prakash *et al.*, "Video-based real-time assessment and diagnosis of autism spectrum disorder using deep neural networks", *Expert Systems*, e13253, 2023.

[302] D. Rane, P. Sharma, M. Singh, and U. Lahiri, "Virtual reality based gaze-sensitive aiming task platform: Role of attention allocation in task performance for individuals with autism and typically developing individuals", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1492–1501, 2023.

[303] S. Raj and S. Masood, "Analysis and detection of autism spectrum disorder using machine learning techniques", *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.

[304] J. B. S, D. Pandian, S. S. Rajagopalan, and D. Jayagopi, "Detecting a child's stimming behaviours for autism spectrum disorder diagnosis using rgbpose-slowfast network", in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 3356–3360.

[305] P. Pennisi *et al.*, “Autism and social robotics: A systematic review”, *Autism Research*, vol. 9, no. 2, pp. 165–183, 2016.