Development of Artificial Intelligence-Based Autism Spectrum Disorder Detection Framework

Thesis submitted in partial fulfillment of the requirements for the award of the degree of

Doctor of Philosophy

by

KAINAT KHAN

(2K21/PhD/CO/501)

Under the supervision of

Prof. Rahul Katarya

Department of Computer Science and Engineering

Delhi Technological University

Delhi, India

2025

Dedicated to

My beloved Parent

CANDIDATE DECLARATION

I hereby declare that the thesis entitled "Development of Artificial Intelligence-Based Autism Spectrum Disorder Detection Framework" submitted to Delhi Technological University, Delhi, in the partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in the Department of Computer Science, is an original work and has been done by myself under the supervision of Prof. Rahul Katarya (Supervisor), Department of Computer Science and Engineering, Delhi Technological University, Delhi, India.

The interpretations presented are based on my study and understanding of the original texts. The work reported here has not been submitted to any other institute for the award of any other degree.

Kainat Khan
Roll No. 2K21/PhD/CO/501
Department of Computer Science and Engineering
Delhi Technological University
Delhi-110042, India

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
(Govt. of National Capital Territory of Delhi)
Shahbad Daulatpur, Main Bawana Road,
Delhi-110042, India

Date:_____

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled "Development of Artificial Intelligence-Based Autism Spectrum Disorder Detection Framework" submitted by Ms. Kainat Khan (Roll No. 2K21/PhD/CO/501) in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy, to the Delhi Technological University, Delhi, India is carried out by the candidate under my supervision and guidance at the Department of Computer Science and Engineering, Delhi Technological University, Delhi, India.

The results embodied in this thesis have not been presented to any other University or Institute for the award of any degree or diploma.

Prof. Rahul Katarya
Department of Computer Science and Engineering
Delhi Technological University
Delhi-110042, India

ACKNOWLEDGMENT

I address my sincere thanks to Almighty God for giving me the inner power to complete my thesis and guide me in every step of my life.

It is an immense pleasure to have the opportunity to express my heartfelt gratitude to everyone who helped me throughout this research. I would like to express my heartfelt gratitude and indebtedness to my supervisor Prof. Rahul Katarya (Dept. of Computer Science & Engineering), for his invaluable and positive guidance, encouragement, and patience. During the research, his motivation and encouragement have inspired me to grow as a scholar and as a person. I am deeply indebted to my supervisor for guiding me in carrying out the research work and morally supporting me in every way during the course's challenging times. His technical expertise, precise suggestions, kind nature, and detailed, timely discussions are wholeheartedly appreciated.

Also, my sincere thank goes to Delhi Technological University for considering my candidature for this course. I am also very thankful to Prof. Prateek Sharma, Vice-Chancellor, Delhi Technological University, Delhi, India, who has been a constant source of enthusiasm. He has always motivated young researchers like me to pursue excellence to achieve higher goals in 3academics and research. Also, my sincere thanks reciprocate to Dr. Vinod Kumar (HoD, Dept. of Computer Science and Engineering), Prof. Rahul Katarya(Chairperson DRC, Dept. of Computer Science and Engineering) for insightful comments and valuable suggestions. Special thanks to my seniors and colleagues of Delhi Technological University, Delhi, India. My sincere thanks to all the professors, faculty, researchers, and nonteaching staff of the Computer Science Department. I would also like to express my gratitude to the Delhi Technological Univerisity (DTU) New Delhi, for providing financial support for the study.

I also wish to take this opportunity to thank all my teachers who have taught me and shaped me into the person I am, aggravated me to be an academician, and have directly indirectly made me capable of succeeding in completing this research work. I am thankful to all my colleagues and friends during my journey as a Ph.D. scholar. The engaging talks, brainstorming, and collaborative teamwork significantly impacted my growth as an independent researcher.

Finally, but most importantly, my heartfelt gratitude is for my parents, who are the motivations behind me; without their blessings, this work could not have been accomplished. I am truly indebted to them.

Kainat Khan
Roll No. 2K21/PhD/CO/501
Department of Computer Science and Engineering
Delhi Technological University
Delhi-110042, India

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social communication, restricted interests, and repetitive behaviors. The condition manifests in early childhood and persists throughout life, impacting an individual's ability to interact with their environment. ASD is highly heterogeneous, with symptoms and severity varying widely across individuals, making its diagnosis and management particularly challenging. Early and accurate identification of ASD is crucial, as timely interventions can significantly improve developmental outcomes and enhance the quality of life for those affected. The traditional approach to diagnosing ASD primarily relies on clinical observations, caregiver reports, and standardized behavioral assessments. While effective in many cases, these methods are often time-consuming, subjective, and dependent on the expertise of clinicians. This reliance on subjective evaluation introduces variability and delays in diagnosis, particularly in regions with limited access to specialized healthcare services. Consequently, there is a growing need for innovative solutions to improve the efficiency and accuracy of ASD detection.

Artificial Intelligence (AI) has emerged as a transformative force in healthcare, offering powerful tools for analyzing complex and diverse datasets. By leveraging AI techniques, it is possible to identify patterns and relationships within data that might not be readily apparent through traditional analysis. Machine learning and deep learning, subsets of AI, have demonstrated significant potential in various domains, including image analysis, natural language processing, and predictive modeling. These capabilities make AI particularly well-suited for addressing the challenges associated with ASD diagnosis. AI-driven approaches offer the advantage of objectivity, scalability, and the ability to integrate multiple data modalities, such as behavioral data, clinical records, and imaging studies. Furthermore, AI can facilitate early detection by identifying subtle patterns indicative of ASD, even in cases that might be missed by conventional diagnostic methods. As a result, AI-based tools have the potential to complement existing clinical practices, enhance diagnostic precision, and expand access to reliable ASD detection in underserved areas. The application of AI in ASD detection is an evolving field, with ongoing research aimed at developing innovative methods to tackle the complexities of the disorder. By combining advances in AI with insights from neuroscience and psychology, researchers aim to create solutions that not only improve diagnostic accuracy

but also offer interpretable results that can support clinical decision-making. Through such interdisciplinary efforts, AI holds promise in transforming the landscape of ASD diagnosis and care, ultimately contributing to better outcomes for individuals and their families. Therefore, this study represents a structured and methodical effort to assess the effectiveness, potential, and applicability of deep learning and computational intelligence techniques in the identification and analysis of ASD.

Objectives: The objectives of this study are structured into four key segments:

• The first objective of the study is to perform a systematic literature review on Autism Spectrum Disorder (ASD), which aims to critically evaluate the existing research, methodologies, and advancements in ASD detection.

• The second objective focuses on developing an intelligent diagnostic model for ASD using deep learning and computational intelligence techniques, aiming to improve the accuracy and efficiency of diagnosis.

• The third objective is to design a multi-modal framework for ASD detection, incorporating various data sources/modalities to enhance the overall performance of the diagnostic model.

• The final objective is to conduct a comparative analysis of the proposed ASD detection model with existing techniques, evaluating its effectiveness, accuracy, and applicability in real-world clinical settings.

Methodology: To accomplish the stated objectives, this study leverages advanced machine learning and deep learning methods, such as evolutionary algorithms, neural networks, attention mechanisms, and transformer-based architectures, due to their significant potential in addressing complex challenges in healthcare. The strategies employed to meet these objectives are as follows:

- To accomplish the first objective, a systematic literature review was conducted, focusing on machine learning techniques applied to Autism Spectrum Disorder (ASD) detection. This review analyzed various studies to identify the most effective models, methodologies, and data sources used in ASD diagnosis.
- For the second objective, two diagnostic models were developed, each utilizing different deep learning and evolutionary approaches. The first model incorporated an adaptive feature fusion technique to enhance the diagnosis process by combining

various data features obtained from particle swarm optimization (PSO) and the Bat algorithm effectively. The second model integrated a white shark optimization algorithm with a deep learning framework utilizing Bi-LSTM to improve the overall accuracy and robustness of ASD detection.

- To address the third objective, a multimodal diagnostic framework was designed that combines various data modalities, such as clinical features and imaging data. This framework employs advanced deep learning techniques, including a multi-head CNN architecture with channel and spatial attention (CBAC) and BERT, to extract and integrate features from diverse modalities for enhanced ASD detection.
- For the fourth objective, a comparative analysis was conducted, evaluating the performance of the above-developed models against existing ASD detection techniques. Key performance metrics, such as accuracy, sensitivity, specificity, and F1-score, were used to compare the effectiveness of the proposed models with current state-of-the-art methods.

Results: The outcomes of the study are as follows:

- A comprehensive review of machine learning techniques for Autism Spectrum Disorder (ASD) detection was conducted. This review analyzed current trends and identified potential future directions in the field, providing valuable insights into existing methodologies and areas for future research.
- A study is conducted to explore bio-inspired techniques for improving ASD diagnosis, with a focus on evolutionary algorithms. The study highlighted the promising potential of these algorithms in enhancing diagnostic accuracy for ASD detection.
- An adaptive feature fusion technique was developed for ASD diagnosis. This hybrid model combined bio-inspired optimization algorithms with feature fusion to effectively integrate various data features, enhancing the accuracy and robustness of the diagnostic process.
- A model was developed by integrating an optimization algorithm with the Bi-LSTM approach. This strategy aimed to improve feature selection, ultimately leading to improved overall performance in the ASD detection system.
- A multi-modal framework was created, integrating sequential (phenotypic information) and visual data (brain MRI) using convolutional block attention component and BERT-

based deep learning architectures. This framework significantly improved ASD detection accuracy and robustness by effectively combining different data types.

- A new approach employing facial images of autistic and non-autistic children combining convolutional networks and vision transformers was developed for the diagnosis of ASD. This model enhanced the processing of visual data, leading to improved diagnostic performance.
- A self-supervised and self-distillation learning approach was explored for ASD classification using facial images. This innovative method aimed to leverage unsupervised learning to improve the classification accuracy in ASD detection.
- A multi-modal diagnostic framework was designed, incorporating various data sources such as clinical and imaging data. This framework leveraged advanced deep learning techniques, including LSTM and transformer-based architectures, to extract and integrate relevant features, improving the diagnostic performance.

TABLE OF CONTENTS

CHAPTER 3: ASD CLASSIFICATION FRAMEWORK USING DEEP LEARNING WITH COMPUTATIONAL INTELLIGENCE TECHNIQUES

CHAPTER 4: ASD DIAGNOSIS USING MULTI-MODALITY ARCHITECTURE

LIST OF ABBREVIATIONS

ACC	Accuracy
AE	Autoencoder
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASD	Autism Spectrum Disorder
ABIDE	Autism Brain Imaging Data Exchange
BERT	Bidirectional Encoder Representations from Transformers
Bi-LSTM	Bidirectional Long Short-Term Memory
CNN	Convolution Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
ECG	Electrocardiography
EOG	Electrooculography
ES	Evolutionary Strategies
FP	False Positive
FN	False Negative
KL	Kullback–Leibler
k-NN	k-Nearest Neighbor
LightGBM	Light Gradient Boosting Machines
LR	Logistic Regression
LSTM	Long Short-term Memory
MAD	Mean Absolute Deviation
MSE	Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MLR	Multiple Linear Regression
MLP	Multi Layer Perceptron
MM	Multi Modal
NB	Naive Bayes
NBM	Multinomial Naive Bayes

NLP	Natural Language Processing
NN	Neural Network
R^2	R-Square
RF	Random Forest
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RoBERTa	Robustly optimized BERT approach
SAE	Stacked Autoencoder
sMRI/fMRI	Structural/Functional Magnetic Resource Imaging
TD	Typically Developed
TP	True Positive
TN	True Negative
SVM	Support Vector Machine
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting

LIST OF TABLES

LIST OF FIGURES

Chapter 1 INTRODUCTION

The emerging field of Artificial Intelligence (AI) in Autism Spectrum Disorder (ASD) detection offers groundbreaking tools to streamline the diagnosis and analysis of ASD. AI-based frameworks aim to enhance the efficiency and accuracy of early detection, significantly improving the support and interventions available to individuals with ASD. The early diagnosis of ASD is pivotal in addressing developmental challenges and improving long-term outcomes for individuals. Data-driven approaches powered by AI have become essential in automating diagnosis processes, shaping clinical practices, influencing policy development, and refining therapeutic programs. The adoption of AI as a foundation for ASD diagnosis highlights its transformative potential in healthcare, serving as the backbone for advancements in the field. ASD detection stands apart from traditional diagnostic methods by its focus on patterns in data, such as neuroimaging, behavioral assessments, and clinical metrics, rather than solely relying on manual evaluations. Over the years, ASD diagnostic frameworks have evolved significantly, with technological innovations enabling the integration of large-scale data analysis and multimodal inputs. Recent advancements in neural networks and deep learning architectures have revolutionized the way clinical and imaging data are processed, leading to more robust and reliable ASD detection frameworks. AI techniques, particularly Machine Learning (ML) and Deep Learning (DL), have emerged as game changers, leveraging diverse data to enhance diagnostic accuracy. The development of AI methodologies tailored to ASD detection has accelerated in the past decade, focusing on clinical tasks such as feature extraction from MRI scans, interpreting behavioral data, and classifying ASD phenotypes. These advancements have substantially benefited clinicians and researchers by providing data-driven insights. Additionally, the application of AI in healthcare is expanding, offering predictive capabilities that enable the estimation of diagnostic outcomes with unparalleled precision. This framework integrates multimodal data such as MRI imaging and meta-features, enhancing the decision-making process in clinical settings. Section 1.1 explores the significance of ASD diagnosis, emphasizing the need for AI-driven approaches. Section 1.2 elaborates on the objectives of developing an ASD diagnosis framework. Subsections provide insights into the principles and uses of the framework. Section 1.3 delves into machine learning (ML), detailing its necessity, types, and applications in healthcare. Section 1.4 discusses the intersection of ASD diagnosis and machine/deep learning, focusing on AI's potential to enhance diagnostic frameworks. Section 1.5 highlights the motivation behind this study, emphasizing the importance of addressing current challenges in ASD diagnosis. Section 1.6 outlines the research objectives, providing a roadmap for the study's focus and direction. Section 1.7 presents the structure of

the thesis, offering an overview of the chapters to guide the reader through the document. Finally, Section 1.8 concludes the chapter with a summary, encapsulating the key points discussed.

1.1. Autism Spectrum Disorder (ASD)

Real-time analysis in the field of ASD diagnosis has emerged as a rapidly growing area of interest among researchers worldwide. Advances in this domain have facilitated the introduction of innovative frameworks for ASD detection and management. The term "autism" originates from the Greek word "autos," meaning "self," reflecting the inward-focused behavior of individuals with ASD. Globally, there is a pressing need to enhance the accuracy and timeliness of ASD diagnoses, given its significant impact on individuals and society. A robust diagnostic framework offers critical insights for developing, implementing, monitoring, and assessing intervention programs aimed at supporting individuals with ASD. According to the World Health Organization (WHO), ASD is a developmental disorder characterized by difficulties in social interaction, communication, and repetitive behaviors. Accurate diagnosis is crucial for planning and evaluating interventions, as well as for understanding the prevalence of ASD in various populations. Early detection is acknowledged as a vital component of effective intervention and care, enabling individuals to access timely support and resources. The primary aim of ASD detection frameworks is to provide clinicians and researchers with meaningful, data-driven evidence to guide decisions and improve the quality of care.

Information derived from ASD diagnostic frameworks is used to identify early signs of the disorder, monitor the effectiveness of intervention strategies, and evaluate emerging trends in ASD prevalence. Diagnostic statistics are vital for understanding the health status of populations, tailoring therapies, and developing strategies to mitigate the challenges faced by individuals with ASD. The significant role of ASD detection in healthcare has motivated research to establish advanced diagnostic methodologies, strengthening the scientific foundation of ASD frameworks. Early identification of ASD symptoms is critical for implementing effective interventions, and data on developmental and behavioral patterns are essential for accurate diagnosis.

The study of ASD detection traces its origins to early developmental psychology and medical investigations into atypical behaviors in children. Historical records of unusual developmental patterns date back centuries, reflecting society’s longstanding interest in understanding such conditions. However, the concept of ASD, as it is understood today, has evolved significantly over time.

Earlier, diagnostic processes relied heavily on observational methods, often leading to delayed or inaccurate diagnoses. The introduction of advanced technologies such as neuroimaging, machine learning, and artificial intelligence has revolutionized the field, allowing for the integrating of multimodal data sources to improve diagnostic accuracy [1]. To distinguish traditional clinical methods from

contemporary frameworks, modern AI-based methodologies focus on leveraging diverse datasets, including imaging and clinical features, to assess ASD comprehensively. This transition has marked a significant milestone in the field, setting the stage for more precise, scalable, and impactful diagnostic solutions.

(a) (b)

Figure 1.1: Facial images of (a) Autistic and (b) Non-Autistic Children

1.2. Objectives of Autism Spectrum Disorder (ASD) Diagnosis Frameworks

AI-based ASD detection frameworks provide the empirical and data-driven insights necessary for informed decision-making and effective interventions. The primary goal of these frameworks is to provide accurate, timely data that can guide clinical actions and improve the support provided to individuals with ASD. The design and implementation of ASD detection systems are influenced by the specific objectives and interventions required for optimal care. For instance, if the goal is to diagnose ASD at an early stage, the framework must be capable of processing data from various sources, such as neuroimaging, clinical assessments, and behavioral data, to deliver precise results quickly. On the other hand, monitoring the progression of the disorder in individuals over time might require systems that track long-term data, such as behavioral changes or intervention outcomes, through continuous or periodic assessments.

Different health objectives related to ASD diagnosis necessitate distinct information systems. The type of diagnostic or monitoring system to be used depends on the specific clinical actions that need to be taken, as well as when and how frequently the data should be collected, analyzed, and used. For example, an AI-driven system designed to predict ASD risk based on early developmental signs might require real-time data analysis, while a system for tracking the long-term outcomes of ASD interventions might gather data over several years.

To design and evaluate an AI-based ASD detection system, it is essential to answer the following questions:

- What constitutes the case definition of ASD in the context of the framework? Is it practical and clinically relevant?
- What are the specific goals of the diagnostic framework, and how do they align with clinical or research objectives?
- Does the system integrate with other diagnostic and health information systems?
- How is data handled? What protocols ensure data privacy, security, and the prevention of delays in data transmission?
- What are the data sources (e.g., MRI scans, behavioral assessments, clinical evaluations)? Who is responsible for reporting and updating data?
- What types of information are collected, and does it meet the needs of healthcare providers and researchers?
- How frequently is data collected (e.g., during regular check-ups, or at significant developmental milestones)?
- How is the data analyzed? Who performs the analysis, and how regularly are updates made?
- How is the information disseminated to relevant stakeholders (e.g., clinicians, researchers, policymakers)? Are reports timely and comprehensive?
- What are the planned applications of the ASD detection framework data?
- What is the target population for the framework (e.g., infants, children, adults)?

Ultimately, the aim of an AI-based ASD detection framework is to monitor the trends in ASD prevalence, symptoms, and treatment efficacy within a population to guide research, management, and prevention strategies. Public health officials and clinicians may utilize the data provided by such frameworks to design more effective diagnostic procedures, interventions, and support systems. However, the application of these frameworks is not limited to merely identifying ASD cases. They also serve to enhance our understanding of the disorder's biological, developmental, and demographic aspects. This deeper understanding can lead to the development of more effective preventive measures, diagnostic tools, and intervention strategies. Additionally, AI-powered ASD detection systems offer a robust foundation for creating and implementing evidence-based clinical practices that are tailored to the specific needs of individuals across diverse populations.

1.2.1. Principles and Uses of Autism Spectrum Disorder Diagnosis Frameworks

The goal of collecting, evaluating, and distributing ASD-related information is to improve diagnostic accuracy and intervention effectiveness. The primary principle of AI-based ASD detection frameworks is that they should be designed and implemented to provide reliable, timely, and actionable information to

clinicians, researchers, and policymakers. In the context of ASD, diagnostic frameworks need to ensure that they are capable of delivering accurate results quickly to support early intervention, which is crucial for effective management of the disorder. Since early-stage ASD can present subtle variations across individuals, maintaining accuracy is essential to avoid misdiagnoses, while also ensuring that the system is efficient and cost-effective to implement at scale. The application of ASD detection data can vary depending on the clinical setting, whether immediate diagnostic decisions, long-term monitoring, or historical data analysis are needed.

The existing research emphasizes principles that are directly relevant to autism spectrum disorder (ASD) detection and management frameworks:

- **Identifying ASD Cases:** Early detection of ASD symptoms or clusters is crucial for timely diagnosis and intervention, enabling support services that mitigate developmental challenges.
- **Analyzing Prevalence and Impact:** Assessing the prevalence, patterns, and societal impacts of ASD within diverse populations helps shape policies and allocate resources effectively.
- **Evaluating Diagnostic and Therapeutic Interventions:** Measuring the efficacy of diagnostic tools and early intervention strategies ensures the adoption of the most effective approaches for managing ASD.
- **Building Awareness and Advocacy:** Effective frameworks foster greater awareness about ASD among the public and healthcare providers, leading to earlier detection, reduced stigma, and better support for affected individuals and their families.
- **Supporting Lifespan Approaches:** Beyond childhood diagnosis, robust frameworks facilitate ongoing monitoring and support across an individual's lifespan, addressing evolving needs in adolescence and adulthood.
- **Advancing ASD Research:** Establishing detection frameworks fosters research into the causes, progression, and treatment of ASD, paving the way for improved understanding, prevention, and intervention strategies.

Incorporating these principles into AI-based ASD detection frameworks can significantly enhance the precision and applicability of the tools, making them valuable not only for individual diagnosis but also for larger-scale epidemiological studies and public health planning. AI in this context helps continuously refine the diagnostic process and identify areas for improvement, ensuring that ASD detection remains at the forefront of medical and technological advancements.

1.2.2. Sources of Data

AI-based ASD detection frameworks rely on a variety of data sources to ensure accurate diagnosis and comprehensive analysis. These data sources are crucial for understanding the complex nature of ASD and

for developing more effective diagnostic tools and interventions. Some of the most commonly used sources of data for ASD detection include:

- **Clinical Reports and Assessments:** Reports from clinicians, including developmental screenings, behavioral assessments, and psychological evaluations, provide essential data for diagnosing ASD. These reports often include data on early childhood development, cognitive and social behaviors, and family medical histories.
- **Neuroimaging Data:** Structural and functional MRI scans, fMRI, and other neuroimaging techniques offer valuable insights into the brain structure and activity of individuals with ASD. These images are crucial for identifying neurobiological markers associated with ASD and for enhancing diagnostic accuracy.
- **Genetic Data:** Genomic studies and DNA sequencing provide insights into the genetic underpinnings of ASD. Data from genetic research help in understanding the hereditary factors contributing to ASD and may also be used to personalize interventions based on genetic profiles.
- **Epidemiological Data:** Population-based studies and surveys offer broad data on the prevalence of ASD across different demographics. This includes data on age, gender, socioeconomic status, and geographic location, which helps in understanding how ASD manifests in various populations.
- **Behavioral and Developmental Data:** Data collected through behavioral observations, parent and teacher questionnaires, and standardized developmental assessment tools are used to track the progress of children with ASD and monitor the effectiveness of interventions.
- **Sensor Data:** Wearable sensors, such as eye trackers, accelerometers, and movement sensors, can provide real-time data on the behavior and physical activity of individuals with ASD. This data is used to monitor physical and social behaviors that are indicative of ASD symptoms.
- **Social Media and Online Data:** Social media platforms and online forums are emerging as valuable sources of data for understanding public perceptions of ASD, gathering patient feedback, and identifying potential trends in ASD diagnosis and treatment. This type of data can also provide insights into the experiences of individuals living with ASD and their caregivers.
- **Special Surveys and Questionnaires:** Surveys such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Spectrum Quotient (AQ) are frequently used to collect specific diagnostic information about individuals at risk for ASD. These tools help to assess various aspects of social and communication behaviors.
- **Educational and School Records:** Data from schools, including performance reports, behavioral evaluations, and teacher observations, contribute valuable information about a child's social interaction skills, communication abilities, and academic performance.

1.2.3.Enhancing the Use of Computer Technology in Autism Spectrum Disorder

In the realm of Autism Spectrum Disorder (ASD) diagnosis, the integration of advanced computer technologies and artificial intelligence (AI) is transforming traditional diagnostic processes and enhancing the accuracy, speed, and personalization of ASD detection frameworks. AI-based systems, utilizing machine learning algorithms and neural networks, can analyze large and complex datasets from multiple sources, thereby improving the understanding and diagnosis of ASD.

Historically, the process of diagnosing ASD has involved clinical assessments, behavioral observations, and diagnostic tools like the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R). While these methods are effective, they can be subjective and time-consuming. The introduction of AI-driven models offers a promising solution by automating the analysis of data such as brain imaging (e.g., MRI scans), genetic data, and behavioral observations. These systems can process and analyze massive datasets at speeds unimaginable through manual methods, improving diagnostic precision and reducing the risk of human error.

For example, the use of deep learning algorithms to process neuroimaging data has shown promise in identifying biomarkers related to ASD, with studies demonstrating the potential of AI models to detect subtle differences in brain structure and connectivity between individuals with ASD and neurotypical individuals. Similarly, models, such as convolutional neural networks (CNNs) and vision transformers, are being applied to analyze MRI scans, helping to detect abnormalities in brain regions associated with ASD, and providing valuable insights that were previously difficult to identify.

Moreover, the use of AI technologies in predictive modeling is advancing. Machine learning models, trained on clinical and demographic data, can predict the likelihood of an individual developing ASD at an early stage, which is crucial for early intervention strategies. For instance, early screening tools powered by AI are increasingly being developed to detect signs of ASD in toddlers or even infants, often before observable behaviors manifest. Furthermore, multimodal frameworks that combine different data sources, such as clinical data, neuroimaging, genetic information, and behavioral data, are gaining traction. By integrating these diverse sources, AI frameworks are able to build a more holistic profile of each patient, improving diagnostic accuracy and facilitating personalized treatment recommendations.

The use of AI in ASD detection also addresses some of the challenges faced by traditional diagnostic methods, such as inter-rater reliability and accessibility. AI-based models can be standardized across different clinics and regions, offering consistent results and reducing disparities in diagnosis across geographical and socio-economic contexts. Additionally, AI models can be deployed in telemedicine settings, enhancing access to diagnostic services in remote areas.

As with any technology, the implementation of AI in healthcare raises important concerns, particularly around data privacy, security, and ethics. Safeguarding patient confidentiality and ensuring that AI

systems are transparent and explainable are critical aspects that must be addressed. Research and development in this area must continue to prioritize these issues while also enhancing the capabilities of AI-driven ASD detection systems.

The future of ASD detection lies in the continued evolution of AI technologies, where their integration into clinical practice promises not only more accurate and faster diagnosis but also improved treatment strategies, early intervention, and better long-term outcomes for individuals with ASD. By harnessing the power of computer technologies and AI, we can significantly enhance our ability to detect, understand, and treat Autism Spectrum Disorder.

1.2.4. Autism Spectrum Disorder and Internet Technology

The pervasive presence of the internet and the vast amounts of digital information it generates have revolutionized numerous fields, including healthcare. This transformation extends to autism spectrum disorder (ASD), where internet technology is enabling novel diagnostic and therapeutic approaches. New concepts such as "digital phenotyping" and "digital health analytics" have emerged, focusing on the use of online data and computational methodologies to analyze behavior and identify patterns indicative of ASD. Digital phenotyping refers to the collection and analysis of behavioral and physiological data from various digital sources, including social media platforms, online interactions, and wearable devices. This approach complements traditional methods of ASD diagnosis by providing a scalable, real-time means to capture early indicators of the disorder.

Internet-based tools have shown promise in identifying ASD markers through the analysis of communication styles, interaction patterns, and user-generated content on online forums and social networks. For example, the linguistic patterns in posts by individuals with ASD, such as repetitive language or unusual sentence structures, have been studied to develop predictive models. These models use deep learning (DL) algorithms to analyze large-scale online behavioral data, enhancing diagnostic accuracy while reducing reliance on subjective clinical evaluations. The accessibility of internet technologies has also played a pivotal role in democratizing healthcare, including ASD diagnosis. Remote telehealth platforms, virtual assessments, and online screening tools allow clinicians to reach individuals in underserved regions, providing equitable access to diagnostic resources. Furthermore, these technologies enable the continuous monitoring of ASD symptoms, improving intervention timelines and developmental outcomes. The integration of internet technology with artificial intelligence (AI) has propelled advancements in this domain. The availability of open-source libraries, faster data processing, and collaborative data labeling have enabled researchers to develop AI models capable of analyzing multimodal datasets, such as neuroimaging data and clinical metadata. These models can uncover complex relationships and hidden patterns that may elude human observation. For instance, computer

vision algorithms applied to neuroimaging data have identified structural and functional brain anomalies associated with ASD. However, the use of internet-based data for ASD diagnosis is not without challenges. Ensuring the quality and reliability of datasets is critical, as noisy or incomplete data can affect model performance. Additionally, ethical concerns surrounding data privacy, informed consent, and the potential for biased predictions highlight the need for transparent and accountable AI systems. Addressing these challenges requires the development of robust computational approaches that balance diagnostic accuracy with ethical considerations. This work seeks to bridge the gap between internet technology, AI, and ASD diagnosis by proposing a novel framework that integrates multimodal data sources. Through advanced fusion techniques and explainable AI, this work aims to enhance diagnostic reliability, interpretability, and accessibility, contributing to the broader adoption of AI-driven ASD detection frameworks.

1.2.5. Popular Public Data Sources for Autism Spectrum Disorder Diagnosis

The development and evaluation of AI-based diagnostic methods for Autism Spectrum Disorder (ASD) heavily depend on the availability of high-quality, well-curated datasets. Public data sources play a crucial role in advancing research by providing standardized datasets that enable the training, testing, and validation of machine learning models. These datasets help researchers identify patterns, train diagnostic algorithms, and improve the accuracy and generalization of AI-based systems. Several public data sources have been widely used in ASD diagnosis, ranging from behavioral assessments to neuroimaging data. Below are some of the most popular public data sources for ASD research:

Table 1.1 Dataset analysis of Publicly available datasets related to autism spectrum disorder

Data Source	Dataset Type					Description
	Images	**Questionnaire**	**Text**	**Gene**	**Time-series**	
ABIDE I [2]	✓		✓			**ASD: Normal =** 539:573
ABIDE II [3]	✓		✓			**ASD: Normal** = 521:593
AGRE [4]				✓		https://www.autismspeaks.org/agre
NDAR [5]	✓	✓	✓	✓	✓	https://catalog.data.gov/dataset/national-database-for-autism-research-ndar
		✓				

Autism Screening [6]		**For Adults:** instances = 704 attributes = 21 **For Toddlers:** instances = 1054 attributes = 18 **For children:** instances = 292 attributes = 21 https://www.kaggle.com/datasets/faizunnabi/autism-screening
NRGR [7]	✓	https://www.nimhgenetics.org/download-tool/AU

1.2.5.1. ABIDE I [45]

Dataset Type: Images, Clinical features

The Autism Brain Imaging Data Exchange (ABIDE) I dataset consists of brain imaging data (structural MRI) and questionnaires collected from 539 individuals diagnosed with ASD and 573 typically developing (TD) participants. The primary focus of the dataset is to analyze the brain's structural and functional differences between these two groups, providing critical insights for the development of neuroimaging biomarkers for ASD.

1.2.5.2. ABIDE II [46]

Dataset Type: Images, Clinical features

ABIDE II extends the ABIDE I dataset and offers additional brain imaging data (including both structural and functional MRI) alongside questionnaires from 521 individuals with ASD and 593 TD individuals. It aims to provide a more comprehensive analysis of the neurodevelopmental variations in ASD, further contributing to the understanding of neurobiological markers associated with the disorder.

1.2.5.3. AGRE [47]

Dataset Type: Gene

The Autism Genetic Resource Exchange (AGRE) dataset primarily includes behavioral and diagnostic questionnaires. It serves as a valuable resource for genetic studies aimed at understanding the hereditary components of ASD. This dataset allows researchers to explore genetic correlations and how they may

influence the development of autism, facilitating the identification of potential genetic markers for early diagnosis.

1.2.5.4. NDAR [48]

Dataset Type: Images, Questionnaire, Text, Gene, Time-series

The National Database for Autism Research (NDAR) is one of the largest and most comprehensive repositories of autism-related data. NDAR includes various types of data such as brain imaging (structural MRI, functional MRI), behavioral questionnaires, genetic data, and time-series data. This extensive dataset is critical for large-scale research to understand the genetic, behavioral, and neurobiological factors contributing to ASD. It also supports longitudinal studies that track the progression of the disorder over time.

1.2.5.5. Autism Screening [49]

Dataset Type: Questionnaire

The Autism Screening dataset is available on Kaggle and includes data from autism screening tests for different age groups: adults, toddlers, and children. This dataset provides clinical features like age, gender, jaundice history and demographic attributes like ethnicity, making it useful for training AI models aimed at ASD diagnosis.

1.1.5.6. NRGR [50]

Dataset Type: Gene

The NeuroGenetics Research Group (NRGR) dataset contains genetic data associated with ASD. This dataset is valuable for studying the genetic underpinnings of ASD and provides insight into how specific genetic variations correlate with ASD symptoms and progression.

These publicly available datasets are instrumental in developing, testing, and refining AI-based diagnostic tools for ASD. By incorporating various data types such as neuroimaging, genetic information, and clinical assessments, these datasets help researchers create more robust and accurate diagnostic models that can assist in early detection and personalized interventions for individuals with autism.

1.2.6. Applications of AI-based Autism Spectrum Disorder (ASD) Diagnostic Methods

AI-based diagnostic methods have gained significant traction in the field of Autism Spectrum Disorder (ASD) diagnosis, offering the potential to improve early detection, enhance diagnostic accuracy, and support personalized treatment approaches. These technologies leverage machine learning, deep learning,

and computer vision to analyze a wide range of data sources, such as behavioral patterns, facial images, medical records, and neuroimaging data. The application of AI in ASD diagnosis has the potential to revolutionize the way healthcare professionals identify and treat individuals with ASD. Some of the key applications of AI-based ASD diagnostic methods include:

1.2.6.1. Early Detection and Screening

AI-based systems can significantly improve early detection of ASD, which is critical for initiating early interventions that can greatly improve long-term outcomes for individuals with ASD. Machine learning models, particularly those that analyze behavioral data or diagnostic questionnaires, can be trained to detect early signs of autism in young children, often before the age of three. For example, AI systems can analyze patterns in parental questionnaires, such as the Modified Checklist for Autism in Toddlers (M-CHAT), to flag children who are at risk of ASD. These systems can be integrated into routine pediatric screenings, enabling quicker identification of children who require further assessment or specialized intervention. Similarly, AI models can process large-scale datasets, such as those from early childhood assessments, to identify subtle signs of ASD that may be overlooked by human observers. This early detection capability can lead to faster referrals to specialists and early therapeutic interventions, which are known to have a significant positive impact on the development of children with ASD.

1.2.6.2. Analysis of Behavioral Data

AI-based methods can be applied to analyze behavioral data collected from various sources, such as interviews, questionnaires, or video recordings of children's interactions. Natural language processing (NLP) techniques can be used to analyze textual data from parent or caregiver reports to identify linguistic patterns or behavioral markers associated with ASD. For instance, children with ASD may exhibit delays in speech development, abnormal conversational patterns, or challenges in understanding social cues. By analyzing these patterns, AI systems can assist in diagnosing ASD with greater precision, detecting subtle features that human clinicians may miss. Moreover, computer vision algorithms can analyze video recordings of children's behaviors and facial expressions. These algorithms can detect anomalies in eye contact, social engagement, and emotional recognition, which are typical indicators of ASD. By processing these visual cues, AI systems can provide valuable insights into a child's developmental progress and offer objective data to support the diagnostic process.

1.2.6.3. Neuroimaging and Biomarker Detection

AI is increasingly being used in the analysis of neuroimaging data, such as MRI and fMRI scans, to identify potential biomarkers for ASD. Deep learning models can be trained to analyze brain images and

identify structural or functional abnormalities in regions typically associated with social cognition, language processing, and executive function. These models can detect subtle differences in brain structure and function that may be indicative of ASD, enabling the development of more objective diagnostic criteria. AI algorithms can also integrate multiple sources of neuroimaging data, such as structural MRI, functional MRI, and electroencephalography (EEG), to provide a comprehensive understanding of the neural underpinnings of ASD. This approach has the potential to lead to the identification of novel biomarkers that can be used to diagnose ASD more accurately and provide insights into the neurological basis of the disorder.

1.2.6.4. Facial Image Analysis

AI-based facial image analysis is another promising application in the diagnostic process for ASD. Computer vision models, particularly convolutional neural networks (CNNs), can be used to analyze facial features and expressions to detect atypical patterns often observed in individuals with ASD. These models can examine subtle differences in facial structure, gaze direction, and emotional expressions that may indicate autism-related traits.

For example, children with ASD may show limited facial expressiveness or differences in their ability to make eye contact. By processing large datasets of facial images, AI algorithms can be trained to distinguish between typical and atypical patterns in facial expressions and use this information as part of a broader diagnostic toolkit. This technology has the potential to be integrated into clinical assessments, offering objective measures to complement traditional diagnostic methods.

1.2.6.5. Personalized Treatment and Intervention

Once ASD is diagnosed, AI systems can play a significant role in creating personalized treatment and intervention plans. By analyzing data from various sources, including behavioral assessments, neuroimaging, and family history, AI models can help tailor interventions to the specific needs of individuals with ASD. This could involve recommending personalized therapies, such as speech or occupational therapy, and monitoring the progress of these interventions over time. AI systems can also provide real-time feedback to clinicians and caregivers, helping to adjust treatment strategies based on an individual's response. For instance, machine learning models can analyze data from wearable devices or mobile apps that track behavioral changes and offer insights into the effectiveness of specific therapies. This personalized approach enhances the precision of interventions, ensuring that individuals with ASD receive the most appropriate care based on their unique needs and progress.

1.2.6.6. Monitoring Long-Term Development

AI-based diagnostic methods can be applied to monitor the long-term development of individuals with ASD. Machine learning models can analyze longitudinal data collected over extended periods, such as behavioral assessments, medical records, and educational performance, to track changes in symptoms and progress. By detecting patterns over time, AI systems can help clinicians identify when interventions are most needed or when adjustments to treatment plans should be made. This ability to track and predict developmental trajectories is particularly valuable in managing ASD, as it allows for the early identification of emerging challenges and the timely adjustment of care plans. Moreover, AI models can support families and caregivers by providing actionable insights into the individual's development, helping them to better understand their child's needs and progress.

1.2.6.7. Improving Diagnostic Accuracy and Reducing Human Bias

AI-based diagnostic methods can help improve the accuracy and consistency of ASD diagnoses by reducing human biases and subjectivity. Traditional diagnostic approaches often rely heavily on clinician experience, which can introduce variability in decision-making. AI systems, however, can process large volumes of data objectively, identifying patterns and anomalies that might not be immediately apparent to human clinicians. This leads to more accurate and consistent diagnoses, particularly when using multimodal data sources such as behavioral data, facial images, and neuroimaging. Furthermore, AI systems can assist in the identification of co-occurring conditions, such as ADHD or anxiety, that are often seen in individuals with ASD. By integrating data from multiple sources, AI models can provide a holistic view of the individual's health, aiding in the diagnosis and management of comorbidities that may complicate the treatment process.

AI-based diagnostic methods hold significant promise for advancing the detection, diagnosis, and treatment of Autism Spectrum Disorder. Through early detection, personalized treatment plans, improved diagnostic accuracy, and the ability to track long-term development, AI has the potential to revolutionize ASD care. However, challenges related to data quality, interpretability, and ethical concerns must be addressed to ensure that these technologies are used responsibly and effectively. As AI continues to evolve, it is expected to play an increasingly critical role in enhancing the lives of individuals with ASD and their families.

1.2.7. Limitations and Challenges of AI-based Autism Spectrum Disorder Diagnostic Methods

This section highlights some of the limitations and challenges encountered when employing AI-based diagnostic methods for autism spectrum disorder (ASD). While AI models have shown promise in

advancing early detection and diagnosis of ASD, several factors hinder their widespread application and effectiveness. Key challenges include:

1.2.7.1. Noise

Noise is a significant challenge in autism data collection, particularly when dealing with autism screening datasets and facial image datasets. Noise refers to irrelevant or misleading information that may distort the analysis and conclusions drawn from the data. In autism research, this can occur when datasets contain extraneous features or mislabeled data that are unrelated to autism characteristics or diagnostic criteria. Furthermore, missing or incomplete data can add another layer of noise, particularly in datasets where participant information, such as demographic details or medical history, is insufficiently recorded. For example, incomplete screening data or facial images with unclear or ambiguous characteristics could lead to misclassification or incorrect analysis of autism traits. To mitigate such noise, preprocessing techniques such as feature selection, data cleaning, and normalization are essential. These methods help eliminate irrelevant data, standardize inputs, and focus on the most relevant features for autism diagnosis. Advanced data processing techniques, including dimensionality reduction, and feature selection, can also be applied to ensure the quality of the data before further analysis. Despite these efforts, continuous refinement and validation of the dataset are necessary to enhance its accuracy and reliability for autism research.

1.2.7.2. Demographic Bias in Autism Research

Demographic biases present significant challenges when working with autism screening datasets and facial image datasets. While such datasets provide valuable insights into autism diagnosis and intervention, critical demographic details such as age, gender, and ethnicity are often underrepresented or inconsistently recorded. This limitation can lead to skewed findings and hinder the generalizability of models developed using these datasets. For instance, autism screening datasets often lack representation from diverse ethnic backgrounds, focusing predominantly on specific demographic profiles. This bias can obscure the variability in autism presentation across different cultural contexts. Similarly, facial image datasets used for autism-related research may overrepresent certain ethnic groups while underrepresenting others, leading to algorithms that perform poorly on underrepresented populations. Another challenge arises from accessibility and participation in data collection. Individuals from marginalized or low-income communities, who may lack access to healthcare services, are often excluded from these datasets. This exclusion can result in models that fail to address the needs of these populations, who are often among the most vulnerable and underserved. Additionally, individuals with severe autism traits or disabilities may be underrepresented due to difficulties in participating in studies, further contributing to demographic bias. Addressing these biases requires intentional efforts to diversify datasets by including

individuals from varied age groups, ethnicities, and socioeconomic backgrounds. This approach will improve the robustness and applicability of autism diagnostic models, ensuring they serve a more inclusive population.

1.1.7.3. Overfitting and Generalization Challenges

AI models, particularly deep learning models, are prone to overfitting, especially when trained on small or non-representative datasets. Overfitting occurs when a model performs well on training data but fails to generalize to new, unseen data. In the context of autism diagnosis, overfitting can lead to models that inaccurately predict or fail to detect autism in diverse real-world settings. This is particularly problematic when datasets lack sufficient variation or are too homogeneous, as is often the case with small clinical samples or limited demographic representation. To address this issue, it is crucial to implement cross-validation techniques and use larger, more diverse datasets for training. Additionally, techniques such as regularization and data augmentation can help prevent overfitting by introducing variability into the model's training process. Ensuring that models can generalize across different population groups, including those with varying severity of ASD symptoms, is essential for the widespread applicability of AI-based diagnostic tools.

1.1.7.4. Interpretability and Trust in AI-based Diagnoses

One of the major challenges in deploying AI-based diagnostic systems in clinical settings is the lack of interpretability of certain machine learning models, particularly deep neural networks. These models are often considered "black boxes" because their decision-making processes are not easily understood by humans. In the case of ASD diagnosis, this lack of transparency can lead to hesitation or mistrust among clinicians, as well as families of individuals with autism, who rely on clear and understandable explanations of diagnosis and treatment recommendations. Improving the interpretability of AI models is crucial for fostering trust and ensuring that these tools are used effectively in clinical practice. Techniques such as explainable AI (XAI), which provide insights into how models arrive at their conclusions, are becoming increasingly important in making AI-based diagnostic tools more accessible and reliable. Ensuring that clinicians and patients understand the rationale behind AI-driven decisions will help improve the acceptance and adoption of these technologies.

1.1.7.5. Ethical and Privacy Concerns

AI-based autism diagnostic methods often rely on sensitive personal data, including medical histories, behavioral data, and facial images. The use of such data raises significant ethical and privacy concerns. Data security and patient confidentiality are paramount, especially when dealing with vulnerable populations such as children with ASD. Additionally, concerns about consent arise when using data from clinical settings, as individuals may not fully understand how their data will be used or may not have the option to opt out of participation in research studies. Ensuring that AI models comply with data protection regulations, such as GDPR or HIPAA, is crucial to safeguarding patient privacy. Furthermore, ethical considerations must be taken into account when designing data collection protocols and algorithms. Researchers must ensure that AI-based methods are developed and deployed in a way that respects the rights and autonomy of individuals, providing transparency about how data is collected, used, and stored.

1.1.7.6. Integration into Clinical Workflow

Even if AI-based autism diagnostic methods achieve high levels of accuracy and reliability, their integration into clinical practice remains a significant challenge. Healthcare professionals may be resistant to adopting AI tools due to concerns about the reliability of the system, the potential for AI to replace human expertise, or simply because of a lack of familiarity with AI technologies. Additionally, implementing AI tools requires appropriate infrastructure, training, and support to ensure that clinicians can effectively use these tools as part of their diagnostic process. To overcome these challenges, AI-based diagnostic methods must be designed with user-friendly interfaces and should complement, rather than replace, the expertise of clinicians. It is also essential to offer training and ongoing support for healthcare professionals to integrate these tools seamlessly into their workflows. This will ensure that AI-driven diagnoses are seen as valuable assets that enhance, rather than undermine, the clinical decision-making process.

By addressing these limitations and challenges, AI-based diagnostic methods for autism can be further refined and optimized, making them more effective, reliable, and accessible for both clinicians and individuals with autism.

1.3. Machine Learning and Deep Learning

Since their inception, humans have leveraged a variety of tools and technologies to perform tasks more effectively. The ingenuity of the human brain has driven the development of sophisticated systems that simplify everyday life by addressing diverse needs across domains such as computing, healthcare, and transportation. Among these groundbreaking innovations are machine learning (ML) and its advanced subset, deep learning (DL).

Machine learning refers to the automated discovery of meaningful patterns in data. These patterns can either enhance our understanding of existing phenomena—such as identifying risk factors for diseases—or predict future outcomes, such as forecasting the spread of infections. Over the last few decades, ML has become a cornerstone in extracting valuable insights from large datasets across various fields. For example, ML algorithms power search engines to deliver the most relevant results, enable anti-spam software to filter unwanted emails, and safeguard financial transactions by detecting fraudulent activities. In daily life, ML supports intelligent personal assistants like Siri and Alexa in interpreting voice commands, while digital cameras utilize it to recognize faces. ML algorithms are also applied in accident prevention systems in automobiles and contribute to scientific domains such as astronomy, medicine, and bioinformatics.

Deep learning, a specialized branch of machine learning, pushes these capabilities further by mimicking the structure and functioning of the human brain through artificial neural networks. Unlike traditional ML algorithms, which often rely on manually engineered features, DL algorithms automatically extract complex features and relationships from raw data. This ability has made DL indispensable for tasks requiring high levels of abstraction, such as image recognition, natural language processing (NLP), and speech synthesis. Applications of DL include autonomous vehicles that interpret their surroundings, medical imaging systems that detect diseases, and language models like ChatGPT that generate coherent text.

The demand for both ML and DL has grown exponentially with the increasing availability of large-scale datasets and computational resources. Advances in hardware, such as GPUs and TPUs, have enabled the processing of vast amounts of data, allowing ML and DL systems to achieve state-of-the-art performance in many domains. For instance, convolutional neural networks (CNNs) have revolutionized computer vision by excelling at tasks such as object detection and facial recognition, while recurrent neural networks (RNNs) and transformers have driven significant progress in sequential data analysis, including machine translation and time-series forecasting. In summary, ML and DL represent transformative approaches to processing and analyzing data. Their ability to derive actionable insights from complex datasets has made them invaluable in a wide range of applications, from everyday conveniences to life-saving technologies. Their continued evolution promises to further redefine how humans interact with and benefit from intelligent systems.

1.3.1. When Do We Need Machine Learning?

The necessity for machine learning (ML) and deep learning (DL) arises when conventional programming approaches fall short of addressing the complexity, scale, or adaptability required by a task. Unlike traditional systems that rely on explicitly defined rules, ML and DL algorithms leverage data-driven

approaches to learn, improve, and adapt. Their utility becomes apparent in scenarios involving highly complex problems, vast amounts of data, and dynamic environments requiring adaptability.

Human-Performed Tasks: Certain tasks that humans perform effortlessly—such as recognizing speech, understanding visual information, and driving—are incredibly challenging to codify into a traditional algorithm. These tasks require a nuanced understanding of patterns, context, and variability, which is often implicit and not easily articulated. For instance, while a human can intuitively distinguish a cat from a dog in an image, explicitly programming the visual rules for this distinction is an overwhelming challenge. ML and DL provide a robust solution by learning from large datasets of examples, enabling systems to achieve near-human accuracy in tasks like image recognition, speech processing, and autonomous navigation. Advanced DL architectures, such as convolutional neural networks (CNNs) and transformers, have proven especially effective in handling these challenges by automatically extracting intricate features from raw data.

Tasks Beyond Human Capabilities: In addition to mimicking human capabilities, ML and DL are indispensable for tasks involving data that exceed human cognitive capacity. Modern datasets in fields like e-commerce, healthcare, genomics, weather prediction, and astronomy are vast, multidimensional, and interlinked, making manual analysis infeasible. For example, in genomics, DL models identify gene interactions from enormous datasets to predict disease risks, while in astronomy, ML is used to classify galaxies and detect anomalies in terabytes of data. These applications uncover hidden patterns, relationships, and insights that would otherwise remain inaccessible, driving innovation across domains. By harnessing the computational power of ML and DL systems, researchers and industry professionals can solve problems on a scale previously unimaginable.

Adaptivity: One significant limitation of traditional programming approaches is their inflexibility. Once designed, a program's behavior is fixed unless it is explicitly reprogrammed to accommodate new scenarios. This rigidity poses a challenge for tasks that evolve over time or vary among users. ML and DL models, on the other hand, are inherently adaptive. They continuously learn from new data, refining their performance to address changing conditions.

Combining Complexity and Adaptivity: Many real-world problems demand solutions that address both the complexity of the task and the need for adaptability. Autonomous vehicles are a prime example, requiring DL models to perform real-time object detection, decision-making in complex and dynamic traffic environments, and adapting to different weather or road conditions. Similarly, in personalized medicine, ML and DL systems analyze genetic, clinical, and environmental data to recommend tailored treatments for individual patients, accounting for the dynamic nature of disease progression and patient response. Furthermore, DL models have extended the reach of ML by achieving breakthroughs in tasks that were once considered out of reach. For instance, generative adversarial networks (GANs) create

realistic images and videos, while transformers like BERT and GPT excel in language modeling and contextual understanding. These advancements have propelled applications like automated content creation, real-time translation, and advanced virtual assistants, showcasing the ability of DL to address both complex and adaptive tasks.

Handling Real-Time Decision-Making: In scenarios requiring rapid, real-time decision-making, ML and DL excel in processing and analyzing data streams almost instantaneously. For example, Autonomous Systems like Self-driving cars rely on DL algorithms to detect objects, predict pedestrian movements, and make split-second decisions to ensure passenger safety. Healthcare monitoring, where continuous patient monitoring using wearable devices generates real-time data that ML models analyze to detect early signs of health issues, such as arrhythmias or seizures, uses ML models to detect early signs of health issues, such as arrhythmias or seizures.

As the volume of digital data continues to grow exponentially, the demand for ML and DL is becoming increasingly urgent. Data from diverse sources—such as social media, IoT devices, medical records, and satellite imagery—contains invaluable information that traditional methods cannot process efficiently. ML and DL algorithms thrive in this environment, enabling organizations to unlock actionable insights, automate decision-making, and optimize processes. Moreover, advancements in computational hardware, including GPUs and TPUs, have significantly reduced the time and cost required for training complex models, making these technologies more accessible.

1.3.2. Types of Machine Learning

Machine Learning (ML) encompasses a variety of techniques designed to address diverse data-related challenges (Figure 1.2). Experts emphasize that there is no universal solution applicable to every problem. The choice of approach depends on several factors, including the nature of the problem, the number and type of variables involved, the selection of a suitable model, and other context-specific considerations.

1.3.2.1. Supervised Learning

Supervised learning is a machine learning approach in which models are trained on labeled data to predict or classify desired outcomes. The dataset is split into two subsets: a training set and a test set. Each entry in the training set includes both an input value and its corresponding target output. The goal of supervised learning is to predict or classify an output variable based on the input data. Algorithms in this approach analyze the training data to identify patterns, which are then applied to the test data to make predictions or classifications. By recognizing relationships between input variables and their corresponding outcomes, supervised learning aims to maximize prediction accuracy. A common example of supervised learning is regression, where a model is trained to predict a continuous outcome based on input data.

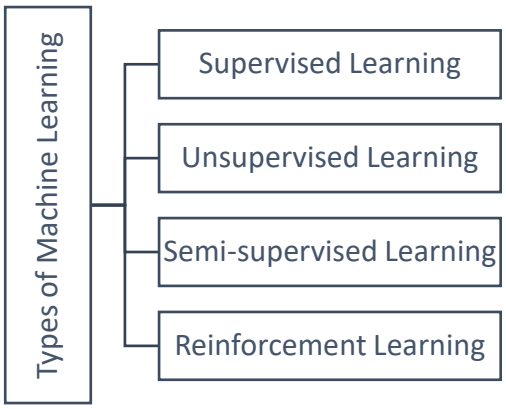


Figure 1.2 Types of Machine Learning Algorithms

1.3.2.2. Unsupervised Learning

Unsupervised learning is a machine learning technique where models analyze and uncover patterns from data that has not been labeled. Instead of being provided with explicit output labels, the algorithm relies on the inherent structure within the input data to identify relationships and trends. This approach is commonly used for tasks like feature reduction and clustering, where the goal is to reduce the complexity of data or group similar data points. In unsupervised learning, the model leverages the natural properties of the data to discover hidden structures without predefined outputs. Techniques like principal component analysis (PCA) are often employed to identify underlying covariance patterns within the data, enabling better understanding and representation of the information.

1.3.2.3. Semi-Supervised Learning

Semi-supervised learning is an approach that bridges supervised and unsupervised machine learning techniques [129]. This method involves a small amount of labeled data alongside a large volume of unlabeled data. Initially, an unsupervised learning technique is employed to identify patterns and structure in the data. These patterns are then used to assist in labeling the unlabeled data. Semi-supervised learning is particularly useful in situations where obtaining labeled data is costly or time-consuming, but unlabeled data are abundant. This hybrid approach is widely applied when accurate predictions are needed, but a large portion of the data lacks outcome labels. It strikes a balance between the accuracy of supervised learning and the scalability of unsupervised learning.

1.3.2.4. Reinforcement Learning

Reinforcement learning (RL) has its roots in early cybernetics and has evolved into a significant area of study in fields such as statistics, psychology, neuroscience, and computer science. Over the past decade, RL has gained substantial attention in the machine learning and artificial intelligence domains due to its practical applications. Unlike other machine learning approaches that rely on labeled data, reinforcement learning involves an agent interacting with a dynamic environment to achieve specific goals. The agent takes actions based on the current state of the environment and receives feedback in the form of rewards or penalties. Through this feedback loop, the agent learns to optimize its actions to maximize long-term rewards. This method is particularly effective in real-time applications, such as robotics, gaming, and autonomous systems, where the agent must learn continuously from its environment. In essence, reinforcement learning enables systems to learn by trial and error, improving their performance over time based on direct experience.

1.3.3. Machine Learning Applications in Healthcare

The primary objective of machine learning in computer science is to enhance the efficiency and reliability of machines. Machine learning is pervasive across various industries and plays a crucial role in numerous real-world applications. It is particularly vital in fields like healthcare, where it contributes to the protection and analysis of medical data. In the healthcare sector, machine learning acts as an extension of a doctor's expertise, functioning as a powerful tool to augment their capabilities. The purpose of machine learning is not to replace medical professionals, but rather to support them in delivering superior care and improving patient outcomes. Some of the applications of machine learning in the healthcare industry are:

1.3.3.1. Identifying and Diagnosing Diseases

Machine learning (ML) approaches play a pivotal role in enhancing the accuracy and speed of medical diagnoses, significantly improving clinical decision-making. By analyzing vast amounts of medical data, including electronic health records (EHRs), imaging data, and genetic information, ML algorithms can identify subtle patterns and anomalies indicative of various diseases. These systems are capable of detecting early signs of conditions such as cancer, cardiovascular diseases, and neurological disorders, providing valuable insights that assist clinicians in making more informed and timely diagnoses. Additionally, ML models can continuously learn and adapt, improving diagnostic accuracy over time as they are exposed to larger datasets, ultimately leading to better patient outcomes.

1.3.3.2. Drug Discovery and Manufacturing

Machine learning is playing a crucial role in the early stages of drug development, emerging as one of the most efficient applications in medicine. This includes advancements in research and development such as precision medicine and next-generation sequencing, which assist in identifying alternative treatment strategies for complex diseases. At present, ML techniques, particularly unsupervised learning, are employed to uncover patterns in data without the need for direct predictions. For instance, Microsoft's Project Hanover leverages ML-driven technologies for various purposes, including the development of AI-based tools for cancer treatment and the personalization of medication combinations for Acute Myeloid Leukemia.

1.3.3.3. Medical Imaging Diagnosis

Computer vision is a transformative technology enabled by both Machine Learning (ML) and Deep Learning (DL). ML techniques employed in computer-aided detection and diagnosis play a crucial role in assisting clinicians with the interpretation of medical imaging data, significantly reducing the time required for analysis. As machine learning becomes increasingly accessible and its analytical capabilities continue to improve, it is anticipated that a wider range of health imaging data will be integrated into AI-driven diagnostic systems, further enhancing the accuracy and efficiency of medical diagnoses.

1.3.3.4. Personalized Medicine

The remarkable performance of machine learning (ML) models in handling complex, large-scale data has led to significant advancements in personalized medicine over the past decade. By integrating individual health data with predictive analytics, personalized therapies are not only more effective but also open new avenues for research and enhanced disease evaluation. Traditionally, clinicians have been limited to selecting diagnoses from a restricted set of options or estimating a patient's risk based on clinical history and genetic information. However, ML is driving substantial progress in the field of medicine. A notable example is IBM Watson for Oncology, which leverages personalized medical data to assist in the development of tailored treatment options. Moreover, the proliferation of advanced gadgets and biosensors with enhanced health monitoring capabilities will further expand the availability of data, facilitating the growth and refinement of such medical systems.

1.3.3.5. Behavioral Modification

Behavioral modification plays a crucial role in preventive healthcare, and with the increasing integration of machine learning (ML) in the healthcare sector, numerous startups have emerged focused on cancer detection, treatment optimization, and various other health-related applications. In addressing mental health challenges, supervised machine learning methods prove to be highly effective. By leveraging deep

learning (DL techniques) such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Artificial Neural Networks (ANN), advanced models can be developed to predict an individual's mental health status based on their facial expressions, physical activities, and body movements.

1.3.3.6. Maintaining Health Records

Managing health records is often a time-intensive task. While technology has made data entry more efficient, many processes still require significant time to complete. The primary role of machine learning (ML) in healthcare is to optimize these operations, reducing effort, time, and costs. For instance, Ciox, a healthcare technology company, leverages ML to enhance the management and exchange of healthcare information. The aim is to improve accessibility to medical data, streamline operations, and boost the accuracy of health records. One of their innovations includes the development of smart charts, which use ML to extract and consolidate medical data from various health records, creating a unified digital profile for each patient.

1.3.3.7. Clinical Trial and Research

Clinical studies are often resource-intensive and can span several years to reach completion. Machine learning (ML) technology offers a valuable solution by predicting clinical trial outcomes, which can lead to reduced drug approval timelines, lower costs, and increased funding opportunities for the development of new treatments [133]. Additionally, ML has been applied to facilitate real-time monitoring of trial participants, optimize sample size determination, and leverage electronic health records to minimize data-related errors.

1.3.3.8. Better Radiotherapy

Radiology is one of the most prominent applications of machine learning (ML) in healthcare. Clinical image analysis involves numerous complex variables that can arise simultaneously, making it a challenging task to mathematically model certain diseases, such as tumors, cancerous lesions, and other abnormalities. However, ML-based approaches, which learn from diverse datasets, significantly enhance the identification and detection of these factors. By leveraging ML, the automation of cancer diagnosis and the identification of healthy physiological structures within organs can be greatly improved. Additionally, ML plays a crucial role in the precise selection of optimal radiation doses, leading to more accurate and efficient treatment planning.

1.4. Autism Spectrum Disorder and Machine/Deep Learning

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder marked by impairments in social communication, restricted interests, and repetitive behaviors. The complexity and heterogeneity of ASD

make it challenging to diagnose, with traditional methods relying heavily on subjective behavioral observations. Early and accurate diagnosis is crucial for effective intervention, as it significantly improves long-term outcomes. In recent years, machine learning (ML) has shown great promise in enhancing the diagnostic process by providing objective, data-driven insights and enabling the identification of subtle patterns that may be overlooked by clinicians. Machine learning offers significant advantages in dealing with the vast and diverse datasets typically associated with ASD research. These datasets include clinical assessments, genetic information, neuroimaging scans, and behavioral data. With the ability to process high-dimensional data efficiently, ML algorithms can reveal complex relationships within these datasets that are difficult for traditional statistical methods to uncover. In particular, ML has proven useful in handling the "curse of dimensionality" that often arises when working with large datasets, offering a more efficient approach to feature selection and data analysis. One of the most exciting applications of machine learning in ASD diagnosis is the analysis of neuroimaging data. Various ML techniques, particularly deep learning algorithms such as convolutional neural networks (CNNs), have been employed to identify structural and functional brain abnormalities that are often associated with ASD. These models can analyze brain scans, such as magnetic resonance imaging (MRI) or functional MRI (fMRI), to detect patterns indicative of ASD, providing an objective tool for clinicians. The use of machine learning in this context allows for the analysis of large, high-dimensional imaging data, revealing neurobiological markers that might not be evident through conventional methods. Another significant area where ML is contributing to ASD research is in the analysis of language and communication patterns. Individuals with ASD often exhibit atypical speech and language use, which can serve as early indicators of the disorder. Natural language processing (NLP) techniques, which focus on the computational analysis of human language, are increasingly being applied to study speech patterns, written texts, and caregiver reports. By analyzing linguistic features, such as sentence structure, word usage, and speech rhythm, ML models can help identify communication impairments that may not be immediately apparent in clinical settings. These tools offer the potential for early, non-invasive screening, even in pre-verbal children, which could lead to earlier diagnoses and interventions. Supervised machine learning algorithms, such as support vector machines (SVM), random forests, and neural networks, have been commonly used in ASD diagnosis, particularly for classification tasks. These models require labeled data, where each instance (e.g., an MRI scan or behavioral assessment) is associated with a known diagnosis. By training on these labeled datasets, the models can learn to differentiate between individuals with ASD and those with typical development. However, unsupervised learning methods are also being explored in ASD research. Techniques like clustering and dimensionality reduction allow for the discovery of hidden patterns and subtypes of ASD without the need for pre-labeled data. These methods are particularly useful for identifying unique phenotypic presentations of ASD, which can help tailor interventions to specific

subgroups of individuals. ML has also been used to predict the long-term developmental trajectories of individuals with ASD. By analyzing early behavioral and clinical data, machine learning models can predict future outcomes, such as language development, social skills, and adaptive functioning. This predictive capability is crucial for creating personalized treatment plans that can be adjusted as the child develops, ensuring that interventions are as effective as possible. Moreover, by identifying at-risk individuals early, these models enable clinicians to intervene before more severe challenges arise, potentially improving quality of life and developmental outcomes. Despite the significant promise of ML in ASD diagnosis and treatment, several challenges remain. One of the primary concerns is the need for high-quality, well-annotated datasets to train ML models effectively. In particular, neuroimaging data and behavioral assessments often suffer from small sample sizes, which can limit the generalizability of the models. Additionally, the "black-box" nature of many ML algorithms, especially deep learning models, raises concerns about interpretability. Clinicians need to understand how these models make decisions to trust and incorporate them into their practice. Efforts are underway to develop more explainable AI models that provide transparent, understandable reasoning behind their predictions, which will help address this issue. In conclusion, machine learning has the potential to revolutionize the diagnosis and treatment of ASD by offering objective, scalable, and personalized tools for clinicians. By integrating various data sources such as neuroimaging, genetic profiles, and behavioral data, ML models can uncover previously hidden patterns, improve early detection, and enable more effective interventions.

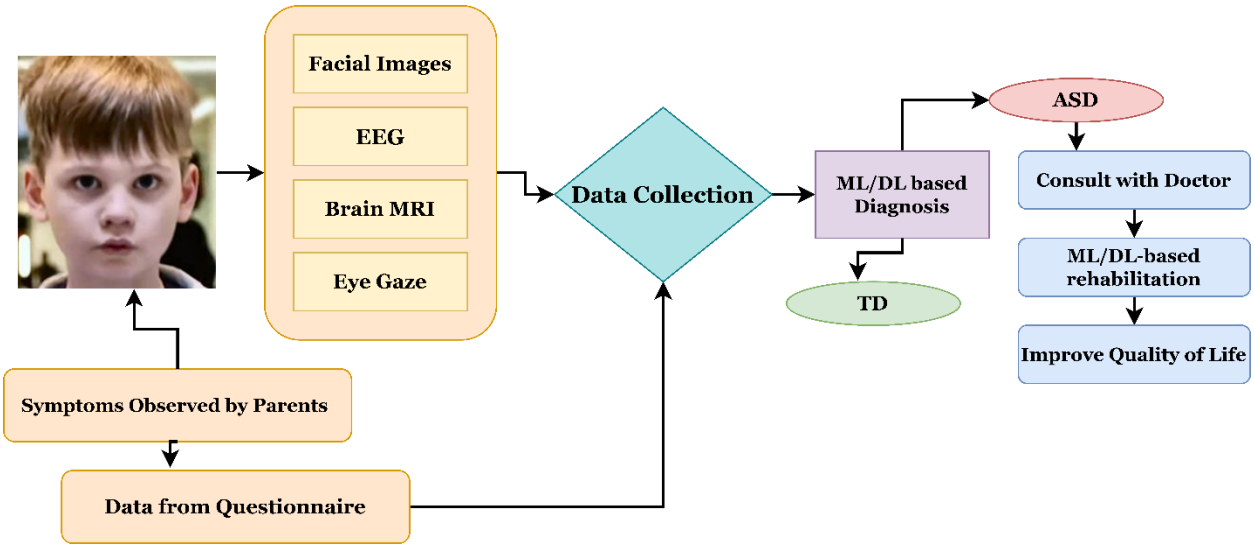


Figure 1.3: Working of a Generic Autism Spectrum Disorder Diagnosis Framework

This thesis aims to contribute to the growing body of work in this area by developing a machine learning-based framework that integrates multimodal data to enhance ASD diagnosis, while also ensuring that the resulting models are interpretable and clinically useful.

1.5. Motivation of Study

The early detection and diagnosis of Autism Spectrum Disorder (ASD) are critical in shaping the developmental trajectory and improving the quality of life for affected individuals. ASD is a neurodevelopmental condition that impacts social interaction, communication, and behavior, with a rising global prevalence. The ability to detect ASD at an early stage significantly enhances the effectiveness of intervention programs, which can help mitigate the challenges faced by individuals and their families. Given the complexity and variability in the presentation of ASD, early and accurate diagnosis remains a challenge within the medical field. For ASD detection, the use of Artificial Intelligence (AI) holds immense promise. Traditional diagnostic methods for ASD typically rely on behavioral assessments and expert evaluations, which can be subjective and resource-intensive. The emergence of AI offers an opportunity to provide objective, reliable, and scalable tools to complement or enhance these methods. Machine learning (ML) and deep learning (DL) techniques can analyze vast amounts of multimodal data from clinical assessments to brain imaging and behavioral data—identifying patterns that may not be immediately apparent to clinicians. Such AI-driven tools can facilitate earlier diagnosis and personalized treatment plans, ensuring better outcomes for individuals with ASD. The challenge lies in developing accurate, interpretable, and generalizable AI models capable of diagnosing ASD across diverse populations and age groups. The complexities associated with ASD, including varying symptom severity, co-occurring conditions, and individual differences, make it a difficult disorder to diagnose. Moreover, the current reliance on limited datasets and the lack of comprehensive, multimodal data further complicate the development of robust diagnostic models. Therefore, the motivation for this study is to address these challenges by leveraging AI techniques to analyze diverse datasets that include clinical data, neuroimaging, and behavioral data. By integrating these multimodal sources, this research aims to enhance the accuracy, interpretability, and generalizability of ASD detection frameworks. The increasing availability of large-scale, publicly accessible datasets, such as the ABIDE (Autism Brain Imaging Data Exchange) dataset, has further fueled the development of AI-driven diagnostic models. These datasets provide a rich source of information that can be used to train machine learning models capable of identifying subtle differences between individuals with ASD and those without, thus improving diagnostic accuracy. However, despite the growing body of research, the challenge remains to develop AI systems that not only excel in accuracy but also ensure that the results are clinically relevant and interpretable for healthcare professionals. As the prevalence of ASD continues to rise globally, there is an

urgent need to implement scalable, cost-effective diagnostic solutions. AI has the potential to revolutionize the way ASD is diagnosed and monitored by offering tools that can be deployed in a variety of settings, from hospitals to remote clinics. The goal of this study is to contribute to this transformation by developing a novel AI-based ASD detection framework that integrates multimodal data sources, addresses current diagnostic challenges, and improves accessibility to diagnostic resources for individuals worldwide. By exploring the integration of AI with clinical and behavioral data, this research aims to enhance the precision and timeliness of ASD diagnosis, ultimately leading to better outcomes for individuals affected by this disorder.

1.6. Research Objectives

The primary aim of this research is to try to overcome restrictions and build novel and innovative methods for inspecting technology-driven techniques. In order to accomplish this aim, the following Research Objectives (ROs) have been established:

For finding the solution to the above sub-queries, the following Research Objectives (ROs) are finalized:

Research Objective 1: To perform a systematic literature review on autism spectrum disorder.

Research Objective 2: To develop an intelligent autism spectrum disorder diagnosis model using deep learning with computational intelligence techniques.

Research Objective 3: To design a multi-modal autism spectrum disorder detection framework

Research Objective 4: To perform a comparative analysis of the proposed work with the existing techniques.

The detailed description of the identified research objectives is as follows:

Research Objective 1: This objective aims to build a comprehensive understanding of the current state of research on Autism Spectrum Disorder (ASD). A systematic literature review (SLR) involves analyzing and synthesizing existing studies on ASD, focusing on key areas such as diagnostic methods, challenges in early detection, the use of artificial intelligence (AI) in healthcare, and multi-modal approaches for ASD diagnosis. The review identifies gaps in existing methodologies and highlights the potential for applying deep learning (DL) and computational intelligence (CI) techniques. The findings serve as a foundation for proposing novel frameworks and models for improved ASD diagnosis.

Research Objective 2: This objective focuses on creating a robust and accurate diagnostic model that leverages advanced DL architectures integrated with CI techniques. The model aims to address challenges such as overfitting, generalization, interpretability, and computational efficiency. Techniques such as feature optimization (e.g., White Shark Optimization, Bat-PSO) and classifiers (e.g., Bi-LSTM, CNN) are employed to enhance model performance.

Research Objective 3: This objective involves developing a multi-modal framework that integrates diverse data modalities, such as clinical data, brain MRI scans, and meta-features, for a comprehensive ASD diagnosis. The framework combines vision transformers and LSTM-based architectures for effectively processing imaging and non-imaging data. A novel fusion mechanism, such as channel and spatial attention-based CBAC, ensures seamless integration of multi-modal inputs. This approach not only improves the accuracy of ASD detection but also addresses the complexities associated with analyzing heterogeneous datasets, offering a holistic and scalable solution for real-world applications.

Research Objective 4: The final objective ensures a thorough evaluation of the developed models and frameworks by comparing them with existing state-of-the-art methods. The comparative analysis includes benchmarking performance metrics, conducting ablation studies, and employing rigorous cross-validation techniques such as leave-one-dataset-out (LODO) and leave-one-site-out (LOSO). Additionally, the analysis highlights the advantages of the proposed methods in terms of accuracy, computational efficiency, and generalization capabilities. This comprehensive evaluation demonstrates the contribution of the research to advancing ASD diagnosis and underscores its potential for clinical implementation.
In this context, for fulfilling the requirement of ROs, Table 1.1 demonstrates the mapping among ROs, and publications.

Table 1.2 Aligning of Research Objectives, and Publications

Research Objectives	List of Publication
RO1. To perform a systematic literature review on autism spectrum disorder	✓ Machine Learning Techniques for Autism Spectrum Disorder: current trends and future directions **[Published]** ✓ Bio-Inspired Techniques in Autism Spectrum Disorder: Comprehensive Survey and Future Trajectories. **[Submitted]**
RO2. To develop an intelligent autism spectrum disorder diagnosis model using deep learning with computational intelligence techniques	✓ AFF-BPL: An adaptive feature fusion technique for the diagnosis of autism spectrum disorder using Bat-PSO-LSTM-based framework. **[Published]** ✓ WS-BiTM: Integrating White Shark Optimization with Bi-LSTM for Enhanced Autism Spectrum Disorder Diagnosis. **[Published]** ✓ S/SD-ASD: Self-Supervised and Self-Distillation Learning Approach for Autism Spectrum Disorder Classification Using Facial Images **[Under Review]** ✓ ASD-CEVT: Convolutional Enhanced Vision Transformer Architecture for the Diagnosis of Autism Spectrum Disorder **[Under Review]**

RO3. To design a multi-modal autism spectrum disorder detection framework	✓ MCBERT: A Multi-Modal Framework for the Diagnosis of Autism Spectrum Disorder [**Published]** ✓ LSTVision: A Multi-Modal Framework for the Diagnosis of Autism Spectrum Disorder utilizing LSTM and Vision Transformer [**Published]**
RO4. To perform a comparative analysis of the proposed work with the existing techniques	✓ S/SD-ASD: Self-Supervised and Self-Distillation Learning Approach for Autism Spectrum Disorder Classification Using Facial Images **[Under Review]** ✓ ASD-CEVT: Convolutional Enhanced Vision Transformer Architecture for the Diagnosis of Autism Spectrum Disorder **[Under Review]**

1.7. Outline of the Thesis

The thesis consists of six chapters describing the entire study in a very concise and precise way. Each chapter is summarised below:

Chapter 2: This chapter presents an extensive review of evolutionary and deep learning techniques applied to ASD diagnosis. It covers machine learning, multi-modality approaches, and self-supervised learning-based methods. The discussion identifies research gaps, highlights limitations in existing methodologies, and sets the foundation for the proposed frameworks. The chapter ends with a summary of the key findings from the literature.

Chapter 3: This chapter describes the methodologies used for ASD classification, including Particle Swarm Optimization, Bat Algorithm, and Adaptive Feature Fusion. It introduces the WS-BiTM framework, detailing its components such as White Shark Optimization, neural networks, and Bi-LSTM. The chapter discusses data preprocessing, feature selection, and performance evaluation. Experimental results of AFF-BPL and WS-BiTM are analyzed, showcasing their efficacy in ASD diagnosis.

Chapter 4: This chapter proposes a multi-modality framework combining CNNs for image analysis and MCBERT for meta-features. It explains components like Multi-Head CNN, CBAC, and BERT, along with preprocessing and classification modules. Experiments demonstrate performance through LOSO tests, comparison with existing works, and ablation studies. Computational complexity and detailed discussions provide insights into the proposed approach's effectiveness.

Chapter 5: This chapter explores various deep learning architectures such as VGG-16, AlexNet, ResNet, and Vision Transformer. It introduces the proposed ASD_CEVT framework, detailing its design and

application. Experiments and evaluations measure the framework's performance and compare it with baseline models. A discussion on the results highlights its practical implications in ASD diagnosis.

Chapter 6: This chapter focuses on self-supervised learning (SSL) and knowledge distillation (KD) techniques for ASD diagnosis. It introduces the use of facial image datasets and advanced architectures like transformers and masked autoencoders. The chapter discusses experimental setups, result visualization, and the implications of SSL and self-distillation in ASD research. A detailed analysis emphasizes the potential of these approaches.

Chapter 7: The final chapter summarizes the research findings and evaluates the limitations of the work. It provides a roadmap for future research directions, suggesting enhancements in AI-based ASD diagnostic methods. Potential advancements include integrating emerging technologies, improving computational efficiency, and extending applications to diverse datasets and populations.

1.8. Chapter Summary

This chapter provides an overview of autism diagnosis, emphasizing the role of artificial intelligence techniques in enhancing diagnostic accuracy. It discusses the integration of multi-modal data, such as clinical and imaging data, in improving the performance of ASD detection systems. The chapter outlines the content and description of each subsequent chapter, highlighting unique concepts and ideas that align with the title and objectives of the thesis. Additionally, it briefly explores autism diagnosis, machine/deep learning methodologies, and publicly available data for ASD detection. The objectives, scope, and motivation for this research are also presented in detail.

Chapter 2 METHODICAL LITERATURE REVIEW

This section explores the state-of-the-art advancements in autism spectrum disorder (ASD). After carefully reviewing the latest and most relevant research, we have identified a group of studies that share common motivations while offering distinct perspectives. This discussion seeks to highlight these works, providing a detailed overview before presenting our proposed methodology. In the last decade, the effectiveness of Deep Learning (DL) and Machine Learning (ML) approaches in diagnosing autism spectrum disorder has been clearly established. A summary of these findings is concisely presented in the Tables below, which offers a comprehensive overview of the existing literature that not only aligns with similar motivations but also varies in its applications.

2.1. Evolutionary and Deep Learning-based ASD Diagnosis works

Table 2.1 shows the literature on deep learning and evolutionary strategies for autism spectrum disorder (ASD) diagnosis demonstrates diverse approaches, focusing on different datasets and optimization techniques. Prasad et al. employed a hybrid sewing training optimization (HSTO) with ZFNet on the ABIDE dataset, achieving a high accuracy of 95.7%, though limited dataset diversity and small sample size hindered generalizability. Similarly, Thanarajan et al. utilized a chaotic butterfly optimizer with LSTM on eye-tracking data, obtaining remarkable accuracy (99.29%) but suggested further optimization and hybrid techniques for improvement. Loganathan et al. achieved comparable performance using Bi-GRU and chaotic optimization with EEG signals but faced challenges related to computational complexity and generalizability. For optimization-focused approaches, Vidyadhari et al. leveraged a fractional social driving training optimizer with a deep quantum neural network on ABIDE, achieving good sensitivity (0.96) but limited interpretability. Kumar and Jayaraj applied resilient fish swarm optimization with CNN and zealous particle swarm optimization with neural networks for ASD classification, noting computational complexities and suboptimal accuracy (77.11%-92.9%). Bhandage et al. used the Adam war strategy with a deep belief network on ABIDE datasets, achieving high specificity (0.935) but highlighted the need for improved accuracy. Some studies incorporated advanced techniques, such as Almars et al., who utilized transfer learning and the gorilla troops framework on autistic facial image datasets, reporting promising results with DenseNet169 but lacking clinical validation. Sree et al. combined jellyfish search and bacterial foraging optimization with gated recurrent units, outperforming existing algorithms but with room for architectural enhancement. Kadry et al. employed whale optimization for MRI slice classification, achieving 98.5% validation accuracy but emphasized the need for broader algorithmic inclusion. Finally, Anurekha and Geetha addressed gene selection for robust feature extraction using deep neural networks, achieving stability in feature selection but lacking

interpretability. Sriramakrishnan et al. implemented pelican and remora optimization algorithms with a deep CNN on ABIDE, attaining a high accuracy of 95.2% but recommending additional metrics for healthcare validation.

Overall, while significant progress has been made in leveraging optimization techniques and deep learning for ASD diagnosis, common challenges include limited dataset diversity, computational complexity, lack of interpretability, and insufficient clinical validation, highlighting areas for further research and development.

Table 2.1 Literature survey performed for the diagnosis of ASD using evolutionary algorithms in association with machine learning/deep learning

Author	**Aim**	**Model**	**Dataset**	**Evolutionary Technique**	**ML/DL**	**Results**	**Limitations**
Prasad et al. [8]	ASD detection using HSTO	HSTO_ZFNet	ABIDE	Hybrid sewing training optimization (HSTO)	Deep learning	Acc: 95.7%; TNR: 92.6%; TPR: 93.7%; FNR: 68.7%; FPR: 75.9%	Limited dataset diversity; small sample size; lack of generalizability and interpretability
Thanarajan et al. [9]	ASD diagnosis based on eye-tracking data	ETASD_CBODL	Eye tracking data	Chaotic butterfly optimizer	LSTM	Acc: 99.29%; Specificity: 99.29% Sensitivity: 99.29%; Precision: 98.78%;	Hybrid techniques can be employed; further optimization can be done
Loganathan et al. [10]	ASD classification and detection using chaotic	-	EEG signals	Chaotic optimization	Bi-GRU	Sensitivity: 98%; F1: 98%; Acc: 98%; MCC: 99%; Precision:	Limited generalizability; computational complexity

	optimization with Bi-GRU					99%	
Vidyadhari et al. [11]	ASD detection using optimization and deep learning	FSDTBO-DQNN	ABIDE	Fractional social driving training-based optimizer	Deep quantum neural network	Acc: 0.90; Specificity: 0.94; Sensitivity: 0.96	Lack of interpretability; accuracy needs to be improved
Kumar and Jayaraj [12]	ASD classification using resilient fish swarm and CNN	RFSO_ECNN	ABIDE	Resilient fish swarm optimization	CNN	Acc: 92.9; TN: 45.15; TP: 47.75%; FN: 3.14%; FP: 3.95%; F1: 93.08	Computational complexity; Interpretability; Generalization to other imaging modalities
Bhandage et al. [13]	ASD classification via optimization and deep belief network	AWSO_DBN	ABIDE I and ABIDE II	Adam war strategy	Deep belief network	Specificity: 0.935; Acc: 0.924; Sensitivity: 0.93	Limited dataset; accuracy needs to be improved
Anurekha and Geetha [14]	Gene selection model to identify robust and stable gene subset in ASD	IHEGS	Six gene expression data	Gene selection	Deep neural network	Provided stable results in terms of feature selection and accuracy	Lack of interpretability
Kumar and Jayaraj [15]	ASD classification using	ZPSO-RMLPNN	ABIDE	Zealous particle swarm	Neural networks	Acc: 77.11; TN: 37.52; TP: 39.58;	Accuracy needs to be improved

	threshold values to assess fMRI images			optimization		FN: 10.86; FP: 12.02; F1: 77.57; FMI: 77.57; MCC: 54.22	
Almars et al. [16]	ASD detection using transfer learning and gorilla troops framework	ASD^2– TL * GTO	Autistic facial image dataset, ASD screening	Gorilla troop optimization	Transfer learning	DenseNet169 outperforms the other employed techniques with a loss value of 0.512	Clinical aspects and interpretability of the model were not addressed
Sree et al. [17]	ASD classification using deep learning and optimization techniques	JSODL_ASDDC	ASD screening dataset	Jellyfish search optimization, Bacterial foraging optimization	Gated recurrent unit	The proposed technique outperforms the existing algorithms	The performance of JSODL_ASDDC architecture can be enhanced
Kadry et al. [18]	Diagnosing normal/ASD MRI slices with improved accuracy	-	MRI slices	Whale optimization	Deep learning	Validation result: 98.5	Results can be generalized by incorporating more algorithms
Sriramakrishnan et al. [19]	ASD detection using optimization techniques and deep learning	CPROA	ABIDE	Pelican and Remora optimization algorithm (POA and ROA)	Deep CNN	Acc: 0.952; Recall: 0.958; F1: 0.963	Clinical validation; lack of interpretability; other evaluation metrics should also be considered while

							working in the healthcare domain

Table 2.2 also shows the integration of evolutionary and deep/machine learning techniques in ASD diagnosis and related fields has significantly advanced the accuracy and efficiency of machine learning (ML) and deep learning (DL) models, yet challenges remain in terms of generalizability, interpretability, and computational cost. Arumugam and Saravanan employed a combination of ShuffleNet_v2, sparse autoencoders, WSO, and gated GRU for skin cancer classification, achieving improved outcomes, though statistical validations were absent. Singh et al. utilized WSO with DenseNet and U-Net for tuberculosis detection, obtaining 94.7% accuracy but identified the need for broader datasets. Hammouri et al. leveraged binary hybrid-sine-cosine WSO for feature selection across 23 medical datasets, achieving ~90% accuracy, indicating room for improvement. In intrusion detection, Alawad et al. combined WSO and K-means algorithms, noting limitations in navigating the search space. Focusing on ASD, Bhandage et al. used Adam War Strategy Optimization with Deep Belief Networks, achieving 92% accuracy but highlighting the potential for generalizing results across datasets. Several studies utilized ensemble and hybrid models. Kang et al. implemented PCA, autoencoders, and LSTM-Conv on ABIDE data for ASD detection, achieving 92.9% accuracy, with suggestions to explore additional feature selection techniques. Tang et al. introduced a two-stage adversarial deep learning model, achieving 80% accuracy, but emphasized the need for dataset generalization. Similarly, Loganathan et al. achieved superior accuracy (99.5%) using chaotic optimization and Bi-GRU on EEG signals. Various approaches were explored by Ahmed et al., who achieved 95% accuracy in detecting ASD through fine-tuning on facial images but called for nature-inspired feature selection. Han et al. designed a multimodal model based on EEG and eye-tracking data, achieving 95.56% accuracy despite high computational costs. Ali et al. and Pavithra and Jayanti employed bidirectional LSTM and PSO with IANFIS, respectively, achieving accuracies exceeding 97%, with suggestions for generalization and exploring bio-inspired methods. For graph-based techniques, Kwon et al. utilized a sparse hierarchical graph framework for brain connectivity in ABIDE, achieving a mean absolute error of 0.96 but lacking parameter normalization. Gaspar et al. achieved 98.8% accuracy on gaze-tracking images using KELM optimized by the Giza pyramid algorithm, but the dataset size was limited. Earlier works, such as Li et al., incorporated clonal selection and evolutionary algorithms for MRI segmentation, but high computation times and low accuracy (84%) were noted. Meanwhile, Sadeghian et al. employed genetic algorithms with KNN for fMRI-based ASD diagnosis, reporting low accuracy (62.59%), and suggested exploring DL classifiers. In summary, evolutionary

techniques have demonstrated potential in enhancing ASD diagnosis and related applications, yet the key challenges include improving accuracy, reducing computational complexity, addressing interpretability issues, and validating results across diverse datasets. Future work can focus on integrating advanced bio-inspired methods and leveraging multimodal approaches for robust outcomes.

Table 2.2 A literature survey performed for autism spectrum disorder incorporating various evolutionary techniques

Author, Year	***Aim***	***Dataset***	***Techniques***	***Results***	***Limitations/Future work***
Arumugam and Saravanan [20], 2024	Automated skin cancer classification (multi-class)	ISIC 2017; HAM1000	Shuffle_Net_v2; Sparse AE; WSO; Gated GRU	Improved detection outcomes on the employed datasets	Lack of interpretability; No statistical test performed
Singh et al. [21], 2024	Tuberculosis severity detection	Sputum images	Dense_net; WSO; MRO; U_net; Adaptive bilateral filter	Acc: 94.7%; TNR: 90.6%; PPV: 89.4%; TPR: 93.3%; NPV: 88%	Need to incorporate more datasets
Hammouri et al. [22], 2024	Feature selection via WSO	23 Medical datasets	Binary hybrid-sine-cosine WSO	Accuracy ~ 90%	Performance can be improved
Alawad et al. [23], 2023	Intrusion detection system via WSO	12 IoT and IDS datasets	WSO; K-means algorithms	The developed architecture performs fairly well on all metrics	Inability of the model to connect with the search space
Bhandage et al. [13], 2023	ASD classification via AWSO + DBN	ABIDE I, ABIDE II	DBN+ Adam war strategy algorithm was used	Acc: 92%, Specificity: 93.5%, Sensitivity: 93%	Accuracy can be improved; results can be generalized on more datasets
Talukdar et al.[24], 2023	Analysis of ASD via ML Techniques	ASD screening dataset	NB, LR, SVM, RF	The highest accuracy was achieved by RF (92.65%)	No feature selection was performed; accuracy can be improved; deep learning methods need to be

					explored
Kang et al.[25], 2023	Recognition of ASD via multi-view ensemble learning	ABIDE	PCA+ Auto encoder + LSTM-Conv	Acc: 92.9%	Accuracy can be improved; other feature selections can be explored
Tang et al.[26], 2023	ASD classification via adverisal deep learning (ADL)	ABIDE	Two-stage ADL model + sliding window concept	Acc: 80%, Specificity: 80%, Sensitivity: 81%	Accuracy can be improved; results can be generalized on more datasets
Loganathan et al. [27], 2023	ASD detection and classification	EEG signals	Hybrid model; chaotic optimization + Bi-GRU	Acc: 99.5%	-
Ahmed et al. [28], 2022	Developed a feature detection system to find children with ASD	Facial image dataset	Worked on InceptionV3, Xception, and Mobilenet models by performing fine-tuning on layers	Achieved maximum accuracy of 95%	Accuracy can be improved; feature selection via nature-inspired techniques can be explored
Kwon et al. [29], 2022	Developed a sparse hierarchical graph framework for brain connectivity	ABIDE	Adopted graph-deep learning model in order to predict ASD severity	MAE score: 0.96	Normalization of parameters was not done
Gaspar et al. [30], 2022	Classification of ASD using optimized KELM	Gaze tracking images	KELM + Giza pyramid construction algorithm	Acc: 98.8%	Less number of data sample
Ajmi N S et al. [1], 2022	Reviewed ML techniques for	-	Discussed various ML models + the	Advised to use graph neural	-

	ASD		general strategy adopted for detecting ASD via ML	network	
Han et al. [31], 2022	Identification of ASD in children using a multimodal approach	EEG and ET data	Designed two-step feature learning with a fusion model based on SDAE	Achieved accuracy of 95.56%	The computational cost was high, accuracy needs to be improved
Sadeghian et al. [32], 2021	ASD diagnosis using genetic algorithm	fMRI images	GA + KNN	Acc: 62.59%	Small data size; low accuracy; deep learning-based classifiers can be adopted
Ali et al. [33], 2021	Classification of ASD using LSTM	EEG dataset	Bidirectional LSTM	Acc: 97.3%	-
Pavithra and Jayanti [34], 2020	Detection of ASD via IANFIS	ISAA	PSO + IANFIS	Acc: 97% Sensitivity: 89%	Results should be generalized on various datasets; other bio-inspired techniques can be explored
Li et al. [35], 2019	Segmentation and function optimization of brain MRI	MRI images	Clonal section + differential evolution + estimation distribution algorithm	Acc: 84%	Low accuracy; high computation time

2.2. Deep/Machine Learning-based ASD Diagnosis works

Table 2.3 has the literature survey on studies employing deep learning (DL) and machine learning (ML) strategies for autism spectrum disorder (ASD) detection and analysis highlighting significant advancements and limitations in the field. Umrani and Harshvadhanan (2024) utilized EEG-based datasets (DEAP and SEED-1V) with a deep CNN guided by an intelligent search optimizer, achieving accuracies of 95.83% and 96.93%, respectively. However, their study was constrained by a smaller

dataset. Similarly, Sandeep and Kumar (2024) applied Mediapipe and ResNeXt to detect pain in autistic children from facial expressions, achieving 93.83% accuracy, but emphasized the need for advanced image processing models and larger datasets. Mouatasim and Ikermane (2023) implemented a control subgradient algorithm (CSA) with DenseNet-121 for ASD diagnosis from facial images, achieving 91% accuracy, which requires further improvement. Kwon et al. (2022) adopted a sparse hierarchical graph framework with deep learning on the ABIDE dataset to predict ASD severity, reporting a Mean Absolute Error (MAE) of 0.96 but lacking parameter normalization. Moridian et al. (2022) reviewed 233 research papers, highlighting the need for larger, multimodal datasets for future ASD research. Wan et al. (2022) developed the FECTS system for emotion identification in Chinese autistic children, achieving 70.22% accuracy. However, the study was limited by a small, low-quality dataset focused solely on Chinese children. Zhang et al. (2022) employed a variational autoencoder to extract functional connectivity features from the ABIDE dataset, achieving 73.2% accuracy, while Chen et al. (2022) proposed the NEGAT method with node-edge features and adversarial training, achieving 74.7% accuracy, both studies underscoring the need for multimodal datasets. Pang et al. (2022) improved classification accuracy by 4.47% using an optimized cascaded classifier for 50 rs-fMRI images but noted room for improvement in accuracy. Wang et al. (2022) proposed the MC-NFE technique for ASD detection, achieving an accuracy of 68.42% and suggesting the adoption of self-supervised techniques. Other notable works include Mason and Happe (2022), who used regression models for quality-of-life prediction, and Sharif and Khan (2021), who achieved 66% accuracy using the VGG16 model, highlighting the challenge of low accuracy in both studies. Earlier studies, such as those by Sherkatghanad et al. (2020), Chaitra et al. (2020), and Lu et al. (2020), explored CNN-based techniques, complex network frameworks, and genetic data combined with rs-fMRI, respectively. Despite achieving moderate accuracies, these studies emphasized the need for multimodal data integration and larger datasets. Wawer et al. (2020) achieved 95% accuracy using text data but faced limitations due to a small dataset. Finally, Pelleriti et al. (2020) called for more research on unsupervised ML with larger datasets, addressing the relatively limited focus on this area in existing studies. Overall, while the application of DL and ML has shown promise in ASD diagnosis and analysis, significant challenges remain, including small datasets, limited multimodal approaches, and the need for advanced feature extraction and optimization techniques.

Table 2.3 A literature review was performed on the studies adopting deep learning/machine learning strategies for autism spectrum disorder

Authors	**Year**	**Objective**	**Dataset**	**Technique**	**Results**	**Limitations**
Umrani and Harshvadhana	2024	Anxiety detection in	DEAP and SEED-1V	Intelligent search optimizer relied on	Accuracy on DEAP = 95.83%;	Smaller data sample size

n [36]		autistic individuals	(EEG-based datasets)	deep CNN	Accuracy on SEED-1V = 96.93%;	
Sandeep and Kumar [37]	2024	Pain detection in autistic children via facial expression	Custom emotion image dataset	Mediapipe and ResNeXt strategies	Accuracy: 93.83%	Data sample size is small, advance image processing models can be employed
Mouatasim and Ikermane [38]	2023	Control subgradient algorithm (CSA) for ASD diagnosis	Facial image dataset	Applied CSA DenseNet-121 CNN model	Accuracy: 91%	Accuracy can be improved
Kwon et al. [29]	2022	Developed a sparse hierarchical graph framework for brain connectivity	ABIDE dataset	Adopted graph-deep learning model in order to predict ASD severity	Achieved MAE score of 0.96	Normalization of parameters was not done
Moridian et al. [39]	2022	Reviewed automated ASD detection via AI	Reviewed 233 research papers	Discussed and compared various existing employed techniques in detail	Highlighted the need for larger datasets and to adopt multi-modal datasets for future research	-
Wan et al. [40]	2022	Developed a framework for emotion identification of Chinese children suffering from	Dataset of 10 Chinese autistic children aged 5-10 years	Built a FECTS system based on different features such as fear, happy which were stored in a cloud-based	Achieved an accuracy of 70.22%	The dataset was too small and of low quality, low recognition rate for some parameters, and

		ASD		evaluation system to analyze features for future		data was from only Chinese children
Zhang et al. [41]	2022	Built a novel feature selection strategy via variational autoencoder	ABIDE dataset	Extracted brain functional connectivity by utilizing MLP trained on variational autoencoder	Achieved accuracy of 73.2%	Accuracy needs to be improved; a multimodal dataset can be adopted
Chen et al. [42]	2022	ASD identification using graph neural network	ABIDE I	Proposed a novel NEGAT method utilizing node-edge features, adversarial training with multimodal MRI data	Gained accuracy of 74.7%	Accuracy needs to be improved; can integrate phenotypic information
Pang et al. [43]	2022	Developed computer-based diagnosis for brain disease	50 rs-fMRI images from ABIDE	Worked on optimized cascaded classifier through sample distribution and improved feature representation	Classification accuracy was enhanced by 4.47%	Accuracy needs to be improved
Wang et al. [44]	2022	Proposed MC-NFE technique for ASD detection using fMRI	ABIDE dataset	Initially classified ASD patients and healthy individuals then extracted features followed by linear SVM	Gained an accuracy of 68.42%	Self-supervised techniques can be adopted to decrease the demand of category labels
Mason and Happe [45]	2022	Predicting QoL through autistic traits	Data of 133 participants (42 autistic;	Estimated regression models and conducted an	Only some parameters adopted by them influence	Small sample size

			91 normal)	exploratory analysis	the QoL	
Sharif and Khan [46]	2021	Developed novel ML ASD detection framework	ABIDE I	Used VGG16 model with 16 layers, softmax function and Adam optimizer	Achieved accuracy of 66%	Low accuracy
Jee et al. [47]	2021	Detection of ASD via orthogonal decomposition & Pearson correlation	Q-Chat	Worked on improving training & time accuracy via dimensionality reduction and ML classifiers such as Linear Regress, KNN, NB, SVM, DT, RF and ANN	LR achieved accuracy of 100% and rest achieved accuracy of nearly 92%	DL techniques can also be explored
Sherkatghanad et al. [48]	2020	Development of automated ASD detection technique	fMRI data from ABIDE	Adopted CNN technique to classify autistic and typical control group and also tested the performance of their model using SVM, KNN, and RF classifiers	Achieved accuracy of 70.22%	Low test accuracy, higher time complexity
Chaitra et al. [49]	2020	ASD prediction via complex network ML framework	ABIDE	Investigated brain network features, RCE-SVM was adopted	Achieved highest accuracy of 70.1%	Low accuracy
Wawer et al. [50]	2020	Explored the limits of automatic means	Data of 74 individuals of ASD and	Worked on Bag-of-words, dictionary-based methods with	Achieved accuracy of 95%	Lower data sample

		of ASD detection from textual data	94 individuals of schizophrenia	machine learning. Used DL methods for inference and text representation		
Pelleriti et al. [51]	2020	Explored the applications of unsupervised ML for ASD	Reviewed 43 research studies	Discussed existing studies and techniques adopted for ASD using unsupervised ML	Highlighted the need of more research using unsupervised ML with a larger ASD dataset	Less research studies were explored
Lu et al. [52]	2020	Classification of ASD based on genetic data and rs-fMRI	Data of 71 individuals taken from NDAR	T-test used for feature reduction, SVM-RFE for optimized feature selection	Achieved accuracy of 83.6%	Accuracy needs to be improved; no integration of multimodal data; lower sample size

2.3. Multi-Modal-based ASD Diagnosis works

Table 2.4 represents the literature on multimodal architectures for the diagnosis of autism spectrum disorder (ASD), other medical conditions, and domains such as behavior analysis highlights a diverse range of methodologies and data modalities. In medical imaging (M1), structural and functional MRI (sMRI, fMRI) dominate, often combined with meta-features (M3) such as clinical data, to enhance the diagnostic accuracy for ASD and other disorders. For instance, Han et al. (2022) utilized EEG and eye-tracking data to design a stacked denoising encoder for ASD diagnosis, achieving an accuracy of 93.56%. Similarly, Du et al. (2022) focused on functional and structural connectivity measures to classify ASD and schizophrenia, achieving 83.08% accuracy. However, these studies are limited by the lack of multimodal feature fusion and reliance on small datasets. Several studies have employed videos (M4) to analyze behavioral traits, such as Song et al. (2023), who proposed a multimodal method to detect responses to a child's name using pose tracking and head pose estimation. Despite achieving ~93.3% accuracy, the approach is constrained by small sample sizes and sensitivity to reaction speed. In speech-related tasks, Passos et al. (2023) integrated graph neural networks with canonical correlation analysis to improve feature learning for energy-efficient speech enhancement, although the study lacked

quantification of energy savings. Multimodal techniques leveraging deep learning architectures, such as transformers and attention mechanisms, have gained prominence. For example, Le et al. (2023) developed a multimodal emotion recognition system combining CNN, ALBERT, and multi-head attention, achieving 85.9% accuracy. Similarly, Herath et al. (2024) proposed an ensemble classifier combining Inception V3, MobileNet, DenseNet, and ResNet50 for ASD diagnosis across ABIDE datasets, achieving 97.82% accuracy. However, these techniques often face challenges related to high computational costs, lack of generalization, and issues with interpretability. In Alzheimer's diagnosis, multimodal approaches have incorporated MRI, PET, and cerebrospinal fluid (CSF) data. Sheng et al. (2024) developed a hybrid framework combining Harris Hawks optimization with kernel extreme learning, achieving an accuracy of 99.2%. However, computational intensity and overfitting remain significant challenges. Similarly, Yu et al. (2024) employed a transformer-based framework with an AUC of 0.993 but reported limitations in validation and generalizability. Other innovative applications include behavior change prediction, drowsiness detection, and aggression detection in surveillance. For instance, Chan et al. (2023) achieved 98% accuracy in behavior change prediction using a combination of sampling techniques, SVM, and feature engineering, though the binary prediction target limits broader applicability. Meanwhile, Jaafar and Lachiri (2023) used 3D CNNs to detect aggression in surveillance videos, achieving a weighted average accuracy of 86.35%, but the results could not generalize across diverse datasets. Despite significant advancements, multimodal approaches face several limitations. Small and imbalanced datasets, lack of generalizability, and high computational costs are recurring issues. Future work should focus on integrating diverse modalities, exploring optimization techniques, and enhancing explainability to ensure the scalability and robustness of these models across broader applications.

Table 2.4 A literature survey performed on the multimodal architectures developed for the diagnosis of ASD, various diseases (Medical), and other domains using machine learning/deep learning, where, ***M₁*:** Image; ***M₂*:** Text; ***M₃*:** Meta-features/Sensor data; ***M₄*:** Videos; ***M₅*:** Audio; ***M₆*:** Signals, denotes different data modalities

Reference	Objective	Techniques	Included Modalities						Target Domain	Outcomes	Limitations/ Future work
			M₁	***M₂***	***M₃***	***M₄***	***M₅***	***M₆***			
[53]	Multimodal ASD diagnosis architecture	Weight learning network; Graph CNN; DeepGCN	✓		✓				Medical	Acc: 77.27%; Pre:77.7%; Recall: 80.96%	Small data size; lack of interpretability ; imbalanced gender ratio
			(ABIDE I)								
[54]	Building multimodal	-	✓		✓				Medical	Consists of 1315 videos	-

	dataset for autism analysis				for social and movement behavior analysis		
[55]	Multimodal technique based on response towards name behavior of children suffering from autism	Human pose tracking; Automatic name detection; Head pose estimation	✓ ✓ Dataset of 30 participants	Medical	Acc: ~93.3%	Small dataset; limited generalizability; dependency on body movements; sensitivity to reaction speed	
[56]	Multimodal drowsiness detection architecture through explainable machine learning	KNN; SVM; RF; SHAP; PDA	✓ (EEG, ECG, EOG)	Medical	Acc: 80.1%; Sen: 70.35; Spec: 82.2%	Small dataset; limited number of features; Deep learning techniques can be explored	
[57]	Multimodal classification technique to analyze the uniqueness of ASD and Schizophrenia	Functional and structural connectivity measures	✓ (fMRI, sMRI)	Medical	Acc: 83.08%	Lack of assessment of symptoms; limited neuro-imaging measures; model-level fusion only	
[58]	Behavior change prediction in students via	Feature engineering; sampling techniques;	✓ ✓	Behavior change	Acc: 98%; Precision: 97%	Prediction target is binary; costly setup	

	multimodal architecture	SVM; NB; DT; RF; KNN; MLP; XGboost				
[59]	Survival prediction via multimodal graph-based framework	Region-based via multimodal module; Embedding module; deep MM graph-based network	✓ ✓ (METABRIC; BASEL dataset)	Medical	Prediction performance: METABRIC = 0.7484; BASEL = 0.7479	Lack of interpretability; spatial simplification impact; less generalization to new data
[60]	Multimodal architecture for energy-efficient speech enhancement	Self-supervised framework integrating graph NN and canonical correlation analysis (CCA)	✓ ✓ (AV ChiMe3 dataset)	Speech enhancement	Proposed framework ensures improved feature learning	Did not quantify the amount of energy saving; biologically realistic neuronal architecture can be developed
[61]	Multilabel and multimodal emotion recognition	Feature extraction (CNN, ALBERT); multimodal fusion (transformers); emotion-level	✓ ✓ ✓ ✓ (IEMOCAP; CMU-MOSEI)	Emotion recognition	The developed framework outperforms existing methods with an accuracy of 85.9%	High computational cost; time-consuming; redundant frames in videos

		embedding (multi-head attention)											
[31]	Proposed multimodal architecture for diagnosing ASD in children	Stacked denoising encoder (SDAE)						✓	(EEG, ET data of 90 individuals)			Medical	Acc: ~93.56%; Sen: ~92.50%; Spec: ~98.0%	High computational cost; advanced NN algorithms can be explored
[62]	Detection of aggression in surveillance	Multiple deep neural networks; 3D-CNN		✓	✓	✓	✓		(Dataset of aggression in trains)			Medical	Unweighted average acc = 85.66%; Weighted average acc = 86.35%	Results cannot be generalized on a huge dataset; the model cannot specify all aggressive situations
[63]	Multimodal ASD identification	Multimodal + multisite ensemble classifier (Inception V3; MobileNet, DensetNet, ResNet50)	✓		✓				(ABIDE-I and II)			Medical	Best acc: 97.82% (improvement of ↑3.25%)	Other data modalities can be incorporated; more number of training images can be added
[64]	Decision support system for ADHD	Seed-based correlation; data augmentation ; CNN	✓		✓				(Eye movement data + fMRI)			Medical	Acc: 82%	Accuracy can be enhanced
[65]	Classification	EEG and	✓					✓				Medical	Best acc: 94%	Small dataset

	of ASD using different modalities	thermographic feature extraction; Naïve Bayes; neural net; logistic regression; random forest	(ADOS-2)			
[66]	Multimodal machine learning-based Alzheimer diagnosis framework	Extreme learning machine; entropy-based polynomial function; attention mechanism	✓ ✓ (ADNI)	Medical	Acc~ 98%	Lack of generalizability; does not address the issue of missing data
[67]	Multimodal transformer-based framework for Alzheimer	Transformers	✓ ✓ (ADNI)	Medical	AUC: 0.993	Lack of generalizability; lack of result validation
[68]	Multimodal hybrid framework for Alzheimer's diagnosis	Harris hawks optimization; kernel extreme learning	✓ (ADNI: MRI + CSF + PET)	Medical	Acc: 99.2%	Parameter sensitivity; computationally intensive; overfitting
[69]	Supervised and self-supervised learning on	Self-attention; latent feature extraction;	✓ ✓ (CINEPS and COEPS datasets)	Medical	The developed model performed significantly	Imbalance in the data ratio;

	multimodal data	cross-modality feature learning										well on various parameters	more data modalities can be considered
[70]	Multimodal Alzheimer's disease diagnosis using brain images	Pyramid attention strategy with GAN	✓						(ADNI dataset: MRI and PET images)		Medical	Acc: 89.9%	Small dataset; accuracy can be enhanced
[71]	Proposed MedVill for Multimodal representation learning	BERT; Multimodal attention strategy	✓	✓					(MIMIC-CXR; Open-I; VQA-RAD)		Medical	MedVill performed well against various considered techniques	Scope on accuracy improvement; Need to work on diverse multi-view studies
[72]	Multimodal tumor segmentation using a mathematical fuzzy framework	Nakagami imaging; Fuzzy fusion; Segmentation	✓						(Two types of images: Binary segmented and FLAIR images)		Medical	Average dice score: 92.78%	High number of training parameters
[73]	Multimodal architecture for diagnosing ASD	3D-ResNet; MLP	✓		✓						Medical	Acc: 74%; Recall: 95%; F1: 0.805	Limited amount of data; Low accuracy; overfitting issues; incorporate optimization techniques

												with large dataset
[42]	Autism identification via adversarial-graph learning networks	Adversarial learning; Garph networks	✓						Medical	Accuracy: 74.7%; Specificity: 77.4%	Accuracy needs to be enhanced; No incorporation of meta-attributes
			(sMRI and fMRI)								

The literature on single-modality architectures for autism spectrum disorder (ASD) diagnosis explores various deep learning and machine learning techniques across different data types such as fMRI, sMRI, and facial images in Table 2.5. In functional MRI (fMRI), Elakkiya and Dejey (2024) proposed two models, MinAutiNet and AutiNet, achieving maximum accuracies of 88.89% and 77.78%, respectively. However, the lack of automatic feature extraction techniques limits their performance. Similarly, N. Li et al. (2024) introduced a multi-level joint learning network leveraging graph networks, achieving 81.5% accuracy, though the model's complexity and many parameters present significant challenges. Tang et al. (2023) adopted an LSTM-based two-stage adversarial approach for multi-site ASD diagnosis, attaining an accuracy and specificity of 0.80 but highlighted the need for improved performance. Kang et al. (2022) developed a multi-view ensemble model incorporating LSTM-Conv architecture and PCA, but with a limited accuracy of 72%, it lacked integration with structural MRI (sMRI). In sMRI-based approaches, Nogay and Adeli (2024) utilized CNN with data augmentation and grid search optimization, achieving an accuracy of 85.42%. However, their model considered only two factors, age and gender, indicating limited scope. Similarly, Mishra and Pati (2023) employed deep CNN with data augmentation and optimization, obtaining a maximum accuracy of 81.35%, suggesting potential for enhancement with advanced techniques. For facial imaging, El Mouatasim and Ikermane (2023) applied a control sub-gradient approach with deep CNN and DenseNet models, achieving high precision, recall, and F1-scores of 98%, 97%, and 97%, respectively. Nevertheless, their single-modality architecture and limited hyperparameter exploration constrain its applicability. Dc et al. (2022) achieved a maximum accuracy of 91.2% using KNN, SVM, and other machine learning algorithms for ASD severity detection but emphasized the potential benefits of adopting deep learning strategies. Finally, in sensor-based methods, Parui et al. (2023) achieved 84.79% accuracy using brain connectivity analysis, though the study underscores the need to enhance accuracy further. Across all studies, limitations such as reliance on single

modalities, high model complexity, and a lack of integration with other data types suggest opportunities for future research to focus on multimodal architectures and optimization techniques to improve diagnostic accuracy and generalizability.

Table 2.5 Literature survey for autism spectrum disorder on single modality architectures

Reference	Objective	Techniques	Data Modality	Achievements	Limitations
[74]	Deep learning integrated activation function for the screening of autism	Developed two models, namely MinAutiNet and AutiNet for processing fMRI	fMRI	Max Accuracy: AutiNet = 77.78%; MinAutiNet = 88.89%	Need for automatic feature extraction techniques: accuracy needs to be improved
[75]	Deep learning-based ASD classification via age and gender factors	CNN, data augmentation, grid search optimization with multiple classifications	sMRI	Max Accuracy: 85.42%	Only two factors were taken into consideration
[76]	Multi-level joint learning network for the brain to diagnose ASD	Graph networks	fMRI	Accuracy: 81.5%	High model complexity; large number of parameters
[26]	Multi-site ASD diagnosis	LSTM; Two-stage adversarial approach	fMRI	Accuracy: 0.80; Specificity: 0.80; Sensitivity: 0.81	Accuracy needs to be improved; single-modality architecture
[38]	ASD diagnosis via facial imaging	Control sub-gradient approach with deep CNN;	Face images	The developed approach with DenseNet enhanced the overall results	Single modality architecture; limited exploration of

		DenseNet model		with Precision = 98%; Recall = 97%; F1-score = 97%	hyperparameters;
[77],	ASD diagnosis via sensor-based and AI approach	Brain connectivity analysis	fMRI	Accuracy: 84.79%	Need to improve accuracy
[78],	ASD classification framework	Deep CNN; data augmentation; optimization	sMRI	Max Accuracy: 81.35%	Accuracy can be enhanced; single modality model
[25]	ASD recognition via multi-view ensemble and multi-site fMRI	LSTM-Conv architecture; SDAE; PCA	fMRI	Accuracy: 72.0%	No incorporation of sMRI; lower accuracy
[79]	ASD severity detection via ML	KNN, DT, SVM, NB, RF; GLCM	Face images	Max Accuracy: 91.2%	Deep learning strategies can be adopted

2.4. Self-supervised learning strategies-based works

The literature in Table 2.6 on self-supervised learning (SSL) techniques reveals their growing impact on various domains, including medical image analysis and image-text multimodal tasks. Kumar and Misra (2024) employed an enhanced MNU2 model with the CAFFE framework to identify masked faces, gender, and age using a facial image dataset, achieving an accuracy of 96.54%. However, further efforts are needed to improve real-time prediction accuracy. Ozbay et al. (2024) utilized autoencoders and transformers to classify kidney tumors using the KAUH and CT dataset, achieving an impressive 99.82% accuracy, though future work should focus on distinguishing between malignant and benign tumors. Tan et al. (2024) combined masked autoencoders and Vision Transformers (ViT) to classify COVID-19-related medical images, achieving a top accuracy of 97.78%. The study highlights challenges with dataset quality and reconstruction. Similarly, Yang et al. (2024) explored self-supervised image quality assessment using masked image modeling and contrastive learning, outperforming AVA dataset benchmarks. The study emphasizes the need for graph-based networks and investigations into mask ratio variations. Bai et al. (2024) utilized SSL techniques with masked autoencoders and ViT for feature extraction in esophageal cancer detection, achieving an accuracy of 93.07% and an AUC of 95.31%.

However, high computational complexity remains a challenge. Ma et al. (2024) proposed joint distillation with disjoint masking for image modeling using an encoder-decoder architecture, resulting in a 3.4% accuracy improvement, though scalability and computational overhead persist as barriers. Chen et al. (2024) focused on multimodal image-text tasks with vision and language encoders paired with a fusion module, achieving effective performance but identifying the need for advanced fusion mechanisms. Qi et al. (2024) applied SwinUNeter with grid-based hierarchical masking for 3D medical image segmentation, outperforming existing SSL methods while noting limitations in handling low-contrast and highly complex images. Liu et al. (2023) enhanced 3D medical image reconstruction using random masking with a Transformer encoder, showing improvements in evaluation metrics but requiring better generalization. Finally, Qi et al. (2024) employed U-Net and masked autoencoders for tumor segmentation in BraTS-2020, achieving superior performance on key metrics but emphasizing the need for broader validation to ensure the findings' generalizability. Collectively, these studies highlight SSL's potential while addressing challenges such as computational demands, dataset quality, and generalization.

Table 2.6 Literature survey conducted on Self-supervised techniques

Reference	***Dataset Used***	***Objective***	***Methodology***	***Key Findings***	***Challenges and Future Directions***
Kumar and Misra, 2024 [80]	Facial images dataset (Sanjaya Subedi)	Identification of masked face, gender, and age	Enhanced MNU2 model with CAFFE framework	Achieved 96.54% accuracy, error rate of 3.46%	Enhance real-time gender and age prediction accuracy
Ozbay et al., 2024 [81]	KAUH and CT dataset	Classify kidney tumors using self-supervised learning (SSL)	Autoencoder and Transformer	Accuracy achieved: 99.82%	Focus on the differentiation between a malignant and benign tumor
Tan et al., 2024 [82]	SARS-COV-CT; COVID-CT	Classify COVID-19-related medical images	Masked autoencoder combined with Vision Transformer (ViT)	Best accuracy: 97.78%	Issues with dataset quality and reconstruction need resolution
Yang et al., 2024 [83]	ImageNet-1K; AVA dataset	Conduct self-supervised image quality assessment	Masked Image Modeling; ViT; Contrastive	Outperformed AVA dataset benchmarks	Explore graph-based networks; investigate the effects of varying

			Learning		mask ratios
Bai et al., 2024 [84]	Whole Slide Images (WSI) dataset of 552 cases	Extract features for esophageal cancer detection	SSL; Masked Autoencoder; ViT	Accuracy: 93.07%, AUC: 95.31%	Address high computational complexity
Ma et al., 2024 [85]	ImageNet-1K	Use joint distillation and disjoint masking for image modeling	Encoder-Decoder with joint distillation	Accuracy enhancement: 3.4%	Reduce scalability challenges and computational overhead
Chen et al., 2024 [86]	ROCO; MedICaT	Develop multimodal models for image-text tasks	Vision and language encoders with fusion module	Effective performance across tasks	Investigate advanced data fusion encoder mechanisms
Qi et al., 2024 [87]	Amos-2022; BraTS-2021	Segment 3D medical images with adaptive masking	SwinUNeter; Grid-based hierarchical masking	Outperformed existing SSL techniques	Address limitations on low-contrast and highly complex images
Liu et al., 2023 [88]	BTCV; LiTS-2017; BraTS-2020	Reconstruct boundaries in 3D medical imaging	Random masking with Transformer encoder	Improvements noted in multiple evaluation metrics	Work needed to enhance generalization
Qi et al., 2024 [89]	BraTS-2020	Perform tumor segmentation via SSL	U-Net; Masked Autoencoder	Achieved better performance on key metrics	Generalize findings for broader validation

2.5. Research Gaps and Limitations

Based on the insights from recent studies, several significant research gaps have been identified within the field of autism spectrum disorder using artificial intelligence:

- **Overfitting and Generalization Challenges:** Conventional ASD diagnostic models often face overfitting, limiting their generalizability across diverse populations and settings. This is particularly problematic in ASD research due to high symptom variability and data source diversity.

- **Interpretability of Complex Diagnostic Models:** Traditional machine learning models in ASD diagnosis often lack transparency, hindering clinical applicability. Interpretability is essential for clinical use, requiring clear insights into feature contributions.

- **Limitations of Single-Modality Approaches**: Most of the ASD diagnostic research has typically relied on single data modalities (e.g., MRI or behavioral data), limiting the potential for a holistic understanding of ASD.

- **Limited Identification of Key Autism-Causing Biomarkers**: Only a few studies focus on identifying significant ASD-causing features. These are vital for uncovering ASD's underlying mechanisms and developing more targeted diagnostic approaches.

- **High Computational Cost and Processing Time:** Computationally demanding models hinder ASD diagnostic models' applicability in clinical practice.

- **Inadequate Use of Deep Learning and Computational Intelligence:** ASD diagnostic research has been dominated by machine learning methods with limited attributes, while deep learning and computational intelligence approaches remain less explored.

- **Lack of Robust Cross-Validation and Statistical Validation:** Basic train-test splits are prevalent in ASD research. However, they lack rigorous validation like cross-dataset or LOSO cross-validation, which are critical for model reliability.

By addressing these gaps, future research can significantly enhance the diagnostic capability of AI systems for autism spectrum disorder.

2.6.Chapter Summary

This chapter presents systematic literature that profoundly discusses evolutionary algorithms, machine/deep learning, and multi-modality studies for autism spectrum disorder. Thus, this chapter presents relevant and up-to-date literature about autism spectrum disorder using AI strategies.

Chapter 3 ASD CLASSIFICATION FRAMEWORK USING DEEP LEARNING WITH COMPUTATIONAL INTELLIGENCE TECHNIQUES

The accurate and timely diagnosis of Autism Spectrum Disorder (ASD) is critical for enabling early intervention and improving developmental outcomes. This chapter introduces a novel classification framework developed using deep learning in conjunction with computational intelligence techniques, specifically tailored for autism screening datasets. The framework addresses key challenges in ASD classification, including the effective selection of features, mitigation of overfitting, and enhancement of model generalizability. By employing cutting-edge optimization algorithms and advanced neural architectures, the proposed methodology harnesses the potential of these datasets to identify meaningful patterns and correlations. The chapter provides an in-depth description of the preprocessing strategies, feature engineering approaches, and model design utilized to analyze autism screening data, offering a robust and scalable solution for accurate ASD classification.

3.1 Proposed Methodologies-: AFF-BPL

3.1.1. Particle Swarm Optimization

Nature-inspired algorithms draw inspiration from various biological systems including beehives, anthills, and swarms of animals like birds and fish [90] [91]. These algorithms explore the interactions among individuals in a population, their interplay, and their interactions with the environment. In the context of PSO, it is inspired by the behavior of a flock of birds in search of food location, making it advantageous for the rest to follow the nearest knowledgeable bird. In this context, each individual in the population represents a bird, having a fitness value within the search space [92][93]. Its objective is to converge towards an optimal solution. The potential solutions are called particles, forming the population. Each particle retains its best solution (P_{best}) evaluated using the fitness function, and the best value from the entire swarm is denoted as g_{best}. The standard PSO process comprises two key steps: changing velocity and updating positions. In the first step, particles adjust their velocity based on P_{best} and g_{best}. In the second step, particles update their position using the new velocity. These operations take place in a D-dimensional place, with each particle represented as $x_i = (x_i^1, x_i^2, x_i^3, \dots x_i^D)^T$. The velocity of i^{th} particle is denoted as $V_i = (V_i^1, V_i^2, V_i^3, \dots V_i^D)^T$, and its best previous position as $P_{best_i}^{t-1}$. The inertia weight (ω) balances the trade-off between exploration and exploitation. Equation 1 and 2 guide the velocity and position updates.

$$V_i^{t+1} = \omega V_i^t + C_1 r_1 (P_{best_i}^t - x_i^t) + C_2 r_2 (g_{best}^t - x_i^t) \quad (1)$$

$$x_i^{t+1} = x_i^t + V_i^{t+1} \quad (2)$$

The algorithm is parameterized by several variables, including D (the dimensionality of the problem), N (number of particles/swarm size), and T (maximum number of iterations).

Random values r_1 and r_2 are employed to avoid local optima, while C_1 and C_2 determine the particle's trajectory, representing self-confidence and swarm confidence factors respectively. Stopping criteria vary based on the specific problem, commonly involving a fixed number of function evaluations or the achievement of an error threshold. Notably, PSO employs P_{best} and g_{best} to update particle positions. The impact of these values, as constants was explored in previous studies. The g_{best} plays a critical role in determining the particle's trajectory and movement. Table 3.1 represents the values of the hyper-parameters used in the work. Figure 3.1 explains the overall flow chart of the proposed architecture.

Table 3.1 Description of hyper-parameters of PSO and BAT

Hyper-parameters	**Description**	**PSO**	**BAT**
Population_size	Swarm size in PSO; Bat population size	50	30
Max_itr	Number of iterations	200	200
C_1	Cognitive component weight	2	-
C_2	Social component weight	2	-
ω	Inertia	0.7	-
A_0	Loudness	-	0.25
r_j	Pulse rate	-	0.5
Velocity	Velocity update	As per equation (1)	As per equation (7)
Position	Position update	As per equation (2)	As per equation (8)

3.1.2. Bat Algorithm

The bat algorithm (BA), draws inspiration from the echolocation ability of microbats (producing loud sounds and capturing the resulting echoes as they rebound from the environment) [94][95]. This optimization technique mimics the bat's foraging behavior and its ability to navigate in low-light conditions [96][97][98]. The bat algorithm is based on the mentioned assumptions:

(1) All bats utilize an echolocation mechanism to discern both prey and obstacles through the received sound frequencies [99][100].

(2) Bats exhibit random flight patterns characterized by their velocity (V_1) at a given position (y_1). Various properties such as frequency (f_1), wavelength (λ), and loudness (A_0) play a role in their behavior.

(3) The loudness transitions from a high positive (A_0) to a low positive value (A_{min}).

Step 1: Initialization of parameters

To examine the optimality of a solution denoted as 'x' within the context of an objective function 'f(x)', for ASD problem; the formulation is:

$$\text{Min } \{f(x) \mid x \in X\},$$

Where f(x) is the objective function

$x = \{ x_i \mid i = 1,2,\ldots, d\}$ is a set of decision variables

$X = \{X_i \mid i = 1,2,\ldots, d\}$ is a possible range of values for every decision variable

d = Number of decision variables

Step 2: Initialization of bat population memory (BM)

BM contains an augmented matrix of size N X d having the set of location vectors of bats (as mentioned in equation 4). The location vectors are generated randomly as:

$$x_i^j = LB_i + (UB_i - LB_i) \text{ X } U(0,1) \quad (3)$$

$\forall_i = 1, 2, \ldots, d$ and

$\forall_j = 1, 2, \ldots, N$

$U(0,1)$ = Uniform random values between the range of 0 to 1

LB and UB = Lower bound and upper bounds

The produced solutions are saved in BM in ascending manner depending upon their f(x) value, where $f(x^1) \leq f(x^2) \leq \ldots\ldots\ldots \leq f(x^N)$

$$BM = \begin{bmatrix} x_1^1 & x_2^1 & \ldots & x_d^1 \\ x_1^2 & x_2^2 & \ldots & x_d^2 \\ \ldots & \ldots & \ldots & \\ .. & & & \\ .. & & & \\ .. & & & \\ x_1^N & x_2^N & \ldots & x_d^N \end{bmatrix} \quad (4)$$

x^{Gbest} , the global best location of bat is stored, where

$$x^{Gbest} = x^1 \tag{5}$$

Step 3: Motion of bats

Every bat x^j flies with velocity V^j influenced by the randomly produced frequency f_j

$$f_j = f_{min} + (f_{min} - f_{max}) \text{ X } U(0,1) \tag{6}$$

$$V_i'^j = V_i^j + (x_i^j - x_i^{Gbest}) \text{ X } f_j \tag{7}$$

$$x_i'^j = x_i^j + V_i'^j \tag{8}$$

Where, $\forall_i = 1, 2, \ldots, d$

$\forall_j = 1, 2, \ldots, N$

The offspring bat's position undergoes continuous updates by incorporating the positional adjustments of the parent bat, which are accompanied by relatively minor increments. This minor increment value occurs when the values at the best global bat location get closer to the parent bat, which in turn becomes near the offspring bat.

Step 4: Intensification of bat population

This specific step serves as the controlled stochastic element within the bat-inspired algorithm. Operating within a defined probability range of pulse rate denoted as "r_j", every subsequent bat location undergoes an update procedure involving a local search strategy that incorporates a random walk centered around the presently identified optimal solutions. The historical bat location represented as "x^{best}", is initially chosen from the pool of current best locations. Subsequently, the update of the new bat location, denoted as "x'^j" is carried out as:

$$x_i'^j = x_i^{best} + \in \hat{A}_j \tag{9}$$

Where $\hat{A}_j$ = mean loudness of all bats

Summarizing step (3) and (4), new bat location $X_i'^j$ can be evaluated as

$$x'^j \leftarrow \begin{cases} x^{best} + \in \hat{A}_j \\ x^j + V'^j \end{cases} \tag{10}$$

Step 5: Updation of bat population memory

For every bat in the bat memory, the new bat location supersedes the current bat location following the mentioned conditions

(a) The value of objective function f(x^{Gbest}) surpasses f(x^{j})

(b) U(0,1) $< A_j$

The pulse rate value r_j and loudness A_j will be updated as:

$$r_j = r_j^0 \left(1 - e^{(-\gamma \times itr)} \right) \quad (11)$$

$$A_j = \alpha A_j \quad (12)$$

Where *itr* represents the generation number in the current time step. As *itr* tends towards infinity, the mean loudness exhibited by the bat tends to converge to zero. Concurrently, the rate of pulse emission steadily approaches its initial emission rate.

$$A_j^{itr} \rightarrow 0, r_j^{itr} \rightarrow r_0^j, \infty \rightarrow 0$$

At last, bats are ranked, and their current best (x^{Gbest}) bat location is determined.

Step 6: Stopping criterion

In this phase, the bat algorithm iterates from step 3 to step 5 until the stipulated termination criterion is satisfied.

3.1.3. Adaptive Feature Fusion Module

The main novelty of AFF-BPL is its capability to choose and dynamically update feature set. This segment explains the concept of adaptive feature fusion module with the mathematical portion mentioned in the algorithm. The employed adaptive feature fusion (ADFF) framework has been devised to overcome the issues and disadvantages inherent in conventional feature selection techniques. Its design revolves around dynamically adapting the fusion strategy to accommodate the underlying characteristics of data and model requirements as mentioned in equation 16. At its core, adaptive feature fusion aims to synergize the strengths of data-driven as well as model-based fusion strategies, fostering more efficient and discriminative feature representations. This segment presents a comprehensive overview of the key components (mentioned in Figure 2) and mechanisms integral to the ADFF framework. Central to this framework is our adaptive fused layer, a specialistic layer seamlessly fused within the main architecture. Adaptive fusion layer gathers data (information/features) from numerous sources, such as different layers of neural networks, and connects them with a blend of model-based and data-driven mechanisms.

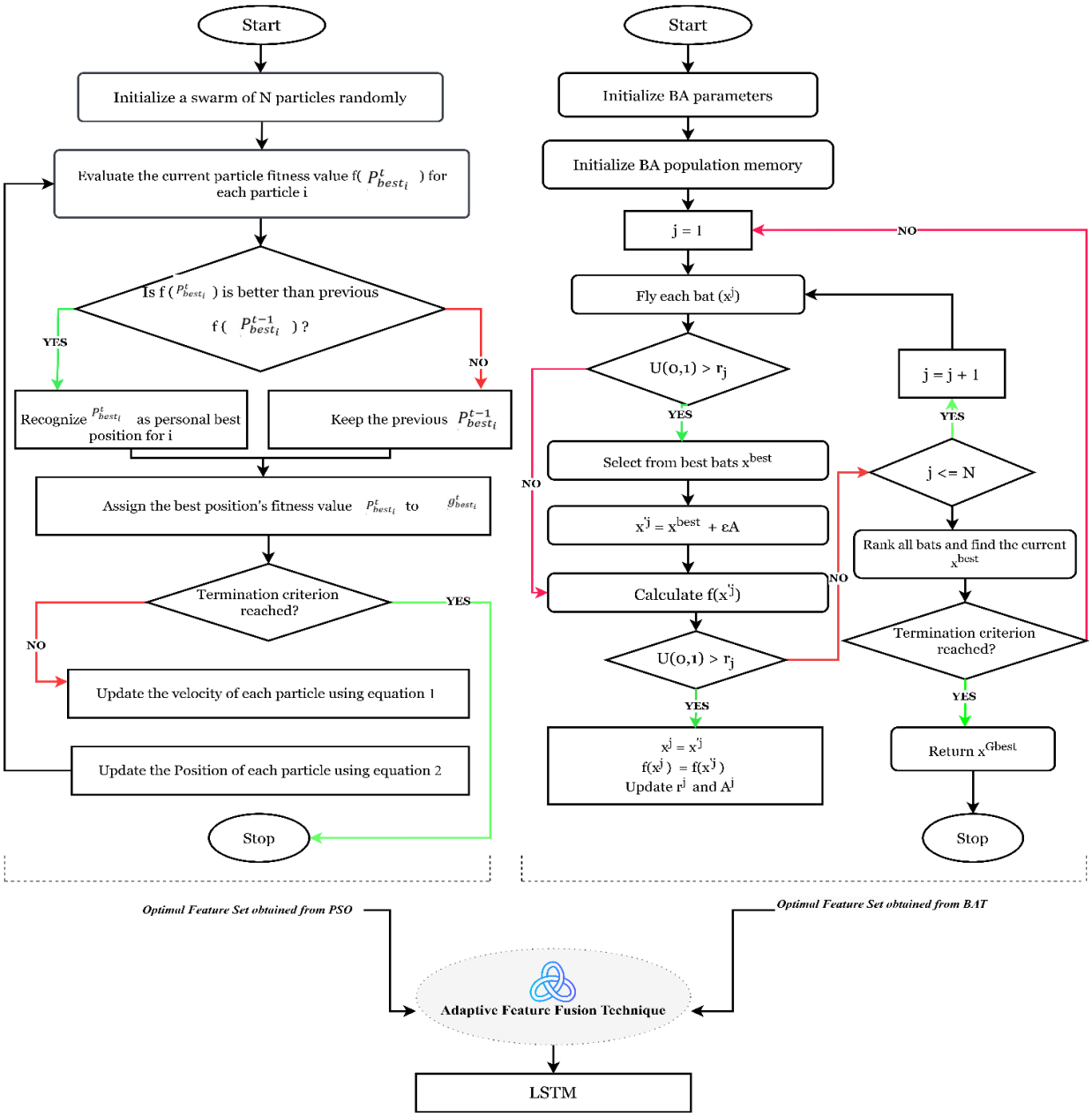


Figure 3.1 . *Overall Architecture of the Proposed Work. The baseline techniques were concurrently used and combined with adaptive feature fusion. Features from the adaptive feature layers is then fed to LSTM.*

In this work, the adaptive feature fusion layer receives features from the PSO and BAT algorithms as mentioned in Figure 3.1 and Figure 3.2. Data-driven strategies, including graph-based algorithms and attention-based strategies, are employed to retain optimal fused weights depending on the input/features and their relationship. Contrary to this, model-based strategies depend on the internal structure of the model in order to guide the fusion process. The fusion mechanism within the adaptive layer is controlled via set of rules (functions/ formulas). These rules are retained throughout the training, that allows the

model to adapt to the certain characteristics of the task and the data. This fusion block enhances adaptability and generalizability. To ensure versatility and scalability, the adaptive feature fusion is designed in accordant with various architectures.

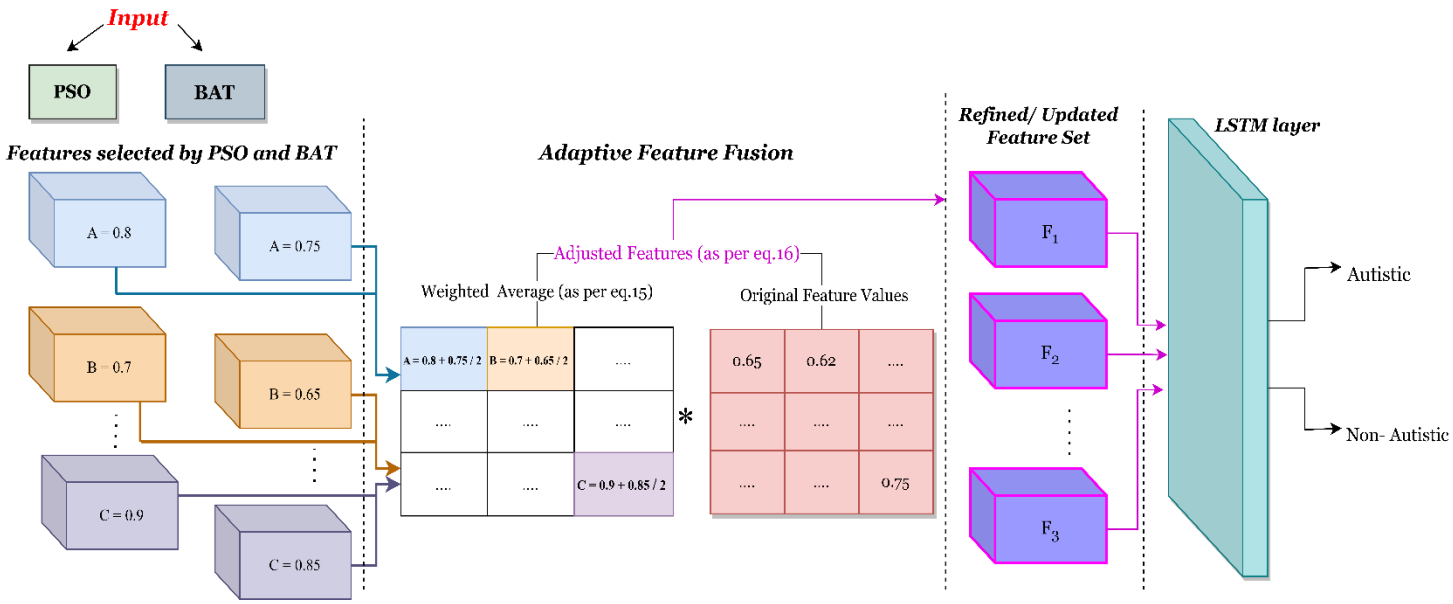


Figure 3.3. Architecture of Adaptive Feature Fusion Block

Normalize the scores obtained from PSO and BAT within the range of [0,1]

For each feature:

Normalize the scores using:

$$\text{Normalized\_Scores}_{\text{PSO}} = \frac{(Score_{PSO} - Min(Score_{PSO}))}{(Max(Score_{PSO}) - Min(Score_{PSO}))} \quad (13)$$

$$\text{Normalized\_Scores}_{\text{BAT}} = \frac{(Score_{BAT} - Min(Score_{BAT}))}{(Max(Score_{BAT}) - Min(Score_{BAT}))} \quad (14)$$

Calculate the weighted average using:

$$\text{Weighted\_avg} = \frac{(Normalized_{Score_{PSO}} + Normalized_{Score_{BAT}})}{2} \quad (15)$$

Merge features based on the calculative adaptive weights: Adjust the features using the calculated adaptive weights: The combined features will represent a weighted aggregation of the most influential features as determined by both PSO and BAT.

$$\text{Adjusted\_feature} = \text{Weighted\_avg} * \text{Original feature} \quad (16)$$

Furthermore, the adaptive feature fusion prevents overfitting and enhances generalization. These techniques can also include the use of dropout and the introduction of auxiliary tasks that encourage the model to learn more robust and discriminative feature representations.

Adaptive feature fusion framework for autism spectrum disorder portrays a novel approach for feature fusion to obtain enhanced performance and generalization capabilities. This framework allows the creation of more adaptable, effective, and robust model that can intelligently address the complex and diverse healthcare problems. Figure 2 shows the architecture of adaptive feature fusion block.

ALGORITHM 1: PSO and BAT algorithm running concurrently for Feature selection

Input: Autism screening dataset with features

Output: Features obtained from PSO and BAT algorithm

Start

Initialize the swarm randomly;

For i = 1 to N do

Initialize the particle's velocity and position using the uniform distribution

V_i^0 and x_i^0 ← random vectors within [LB, UB]^D ;

$P_{best_i}^0 \leftarrow x_i^0$ Initialize P_{best} to its initial position

end for

Initialize g_{best} to position with the minimum fitness value

t ← 1;

while t <= T do

for i = 1 to N do

r_1 , r_2 ← two independent vectors randomly generated from [0, 1]^D ;

Apply equation (1);

Apply equation (2);

if f(x_i^t) < f($P_{best_i}^{t-1}$) then

f($P_{best_i}^t$) ← f(x_i^t)

end if

end for

Update the swarm's overall best position to find g_{best}^t;

t ← t + 1;

end while

End

Start

Initialization of BA parameters

for j = 1 to N do

for i = 1 to d do

$x_i^j = LB_i + (UB_i - LB_i) \times U(1, d)$

end for

end for

find x^{Gbest},

while *itr* < Total iterations do

for j = 1 to N do

$f_j = f_{min} + (f_{min} - f_{max}) \times U(0, 1)$

for i = 1 to d do

$V'^j_i = V_i^j + (x_i^j - x_i^{Gbest}) \times f_j$

$x'^j_i = x_i^j + V'^j_i$

end for

if U(0,1) > r_j then

for i = 1 to d do

$x'^j_i = x_i^{best} + \epsilon \hat{A}_j$

end for

end if

if U(0, 1) < A_j and $f(x'^j) < f(x^{Gbest})$ then

$x^j = x'^j$

$f(x^j) = f(x'^j)$

$A_j = \propto A_j$

$r_j = r_j^0 (1 - e^{(-\gamma itr)})$

end if

end for

Update x^{Gbest}

end while

End

ALGORITHM 2: Proposed Adaptive Feature Fusion Algorithm

Input: Autism screening dataset with features

Feature importance scores from PSO and BAT

Output: Modified feature set with adaptive feature fusion

Classification of autistic vs non-autistic individuals

1. Calculate feature importance scores using PSO and BAT for each feature:
   Implement PSO and BAT: Utilize PSO and BAT algorithms to perform feature selection concurrently on the dataset.
   Obtain feature importance scores: Store the feature importance scores from PSO and BAT for every feature in the dataset.

2. Normalize the scores obtained from PSO and BAT within the range of [0,1]
   For each feature:
   Normalize the scores using:

$$\text{Normalized\_Scores}_{\text{PSO}} = \frac{(Score_{PSO} - Min(Score_{PSO}))}{(Max(Score_{PSO}) - Min(Score_{PSO}))}$$

$$\text{Normalized\_Scores}_{\text{BAT}} = \frac{(Score_{BAT} - Min(Score_{BAT}))}{(Max(Score_{BAT}) - Min(Score_{BAT}))}$$

3. Calculate the weighted average using:

$$\text{Weighted\_avg} = \frac{(Normalized_{Score_{PSO}} + Normalized_{Score_{BAT}})}{2}$$

4. Merge features based on the calculative adaptive weights
   Adjust the features using the calculated adaptive weights: The combined features will represent a weighted aggregation of the most influential features as determined by both PSO and BAT.

   Adjusted_feature = Weighted_avg * Original feature

5. Prepare the modified features with adaptive feature fusion as the input for LSTM.

6. Train and evaluate the LSTM model using the updated features.

3.2. Introduction to WS-BiTM-Related Methodologies

3.2.1. White Shark Optimization

WSO lies in the category of bio-inspired techniques (metaheuristic optimization) influenced by the hunting strategy of white sharks in order to provide solutions to complex optimization problems [101]. WSO can be employed for feature selection in various data mining and machine learning tasks [102][103][104]. Its objective is to choose a feature subset to enhance the performance of a classification task. The general strategy of the white shark optimizer is expressed below:

(a) Initialization: initialize the population of white sharks (1)

Every shark shows a potential solution.

Every shark is given a position in search space, which represents a feature subset.

(b) Objective function: describe an objective function that evaluates fitness/quality based on the selected feature set.

Set max_no_of_iterations and termination criterion (2)

(c) Hunting strategy: hunting behavior involves

Exploration – explores new sections of search space in order to find a potential solution

Exploitation – shark focuses on the region with high quality and optimal solutions, directed by the fitness value

(d) Update: after every iteration, update the shark's position. To do so, various mathematical operators are applied depending on the problem statement. This helps the population to evolve and converges it towards optimal solutions.

(e) Evaluation: compute the fitness of every shark in the population

(f) Termination: WSO's termination criterion determines when to stop the overall optimization process.

The termination criterion can either be a satisfactory solution or the maximum number of iterations.

(g) Output: select the best feature subset from the final population

The detailed working strategy adopted for feature selection via WSO is mentioned in section 4.2.

3.2.2.Neural Networks

Neural networks (NN) learn from the existing data and handle complex problems efficiently. NN is a set of numerous artificial neurons that function as simple processing components [105]. The framework of NN is based on a weighted graph between the neuron input and the output, with artificial neurons acting as nodes and directed edges. NN has three layers, node layer (containing input), hidden layer (one or more), and output layer [106]. The total number of neurons is equal to the total number of features in the data [107][57]. The output obtained from the input layer is provided to the next layer i.e., the hidden layer. The number of hidden layers relies on the model size and the size of the dataset. Hidden layers may have different sum of neurons, that are generally more significant than the number of features. The output obtained from each layer is computed via matrix multiplication of the preceding layer's output with learnable weights, followed by the addition of learnable biases and activation functions. These are important for a system to be non-linear [108]. The output of the hidden layer is passed through operations such as ‘softmax' or 'sigmoid', which is responsible for converting the output of each class into the likelihood score. The obtained information is fed to the architecture and the associated output of each layer is gathered. This level is known as “feed-forward” [109][110].

Feed-forward: compute the error using the error function (cross-entropy, square loss); from this point, backpropagate to the model by determining the derivatives.

3.2.3.CNN

Convnets are neural networks with shared parameters. A convnet is a multi-layered feed-forward NN made up of a sequence of layers, each of which is capable of changing from one volume to another via a differentiable function. They may learn hierarchy because of its sequential design. Convnets employ a number of different layers with varying functionality (a convolutional layer, a pooling layer, and a fully connected layer) [106][111]. The convolutional layer calculates the dot product of two matrices, one of which is a ‘set of learnable parameters’ known as ‘kernel’ and the other is the restricted section of the receptive field. Convolution makes use of three significant concepts i.e., sparse interaction, parameter sharing, and equivariant [112].

Sparse interaction specifies that we have to choose key parameters, which not only minimizes the model’s memory demand but also improves its statistical efficiency.

Parameter sharing: weights employed to one input are similar to weights used elsewhere when computing output.

Equivariant: it states that if we modify the input somehow, the output will transform in the same manner.

The activation layer introduces the concept of nonlinearity to the network by adding an 'activation function to the previous layer's output. It applies an element-wise activation function to the convolution layer's output [100]. ReLu, tanh, and LeakyRelu are some of the types of activation functions [107][113]. The pool_layer changes network outcome at particular points by computing a summary statistic of surrounding outputs. The strategy helps to decrease the spatial size of the representation, which in turn reduces the desired number of weights and computations [114]. The pooling function is applied separately on each slice of the representation. To map the representation between the input and output, a fully connected (FC) layer is employed [115]. The output obtained from the FC layer is given to a logistic function for the classification task.

3.2.4.LSTM

LSTM networks are advanced recurrent neural nets (RNN), which were introduced to overcome the issue of vanilla RNN i.e., the long-term dependency problem [116][117]. LSTM works via a chain-like structure having four neural nets with discrete memory blocks known as cells. LSTM has feedback connections that allow the analysis of the complete data sequences without handling each point in the sequence separately (as mentioned in Fig. 3.3), but rather by preserving important knowledge about earlier data points in the sequence to aid in the processing of new incoming data points [118].

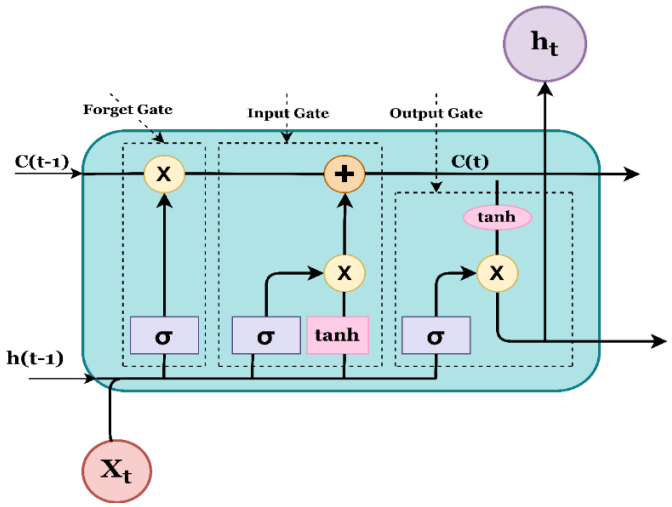


Fig 3.3 The architecture of the LSTM unit

LSTM employs various 'gates' that are responsible for handling how information in a data sequence enters, is stored in, and exits the network. An LSTM incorporates three basic gates: a forget gate (determine which bits of the network's cell state are significant, given both the previous hidden state and the current input data), an input gate (identifies the new information to be added to the cell state), and an output gate (extract significant data). The hidden state is termed as ‘short-term memory’, whereas the cell state is stated as ‘long-term memory’ [119].

Our proposed architecture focuses on the diagnosis of ASD using a novel WSO- BiLSTM-based network. The work begins with the data pre-processing, followed by feature selection via WSO (incorporates the mathematical explanation behind the working of WSO), classification through the Bi-LSTM network, and comparison of proposed work with baseline techniques i.e., NN, LSTM, CNN. A detailed explanation of the work is mentioned below:

3.2.5. Data Pre-processing

The data pre-processing step is a crucial task for ASD diagnosis. The dataset employed incorporates categorical, binary, and continuous attributes. Since the screening dataset includes a few categorical and non-contributing attributes, we need to pre-process the dataset. Pre-processing defines the transformation performed on the datasets before giving them to the model. Missing values were handled in this step to make the dataset suitable for analysis. To handle categorical attributes, label encoding was performed, which transforms the categorical values/labels into numeric format thus making it machine readable. For example, the 'sex' column, 'ASD traits', etc. were chosen for binary label encoding as they have only two classes. The 'Ethnicity' column incorporates eleven different classes and on them, one hot encoding was performed.

3.2.6. WSO: Feature Selection

The presence of irrelevant or less significant features in the dataset reduces the overall accuracy of the developed architecture, making the model learn those unnecessary features. This problem is stated as the optimization problem. In order to deal with this problem, the optimal solution or the optimal feature set from the ASD screening data needs to be extracted. To do so, we employed the white shark optimization technique.

WSO algorithm mimics the foraging behavior of white sharks. WSO incorporates a population of white sharks, in which every shark represents a candidate solution. WSO aims to solve optimization problems like feature selection. For this objective, the technique moves with a series of iterations mimicking the hunting behavior of white sharks until a potential and satisfactory solution is found. The hunting behavior of white sharks relies on three strategies,

(1) Velocity of shark for catching prey (target)
(2) Searching towards the best optimal food (optimal solution)
(3) Movement of the rest of the sharks who are close to the food source

The population of white sharks is given as:

$$W_z^y = lb_z + n(up_z - lb_z) \quad (3)$$

Where W_z^y denotes the initial parameter of y_{th} white shark in the z_{th} dimension.
up_z and lb_z denotes the upper bound and lower bound in the z_{th} dimension. n represents a random number between [0,1].
The velocity of white shark to find the prey (target) rely on the motion of sea waves and is represented by

$$Vl_{k+1}^y = \mu\left[Vl_k^y + R_1\left(W_{gbestk} - W_k^y\right) \times U_1 + R_2\left(W_{best}^{Vl_k^y} - W_k^y\right) \times U_2\right] \quad (4)$$

Where k = 1,2,3,…, P is index of white shark having population size P. The updated velocity of the y_{th} shark is shown as Vl_{k+1}^y in the $(k+1)_{th}$ step. The initial speed of y_{th} shark in k_{th} step is shown as Vl_k^y
W_{gbestk} represents the global best position obtained by any y_{th} shark in k_{th} step. W_k^y denotes the initial position of the y_{th} shark in k_{th} step.

The best position of y_{th} shark and index vector for obtaining best position are shown by $W_{best}^{Vl_k^y}$ and vc^i.
U_1 and U_2 represents the generation of a uniform random number in the interval [1,0].

R_1 and R_2 in the equation show the force of shark in order to handle the effect of W_{gbestk} and $W_{best}^{Vl_k^y}$ on W_k^y.
μ denotes the shark's convergence factor. The white shark index vector is shown by

$$vc = [a \times rand(1,a)] + 1 \quad (5)$$

$rand(1,a)$ denotes a random number vector achieved through uniform distribution between the interval [0,1].

R_1 and R_2 can be further expanded as:

$$R_1 = R_{max} + (R_{max} - R_{min}) \times e^{-(4m/M)\text{^}2} \quad (6)$$

$$R_2 = R_{min} + (R_{max} - R_{min}) \times e^{-(4m/M)\text{^}2} \quad (7)$$

m and M denote the initial as well as the maximum sum of iterations. R_{min} denotes current velocity and R_{max} denotes the sub-ordinate velocity of white sharks.
The convergence factor μ is shown as

$\mu = 2 / | 2- \tau - \sqrt{\tau^2 - 4\tau} |$ (8)

In which τ states the acceleration coefficient.

The process of updating the white shark's position is

$$W_{k+1}^{y} = W_k^{y} . \neg W_0 + up.h + l_0.d ; \quad rand < QS \tag{9}$$

$$W_k^{y} + Vl_k^{y} / fq ; \quad rand \geq QS \tag{10}$$

The $\neg$ denotes the negation operator, h and d specifies binary vectors. The lower search space is shown by l_0.

$$h = \text{sgn}(W_k^{y} - up) > 0 \tag{11}$$

$$d = \text{sgn}(W_k^{y} - 1) > 0 \tag{12}$$

$$W_0 = \oplus (h, d) \tag{13}$$

$$fq = fq_{min} + \frac{fq_{max} - fq_{min}}{fq_{max} - fq_{min}} \tag{14}$$

W_0 represents the logical vector and fq denotes the frequency by which the sharks move. Whereas fq_{max} and fq_{min} define maximum and minimum frequency respectively.

The increment in force at every iteration is

$$QS = \frac{1}{h_0 + e^{(\frac{k}{2} - k)/ h_1}} \text{ , QS shows the weights of the features} \tag{15}$$

The equation of the best optimal solution is given a

$$W_{k+1}^{\prime y} = W_{gbestk} + n_1 \overrightarrow{Dis_w} \, sgn(n_2 - 0.5) n_3 \; < Str \tag{16}$$

Where updating the position following the food source of y_{th} shark is expressed as $W_{k+1}^{\prime y}$.

In order to update the search direction, $sgn(n_2 - 0.5)$ produces -1 or 1. The optimal food source, shark distance $\overrightarrow{Dis_w}$ with the strength of white shark to follow other sharks near to food source Str is represented as

$$\overrightarrow{Dis_w} = |rand \times (W_{gbestk} - W_k^{y})| \tag{17}$$

$$Str = | 1 - e^{(h_2 \times k/K)} | \tag{18}$$

The best initial optimal results are kept constant, whereas the position of the rest of the sharks is modified as per the two constant optimal results.

$$We^{y} = \frac{1}{p-1} * \left(\frac{\sum_{Y=1, Y \neq j}^{p} fit^{z}}{\sum_{Y=1}^{p} fit^{z}} \right) \tag{19}$$

Where fit^z defines the fitness of each feature, which can be further expanded to

$$We^y = \frac{1}{p-1} * \frac{[\, fit^1 + fit^2 + \cdots + fit^{z+1} + \cdots + fit^{p-1} + fit^p \,]}{fit^1 + fit^2 + \cdots + fit^{z-1} + fit^z + fit^{z+1} + \cdots + fit^{p-1} + fit^p} \quad (20)$$

The selected features are given as *Sel*(y = 1,2,……..p). The output of WSO is expressed as *(sel)* = { $sel^1, sel^2, \ldots sel^p$ }, which denotes a new sub-part of features in the dataset. At last, the feature selection process with WSO generates a feature subset having optimal features.

Mathematical Explanation

Assuming the number of features in the dataset = 5

Step 1: Initialization

Initialized the shark_population size = 30

Step 2: Initial position of sharks

Using equation 1, $lb_z = 0;\ up_z = 1;\ n = [0,1]$

For simplicity, initialize three sharks

Shark 1: $W_1^1 = 0.2, W_2^1 = 0.7, W_3^1 = 0.3, W_4^1 = 0.5, W_5^1 = 0.9$

Shark 2: $W_1^2 = 0.6, W_2^2 = 0.2, W_3^2 = 0.8, W_4^2 = 0.4, W_5^2 = 0.7$

Shark 3: $W_1^3 = 0.3, W_2^3 = 0.9, W_3^3 = 0.4, W_4^3 = 0.6, W_5^3 = 0.2$

Step 3: Velocity update (calculated using equation 4)

μ is calculated using equation 8

$\mu = 2 / |\, 2 - 4.12 - \sqrt{4.12^2 - 4.4.12}\,|$

Expanding and calculating

$\mu = 2 / |\, 2 - 4.12 - \sqrt{0.4944}\,|$

$\mu \sim 0.708$

U_1 and U_2 are between [0,1]

R_1 and R_2 are calculated using equation 6 and 7

For the first iteration (*m= 1* and *M= 100*)

$R_1 = 1.5 + (1.5 - 0.5) \times e^{-(4X1/100)^{\wedge}2}$

$R_1 = 1.5 + 1 \times e^{-0.0016}$

$R_1 = 1.5 + 1 \times 0.9984$

$R_1 \sim 2.4984$

$R_2 = 0.5 + (1.5 - 0.5) \times e^{-0.0016}$

$R_2 = 0.5 + 1 \times 0.9984$

$R_2 \sim 1.4984$

Calculating for Shark 1 in the first iteration

$Vl_1^1 = 0$ (assuming initial velocity is 0)

W_{gbestk} (global best position of sharks, assuming initially it is Shark's 2 position)

$W_{best}^{Vl_k^y}$ (best position of current shark)

Using random values of U_1 and U_2 as 0.5

$Vl_2^1 = 0.708[0 + 2.4984(0.6 - 0.2) \times 0.5 + 1.4984(0.7 - 0.2) \times 0.5]$

$Vl_2^1 = 0.708 \times 0.87428$

$Vl_2^1 \sim 0.6192$

Step 4: Position update

Using equation 9 and 10

$QS = 0.3$

$Vl_2^1 \sim 0.6192$

$W_2^1 = 0.2 + \frac{0.6192}{0.41}$

$W_2^1 \sim 1.71$

Since W_2^1 exceeds the upper bound of 1, we set it to 1.

So, the updated position of Shark 1 for feature 1 becomes 1.

Step 5: Iteration

This process continuous for 100 iterations. At each step, velocities and positions are updated based

on the equations provided, and the sharks converge towards the optimal set of features.

Step 6: Feature Selection

After running for 100 iterations, the positions of the sharks represent the importance of each feature. The features with higher importance (closer to 1) are considered important.

Assume after several iterations, Shark 1 has the following positions:

$W_1^1 = 0.95$

$W_2^1 = 0.4$

$W_3^1 = 0.92$

$W_4^1 = 0.3$

$W_5^1 = 0.85$

We set threshold to select features (for e.g., 0.5), based on this the selected features are 1, 3, and 5 as they have values above 0.5

3.2.7.Bi-LSTM

A Bidirectional Long Short-Term Memory (Bi-LSTM) network is an advanced type of Recurrent Neural Network (RNN) designed to capture dependencies in sequence data from both forward and backward directions. This architecture allows the model to consider both past (left context) and future (right context) information at any point in the sequence, enhancing its ability to understand the context and improve performance on various tasks such as classification.

An LSTM cell, the fundamental building block of an LSTM network, contains several components designed to control the flow of information: the input gate, the forget gate, and the output gate. These gates regulate the information passing through the cell, allowing the network to maintain long-term dependencies.

1. Forget Gate: The forget gate decides what information from the previous cell state C_{t-1} should be discarded. It uses a sigmoid activation function to produce a value between 0 and 1 for each number in the cell state C_{t-1} :

$$f_t = \sigma\left(W_f \cdot [h_{t-1},\ x_t] + b_f\right) \quad (21)$$

where:

- f_t is the forget gate vector at time t
- σ is the sigmoid function,

- W_f is the weight matrix for the forget gate,
- $h_{t-1},\ x_t$ is the concatenated vector of the previous hidden state and the current input,
- b_f is the bias term for the forget gate.

2. Input Gate: The input gate controls the updating process of the cell state. It determines which values will be updated and how much of the new information should be added to the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1},\ x_t] + b_i) \tag{22}$$

The candidate cell state $\tilde{c}_t$ is computed using the tanh activation function, which produces values between -1 and 1:

$$\tilde{c}_t = \ tanh(W_C \cdot [h_{t-1},\ x_t] + b_C) \tag{23}$$

where:

- i_t is the input gate vector at time t,
- W_i and W_C are the weight matrices for the input gate and the candidate cell state, respectively,
- b_i and b_C are the bias terms for the input gate and candidate cell state, respectively.

3. Cell State Update: The new cell state C_t is a combination of the old cell state C_{t-1}, modulated by the forget gate, and the candidate cell state $\tilde{c}_t$, scaled by the input gate:

$$C_t = \ f_t \cdot \ C_{t-1} + \ i_t \ \cdot \ \tilde{c}_t \tag{24}$$

This equation ensures that important information is carried forward through time, while irrelevant information is discarded.

4. Output Gate: The output gate determines the hidden state h_t, which is used for the next time step and for any required output. The output gate uses the sigmoid function, and the hidden state is modulated by the tanh function applied to the cell state:

$$o_t = \sigma(W_o \cdot [h_{t-1},\ x_t] + b_o) \tag{25}$$

$$h_t = \ o_t \cdot \tanh\ (C_t) \tag{26}$$

where:

- o_t is the output gate vector at time t,
- W_o is the weight matrix for the output gate,
- b_o is the bias term for the output gate,
- h_t is the hidden state vector at time t.

In a Bi-LSTM network, these LSTM equations are applied in two parallel layers. One LSTM layer processes the input sequence in the forward direction from $t = 1$ to $t = T$, while the other LSTM layer processes the sequence in the backward direction from $t = T$ to $t = 1$. The hidden states from both directions at each time step t are concatenated to form the final output:

$$\vec{h}_t = LSTM_{forward}(x_t, \vec{h}_{t-1}) \tag{27}$$

$$\leftarrow h_t = LSTM_{backward}(x_t, h_{t-1}) \quad (28)$$

$$h_t = [\overrightarrow{h_t}; \leftarrow h_t] \quad (29)$$

Here, $\vec{h}_t$ represents the hidden state from the forward LSTM at time step t, and $\leftarrow h_t$ represents the hidden state from the backward LSTM at the same time step. The concatenated hidden state h_t combines information from both directions, allowing the model to capture context from the entire sequence, both past and future.

For classification tasks, the concatenated hidden states h_t are typically passed through additional layers, such as fully connected (dense) layers, to perform the final classification as illustrated in Fig 3.1. The output layer often uses a softmax activation function to produce probability distributions over the possible classes:

$$y_t = softmax[W_{out} \cdot h_t + b_{out}] \quad (30)$$

where:

- y_t is the output vector representing the probability distribution over classes at time t,
- W_{out} is the weight matrix for the output layer,
- b_{out} is the bias term for the output layer.

By utilizing both forward and backward LSTM layers, Bi-LSTM networks provide a comprehensive understanding of the data, making them effective for various tasks.

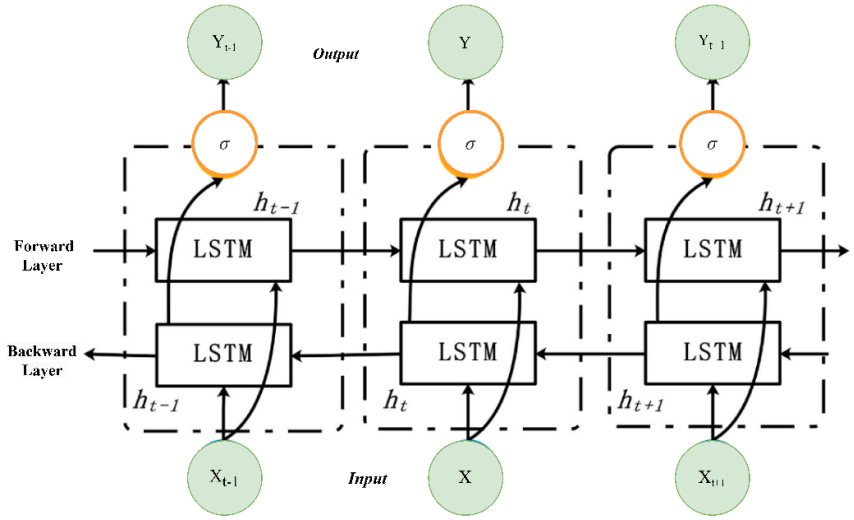


Fig 3.4 The architecture of the Bi-LSTM module

In a questionnaire dataset, while responses are not sequential in a temporal sense, they can be contextually related. Bi-LSTM's ability to process data bidirectionally allows it to understand how responses might influence each other. For example, the answer to one question might provide context for interpreting answers to subsequent questions, and vice versa. By considering both previous and future responses, Bi-

LSTM helps in capturing dependencies and interactions that might be missed by unidirectional models. This bidirectional approach is particularly useful in understanding complex patterns and nuances in ASD features that could be crucial for accurate diagnosis.

In our work, we also applied dropout regularization within the Bi-LSTM layers to prevent overfitting. Dropout randomly deactivates a fraction of the neurons during training, which helps the model generalize better by reducing its reliance on any single neuron. Additionally, during training, we monitored the model's performance on the test set and implemented early stopping. The training was halted when the validation performance no longer improved, thereby preventing the model from overfitting to the training data.

3.2.8. Why choose the white shark optimization technique for autism spectrum disorder?

A bio-inspired meta-heuristic white shark optimization technique is superiorly fitted for the proposed architecture because of the following reasons.

- White shark optimization is explored in various other fields for cracking numerous complex problems. The behavior of WSO algorithm is exciting to explore in the field of ASD as well. The motivating and key idea behind this is the intelligent and social behavior of white sharks which is somewhat similar to human behavior. The sensing capability, understanding of the environment, navigation through the complex ocean (search space), and problem-solving behavior are robust and sturdy as compared to humans. Another appealing behavior of white sharks is the strategy of search and ambush tactics. Similarly, in the context of ASD, this can be associated with the influence of related features on classification and prediction tasks.

Algorithm of the proposed WS-BiTM

Input: Pre-processed dataset having various features

Output: Optimized feature set and classification outcome

Initialize White Shark population

Generate initial positions for WSO

For Initial population

    Initialize the velocity

    Evaluate the position

While (m < M) **do**

  Update the parameters

  for i=1 to n **do**

$$Vl_{k+1}^{y} = \mu\left[Vl_{k}^{y} + R_1\left(W_{gbestk} - W_{k}^{y}\right) \times U_1 + R_2\left(W_{best}^{Vl_{k}^{y}} - W_{k}^{y}\right) \times U_2\right]$$

end for

for i=1 to n **do**

if $rand < QS$ **then**

$$W_{k+1}^{y} = W_{k}^{y} . \neg W_0 + up.h + l_0.d$$

else

$$W_{k}^{y} + Vl_{k}^{y} / fq$$

end if

end for

for i=1 to n **do**

if $rand < Str$

$$\overrightarrow{Dis_w} = |rand \times \left(W_{gbestk} - W_{k}^{y}\right)|$$

if i==1 **then**

$$W_{k+1}^{y} = W_{gbestk} + n_1 \overrightarrow{Dis_w}\, sgn(n_2 - 0.5)$$

else

$$W_{k+1}^{\prime y} = W_{gbestk} + n_1 \overrightarrow{Dis_w}\, sgn(n_2 - 0.5)$$

end if

end if

end for

Adjust the position of white sharks

Evaluate/update the new positions

K = k+1

Return the optimal solution

Final feature set *(sel)*

Split the dataset D into training and test sets $\mathbf{D_{train}}$ and $\mathbf{D_{test}}$ using the optimized feature set ***(sel)***

Input features into BiLSTM layer X

Bi-LSTM layer with forward and backward cells

$$\vec{h}_t = LSTM_{forward}(x_t, \vec{h}_{t-1})$$

$$\leftarrow h_t = LSTM_{backward}(x_t, h_{t-1})$$

Concatenate the outputs of forward and backward cells

$$h_t = [\vec{h}_t; \leftarrow h_t]$$

Add the fully connected dense layer

$y_t = softmax[W_{out} \cdot h_t + b_{out}]$

Loss function: categorical cross entropy; Optimizer: Adam

Train

Model.fit(D_{train},epoch = 100, batch_size = 64)

Evaluate and Return the values obtained

Performance_metrics = model.evaluate(D_{test})

- The velocity of the white shark shows the motion and the direction of exploration in the search space. The velocity defines how quickly the algorithm will converge toward the optimal solution. For ASD, it is associated with the exploration of a combination of features to find the most contributing indicators for the prediction of ASD. By dynamically updating the velocities, the technique effectively explores the feature space. Parameters like W_{gbestk}, W_k^y, R_1, R_2, U_1, U_2, and μ compute the magnitude and direction of velocity updates for every shark, and adjustment of these parameters strikes a balance between exploration and exploitation, which is necessary, enabling efficient search.

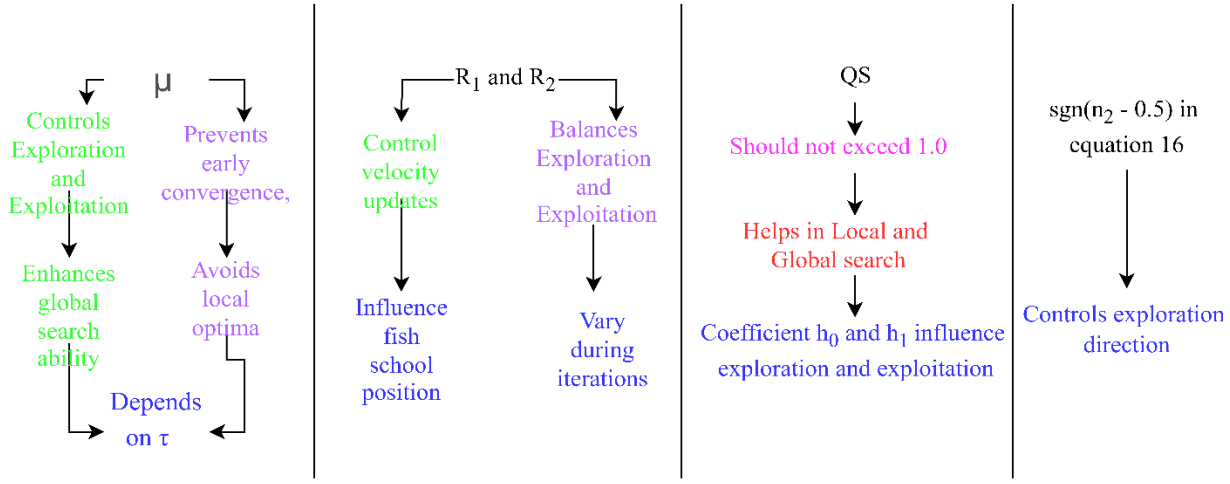


Fig 3.5 Concept map illustrating the exploration/exploitation abilities of WSO

- WSO leverages parallel processing capability by computing the fitness score of multiple solutions simultaneously, thus decreasing the computational time. Fig 3.5 illustrates the concept map behind the selection of WSO parameters and showcases the exploration/exploitation ability of WSO.

- WSO employs a population of white sharks, and this population-based approach allows exploration of diverse search space and prevents premature convergence to the suboptimal solution, which in turn improves the ability of the technique to identify near-global/global optimal solutions.

3.3. Experiments and Results

This module represents the analysis of the experimental outcomes obtained from the proposed AFF-BPL and WS-BiTM architectures. The analysis of results is segregated into three sub-sections to effectively portray the outcomes of the proposed work. Section 3.4.1 explains the results of the proposed work on three datasets. Section 3.4.2 compares the results of the proposed work with state-of-the-art techniques on three datasets. Section 3.4.3 provides an ablation study conducted by taking various cases into consideration.

3.3.1. Performance Evaluation Parameters

The efficacy of the WS-BiTM is evaluated using five key metrics: Sensitivity, Precision, Accuracy, Specificity, and F1.

3.3.2. Dataset Description

To facilitate the ASD diagnosis work, three ASD datasets incorporating individuals of various age groups were used: the 'Toddler' dataset, the 'Children' dataset, and the 'Adult' dataset [120] . All three datasets were obtained from the public platform known as 'Kaggle'. The dataset includes a set of questionnaires along with personal information. The toddler dataset contains 1054 instances, adults have 704 instances and children include 292 instances.

Table 3.2 Dataset description of common features

Feature Id	Feature Description
1.	Age
2.	Gender
3.	Ethnicity
4.	Jaundice History
5.	PDD with Family Members
6.	Who is completing the test?
7.	Country
8.	Whether the users have used the screening app
9.	Type of screening method
10-19.	Answer of the questions
20.	Screening Score

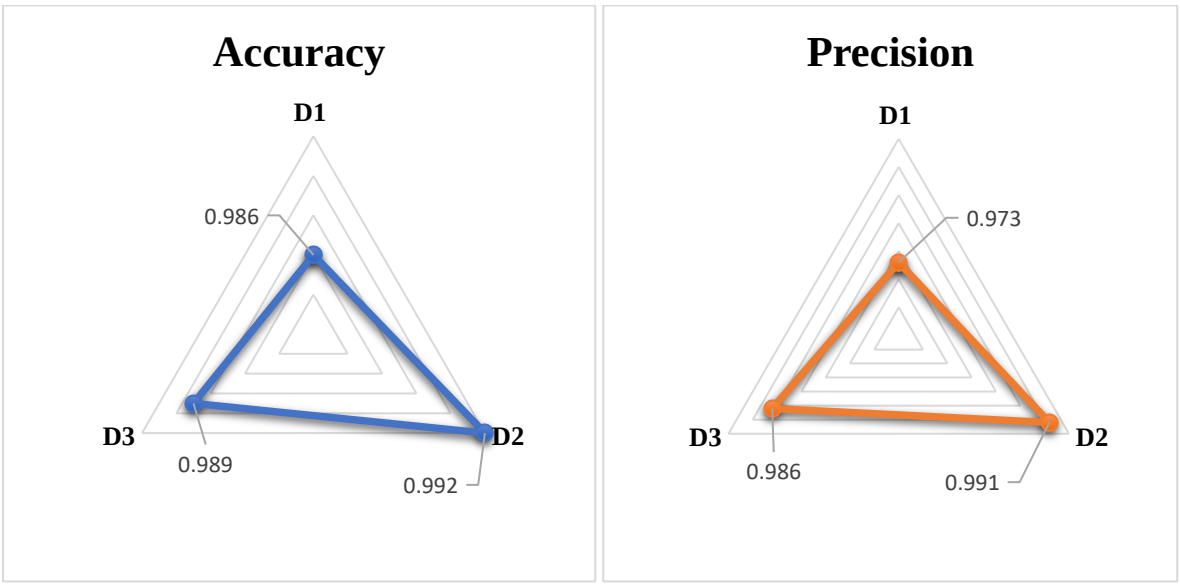
3.3.3. Result Analysis: Comparison of Proposed work on Three autism datasets on AFF-BPL

Table 3.3 shows the experimental results of AFF-BPL obtained on various parameters on three autism screening datasets. As shown in the table, the AFF-BPL architecture performed fairly well on all three datasets.

Table 3.3 Comparison of AFF-BPL results evaluated on five parameters on three datasets.

Dataset	Accuracy	Precision	Sensitivity	Specificity	F1-score
Adult (D1)	0.986	0.973	0.973	0.975	0.976
Toddlers (D2)	0.992	0.991	0.984	0.986	0.990
Children (D3)	0.989	0.986	0.986	0.984	0.989

Figure 3.6 portrays the radar plots by comparing the accuracy, precision, specificity, sensitivity, and f1-socres across datasets. In the radar plots, D1 stands for the 'Adult' dataset, D2 stands for the 'Toddlers' dataset, and D3 stands for the 'Children' dataset.



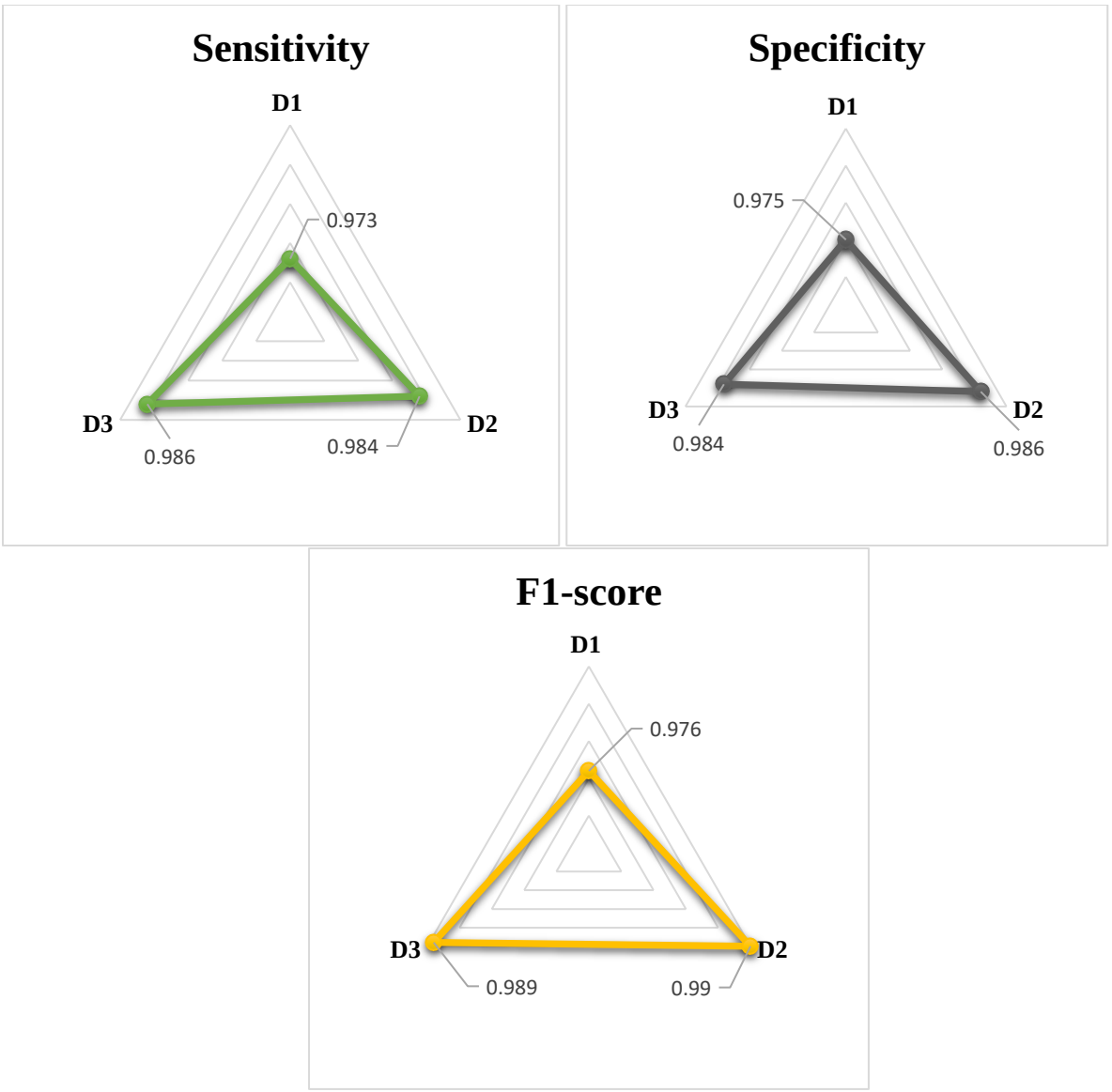


Figure 3.6. Radar plots of Accuracy, precision, sensitivity, specificity, and F1 score on Adult, Toddler, and Children dataset. Here, D1 stands for Adult dataset, D2 stands for Toddler dataset, and D3 stands for Children dataset

3.3.3.1. Comparison with State-of-the-art Techniques

Table 3.4 compares the experimental results obtained by employing the proposed architecture and the state-of-the-art techniques on the 'Adult' dataset. We evaluated the efficacy of the proposed with neural network (NN), convolution neural network (CNN), and long short-term memory (LSTM) network.

Table 3.4 Comparison of the proposed with three state-of-the-techniques on the Adult dataset

Model	Accuracy	Precision	Sensitivity	Specificity	F1-Score
NN	0.769	0.825	0.857	0.918	0.75

CNN	0.958	0.942	0.934	0.95	0.94
LSTM	0.937	0.921	0.912	0.84	0.95
Proposed Model	0.986	0.973	0.973	0.975	0.976

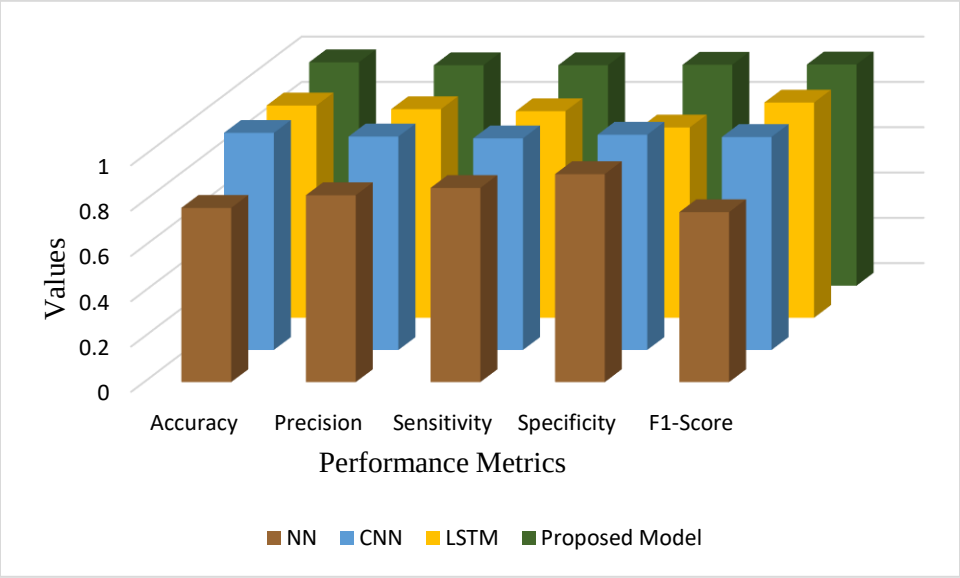


Figure 3.7 Diagrammatic representation of comparative analysis of the proposed work with state-of-the-art techniques on Adult dataset

Figure 3.7 illustrates the comparison of the proposed work with neural network, convolution neural networks and long short-term memory on the Adult dataset in the form of bar plots.

Table 3.5 Comparison of the proposed with three state-of-the-techniques on the Toddler dataset

Model	Accuracy	Precision	Sensitivity	Specificity	F1-Score
NN	0.797	0.80	0.80	0.815	0.80
CNN	0.936	0.932	0.918	0.917	0.932
LSTM	0.946	0.936	0.925	0.921	0.93
Proposed Model	0.992	0.991	0.984	0.986	0.990

Table 3.5 compares the experimental results obtained by employing the proposed architecture and the state-of-the-art techniques on the 'Toddler' dataset. The outcomes highlight the improvement in results after employing our proposed approach.

Figure 3.8 represents bar plots to effectively show the comparison of proposed work with neural network, convolution neural network and long short-term memory on Toddler dataset.

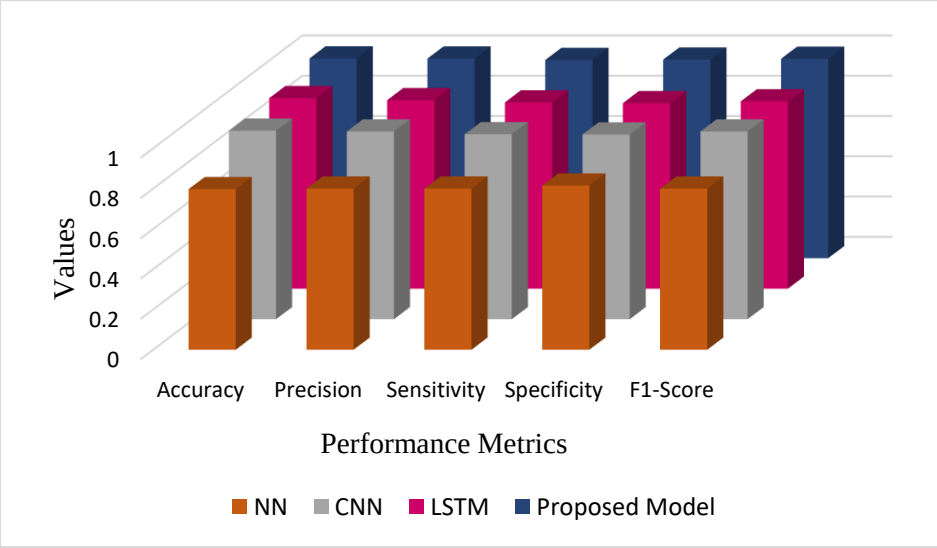


Figure 3.8 Diagrammatic representation of comparative analysis of the proposed work with state-of-the-art techniques on Toddler dataset

Table 3.6 compares the experimental values obtained by employing the proposed architecture and the state-of-the-art techniques on the 'Children' dataset. The outcomes highlight the improvement in results after employing our proposed approach.

Table 3.6 Comparison of the proposed with three state-of-the-techniques on the Children dataset

Model	Accuracy	Precision	Sensitivity	Specificity	F1-Score
NN	0.752	0.812	0.84	0.90	0.84
CNN	0.912	0.906	0.90	0.924	0.90
LSTM	0.926	0.917	0.91	0.925	0.91
Proposed Model	0.989	0.986	0.986	0.984	0.989

Figure 3.9 showcases bar plots to effectively show the comparison of proposed work with neural network, convolution neural network, and long short-term memory on Children dataset. To further validate the statistical performance results of the AFF-BPL, we performed a paired t-test. It is noteworthy that the paired t-test was used to achieve our objectives via a pairwise comparison between the employed and the proposed model to showcase the statistical evidence.

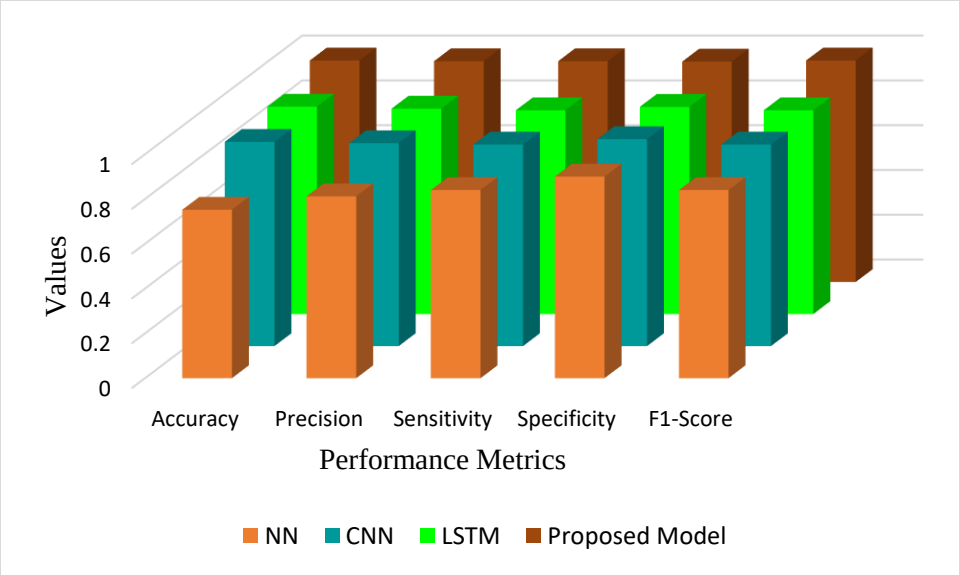


Figure 3.9 Diagrammatic representation of comparative analysis of the proposed work with state-of-the-art techniques on Children dataset

Specifically, we set the p-value (significance level) to be 0.05 and if the obtained result is less than the significance value, we can reject the null hypothesis and represent that there is a significant difference in the performances. The obtained statistics and p-values are showcased in Table 3.7.

Table 3.7 Experimental outcomes of statistical performance on three datasets namely Adult(D1), Toddlers(D2), and Children(D3)

Metrics	**Models**	**Adult (D1)**		**Toddlers (D2)**		**Children (D3)**	
		Statistics	***P-value***	***Statistics***	***P-value***	***Statistics***	***P-value***
Accuracy	NN	*4.93*	*0.00789*	*41.41*	*$2.03e^{-6}$*	*6.06*	*0.00375*
	CNN	*12.61*	*0.000228*	*22.90*	*$2.15e^{-5}$*	*17.57*	*$6.16e^{-5}$*
	LSTM	*4.17*	*0.01404*	*14.62*	*0.000127*	*22.25*	*$2.42e^{-5}$*
Precision	NN	*3.41*	*0.027*	*47.67*	*$1.16e^{-6}$*	*8.61*	*0.001*
	CNN	*8.05*	*0.00129*	*29.22*	*$8.17e^{-6}$*	*17.33*	*$6.50e^{-5}$*
	LSTM	*12.44*	*0.00024*	*29.50*	*$7.87e^{-6}$*	*25.21*	*$1.47e^{-5}$*
Sensitivity	NN	*5.08*	*0.00708*	*16.31*	*$5.96e^{-7}$*	*10.77*	*0.00042*
	CNN	*12.54*	*0.000233*	*39.39*	*$2.48e^{-6}$*	*15.54*	*0.0001*
	LSTM	*4.96*	*0.00771*	*37.12*	*$3.14e^{-6}$*	*21.35*	*$2.84e^{-5}$*
Specificity	NN	*3.59*	*0.02299*	*42.38*	*$1.85e^{-6}$*	*13.51*	*0.00017*
	CNN	*8.82*	*0.000912*	*34.24*	*$4.34e^{-6}$*	*14.40*	*$1.60e^{-9}$*
	LSTM	*9.07*	*0.000820*	*30.01*	*$7.34e^{-6}$*	*16.49*	*$3.26e^{-8}$*
	NN	*2.63*	*0.04822*	*49.25*	*$1.02e^{-6}$*	*11.42*	*0.00034*

F1-score	CNN	*22.77*	*$2.20e^{-5}$*	*26.93*	*$1.13e^{-5}$*	*17.54*	*$6.20e^{-5}$*
	LSTM	*3.02*	*0.03913*	*37.52*	*$3.01e^{-6}$*	*25.33*	*$1.44e^{-5}$*

3.3.3.2. Ablation Study

This segment showcases an ablation study to validate the proposed architecture's contribution toward feature selection in AFF-BPL architecture. The AFF-BPL is trained on three autism screening datasets namely, Adult, Children, and Toddlers. The mentioned study cases are evaluated:

Case A: With only BAT for feature selection

Case B: With only PSO for feature selection

Case C: Concurrently using PSO and BAT without adaptive feature fusion

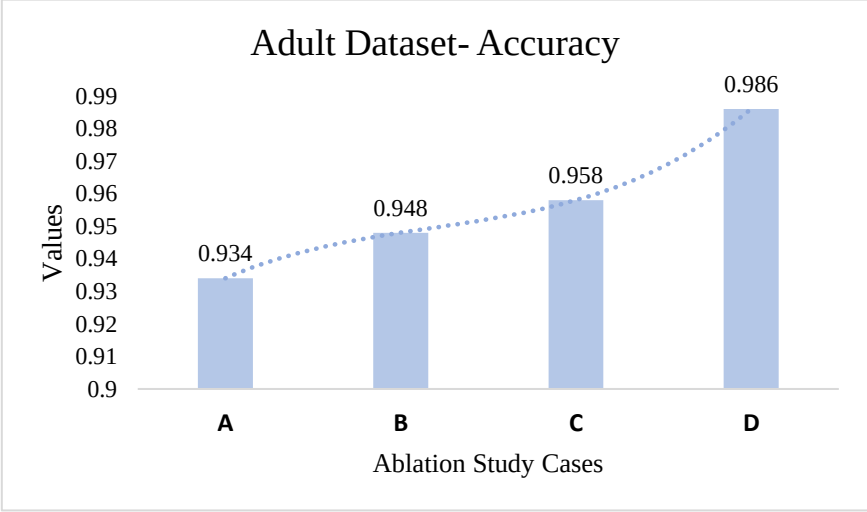
Case D: Concurrently using PSO and BAT with adaptive feature fusion

Table 3.8 Ablation study values for the proposed architecture. The architecture is trained on three datasets. Case A contains scores obtained by using only PSO. Case B incorporates scores obtained by using only BAT. Case C involves concurrently using BAT and PSO without adaptive feature fusion. Case D involves concurrent use of BAT and PSO with adaptive feature fusion technique.

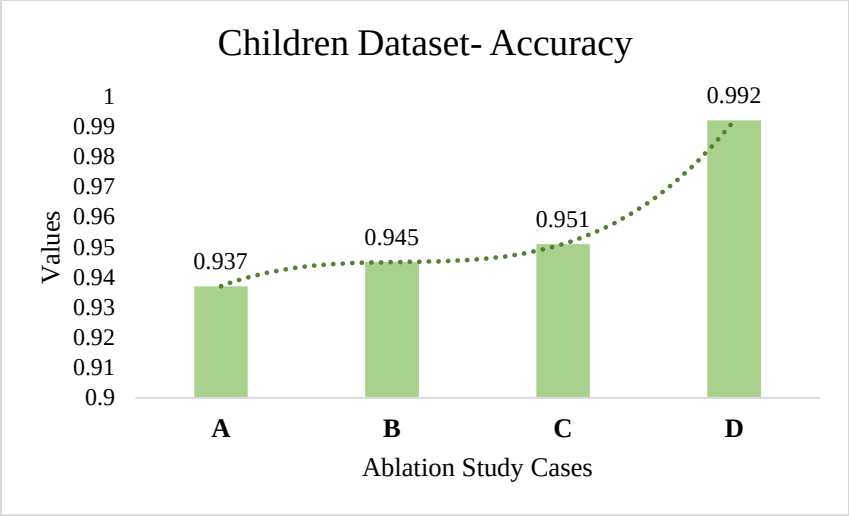
Dataset	**Ablation Study**	**Case**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1- score**
Adult	With only BAT for feature selection	A	0.934	0.901	0.932	0.931	0.921
	With only PSO for feature selection	B	0.948	0.949	0.952	0.951	0.950
	Concurrently using PSO and BAT without adaptive feature fusion	C	0.958	0.960	0.952	0.952	0.955
	Concurrently using PSO and BAT with adaptive feature fusion	D	0.986	0.973	0.973	0.975	0.976
Children	With only BAT for feature selection	A	0.937	0.924	0.931	0.928	0.927
	With only PSO for feature selection	B	0.945	0.948	0.950	0.949	0.949
	Concurrently using PSO and	C	0.951	0.954	0.957	0.952	0.953

	BAT without adaptive feature fusion						
	Concurrently using PSO and BAT with adaptive feature fusion	D	0.992	0.991	0.984	0.986	0.990
Toddlers	With only BAT for feature selection	A	0.930	0.912	0.925	0.922	0.920
	With only PSO for feature selection	B	0.946	0.947	0.950	0.949	0.948
	Concurrently using PSO and BAT without adaptive feature fusion	C	0.959	0.962	0.956	0.953	0.958
	Concurrently using PSO and BAT with adaptive feature fusion	D	0.989	0.986	0.986	0.984	0.989

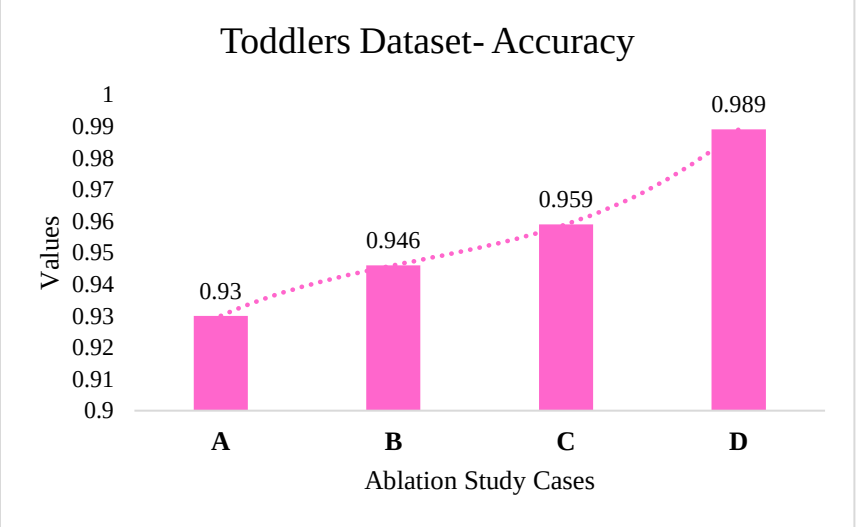
Table 3.8 shows the comprehensive values derived from the conducted ablation study. Specifically, case D, characterized by the incorporation of concurrent use of PSO and BAT with adaptive feature fusion, surpasses the performance of all other cases in the analysis.



(a)



(b)



(c)

Figure 3.10 (a), (b), and (c). Ablation study cases for AFF-BPL architecture. The architecture is trained on three datasets. Case A contains scores obtained by using only PSO. Case B incorporates scores obtained by using only BAT. Case C involves concurrently using BAT and PSO without adaptive feature fusion. Case D involves the concurrent use of BAT and PSO with an adaptive feature fusion technique. Increasing values of accuracy highlight the significance of the proposed architecture.

Figure 3.10 (a), (b), and (c) illustrates the impact of the enhancement made in the techniques. The increasing values of accuracy on every dataset highlight the significance of developed architecture.

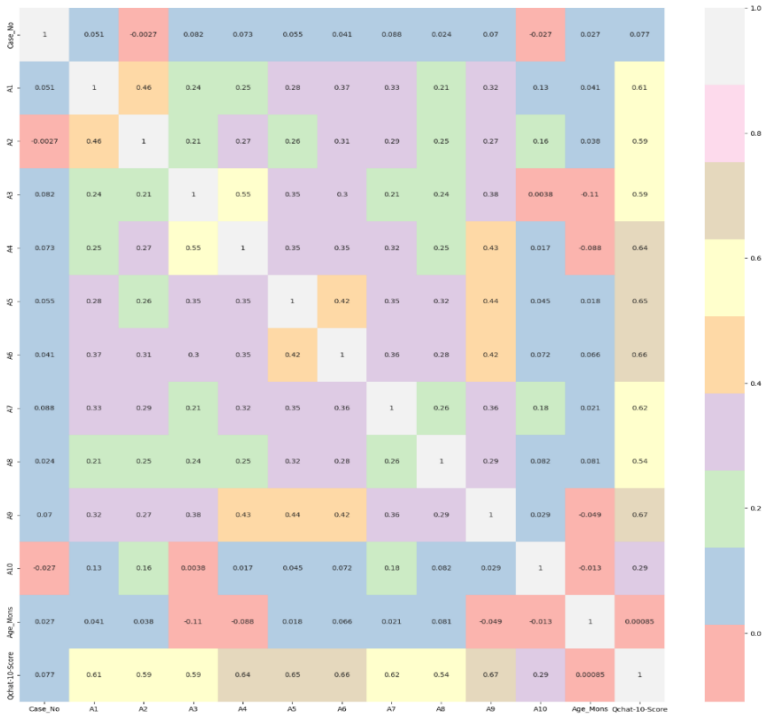
3.3.3.3.Computational Complexity Analysis

The complexity of the model can be estimated by the number of trainable operations/parameters, which depend on various hyper-parameters in each component: BAT, PSO, adaptive feature fusion, and LSTM.

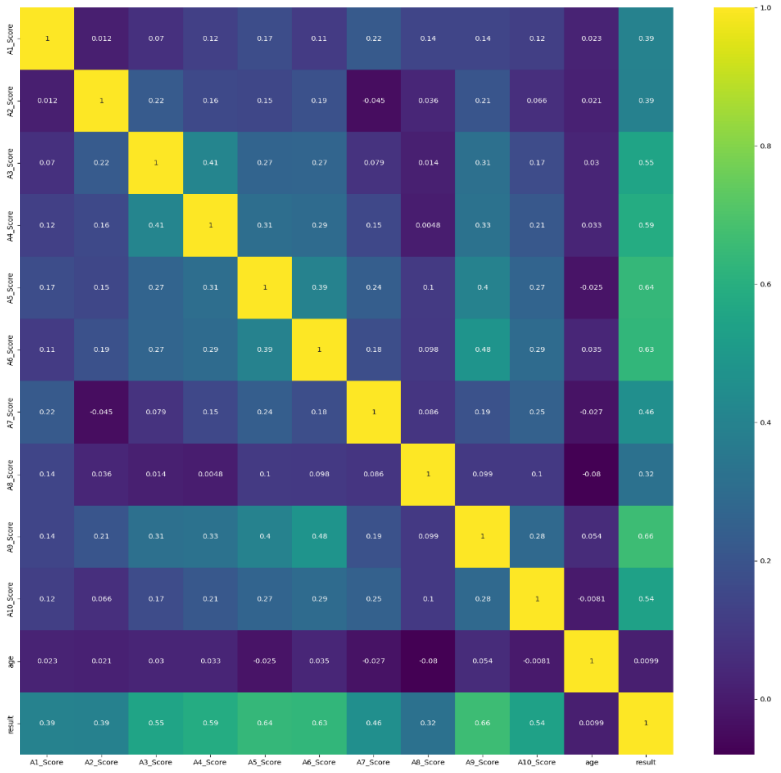
These parameters include the number of bats, particles initialized, the number of input features, and the number of LSTM units. The PSO and BA algorithms, both employed for feature selection, have a complexity of $O(N \cdot D \cdot T)$, where N is the number of particles or bats, D is the dimensionality of the feature space, and T is the number of iterations. The Adaptive feature fusion (ADFF) process, which normalizes and merges the feature importance scores from PSO and BA (as per equation 15 and 16), operates with a complexity of $O(D)$ due to the linear nature of normalization and averaging computations across the feature set. Finally, the LSTM classification phase, responsible for classifying individuals based on the fused feature set, has a complexity of $0(T' \cdot D_{LSTM}^2)$, where T' represents the number of features fed to the LSTM and D_{LSTM}^2 denotes the dimensionality of the LSTM's hidden layers. Overall, the total complexity of the system is dominated by the PSO/BA feature selection step and the LSTM classification step, making it $O(N \cdot D \cdot T) + 0(T' \cdot D_{LSTM}^2)$.

3.3.4. Results Analysis on WS-BiTM

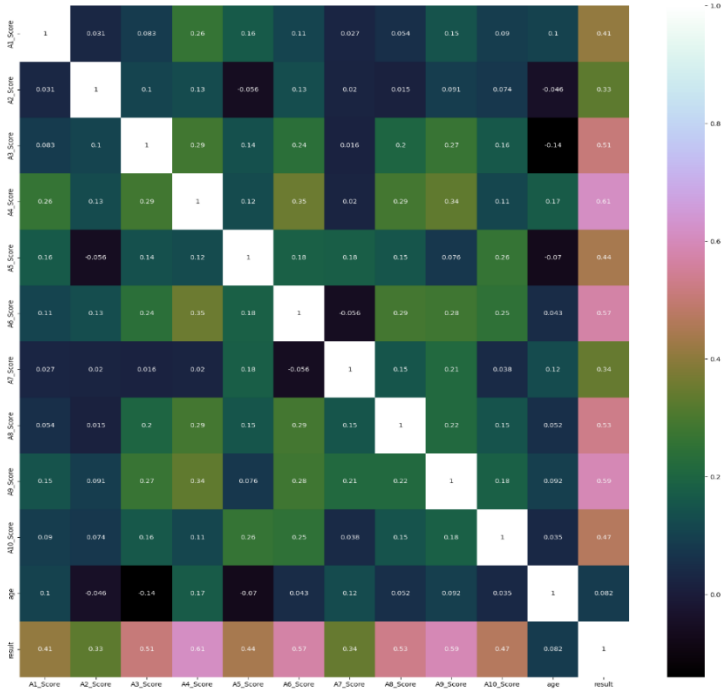
This segment explains the performance evaluation matrices and results obtained on them performing feature selection, and classification via Bi-LSTM. Experimental outcomes obtained by comparing the baseline techniques with the proposed architecture, paired t-test, and ablation study cases are briefly described in this section.



(a)



(b)



(c)

Fig 3.11 (a) Toddlers, (b) Adult, (c) Child illustrates the correlations and pattern in all three datasets

3.3.4.1. Parameter Settings and Feature Selection

The parameter values used for WSO are mentioned in Table 3.9. The shark_population_size defines the number of candidate solutions (sharks) evaluated at each iteration. A population size of 30 strikes a balance between exploration (searching broadly in the solution space) and exploitation (refining solutions) while maintaining computational efficiency. Population sizes between 20 and 50 are commonly used for optimization to avoid overloading computational resources.

Table 3.9 Parameter settings used for WSO

Parameters	**Values**
Shark_population_size	30
Num_iterations	100
R_{min}	0.5
R_{max}	1.5
fq_{max}	0.75
fq_{min}	0.07
up_z	1
lb_z	0
τ	4.12
h_0	6.25
h_1	100
h_2	0.0005
Dimensions	Num_of_features

Table 3.10 represents the features selected by WSO on three different datasets - $\mathbf{D_T}$, $\mathbf{D_C}$, and $\mathbf{D_A}$. The dataset comprises numerous features, but the table consists of the most important features, including demographic information (e.g., Age, Gender/Sex, Ethnicity) and domain-specific attributes (e.g., A1–A10, QChat). Some features were consistently selected across all three datasets, implying their importance and generalizability: A1 – A10, this feature group, likely representing critical domain-specific attributes, was selected in all datasets (D_T, D_C, and D_A), indicating it plays a significant role regardless of the dataset used. These features likely capture key patterns that are relevant across various contexts. Age is another critical feature, selected in all datasets. The fact that "Age" is universally chosen suggests it has a strong correlation with the classification or outcome prediction task across datasets. Gender/Sex, Ethnicity, and Class ASD are the demographic features that are also consistently selected across datasets,

highlighting their importance in the model's decision-making process. It shows that the WSO algorithm finds these attributes to be essential for achieving high accuracy in classification, potentially due to their relevance in understanding the factors behind ASD.

Table 3.10 Feature selection using WSO across three datasets (D_T, D_C, and D_A), demonstrating both consistent and dataset-specific feature importance.

Feature Name	**Datasets**		
	D_T	**D_C**	**D_A**
A1 – A10	✓	✓	✓
Age	✓	✓	✓
QChat	✓	×	×
Gender/Sex	✓	✓	✓
Ethnicity	✓	✓	✓
Country of Origin	×	✓	✓
Used App	×	×	✓
Class ASD	✓	✓	✓

3.3.4.2. Sensitivity Analysis of Hyper-parameters

In this section of sensitivity analysis, we investigate the impact of various hyperparameters on the performance of the White Shark Optimization (WSO) algorithm for feature selection. Specifically, we examine four parameters namely shark population size, number of iterations, and bounds for the optimization process. By systematically varying these hyperparameters within specified ranges, we aim to determine their influence on the key performance metric, i.e., classification accuracy.

Table 3.11 Performance comparison after conducting sensitivity analysis of hyper-parameters

	Hyper-parameter				Accuracy		
	Shark_population_size	*Num_iterations*	R_{min}	R_{max}	D_T	D_C	D_A
Tested Values	15	50	0.25	1.0	0.804	0.812	0.801
	30	**100**	**0.5**	**1.5**	**0.976**	**0.964**	**0.962**
	50	150	0.75	2.0	0.915	0.912	0.912
	60	200	0.5	1.5	0.922	0.932	0.936

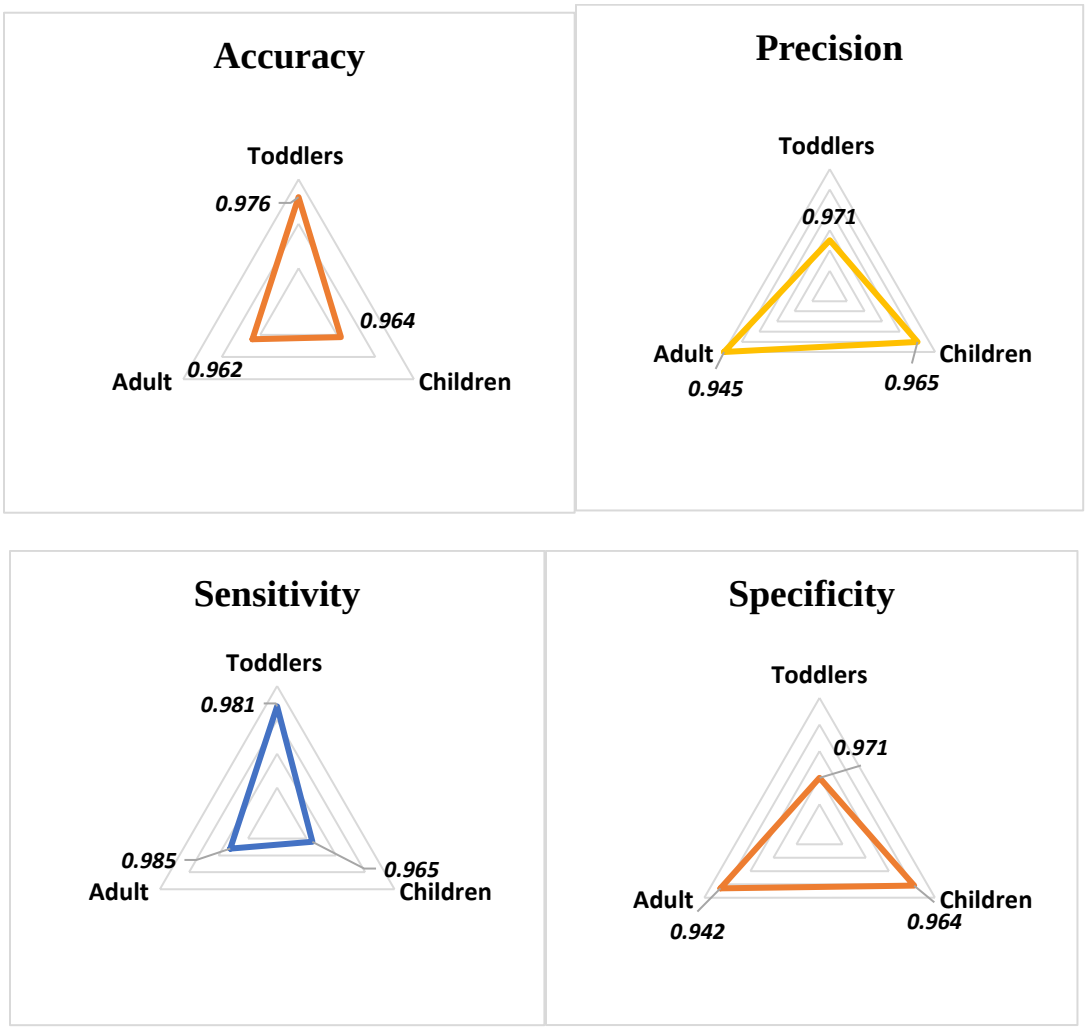
This analysis not only reinforces the rationale behind our chosen hyperparameter values but also provides insights into the robustness of the WSO approach. The best accuracy obtained on their specific parameter value on all three datasets is highlighted in Table 3.11.

3.3.4.3. Experimental Outcomes of WS-BiTM on Three Autism Datasets

Table 3.12 Comparison of WS-BiTM outcomes analyzed on five parameters on three datasets.

Dataset	Accuracy	Precision	Sensitivity	Specificity	F1-score
Toddlers (D_T)	0.976	0.971	0.981	0.971	0.975
Children (D_C)	0.964	0.965	0.965	0.964	0.964
Adult (D_A)	0.962	0.945	0.985	0.942	0.964

Table 3.12 enumerates the experimental simulation outcomes of WS-BiTM on three datasets. As specified in the table, WS-BiTM performed exceptionally well on all three datasets.



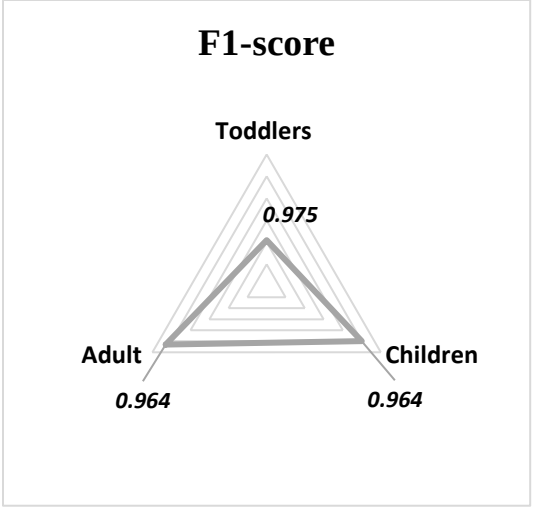
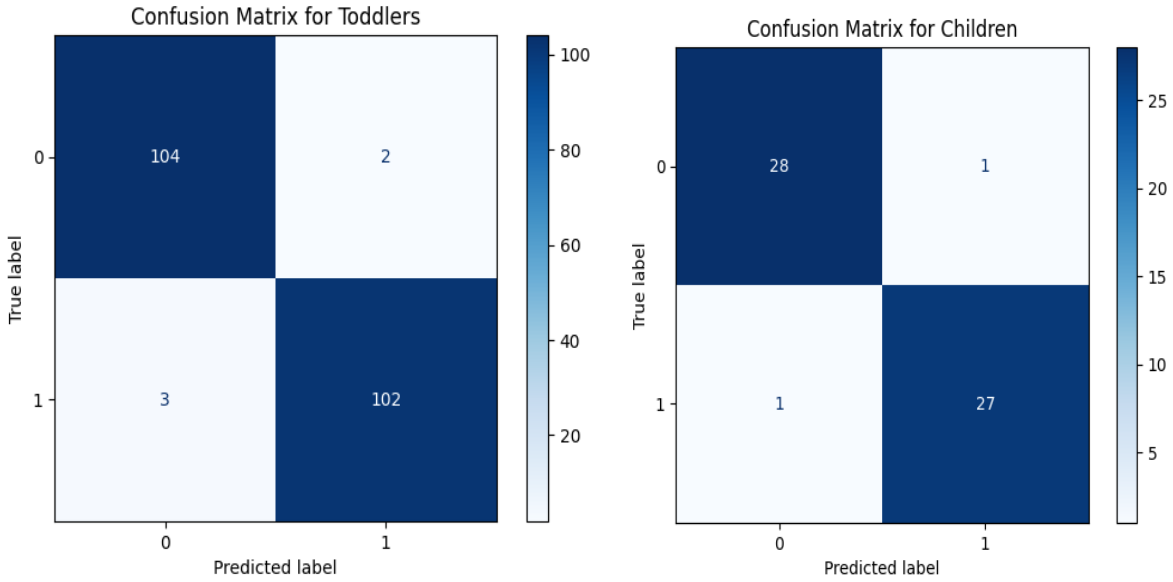


Fig 3.12 Radar plots Accuracy, precision, sensitivity, specificity, and F1 score obtained on Children, Toddlers, and Adults datasets. The value obtained on all three datasets is presented on different color-coded radar plots for every performance metric.

Fig 3.12 illustrates radar plots of five performance measures employed to showcase the effectiveness of our work on the employed three datasets. The confusion matrices in Fig 3.13 for the test set datasets: Toddlers, Children, and Adults; highlight the classification model's performance. In the Toddlers dataset, the model demonstrated high accuracy with a substantial number of true positives and true negatives while maintaining a low rate of false positives and false negatives. Similarly, the Children dataset exhibited robust classification capabilities, with effective differentiation between autistic and non-autistic.



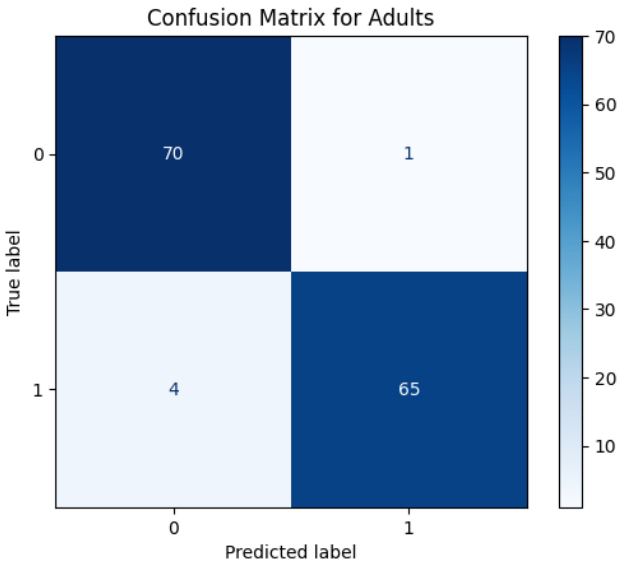


Fig 3.13 Confusion matrices of all three datasets on the test set

The Adult dataset also showed effective performance, with the model successfully identifying most instances while exhibiting minor misclassifications. Overall, these matrices illustrate the model's effectiveness across various age groups within the test set.

3.3.5. Comparison with State-of-the-Art Approaches

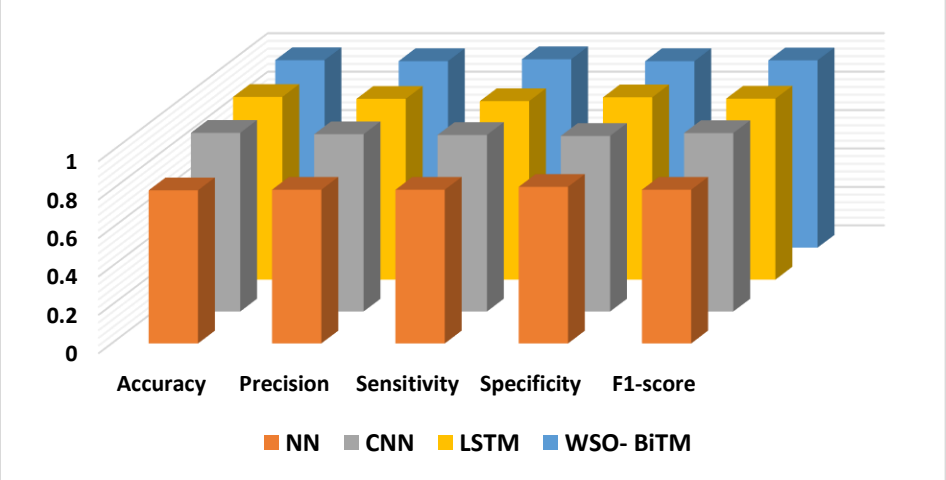
For the 'Toddlers' dataset (Table 3.13, denoted as $\mathbf{D_T}$), we analyzed that the performance values on different matrices for NN are 0.797, 0.80, 0.80, 0.815, 0.80, whereas for CNN the values are 0.930, 0.923, 0.919, 0.914, 0.929. LSTM performed fairly well with values of 0.951, 0.941, 0.929, 0.949, 0.942. The integrated WS-BiTM demonstrates better performance with values 0.976, 0.971, 0.981, 0.971, 0.975. As observed, WS-BiTM achieved the highest accuracy. Hence for the 'toddlers' dataset, the developed architecture surpassed all baseline techniques considered.

For the 'Children' dataset (denoted as $\mathbf{D_C}$), we observed that the performance values on different metrics for NN are 0.779, 0.835, 0.87, 0.93, 0.86, whereas for CNN the values are 0.912, 0.906, 0.90, 0.924, 0.90. In this case also, LSTM performed fairly well with values 0.926, 0.917, 0.91, 0.925, 0.91. The integrated WS-BiTM shows better performance with values 0.964, 0.965, 0.965, 0.964, 0.964. As observed, WS-BiTM achieved the highest accuracy. Hence for the 'Children' dataset also, the developed architecture outperforms all baseline techniques considered.

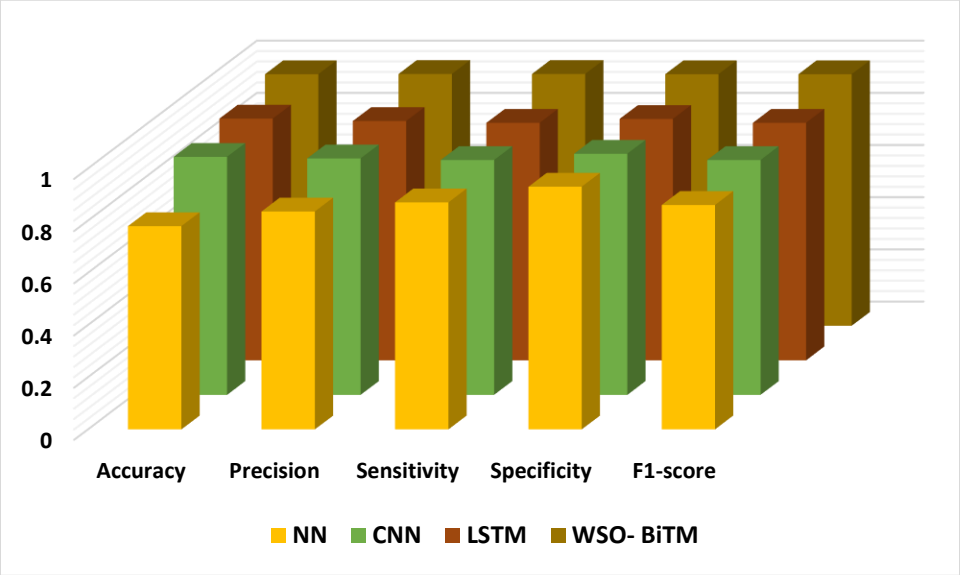
Table 3.13 Comparison of evaluation parameter values on various parameters obtained from three datasets. Here $\mathbf{D_T}$ $\mathbf{D_C}$, $\mathbf{D_A}$ signifies values on Toddlers, Children, and Adult dataset respectively

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score
Toddlers (D_T)					
NN	0.797	0.80	0.80	0.815	0.80
CNN	0.930	0.923	0.919	0.914	0.929
LSTM	0.951	0.941	0.929	0.949	0.942
WSO- BiTM	***0.976***	***0.971***	***0.981***	***0.971***	***0.975***
Children (D_C)					
NN	0.779	0.835	0.87	0.93	0.86
CNN	0.912	0.906	0.90	0.924	0.90
LSTM	0.926	0.917	0.91	0.925	0.91
WSO-BiTM	***0.964***	***0.965***	***0.965***	***0.964***	***0.964***
Adult (D_A)					
NN	0.769	0.825	0.857	0.918	0.75
CNN	0.958	0.942	0.934	0.95	0.94
LSTM	0.937	0.921	0.912	0.84	0.95
WSO-BiTM	***0.962***	***0.945***	***0.985***	***0.942***	***0.964***

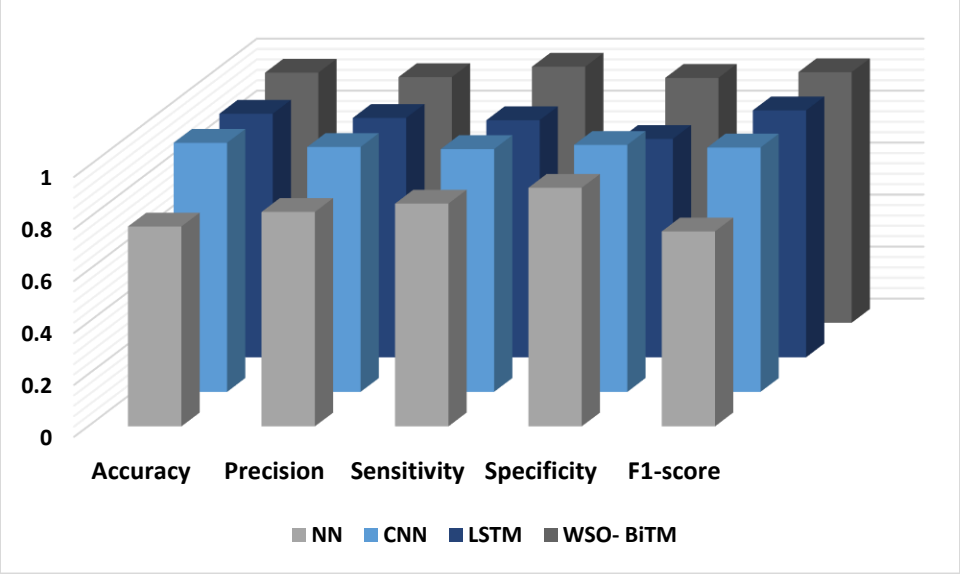
For the 'Adults' dataset (denoted as $\mathbf{D_A}$), we examined that the performance values on different parameters for NN are 0.769, 0.825, 0.857, 0.918, 0.75, whereas for CNN the values are 0.958, 0.942, 0.934, 0.95, 0.94. Values achieved by LSTM are 0.937, 0.921, 0.912, 0.84, 0.95.



(a)



(b)



(c)

Fig 3.14 Graphical representation of model comparison on various parameters on three datasets, where (a), (b), and (c) represent statistics of toddlers, children, and adult datasets respectively

The integrated WS-BiTM shows better performance with values 0.962, 0.945, 0.985, 0.942, 0.964. As observed, WS-BiTM achieved the highest accuracy. At last, for the 'Adults' dataset, the developed architecture serves promising results compared with all baseline techniques. Fig 3.14 (a), (b), and (c) graphically represent the comparison of performance of all the employed techniques across the three datasets. We examined that WS-BiTM achieved the highest accuracy across all the datasets. Table 3.14 comprehensively evaluates various classification algorithms, focusing on their accuracy and computational time across three datasets: Adults, Toddlers, and Children. The neural network (NN) classifier achieved accuracies of 76.9%, 79.7%, and 77.9%, respectively, with a computation time of approximately 127.6 seconds, indicating its relatively lower performance compared to more complex models. The convolutional neural network (CNN) demonstrated significantly higher accuracy, reaching 95.8% for the 'Adult' dataset, 93.0% for 'Toddlers', and 91.2% for 'Children', albeit at a considerably greater computation time of 520.4 seconds. Long Short-Term Memory (LSTM) networks also performed well, achieving accuracies of 93.7%, 95.1%, and 92.6%, but required even more computation time, approximately 1070.5 seconds. Notably, the proposed WS-BiTM model outperformed all other algorithms, with accuracies of 96.2% for 'Adult', 97.6% for 'Toddlers', and 96.4% for 'Children' data, albeit at the highest computation cost of around 2800.6 seconds. This analysis underscores the trade-off between accuracy and computational efficiency, highlighting the superiority of the proposed model in accuracy while also necessitating significantly greater computational resources.

Table 3.14 Comparison of algorithm's accuracy and computational time

Algorithm/Classifier	**Accuracy**			**Computation time (Approx)**
	Adult	**Toddlers**	**Children**	
NN	0.769	0.797	0.779	127.6
CNN	0.958	0.930	0.912	520.4
LSTM	0.937	0.951	0.926	1070.5
Proposed: WS-BiTM	0.962	0.976	0.964	2800.6

3.3.4.4. Statistical Significance: Paired t-Test

In this segment, we sought to further validate the statistical significance of the WS-BiTM model by conducting a paired t-test. This statistical method enabled us to compare the proposed WS-BiTM model against various employed models, thereby providing robust statistical evidence through pairwise comparisons. Specifically, we utilized the paired t-test to evaluate the performance of different models such as Neural Networks (NN), Convolutional Neural Networks (CNN), and Long Short-Term Memory

(LSTM), relative to the WS-BiTM model. Our objective was to ascertain whether statistically significant differences existed in performance metrics, including Accuracy, Precision, Sensitivity, Specificity, and F1-score, across three distinct datasets: Adults, Toddlers, and Children. In our analysis, each pair in the paired t-test consisted of the performance metrics (e.g., Accuracy, Precision) of the same model evaluated on identical datasets. We established a significance level (p-value) of 0.05; a p-value below this threshold would lead us to reject the null hypothesis, indicating a significant difference in performance between the models. The statistical results and corresponding p-values obtained from this analysis are summarized in Table 3.15.

Table 3.15 Statistical results showcasing the significance in model performance across all datasets

Metrics	**Models**	**Adult**		**Toddlers**		**Children**	
		Statistics	***P-value***	***Statistics***	***P-value***	***Statistics***	***P-value***
Accuracy	NN	*3.97*	*0.00734*	*36.22*	*$3.41e^{-6}$*	*5.54*	*0.00299*
	CNN	*10.74*	*0.000198*	*19.42*	*$3.22e^{-5}$*	*12.02*	*$5.76e^{-5}$*
	LSTM	*5.18*	*0.01340*	*11.35*	*0.000119*	*17.18*	*$3.13e^{-5}$*
Precision	NN	*4.88*	*0.034*	*42.05*	*$2.86e^{-6}$*	*9.98*	*0.0026*
	CNN	*9.70*	*0.00119*	*21.14*	*$7.81e^{-6}$*	*13.05*	*$4.58e^{-5}$*
	LSTM	*11.36*	*0.00031*	*19.26*	*$6.21e^{-6}$*	*19.28*	*$2.33e^{-5}$*
Sensitivity	NN	*4.27*	*0.00698*	*14.45*	*$5.91e^{-7}$*	*11.66*	*0.00037*
	CNN	*10.24*	*0.000255*	*30.14*	*$3.69e^{-6}$*	*14.71*	*0.00062*
	LSTM	*4.79*	*0.00664*	*31.06*	*$2.98e^{-6}$*	*20.12*	*$3.24e^{-5}$*
Specificity	NN	*4.66*	*0.03311*	*36.47*	*$1.72e^{-6}$*	*12.40*	*0.00044*
	CNN	*7.47*	*0.000899*	*29.13*	*$4.22e^{-6}$*	*13.57*	*$2.65e^{-9}$*
	LSTM	*10.42*	*0.000746*	*28.89*	*$2.83e^{-6}$*	*15.69*	*$2.03e^{-8}$*
F1-score	NN	*3.20*	*0.03791*	*33.11*	*$1.54e^{-6}$*	*12.28*	*0.00019*
	CNN	*17.02*	*$2.30e^{-5}$*	*18.62*	*$1.69e^{-5}$*	*14.39*	*$4.19e^{-5}$*
	LSTM	*3.65*	*0.02919*	*34.25*	*$3.17e^{-6}$*	*19.04*	*$2.87e^{-5}$*

3.3.4.5. Comparison with Other Bio-inspired Techniques for Feature Selection

This section compares the feature selection performance of the White Shark Optimization (WSO) technique against the classical Particle Swarm Optimization (PSO) method across multiple datasets. A thorough examination of hyperparameters associated with PSO, such as swarm size, number of iterations,

and cognitive and social component weights provides a foundation for understanding the algorithm's operational framework.

Table 3.16 Hyperparameter values of PSO for feature selection

Hyper-parameters	Description	PSO
Population_size	Swarm size in PSO; Bat population size	50
Max_itr	Number of iterations	100
C_1	Cognitive component weight	2
C_2	Social component weight	2
ω	Inertia	0.7
Velocity	Velocity update	As per equation (1)
Position	Position update	As per equation (2)

Subsequently, we present the results of both techniques in Table 3.17, highlighting key performance metrics including accuracy, precision, sensitivity, specificity, and F1-score. The findings reveal the efficacy of WSO in enhancing feature selection capabilities, ultimately demonstrating its superiority over PSO in optimizing classification performance across diverse datasets.

Table 3.17 Comparative representation of performance parameters by using PSO and WSO for feature selection across three datasets

Technique	Dataset	Accuracy	Precision	Sensitivity	Specificity	F1-score
PSO	Dataset 1: $\mathbf{D_T}$	0.936	0.932	0.928	0.907	0.931
WSO		**0.976**	0.971	0.981	0.971	0.975
PSO	Dataset 2: $\mathbf{D_C}$	0.902	0.916	0.915	0.911	0.914
WSO		**0.964**	0.965	0.965	0.964	0.964
PSO	Dataset 3: $\mathbf{D_A}$	0.914	0.907	0.901	0.926	0.911
WSO		**0.962**	0.945	0.985	0.942	0.964

3.3.4.6. Ablation Study

This section presents an ablation study to validate the contribution of the proposed WS-BiTM architecture. The WS-BiTM is trained using three screening datasets: 'Toddlers', 'Adults', and 'Children'. The work evaluates the performance across the cases mentioned below:

Case 1: Without WSO for feature selection (randomly selected features)

Case 2: With using only LSTM for classification

Case 3: Concurrently using WSO and Bi-LSTM (proposed work)

Table 3.18 Ablation outcomes with the proposed WS-BiTM architecture. The architecture is trained via three datasets. Case 1 incorporates scores obtained without WSO. Case 2 studies the incorporation of using only LSTM for classification. Case 3 involves the concurrent use of WSO and Bi-LSTM.

Dataset	**Ablation Study**	**Case**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-score**
Toddlers	Without WSO for feature selection	1	0.902	0.90	0.912	0.911	0.913
	With using only LSTM for classification	2	0.933	0.932	0.931	0.932	0.925
	Concurrently using WSO and Bi-LSTM	3	***0.976***	***0.971***	***0.981***	***0.971***	***0.975***
Adult	Without WSO for feature selection	1	0.904	0.903	0.911	0.912	0.914
	With using only LSTM for classification	2	0.938	0.930	0.932	0.932	0.935
	Concurrently using WSO and Bi-LSTM	3	***0.964***	***0.965***	***0.965***	***0.964***	***0.964***
Children	Without WSO for feature selection	1	0.904	0.912	0.902	0.901	0.901
	With using only LSTM for classification	2	0.938	0.930	0.932	0.932	0.935
	Concurrently using WSO and Bi-LSTM	3	***0.962***	***0.945***	***0.985***	***0.942***	***0.964***

The findings from the experimental simulations highlight the significant contribution of the WS-BiTM network in enhancing the reliability and accuracy of ASD diagnosis. The obtained results lay the foundation for future research advancements and development in the domain of ASD.

3.3.4.7. Leave-One-Dataset-Out (LODO) Cross Validation

To ensure the generalization of the proposed WS-BiTM model across different populations and verify that the model does not overfit a particular dataset, we performed Leave-One-Dataset-Out cross-validation (LODO-CV). This approach allows us to evaluate the model's performance when trained on two datasets and tested on a third, independent dataset, thus simulating real-world scenarios where the model is applied to unseen data. We conducted the LODO-CV using three datasets as mentioned in the above sections i.e., Toddler, Adult, and Children. For each experiment, the model was trained on two datasets and evaluated on the left-out dataset. This procedure was repeated until each dataset had been used as a testing set. We have constructed three cases based on our datasets as mentioned below.

Case 1: Train on Toddler and Adult, Test on Children: In this case, the WS-BiTM model was trained on the combined Toddler and Adult datasets, which consisted of a total of 1758 instances. The model was then evaluated on the Children dataset, which contained 292. The results showed that the model achieved an accuracy of 97% on the Children dataset, indicating its ability to generalize to this population.

Case 2: Train on Toddler and Children, Test on Adult: The second experiment involved training the WS-BiTM model on the combined Toddler and Children datasets, with a total of 1346 instances. The Adult dataset, containing 704 instances, was used for testing. In this configuration, the model achieved an accuracy of 95.9%, demonstrating good generalization to the adult population.

Case 3: Train on Adult and Children, Test on Toddler: For the final experiment, the model was trained on the combined Adult and Children datasets, comprising a total of 996 instances, and tested on the Toddler dataset, with 1054 instances. The performance metrics on the Toddler dataset showed an accuracy of 93.5%, indicating fair model performance even when tested on a larger population.

Table 3.19 Leave-One-Dataset-Out cross-validation

Training Datasets			**Test Dataset**	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
Toddler	**Adult**	**Children**						
✓	✓		Children	0.97	0.968	0.967	0.967	0.969
✓		✓	Adult	0.959	0.957	0.954	0.957	0.956
	✓	✓	Toddlers	0.935	0.933	0.936	0.934	0.934

The results of the LODO-CV experiments are summarized in Table 3.19. Across all three configurations, the WS-BiTM model demonstrated consistent performance in two cases. In case 3, there is a drop in accuracy when compared with other cases due to the larger number of test instances and lower training instances. Fig 3.15 shows the graphical illustration of the LODO-CV in these cases. Cross-dataset

validation confirmed that the model effectively generalizes across different populations and is not overfitting to any particular dataset.

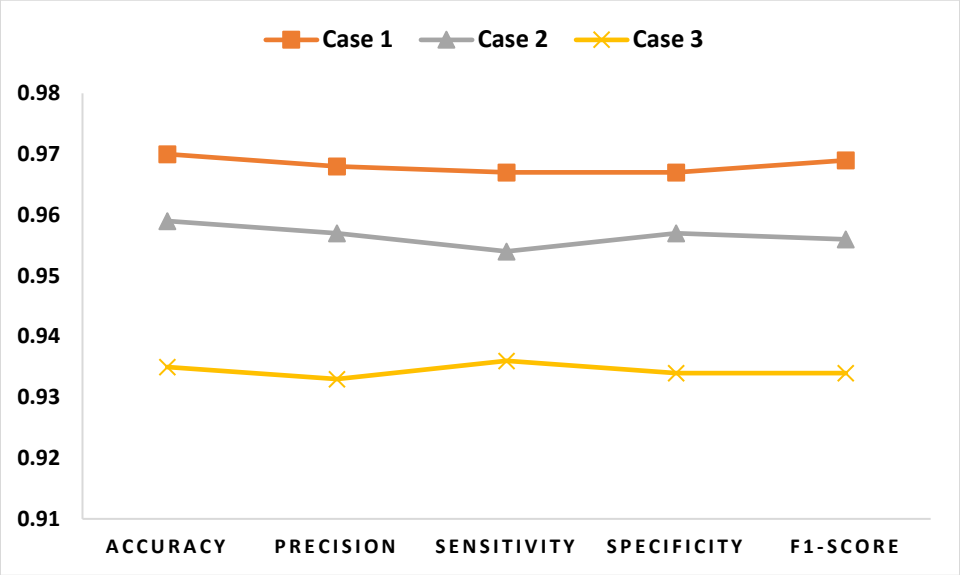


Fig 3.15 Graphical representation of the LODO-CV across three cases

3.4. Chapter Summary

This chapter emphasizes the integration of machine learning, nature-inspired algorithms, and deep learning techniques to enhance the diagnostic process for ASD. Traditional methods often face challenges such as insufficient feature representation and a lack of semantic understanding, limiting their effectiveness. To overcome these challenges, this research introduces two novel approaches: AFF-BPL (Adaptive Feature Fusion with Bat-PSO-LSTM) and WS-BiTM (White Shark-BiLSTM-based network). Both methods demonstrate significant advancements in ASD diagnosis. The AFF-BPL model employs adaptive feature fusion to integrate bio-inspired optimization with deep learning, achieving exceptional diagnostic accuracy across three distinct autism datasets (Toddlers, Children, and Adults). Experimental results highlight its superior performance, attaining accuracies of 0.992, 0.989, and 0.986, respectively, on the toddler, children, and adult datasets. Similarly, the WS-BiTM model combines White Shark optimization with BiLSTM to address the complexities of ASD diagnosis, achieving an accuracy of 97.6%, thereby outperforming baseline models such as NN, LSTM, and CNN. The findings underscore the potential of combining bio-inspired techniques with deep learning methodologies to enhance feature representation. By harnessing the synergy of computational intelligence and machine learning, this work anticipates significant advancements in the efficacy and accuracy of ASD diagnosis, contributing to improved healthcare outcomes.

Chapter 4 ASD DIAGNOSIS USING MULTI-MODALITY ARCHITECTURE

The diagnosis of Autism Spectrum Disorder (ASD) has remained a challenging task due to the complex and heterogeneous nature of the disorder. In recent years, the integration of multi-modal data sources, such as neuroimaging and clinical features, has shown promising potential for improving diagnostic accuracy. This chapter presents the development and evaluation of a novel multi-modality architecture for ASD diagnosis, leveraging the complementary strengths of structural MRI data and non-imaging meta-features. By combining advanced computational techniques, such as convolutional neural networks (CNNs), attention mechanisms, and transformer-based models, the proposed framework provides a robust and interpretable solution for ASD detection. The chapter elaborates on the design, implementation, and performance of the architecture, highlighting its ability to capture nuanced patterns from diverse data modalities and its superior performance compared to existing approaches. Section 4.1 provides an overview of the chapter, introducing the role of multi-modality architectures in ASD diagnosis. Section 4.2 discusses the use of CNNs for image analysis, while Section 4.3 presents the proposed MCBERT architecture, integrating CNNs, attention mechanisms, and BERT for enhanced diagnosis. Section 4.4 outlines the experiments and results, including dataset details, evaluation metrics, comparative analysis, and generalization testing. Section 4.5 examines the computational complexity of the approach. Section 4.6 provides a discussion of the findings and their significance. Finally, Section 4.7 summarizes the chapter's contributions and insights.

4.1. Overview

Within the domain of neurodevelopmental disorders, autism spectrum disorder (ASD) emerges as a distinctive neurological condition characterized by multifaceted challenges. The delayed identification of ASD poses a considerable hurdle in effectively managing its impact and mitigating its severity. Addressing these complexities requires a nuanced understanding of data modalities and the underlying patterns. Existing studies have focused on a single data modality for ASD diagnosis. Recently, there has been a significant shift towards multimodal architectures with deep learning strategies due to their ability to handle and incorporate complex data modalities. In recent developments, non-invasive brain imaging has provided a more comprehensive understanding of the neural circuitry linked with neural developmental disorders [121]. Notably, fMRI enables the visual evaluation of the functional characteristics of the brain. This offers precise insights into various neurological disorders [122]. For example, in the diagnosis of ASD, rather than solely depending on observational methods and patient interactions, physicians leverage neuroimages to detect anomalies in brain activity. This approach

enhances the efficiency and precision of identifying differences in neural pathways among patients. Developing an architecture that takes both images and its meta-data (multi-modal) into account, can enhance the efficacy and association between both modalities. This multi-modal strategy can prove to be effective in diagnosing ASD.

There has been a substantial accumulation of non-imaging (meta-features) datasets. Elements like gender, behavioral characteristics, patient history, and genetic sequences significantly influence disease diagnosis. The integration of non-imaging and imaging data through multimodal architecture is crucial for enhancing the efficacy of algorithms. Nonetheless, non-imaging/meta-features exhibit high dimensionality, constraining the representational capabilities of conventional machine-learning approaches [123]. Deep learning strategies present an avenue for efficiently amalgamating multimodal data to facilitate the diagnosis of autism spectrum disorder (ASD). The research framework built in this paper primarily concentrates on adopting deep-learning image and text-processing techniques to establish a multi-modal framework for diagnosing ASD. This is achieved by employing various techniques and fusing their outputs at the end. In the developed architecture, we introduced blocks/ components for image and meta-features modalities to diagnose ASD.

4.2. Convolution Neural Network (CNN)

In contemporary deep learning for image recognition/classification, CNNs stand out as a prominent neural network architecture. The architecture of CNN is structured into three layers: (i) the entry layer, (ii) the hidden (latent) layer, and (iii) the output layer. The hidden layers, alternatively termed pooling, or completely connected layers, play a major role in the overall architecture [29] [116].

The Convolutional Layer: The convolutional strategy is applied recurrently within this layer to induce changes in the output function. Comprising the neuronal maps, also known as the “filter/feature maps” or “characteristic maps”, the discrete convolution of receptors quantifies neural activity (Figure 4.1, block A). This process involves computing the overall neural weights of the input and activation function assignments [40]. Figure 4.1 provides a visual representation of a generic discrete convolutional layer.

Max Pool Layer: The max pool layer forms a multitude of grids from the segmented convolutional layer’s output. Sequential matrices are created using the maximum grid value. Operators are employed to derive the average or maximum value for each matrix. Figure 1, block B, illustrates the construction of the max pool layer [50].

Fully Connected Layer (FCL): Constituting 90% of the entirety of the structural elements of the CNN, the FCL allows the transmission of the input across the network with a pre-configured vector length. Data is

transformed within this portion before grading. The convolutional layer is also transformed to conserve information integrity. Neurons from every preceding layer are utilized in these FCLs, serving as the network's ultimate layer [124].

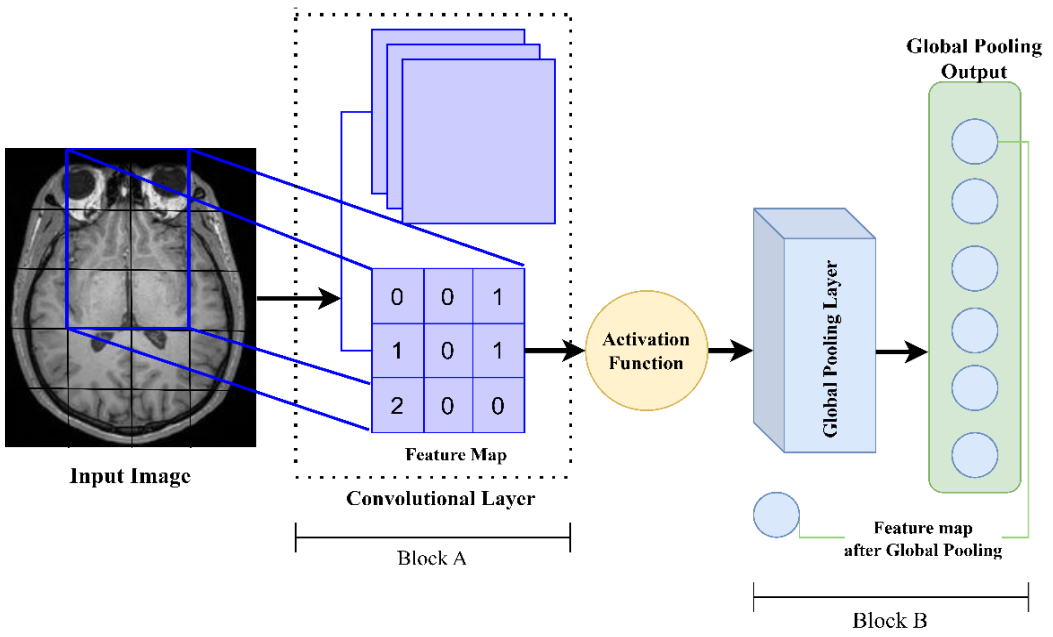


Figure 4.1 Visual representation of the workflow of generic convolution neural network

4.3. Proposed Architecture

4.3.1. Multi-Head CNN

In this research paper, we introduced a three-headed convolutional neural network specifically crafted to extract pertinent patterns from input images. The convolutional layer comprises multiple convolutional filters that, through convolution operations, generate the output feature map (mainly explained in the above section) from input images. Within the convolutional layers, the obtained feature maps via the preceding layer undergo convolution via various kernels [125]. Additionally, bias is incorporated to augment the outcome of the convolution operation, which subsequently passes via an activation function, giving rise to the feature maps for the subsequent layers. Mathematically, the mth feature map at the lth layer of the eth head of the multi-head CNN is represented as a matrix, with the value at the kth row denoted as $R_{lm}^{k,e}$. The calculation of this value follows the formula presented in equation (1).

$$R_{lm}^{k,e} = f_{Relu}(f_{conv2d}^{e}(R_{l-1}^{k+j})), \quad \forall e = 1,2,3 \tag{1}$$

Here, f_{Relu} denotes the activation function that replaces all negative values with 0 (zero) in the feature map, while f^{e}_{conv2d} represents the convolution function of the eth head in our multi-head CNNs, s articulated in equation (2).

$$f^{e}_{conv2d}(R^{k+j}_{l-1}) = b_{lm} + \sum_{i} \sum_{j=0}^{\eta^{e}_{l}-1} W^{je}_{lmi}\, R^{k+j,e}_{(l-1)} \tag{2}$$

b_{lm} represents the bias for a particular feature map, where i is the index of the feature map at the ($l-1$) layer. Additionally, W^{je}_{lmi}signifies the weight matrix present at the position j of the convolution kernels, and η^{e}_{l} represents the length of the kernel of the eth head in our multi-head CNN. A crucial element in the developed multi-head CNN architecture is the pooling layer. This layer plays a pivotal role in reducing the parameter count and computations by decreasing the spatial size of the feature representation. Among the various pooling techniques, max pooling is the most popular and widely utilized method.

$$P_{hlm} = max_{(y,\ z)\ \in\ \mathbb{R}_{l,m}}\ v_{hyz} \tag{3}$$

Here P_{hlm} represents the pool operation of the hth feature maps. v_{hyz} signifies the component at position (y, z) enclosed by the pool region $\mathbb{R}_{l,m}$. This region shows a receptive field around (l, m).

4.3.2. Convolution Block Attention Component (CBAC)

In the developed architecture, each head incorporates two CBACs to optimize training performance by accentuating both spatial and channel features with brain MRI images. The CAC network empowers the MCBERT framework to concentrate on crucial channel features while disregarding others. These channel features contain intricacies intrinsic to the individual color or feature channels, delineating distinct aspects such as textual nuances and color variations. This channel-level scrutiny is necessary for capturing fine-grain details by facilitating a comprehensive characterization of image content. To assess the importance of every channel, diverse weight information is utilized to various feature channels and feature dimensions of the visual data. SAC enables the architecture (Figure 4.2, block A) to prioritize spatial dimension information on the feature map. The features encapsulate the spatial relationships, structural configurations, and overall layout of the image, presenting a holistic perspective on the contextual arrangement of visual elements. The analysis of spatial features is pivotal for decoding the spatial semantics and intrinsic geometry embedded within the visual data. For feature extraction, CBAC sequentially extracts a 1-D channel attention map $A_C \in \mathbb{R}^{c\times 1\times 1}$ and a 2-D spatial attention map $A_s \in \mathbb{R}^{1\times H\times W}$ from the provided intermediate feature map $I \in \mathbb{R}^{c\times H\times W}$ of the MRI visual data. The comprehensive attention mechanism is articulated in equations (4) and (5).

$$I' = A_C(I) \otimes I \quad (4)$$

$$I'' = A_s(I) \otimes I' \quad (5)$$

In this context, the symbol $\otimes$ denotes element-wise multiplication, producing the refined feature I''. The channel attention features undergo compression along the spatial dimension, and reciprocally. The CAC network, illustrated in Figure 4.3, augments the significance of relevant information while diminishing the weight of unnecessary details in the feature channel. Consequently, the developed module accentuates channels within the MRI images.

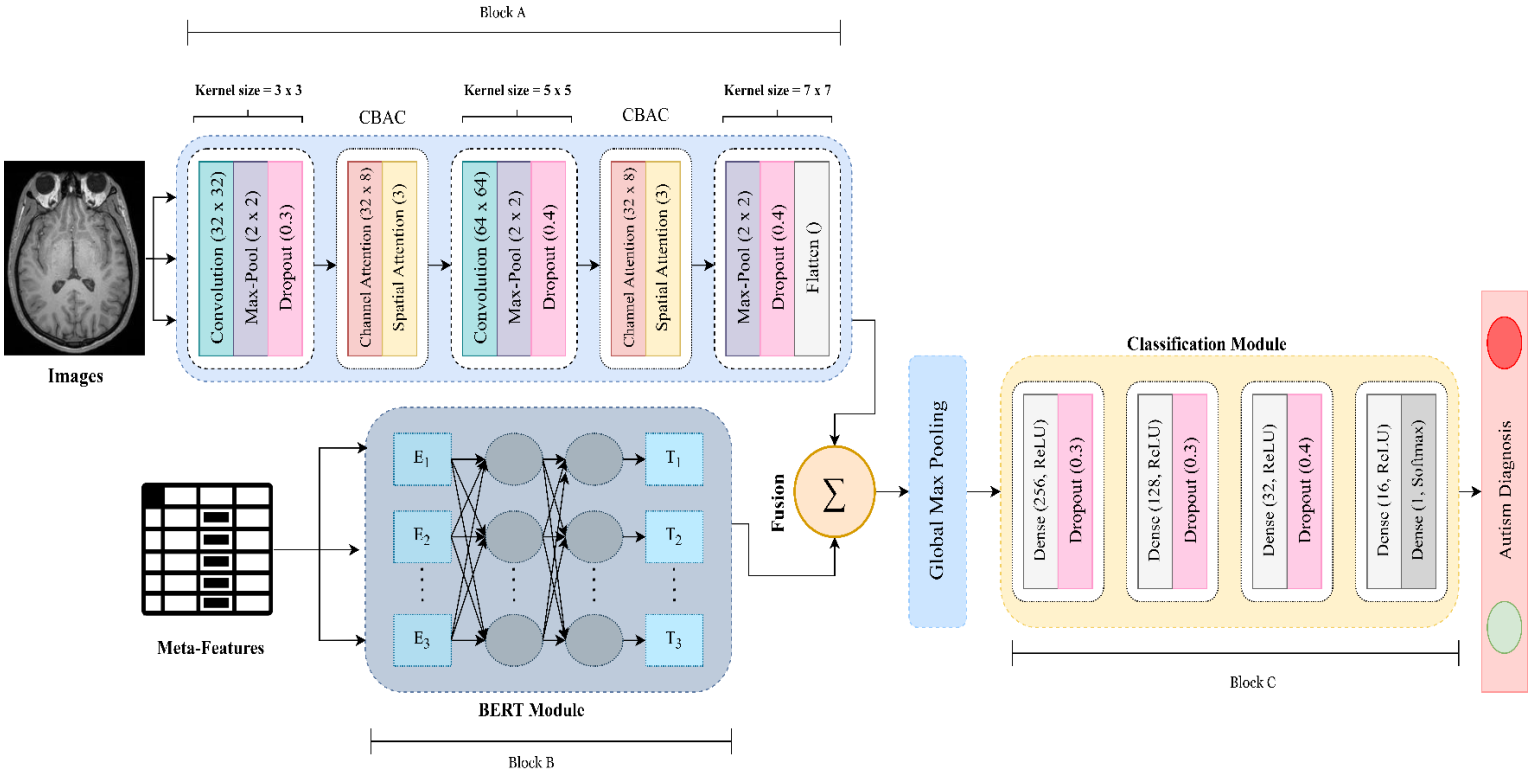


Figure 4.2 Multimodal architecture of MCBERT incorporating convolutional layers with channel block attention component (Block A) for image modality, a BERT module (Block B) for the meta-features, fusing the output of the block A and block B, and passing it through global max pooling and the final classification module (Block C) to diagnose ASD.

Within the CAC, average-pooled patterns and max-pooled attributes are extracted (as in Figure 3) from the aggregated feature map, employing both average-pooling and max-pooling operations on spatial information [126][127]. These high-level patterns undergo processing in a shared multi-layer perceptron (MLP) model, featuring a hidden layer. The outcome of the shared network traverses a pipeline involving additional max-pooling and average-pooling operations, coupled with a non-linear activation function

(ReLU) [128], to generate the channel attention map $A_C \epsilon \mathbb{R}^{c\times1\times1}$. The utilization of two pooling operations enhances the extraction of high-level features [129]. The mathematical calculation of channel attention is expressed in equation (6), where σ represents the sigmoid function.

$$A_C(I) = \sigma\left(MLP\left(AvgPool(I)\right)\right) + \left(MLP\left(MaxPool(I)\right)\right) \quad (6)$$

$$A_s(I') = \sigma\left(f^{7\times7}\left([AvgPool(I');\ MaxPool(I')]\right)\right) \quad (7)$$

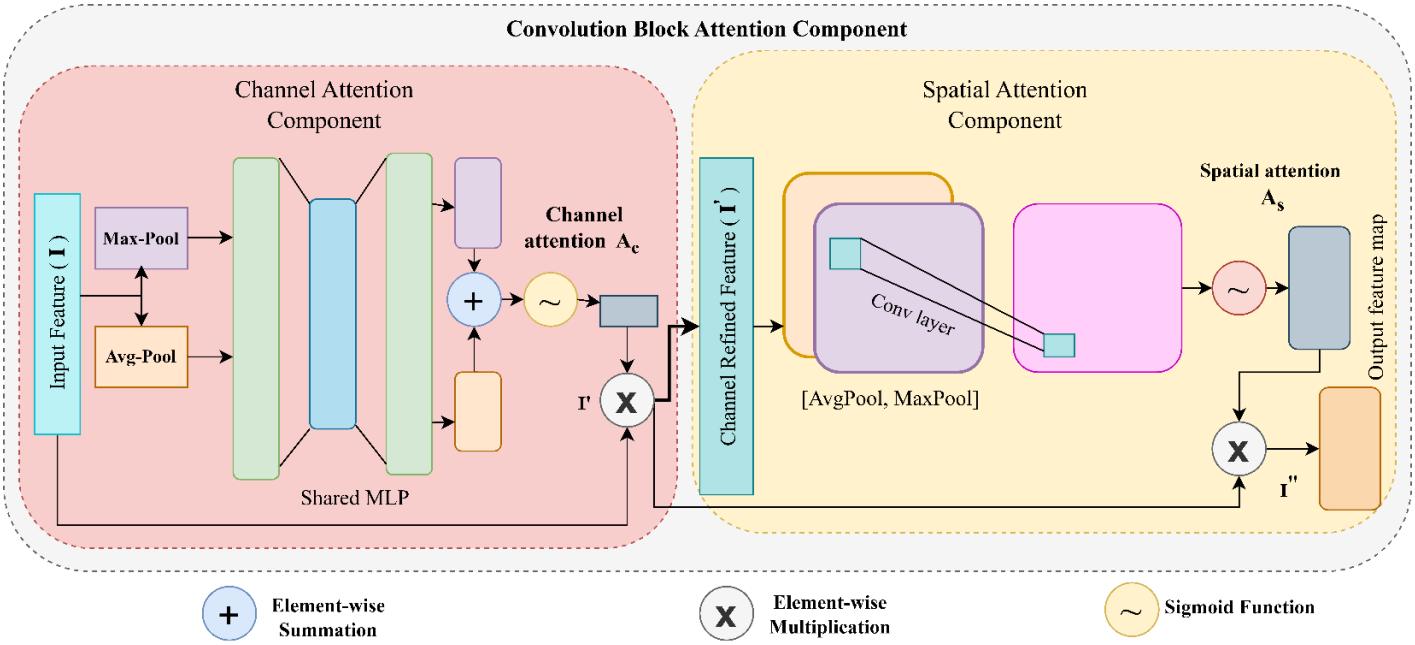


Figure 4.3 Detailed architecture of convolution block attention component (CBAC) incorporating the visual representation of channel attention component (CAC) and spatial attention component (SAC)

Moving to the SAC network, depicted in Figure 4.3, enhances the spatial dimension features in the feature map through feature filtering on pixels at different positions within the same spatial dimension, assigning weights to significant features. SAC executes average-pooling and max-pooling operations on the feature map I' along the channel dimension, producing two feature maps that are subsequently fused and convolved by a 7×7 kernel size. This convolution operation yields the final spatial attention map $A_s \epsilon \mathbb{R}^{1\times H\times W}$. Where 7×7 denotes the convolution operation with a filter size of 7×7. The mathematical formulation of SAC is provided in equation (7). The extracted feature maps from the multi-head CNN, refined through CBAC, are then combined with patient meta-features for further analysis.

4.3.3. Bidirectional Encoder Representations from Transformers

BERT or Bidirectional encoder representations from transformers stand out as an efficient and revolutionary model for feature extraction in various tasks. For the meta-features (patient information),

we employed the BERT model, a pre-trained language representation model (Figure 2, Block B). The employed BERT model transforms the input patient data into vector representations. These vectors capture both inter-feature relationships and sentence-level features from the patient information. Its primary function is to transform input into vectors [130]. In contrast to conventional language pre-training models, BERT incorporates two tasks for model pre-training. Consequently, the word vectors produced by BERT not only convey inter-word features but also encompass features at the sentence level [131]. The pivotal component in BERT is the Bi-transformer, utilizing a self-attention mechanism and fully connected (FCL) layer to model input, diverging from the use of recurrent neural networks and CNN for feature extraction. The self-attention mechanism, paramount in transformers, computes relationships between the data, adjusting the weight of importance based on these relationships. Thus, each word's vector not only signifies its meaning but also provides insights into relationships with other features [132]. The BERT module's output is then used as input for the multi-head self-attention mechanism, as described in Equation (8). The computational process is depicted in equation (8), where Q denotes the query vector, K is for the representation vector, the value vector is V, and the input vector dimension is denoted by d_k. Here, the input vectors are derived from the meta-features encoded by BERT, and the attention mechanism computes the relationships between the features, adjusting their weights based on their relevance.

$$Attention(Q,K,V) = softmax\left(\frac{Qk^T}{\sqrt{d_k}}\right)V \quad (8)$$

The transformer also incorporates a multi-attention mechanism as the self-attention mechanism alone is limited to capturing information in a single dimension. Initially, the vectors Q, K, and V undergo linear mapping h times. Finally, the resulting attention matrices are concatenated, enabling the acquisition of multi-dimensional information. The formula describing the process is as follows:

$$Multihead(V,Q,K) = concat(\ head_1,\ head_h\ ) \quad (9)$$

$$head_i\ = attention(\ QW_i^Q, KW_i^K, VW_i^V) \quad (10)$$

4.3.4.Pre-processing:

To address the distinct statistical properties in our multimodal data during training, we adopted a standardization and normalization approach for meta-features (non-imaging data). Specifically, dictionaries are crafted from the data, encompassing age, site, and gender information for each sample. Age values fall within the range of (6,64), while the 17 sites are encoded as (0, 1, ……,15, 16). And gender is represented as (0,1). A normalization process is applied to both sites and ages, transforming their values to lie within the standardized interval of (0, 1). This meticulous preprocessing ensures that the

non-imaging input data is appropriately rescaled and ready for integration with the multi-modal data, promoting improved convergence and effectiveness during the training process.

```
Function normalize_meta_features(u):
   Create an empty dictionary for normalized data
      Normalized_data = {}
   For data_type in ['site', 'age', 'gender']:
      Extract the data for the current_type
         X = u[data_type]
   Normalize the data to the range [0, 1]
         X_normalized = (X - min (X)) / (max (X) – min (X))
   Standardization of the data
         X_standardized = (X – np.mean (X)) / np.std (X)
   Add the normalized data to the dictionary
         Normalized_data [data_type] = X_normalized

Return normalized_data
```

Due to the inherent limitation of the self-attention mechanism in capturing the sequential order of input, BERT introduces position embedding and segment embedding to discern between adjacent sentences. Within the BERT framework, each input variable in its input sequence is derived through the summation of a word vector, a position vector, and a segment vector. The ultimate word vector is produced via a process of deep bidirectional coding, following which it is sent into the classification module mentioned in the section below.

4.3.5. Classification Module

The combined output vectors generated by the multi-head CNN (Block A) and the BERT module are integrated and fed into the classification module for the diagnosis of ASD through the utilization of global max pooling (GMP). The GMP layer efficiently selects the most salient features and produces a feature map for both the target classes, contributing to the reduction of trainable parameters. Following this, fully connected layers are employed, incorporating neurons with ReLU activation function, specifically 256, 128, 32, and 16 neurons in each layer. To address potential overfitting, dropout layers are strategically

inserted in conjunction with these fully connected (FCL) layers. The final stage involves applying the softmax activation function to calculate the class score for both target classes, determining the correct diagnosis result with the high probability score. The representation of the softmax function during diagnosis is formulated in equations (11) and (12). Where ϕ denotes the output features from the preceding FCLs. During training, the cross-entropy loss function ($\mathcal{L}$) is employed to minimize the loss value, as depicted in equation (13). Here, y_i signifies the actual classes, and $\hat{y}_i$ indicates the outcomes through the developed architecture.

$$P = softmax(\phi) = \frac{\exp(\phi)}{f_0^1 \exp(\phi)} \quad (11)$$

$$\hat{y} = argmax(P) \quad (12)$$

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log (1 - \hat{y}_i)] \quad (13)$$

N signify the data samples.

4.4. Experiments and Results

In this segment, we explained the experimental outcomes obtained from the multimodal autism spectrum disorder (ASD) diagnosis architecture to present the efficacy of the developed architecture. This segment incorporates sub-sections giving brief descriptions of the experimental configuration, the ASD multimodal dataset employed, performance evaluation metrics considered, the quantitative analysis of results, and the leave-one-site-out-classification test. Furthermore, we conducted a comparative analysis to contrast the findings of our work with various existing state-of-the-art approaches.

4.4.1.Experimental Configuration

All experiments in this study were conducted on a laptop with an Intel Core i5 10th Generation processor, 8GB of RAM, 512GB of storage, and running the Windows 11 operating system. The system was also equipped with an NVIDIA GTX 1650 graphics card with 4GB of VRAM, which was utilized to enhance computational performance, particularly during model training. The experiments were implemented using Python, and several libraries were employed for data analysis and model development. Numpy was used for numerical computations and matrix operations, while Pandas handled data manipulation tasks, including loading and preprocessing datasets. For visualization, Matplotlib and Seaborn were used to plot training results and statistical graphics, respectively. Scikit-learn was applied for model evaluation and computation of performance metrics. These tools and frameworks formed the core of the experimental setup and were integral to the development and evaluation of the proposed model.

4.4.2.Dataset Description

Our research conducted experiments on ABIDE-I, a publicly available data repository. This multi-modal data is gathered from 1112 participants across 17 sites worldwide [133]. This multimodal dataset comprises (a) MRI scans and comprehensive (b) phenotypic information (as in Table 5) for each subject. These phenotypic measures included demographic information (age, gender), and count of autism spectrum disorder (ASD) to TD participants. We selected these specific measures as they are clinically relevant to understanding ASD, capturing all dimensions essential for a comprehensive analysis. Table 3 represents the phenotypic measure summary of the employed ABIDE dataset. The inclusion of these phenotypic measures complements the imaging data, allowing us to address the heterogeneity observed in ASD. This multimodal approach aligns with our objective of developing a model that integrates both neuroimaging and non-imaging data for better diagnostic accuracy. Our study focuses on resting-state structural MRI (rs- MRI) scans for the imaging part. To maintain data quality and maintain methodological comparability, we meticulously excluded data with missing series (non-imaging part), incomplete brain coverage, and other scanning artifacts. Our analysis ultimately focused on 875 participants, including 403 participants diagnosed with autism spectrum disorder (ASD) and 472 typically developed (TD). Table 6 describes the dataset description with a training and test split ratio of 80:20.

Table 4.1 Phenotypic measure summary of the ABIDE-I dataset

Site	ASD	TD	Male count	Female count	Average age
CMU	14	13	21	6	26
Caltech	19	18	29	8	27
Leuven	29	34	55	8	18
KKI	20	28	36	12	10
NYU	75	100	139	36	15
MaxMun	24	28	48	4	25
OLIN	19	15	29	5	16
OHSU	12	14	26	0	10
SBL	15	15	30	0	34
PITT	29	27	48	8	18
Stanford	19	20	31	8	9
SDSU	14	22	29	7	14
UCLA	54	44	86	12	13
Trinity	22	25	47	0	16
USM	46	25	71	0	22

UM	66	74	113	27	14
Yale	28	28	40	16	12

Table 4.2 Dataset description of the employed ABIDE-I dataset

Images with ASD	14,105		
		Total = 30,625	**Training: Testing (80:20)**
Images of typically developed (without ASD)	16,520		24,500: 6,125

4.4.3. Performance Evaluation Metrics

The efficacy of the MCBERT model is evaluated using the three primary metrics, namely sensitivity, accuracy, and specificity. Table 7 shows the metrics and their respective formulas.

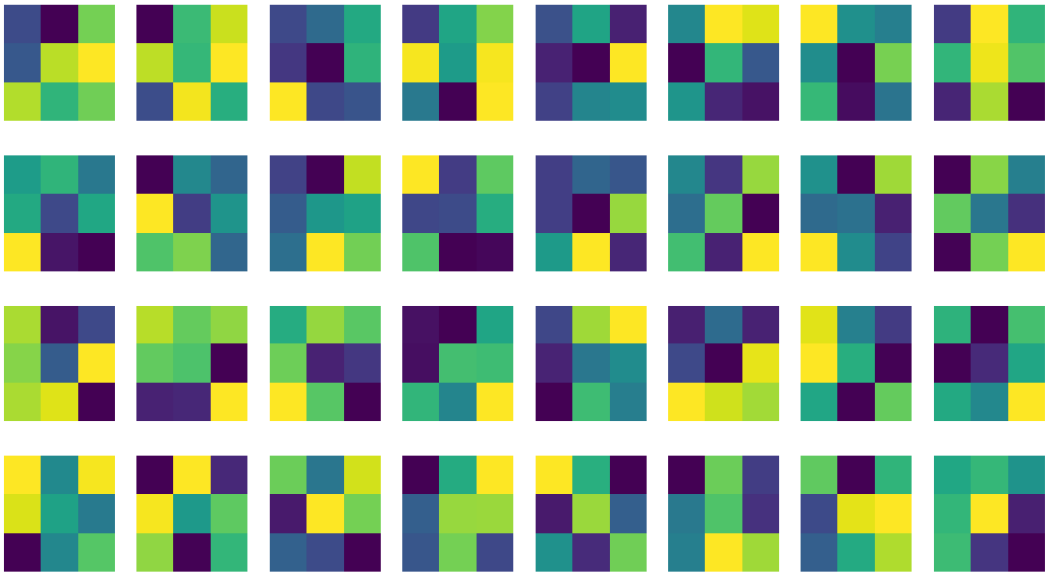
Table 4.3 Key classification metrics employed to evaluate the proposed work

Performance Metric	Formula	Value Range	Cases Assumed
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	[0,1]	***TP***: Autistic individuals identified as autistic individuals. ***TN***: Non-autistic/Healthy individuals identified as non-autistic ***FP***: Non-autistic/Healthy individuals identified as autistic ***FN***: Autistic individuals identified as non-autistic
Sensitivity (Recall)	$\frac{TP}{TP + FN}$		
Specificity	$\frac{TN}{TN + FP}$		

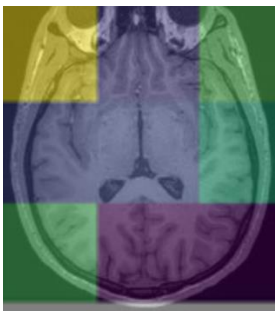
4.4.4. Result Analysis

The performance of our developed approach, which leverages BERT for meta-feature extraction and a multi-head CNN for image feature extraction, followed by the fusion of their outputs and passing them to

a classification module, yielded promising results. Through the methodology, we aimed to enhance the diagnostic process by incorporating rich contextual embeddings from the BERT module and extracting spatial and channel-specific features through the multi-head CNN. We conducted experiments for approximately 100 epochs to both train and assess the performance of the developed MCBERT architecture. The use of BERT for meta-feature extraction allowed the model to capture complex semantic relationships between patient metadata and brain MRI data, which proved beneficial in enhancing diagnostic accuracy. The output from the multi-head CNN is illustrated in Figures 4.4 and 4.5.



(a)



(b)

Figure 4.4 (a), (b). The output obtained via the activation pattern learned in the initial convolutional layer when the brain images are passed by block A

Figure 4.4 demonstrates the activation patterns captured by the initial convolutional layer of the multi-head CNN, which includes our Convolution Attention Block (CAB) module, during the processing of brain MRI images where each matrix represents the output from distinct filters of the initial convolutional

layer, visualizing how different regions of the brain MRI images are processed by the CNN. The inclusion of multiple matrices serves to illustrate the model's ability to focus on various diagnostic regions simultaneously, enabling it to capture complementary information for accurate classification. The activation maps are represented using a color gradient, where blue indicates low activation, green denotes medium activation, and yellow represents high activation levels. These color-coded maps show how the CNN's filters respond to specific regions of the input image. Early convolutional layers in CNNs generally focus on detecting fundamental image features, such as edges or anatomical structures. In Figure 4.4 (a) and (b), the high activations (yellow) in certain regions highlight key anatomical boundaries, edges, or structures that are likely to hold diagnostic significance, such as cortical boundaries or ventricles. These regions are critical for constructing hierarchical representations of the data as it passes through deeper layers of the network.

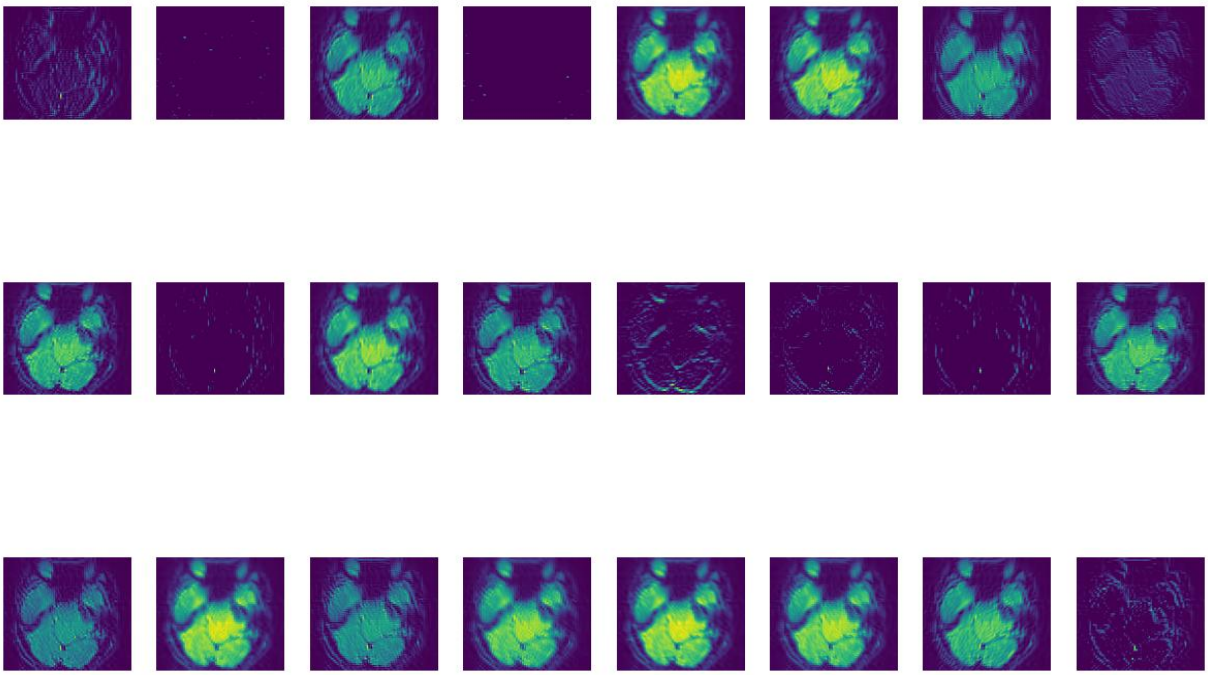


Figure 4.5 Visual representation of the feature maps/feature extraction capability of multi-head CNN when applied to brain images

Conversely, the medium (green) and low (blue) activation areas correspond to regions with less prominent or diagnostic features. Over approximately 100 epochs, the model refined its feature extraction process. Early in the training, activations in the initial layers are more generalized, but as training progresses, these patterns become more focused, allowing the network to identify the most informative features for the diagnostic task. The activation maps presented in Figure 4.5 provide visual evidence of this learning process, showing how the model transitions from emphasizing simple, low-level patterns, such as edges, to capturing more complex features that contribute to the improved classification accuracy

of the MCBERT architecture. This hierarchical feature extraction is typical of CNNs, where initial layers detect basic patterns, and deeper layers identify more abstract, high-level features. The observed activation patterns are crucial for understanding how the multi-head CNN processes spatial and channel-specific features from the MRI data. By activating different filters in response to specific brain structures or abnormalities, the network demonstrates its ability to capture spatial relationships between anatomical structures, which directly supports the effectiveness of our approach in improving diagnostic accuracy. Additionally, the presence of distinct activations in key regions of the MRI images suggests that the model is effectively identifying features associated with autism. Figure 4.5 further illustrates the contributions of each of the three heads in the multi-head CNN in the initial learning layers. Each head processes the MRI image through distinct convolutional paths, allowing the model to extract diverse, complementary features from different regions of the input, such as anatomical structures, textures, and abnormalities. After passing through convolutional and global max pooling layers, the extracted feature maps are reduced to fixed-size vectors, preserving critical information for classification. The diverse activation patterns across the three heads indicate that the multi-head CNN is capable of focusing on various aspects of the MRI images, which is essential for identifying subtle differences in brain structures that could be associated with ASD. These visualizations effectively demonstrate the hierarchical nature of feature extraction in the multi-head CNN, where early layers detect simple patterns like edges, while deeper layers capture more complex, high-level features.

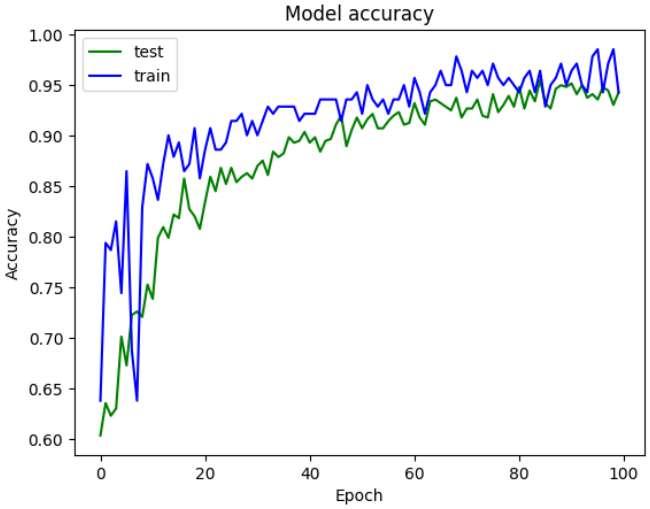


Figure 4.6 (a) Epoch vs Accuracy curve of MCBERT

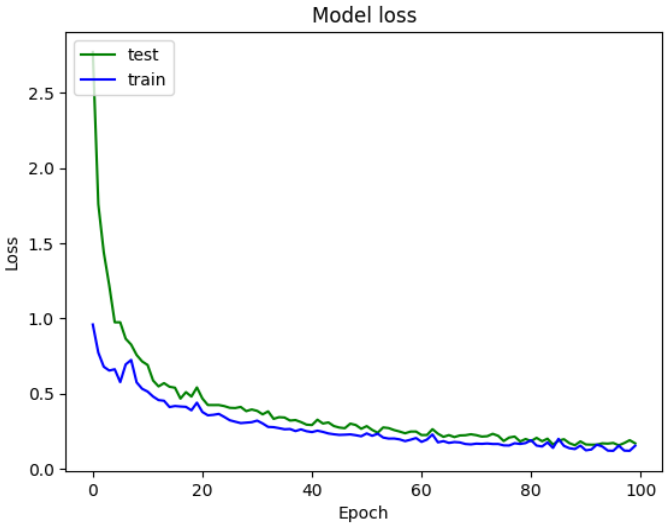


Figure 4.6 (b) Epoch vs Loss curve of MCBERT

The figure provides valuable insights into how the convolutional layers of the CAB module interact with the MRI data and highlights the model's ability to focus on important diagnostic features. This contributes to the overall success of the MCBERT architecture in accurately classifying MRI images, demonstrating the robustness of the proposed approach in improving neuroimaging-based diagnosis. Upon analyzing the experimental outcomes, we observed significant improvement in accuracy as compared to existing techniques (comparison with existing techniques is mentioned in the further sub-sections). Figure 4.6 (a) represents the accuracy vs epoch curve demonstrating a diagnosing accuracy of 93.4%. Figure 4.6 (b) represents the loss curve of the developed architecture.

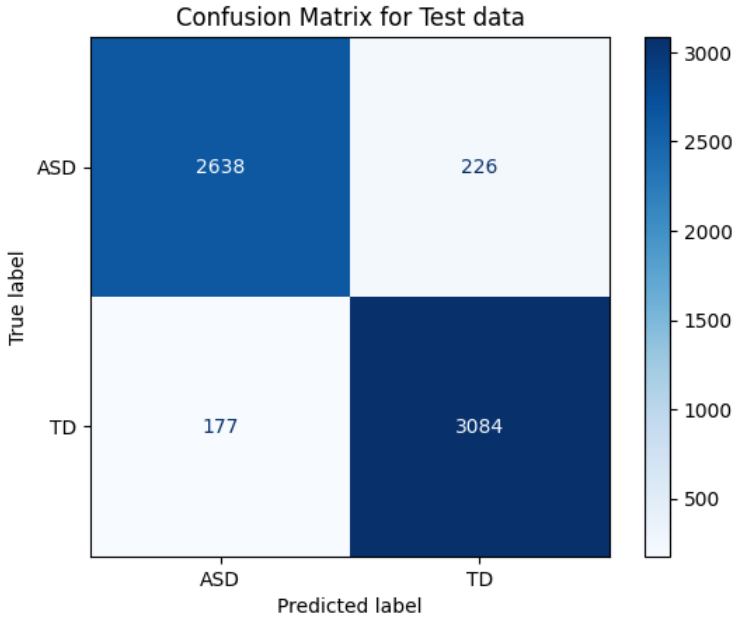


Figure 4.7 Confusion matrix obtained on the test set

Overall, our results demonstrate the effectiveness of the developed multi-modal MCBERT, a BERT, multi-head CNN, and its seamless integration with the classification module for the diagnosis of autism spectrum disorder. This approach not only showcases the power of utilizing pre-training architectures but also showcases the potential of deep learning in feature extraction and further advancement in the healthcare domain. Figure 4.7 illustrates the performance of the developed model through the confusion matrix obtained from the test data. The test set consists of 6,125 images.

4.4.5. Comparison with existing works

This section highlights a comparison of various methods used for autism spectrum disorder (ASD) diagnosis on the ABIDE dataset, focusing on the performance of the proposed MCBERT model against other state-of-the-art techniques. The quantitative results of this comparison are summarized in Table 8. [78] proposed an sMRI-based ASD detection framework using an ensemble of deep convolutional neural networks (DCNN) combined with different optimizers (Adam, Nadam, and RMSProp). After filtering out noisy slices, the study utilized raw sMRI scans from the ABIDE dataset without advanced preprocessing. The ensemble of optimizers aimed to enhance model performance by improving robustness. They tested the model using three data splits (70:30, 80:20, and 90:10) and achieved accuracies of 77.58%, 77.66%, and 81.35%, respectively. [26] proposed a two-stage adversarial learning model to address the challenges associated with multi-site ASD classification using resting-state functional magnetic resonance imaging (rs-fMRI). Their approach begins with the sliding window sampling technique, which preserves spatial and temporal information from rs-fMRI data. This is followed by an adversarial learning model that extracts site-shared features, effectively tackling the issue of site heterogeneity common in multi-site studies. The model is then fine-tuned to extract disease-related features, specific to ASD classification. The data used in this study was sourced from the ABIDE dataset. For model evaluation, the authors employed ten-fold cross-validation, with the dataset randomly split into training (81%), validation (9%), and test sets (10%). [53] developed a WL-DeepGCN framework that combines fMRI data and non-imaging demographic information for ASD diagnosis. The model uses a weight-learning network to define graph edge weights in the latent space, and residual connections in the GCN to avoid gradient issues. An edge-drop strategy reduces overfitting by sparsifying node connections. The study applied a nested 10-fold cross-validation on the ABIDE-I dataset to ensure robust evaluation, avoiding feature peeking and overfitting. Recursive feature elimination (RFE) was used for feature selection. [134] conducted a comprehensive review of different brain networks and their functional connectivity to distinguish between individuals with ASD and TD participants. The study utilized 871 rs-fMRI samples from the ABIDE repository. The authors employed bootstrap analysis of stable clusters (BASC) as the most predictive brain parcellation technique, aiming to find the optimal method for classifying ASD. The

methodology involved exploring eight different brain parcellation techniques, which included structural, functional, and data-driven approaches, to identify the best brain atlas for ASD classification. Additionally, three functional connectivity metrics, correlation, partial correlation, and tangent space, were evaluated to assess their stability and efficiency. The study found that the correlation metric was the most stable among the metrics. In terms of machine learning models, the paper compared four supervised learning algorithms: kernel Support Vector Machine (kSVM), which was identified as the optimal classifier for the task, outperforming others. The experiments used 5-fold cross-validation, repeated 10 times to ensure the reliability and stability of the results. [135] worked on an unimodal ASD identification architecture incorporating Inception V3 model with CNN. They worked on fMRI employed from the ABIDE dataset. They took three imaging features into account i.e., epi images, glass brain images, and stat map images. In another work by [63], they focused on developing multimodal ASD architecture incorporating transfer learning with deep ensemble learning. They developed a multimodal-multisite ensemble classifier to diagnose ASD from fMRI and phenotypic information from the ABIDE dataset. They tested their model on various parameters and presented a detailed analysis of their work. [136] focused on using a combination of functional and structural MRI data for the classification of ASD patients versus control participants. The key features used included functional connectivity patterns among brain regions from fMRI and volumetric correspondences of gray matter volumes from sMRI. Their classification network was built using stacked autoencoders trained in an unsupervised manner, combined with multilayer perceptrons (MLP) trained in a supervised manner. The study analyzed data from 817 cases in the ABIDE-I dataset, involving 368 ASD patients and 449 controls. The evaluation methodology involved 10-fold cross-validation, wherein each fold, 10% of the data was used for testing, while 90% was used for training and validation (split into 70% for training and 30% for validation). Additionally, they conducted leave-one-site-out cross-validation to assess the model's performance. This paradigm, alongside reporting of accuracy, sensitivity, and specificity, provided a thorough quantitative and qualitative comparison with other state-of-the-art methods.

Table 4.4 Comparison with existing works conducted for ASD on the ABIDE dataset

Reference	**Methodology**	**Dataset**	**Modalities Incorporated**	**Best accuracy**	**Sen**	**Spec**
[78]	Optimizer + Deep CNN	sMRI	1	77.58%	78.16%	76.99%
[137]	Graph NN + Ensemble technique	Phenotypic + HO	2	73.13%	76.00%	69.00%

[26]	Adversarial learning + LSTM	fMRI	1	80.00%	81.00%	80.00%
[53]	Weight learning + Graph CNN + Deep CNN	Phenotypic + HO	2	77.27%	80.96%	-
[135]	Inception V3	fMRI	1	98.35%	-	-
[63]	Inception V3 + ResNet50 + DenseNet + MobileNet	fMRI + Phenotypic	2	97.82%	-	-
[134]	Feature extraction via function connectivity matrix	fMRI	1	69.43%	64.57%	73.61%
[136]	MLP + Autoencoder	sMRI + CC200	2	85.06%	81.00%	89.00%
[77]	Correlation matrix + Graph Theory	fMRI	1	84.79%	89.63%	78.96%
MCBERT (Proposed)	***Multi-Head CNN + BERT***	***Phenotypic + sMRI***	***2***	***93.4%***	***92.1%***	***94.5%***

[77] utilized rs-fMRI data from the ABIDE-I dataset to propose an approach for diagnosing ASD. The study focuses on constructing functional connectivity networks from the rs-fMRI time-series data, calculating correlation matrices that represent interactions between brain regions. The ABIDE-I dataset, consisting of 1,112 individuals (539 ASD and 573 typically developing controls), served as the basis for the experiments. The authors tested 11 classification algorithms, including linear support vector machines (SVM) and 2D CNN, and identified these as the best-performing methods across all atlases. Additionally, the authors also performed stratified 10-fold and 3-fold cross-validation on the best classifiers (Linear SVM and 2D CNN), observing consistent accuracy across these methods. Most of the existing works in the literature have utilized deep learning frameworks. Some approaches focus on single-modality data, such as structural MRI (sMRI) or functional MRI (fMRI), while others combine multiple data modalities, like phenotypic information and neuroimaging data. Our proposed MCBERT model, which integrates a

Multi-Head CNN and BERT architecture, operates on multimodal inputs, specifically phenotypic data and sMRI. As shown in Table 4.4, the MCBERT model outperforms other methods in terms of accuracy, sensitivity, and specificity, which demonstrates its effectiveness in ASD diagnosis. The model's ability to handle both phenotypic and sMRI data contributes to its robust performance, leading to higher classification accuracy compared to the methods that rely on single-modality data.

4.4.6.Leave-one-site-out (LOSO) cross-validation test

In this study, the primary experimental paradigm utilized was leave-one-site-out (LOSO) cross-validation. This method was chosen to evaluate the generalization ability of the MCBERT model across different screening sites within the ABIDE-I dataset, which includes data from 17 different sites. For each LOSO iteration, one site was selected as the test set, while the remaining sites were split into training and validation sets. This setup allowed the model to be tested on unseen data from various sites, highlighting its adaptability to site-specific variations in the dataset. Each site/data was trained and tested under identical conditions, and performance metrics, including accuracy, specificity, sensitivity, and AUC, were recorded for each site. Table 4.5 presents the performance outcomes of the LOSO test, demonstrating the robustness of MCBERT in generalizing across different sites. The mean accuracy of MCBERT across all sites was determined to be 83.64%. Notably, four sites UM, STANFORD, PITT, and MAX_MUN exhibited lower performance compared to the mean values of the evaluated metrics. This observation underscores the presence of site-specific variability and a lack of homogeneity in the dataset. Despite these variations, the high global mean values attest to the effectiveness of the MCBERT architecture.

Table 4.5 Quantitative performance analysis of the MCBERT model on the LOSO test using the ABIDE-I dataset

Site	Accuracy	Specificity	Sensitivity	AUC
CMU	91.00	95.00	87.00	84.00
CALTECH	88.50	88.00	86.00	85.00
MAX_MUN	79.30	80.00	78.70	76.00
LEUVEN	85.00	86.00	84.01	87.00
KKI	86.08	84.03	91.03	90.03
OHSU	84.00	83.79	83.46	77.31
NYU	86.02	80.81	82.55	88.61
OLIN	89.50	87.62	91.52	89.01
SDSU	87.00	80.34	81.53	81.34
PITT	76.09	75.50	75.00	74.50

SBL	89.60	86.02	89.04	87.57
STANFORD	78.00	78.42	77.67	86.40
UCLA	80.50	82.09	78.52	79.66
TRINITY	81.07	82.41	80.54	82.24
USM	87.30	87.80	86.80	90.30
UM	75.40	76.03	75.51	76.06
YALE	87.50	82.47	82.47	79.84
Mean	**83.64**	**81.96**	**80.81**	**83.86**

4.4.7.Ablation Study

This section presents an ablation study to validate the contribution of the proposed MCBERT architecture on the multimodal ABIDE-I dataset. In Case A (baseline model without attention mechanisms) both the channel attention component (CAC) and spatial attention component (SAC) are removed from the multi-head CNN architecture. The model relies solely on core convolutional layers without attention mechanisms to process the input data. This case aims to quantify the contribution of attention mechanisms by comparing the model's performance with a standard CNN, allowing us to isolate the effect of the attention modules on classification performance. For case B (without channel attention) the channel attention component (CAC) is disabled while the spatial attention component (SAC) remains active. This setup focuses on analyzing the spatial features of the MRI data. This experiment aims to evaluate the importance of channel-specific attention. It provides insight into whether focusing on channel-specific features significantly impacts the model’s ability to classify ASD. Similarly, for case C (without spatial attention) the spatial attention component (SAC) is removed, while the channel attention component (CAC) remains active. This setup assesses the model's performance when spatial patterns are not specifically highlighted. The focus here is to understand the role of spatial attention in identifying relevant spatial features from MRI images and determine its contribution to the model's overall performance. For the case D (without the BERT module) the BERT module, which processes meta-features, is removed entirely. The model uses only the multi-head CNN to process the MRI image data without leveraging any meta-feature information. The purpose of this experiment is to evaluate how much of the model's success is attributable to the BERT-processed meta-features. It will show whether the image data alone is sufficient to achieve high diagnostic accuracy or if meta-features play a crucial role.

Table 4.6 Ablation outcomes with the proposed MCBERT architecture.

Dataset	Case	Case description	Accuracy	Specificity	Sensitivity
ABIDE-I	A	Without attention mechanism i.e. CAC and SAC	72.9	81.2	81.8
	B	Without channel attention (CAC)	85.3	84.8	85.1
	C	Without spatial attention (SAC)	83.4	82.6	83.2
	D	Without BERT module (Image-only model)	78.3	78.2	78.3
	E	BERT module only	73.1	72.6	84.9
	F	Without global max pooling (GMP)	91.6	90.1	91.2
	G	Complete architecture (MCBERT)	***93.4%***	***92.1%***	***94.5%***

In case E (BERT module only) the multi-head CNN is removed it explores the performance of the model when only meta-features are used for classification, without the additional information provided by the MRI images. It allows an assessment of the relative value of meta-feature data compared to image data in ASD classification. For case F (without global max pooling) the global max pooling (GMP) layer is removed from the architecture and replaced with the average pooling. The goal here is to determine the significance of the GMP layer in selecting the most salient features before the fully connected layers. It helps assess whether the GMP layer plays a critical role in the final classification performance by maximizing key features. The work evaluates the performance across the cases mentioned in Table 4.6. At last, case G refers to the performance of the complete architecture i.e., MCBERT.

4.5. Computational Complexity

The computational complexity of the proposed model can be described in terms of the dominant operations in its architecture, including convolutional layers, attention mechanisms, and BERT. The convolutional layers, responsible for processing image data, contribute a complexity of $O(N^2)$, where N is the spatial dimension of the input (MRI images). This quadratic complexity arises from input size, number of channels, and filter sizes in the convolution operations. The channel and spatial attention

mechanisms, which operate on feature maps, also scale linearly with the number of channels and spatial dimensions but remain dominated by the $O(N^2)$ behavior. Additionally, the BERT component, used for processing meta-features, introduces a complexity of $O(L^2)$, where L is the sequence length, reflecting the quadratic nature of the self-attention mechanism. As a result, the overall computational complexity of the model is approximately $O(N^2 + L^2)$, with the convolutional layers typically dominating for large image inputs, while the BERT module adds significant complexity depending on the length of the meta-feature sequences. This combined quadratic complexity is characteristic of deep learning models utilizing both convolution and attention mechanisms.

Generally, BERT requires high computational demands, but several strategies could be employed to reduce the computational needs. One approach is to use model compression techniques such as pruning and quantization, which can reduce the number of parameters without significantly impacting model performance. Additionally, lighter versions of BERT, such as DistilBERT or ALBERT, could be considered, as they retain most of the model's accuracy while offering reduced complexity. Furthermore, implementing mixed-precision training or utilizing distributed training frameworks may also help optimize computational resource usage. These techniques, in combination, can effectively reduce the overall processing load while maintaining the efficacy of the models.

4.6. Discussion

4.6.1. Study contributions

In this study, we proposed a novel multimodal architecture, MCBERT, for diagnosing autism spectrum disorder (ASD) by integrating brain MRI images and meta-features such as gender, behavioral characteristics, and patient history. Our model fuses a Multi-Head CNN (MCNN) with bidirectional encoder representations from transformers (BERT) to capture both spatial and channel attributes from the image modality, while efficiently handling the high dimensionality of meta-features. The results demonstrated that MCBERT achieves high diagnostic accuracy, with an overall accuracy of 93.4%, surpassing other state-of-the-art systems.

The key contributions of this study include the development of a novel fusion technique that integrates multimodal data for ASD diagnosis. By combining the strengths of CNN for processing brain MRI images and BERT for extracting meaningful information from meta-features, we were able to achieve superior performance. Additionally, the incorporation of convolutional block attention components (CBAC) enhances the model's ability to capture spatial and channel attributes, further improving its diagnostic power. The use of leave-one-site-out (LOSO) cross-validation provided a rigorous assessment

of the model's ability to generalize across different data sites, which is crucial for ensuring the robustness of ASD diagnostic models in real-world clinical settings.

4.6.2. Challenges and future directions

Despite these contributions, the proposed MCBERT architecture has several limitations that should be acknowledged. Firstly, the model only utilizes structural MRI (sMRI) data and does not include functional MRI (fMRI) data, which captures brain activity and could provide deeper insights into the neural mechanisms associated with ASD. By focusing solely on sMRI, the model may overlook important functional abnormalities that are often present in individuals with ASD. Furthermore, the study is limited to the ABIDE-I dataset, which constrains the ability to generalize the findings to other datasets or populations. The diversity of ASD manifestations across different groups means that relying on a single dataset could limit the model's applicability in broader clinical settings. Additionally, the reliance on pre-existing meta-features, which are not universally standardized across datasets, introduces potential variability in model performance when applied to new data.

Looking ahead, there are several promising directions for future work. One of the key areas for expansion is the inclusion of fMRI data in conjunction with sMRI, which would allow for a more comprehensive analysis of both structural and functional aspects of the brain. Exploring the combination of these two imaging modalities could enhance diagnostic accuracy and provide a more detailed understanding of ASD's underlying neural mechanisms. Additionally, extending the model to the ABIDE-II dataset and other large, multimodal datasets would enable further validation of the model’s generalizability across different populations. The incorporation of advanced hybrid networks, combining convolutional and transformer-based architectures, could also lead to improved performance in early ASD detection. These advancements hold the potential to refine ASD diagnosis and contribute to the development of personalized treatment plans based on a more thorough understanding of each individual's neurodevelopmental profile.

4.6.3. Real-world applicability

In terms of real-world deployment, the MCBERT model shows strong potential for integration into clinical workflows, provided that certain developments are made. Future work should focus on adapting the model to handle large-scale, real-time clinical data, ensuring that it meets regulatory standards and is interpretable by medical professionals. This could involve refining the model’s output to provide clear, actionable insights that clinicians can easily integrate into their decision-making process. Additionally, integrating the model into existing hospital information systems or diagnostic software platforms would help streamline its adoption in clinical practice. By focusing on these areas, MCBERT could be positioned as a supportive diagnostic tool for healthcare professionals, facilitating more efficient and accurate ASD diagnoses. The model’s multimodal approach, which combines brain imaging with meta-

features, offers a robust framework that is aligned with the growing trend of precision medicine and personalized healthcare.

Another critical step toward real-world application is the validation of the model's results with input from medical experts. Collaboration with neurologists, radiologists, and other healthcare professionals specializing in ASD is essential for establishing the clinical credibility of the model. Future studies should focus on comparing the model's predictions with expert diagnoses to ensure its reliability in a clinical setting. Expert feedback could also be invaluable in refining the model further, particularly in cases where subtle patterns in the data might lead to misclassification. This validation process would not only enhance the model's accuracy but also foster trust among healthcare providers, increasing the likelihood of its integration into routine clinical practice.

In summary, the MCBERT architecture is well-positioned to be adopted as a real-world clinical tool. With further validation and refinement, particularly in the areas of regulatory compliance, scalability, and expert validation, the model could play a significant role in improving the early diagnosis of ASD. These advancements would ultimately contribute to better patient outcomes, supporting early interventions and more personalized treatment plans for individuals with ASD. By addressing these limitations and expanding the scope of the research, future work aims to push the boundaries of ASD diagnosis through more advanced multimodal deep learning techniques.

4.7. Chapter Summary

In this work, we propose a novel fusion technique for diagnosing Autism Spectrum Disorder (ASD) in a multimodal setting, integrating information from brain MRI images and associated meta-features. This is achieved by combining the outputs of a multi-head CNN and Bidirectional Encoder Representations from Transformers (BERT). To enhance feature fusion, we utilize a convolutional block attention component (CBAC) for extracting spatial and channel attributes. Additionally, a BERT-based architecture is designed to efficiently handle meta-features, enabling the extraction of key attributes from this modality. The proposed fusion approach demonstrates significant improvements in feature integration and predictive performance for ASD diagnosis, validated against baseline methods.

Chapter 5 APPLICATIONS OF DEEP LEARNING FOR ASD DIAGNOSIS

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by a wide range of symptoms and behavioral patterns, often requiring early and accurate diagnosis for effective intervention. In recent years, deep learning has emerged as a transformative approach in medical research, leveraging its ability to process and analyze complex, high-dimensional data with remarkable precision. This chapter delves into the applications of deep learning techniques for ASD diagnosis, highlighting their potential to address critical challenges such as early detection, personalized treatment planning, and the integration of diverse data modalities. By exploring advanced architectures, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models, this chapter aims to illustrate how deep learning facilitates breakthroughs in understanding and diagnosing ASD. Through case studies, comparative analyses, and novel methodologies, the chapter underscores the critical role of deep learning in shaping the future of ASD diagnosis and care. Section 5.1 introduces the chapter, highlighting the role of deep learning in ASD applications. Section 5.2 provides background information on key deep learning models, including VGG-16, AlexNet, ResNet, and Vision Transformers. Section 5.3 presents the proposed ASD_CEVT architecture, designed for enhanced ASD diagnosis. Section 5.4 details the experiments and results, covering evaluation measures and result analysis. Section 5.5 discusses the findings and their implications. Finally, Section 5.6 concludes the chapter with a summary of the key contributions.

5.1.Overview

Autism spectrum disorder is one of the most complex neuro-developmental conditions that hinders one's social communication. Autism is called a "spectrum" disorder because the intensity and the type of symptoms differ greatly. Individuals suffering from ASD require special treatment and care [138] [139]. The broad range of characteristics linked with ASD makes diagnosis difficult and its causes are multifactorial [140] [141]. ASD influences people of every racial, socio-economic, and ethnic background [142][13]. ASD sufferers face communication difficulties and they cannot convey themselves via words, facial expressions, or gestures [27][143]. Although ASD is a lifetime illness, researchers have proven that early diagnosis and advanced medical care can increase the chances of better mental health [144]. The prevalence of ASD is a global concern and as per reports, one in every 54 children (in the USA) has been detected with ASD [145][146]. Boys have more chances of being diagnosed with ASD than girls. The pediatrics have recommended that everyone should have an early screening test for ASD as a routine health checkup in order to identify whether they should look for an advanced clinical diagnosis or not.

World Health Organization report of June 2021, discloses the epidemiological statistics showing that one in 160 children are affected by ASD worldwide, whereas the statistics of ASD in middle and low-income countries still remain unknown [147]. The lack of appropriate tools, medical tests, and treatment makes the diagnosis of ASD difficult. ASD is a disorder that influences various sections of the brain and is caused by a problem called polymorphism (a genetic effect caused by human gene interaction). These gene variants affect biological processes, like synaptic connectivity, and brain development, which are necessary for brain functioning. Consequently, further research is needed to unleash the intricate factors that contribute to the development of ASD [148][149].

Medical experts diagnose ASD on the basis of neurophysiological signs generated, as there is no diagnostic procedure that can detect ASD at any time [124][150]. There is no particular medication for ASD. The advancements in cost-effective, reliable, and easy-to-use screening tools are necessary due to the increasing number of ASD cases. To diagnose ASD efficiently, physicians require a child's progressive history and the presence of ASD biomarkers. Some of the essential biomarkers that are being explored by researchers are eye-tracking, neurophysiological, functional/anatomical brain characteristics, genetic, and behavioral [151]. The face is one of the crucial bio-markers as the nervous system takes and processes data via facial elements directly. The potential to classify different facial features and expressions are crucial characteristic that will aid in classifying brain asymmetry (neuro-developmental disorder) [152][78]. Various conventional screening techniques are used to diagnose ASD such as interviews where the severity is assessed by various questionnaires (Q-chat, AQ-10, ADOS-2, ADI-R). These techniques are cost-effective, easy, and lead to reliable diagnosis. The limitation of these techniques is 'bias'. Some other modalities, like functional magnetic resource imaging (fMRI), blood tests, and electroencephalograms (EEG) are used for diagnosis depending upon the physician [153][123]. However, most of these methods come with higher costs, and may not be accessible for people living in low-income areas. To address the requirement for better ASD diagnostic tools, medical professionals and researchers are working on creating widely accessible tools. The investigation of early bio-markers is never-ending. The extraction of facial characteristics as a physical element to diagnose autism is amongst the most trending, fairly new, and rapidly evolving areas of autism. Due to its distinctive characteristics, facial image recognition might be the most accurate technique of diagnosis. Recent research portrays the capability of deep neural networks (DNN), specifically the significance of convolutional neural network models in disease diagnosis. Because of its remarkable capability to understand via automatically fetching the hidden characteristics (features) from a huge pool of images, CNNs are extensively acquired feature extractors for image classification and object detection. CNN proves to be exceptionally efficient and appropriate, training a CNN model takes a considerable number of computational resources and time. Thus, rather than developing from scratch, it is favorable to work with pre-trained models that have been

built using a large pool of data and supercomputers. Transfer learning is an approach that associates parameters and weights of the pre-trained models to improve and modify the ultimate outcome depending upon the desired tasks, which in turn enhances classification accuracy. In this work, we worked on 2D face images using CNN and a vision transformer to extract features (attributes) from images. We proposed an optimal transfer learning-based framework to detect autism precisely. We worked on a variety of face images of autistic and non-autistic individuals with various deep learning and optimized pre-trained models for analysis. Our work is primarily an image classification task, in which a well-trained transfer learning model detects autism when exposed to an input facial image.

5.2.Background

5.2.1. VGG16

VGG (Visual Geometry Group) is a well-known deep learning network in the domain of computer vision, prominent for its exceptional achievements in object classification and detection tasks, capable of classifying thousands of images of numerous categories. Figure 5.1 illustrates the workflow of VGG16 that has sixteen weight layers (learnable parameters) and incorporates 13 convolutional layers, 5 max_pooling layers, and 3 dense layers, making it a deep architecture responsible for learning crucial features from images [154][155]. Instead of using a massive amount of hyperparameters, VGG16 has a convolution layer of '3X 3 filter' with stride 1 and a max_pool layer of '2X 2 filter' with stride 2.

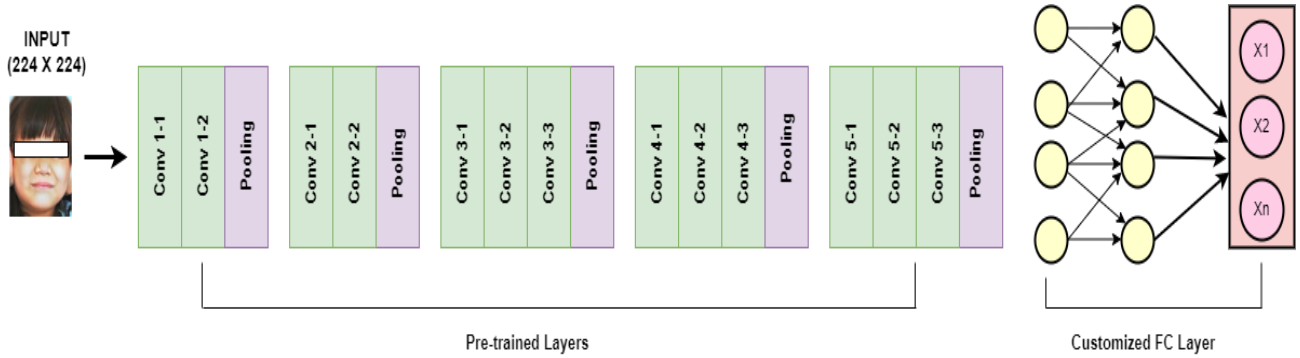


Figure 5.1 The general architecture of VGG16 comprising various pre-trained layers having convolution layer (in green) to process information and extract features from the facial images and customized fully connected layers to produce output

The convolutional layers are the essential building blocks capable of capturing hierarchical depictions of the input data. The selection of stride and filter size promotes the effective extraction of feature maps while managing a high-resolution feature map. The pooling layer optimizes the spatial dimensions of the feature map. It helps in making the network recognize objects irrespective of the location in the input image. The fully connected layers are responsible for giving final predictions from the learned

characteristics. The softmax activation function converts final scores into probabilities, showing the likelihood of an image belonging to each class [156].

5.2.2. AlexNet

Alexnet is one of the most used CNN with a substantial count of 60 million parameters. Consisting a total of eight layers, the network includes five initial convolutional layers, followed by three fully connected layers [157]. Figure 5.2 illustrates the workflow of AlexNet, where every convolutional layer is equipped with an activation function, which introduces non-linearity and facilitates the architecture's ability to learn intricate features from the image. The initial 2 conv layers are seamlessly connected with overlapping max_pooling layers, meticulously combined to extract the most significant features from the input image. This strategic incorporation of max_pooling helps reduce the spatial dimensions of the data, promoting translation invariance and capturing essential patterns regardless of their location within the image.

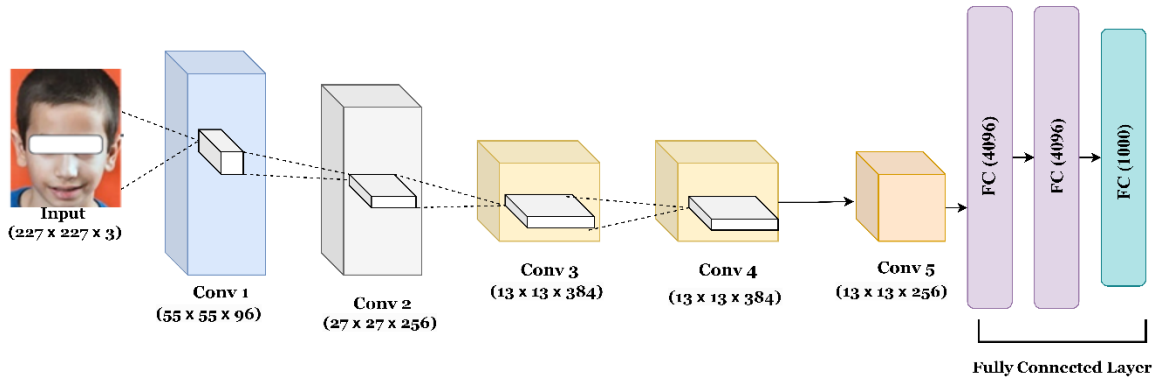


Figure 5.2 General architecture of AlexNet illustrating the sequence and size of convolutional, and fully connected layers to produce final predictions

The last three conv layers form a direct linkage with a fully connected layer, enabling a seamless flow of information from the conv feature maps to the dense layers [158][159].

Output(conv layer) + output(FC layer) are connected to ⟶ ReLU function

The integration of fully connected layers at the end of architecture fosters high-level feature representation contributing to more robust and discriminative classification capabilities [160][31].

5.2.3. ResNet

Residual network 18 is an 18-layer deep convolutional neural network to overcome the challenge of training deeper networks efficiently by incorporating the strategy of skip connections (known as shortcuts). A residual unit incorporates multiple convolutional layers, followed by skip connections that

sum the input with the output of the layers [161][162]. Figure 5.3 shows the diagrammatic representation of how the skip connection in Resnet works, which serves as a base for our proposed architecture. The main motive of these skip connections is to mitigate the problem known as the 'vanishing gradient' (the gradients of the loss function diminish as they backpropagate through various layers), which hinders successful training of deep networks and leads to suboptimal convergence. Resnet addresses this issue through the strategic introduction of skip connections, enabling the layers to bypass certain layers and facilitate the direct flow of gradients across various depths. Skip connection takes the activation from (n-1)th convolution layer and adds it to the output of (n+1)th layer and then applies ReLU function on this sum, hence skipping the n layer. This facilitates efficient gradient propagation during training, allowing the network to learn meaningful features from the data [25][24]. The model incorporates various residual units which propagate in both backward and forward directions using identity mapping. Propagation can happen among blocks having a high rate of accuracy with respect to classification performance. These residual mappings make training more generalized and more accessible. Resnet18 has nearly '11 million' trainable parameters. Resnet models are over 100 layers deep and they exhibit exceptional classification accuracy.

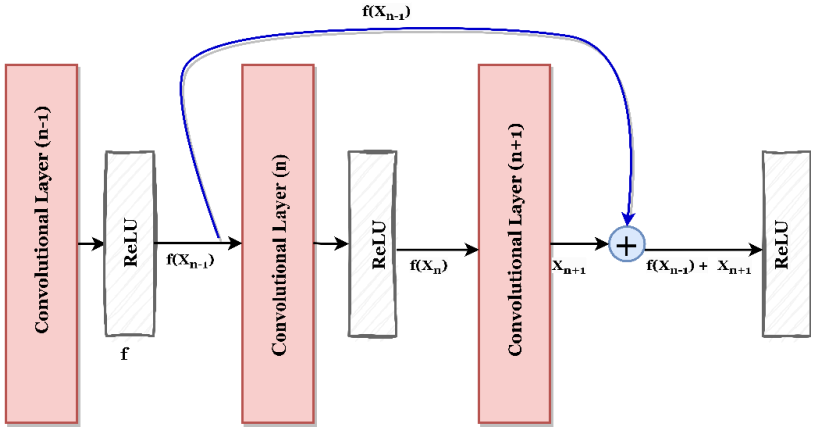


Figure 5.3 Diagrammatic representation of the working of skip connections in ResNet that takes the activation from (n-1)th convolution layer and adds it to the output of (n+1)th layer and then applies ReLU function on this sum

5.2.4. Vision Transformer (ViT)

ViT is a cutting-edge neural network tailored for various computer vision applications, specifically image classification. ViT draws inspiration from transformer networks (originally designed for natural language processing tasks), they undergo significant enhancements to optimize their efficacy in image processing. Notably, the primary distinction lies in how images are represented within the network [27][163]. In

contrast to NLP tasks, where text is represented as a sequence of words, ViT handles images as a sequence of smaller patches (each typically measuring 16 X 16 pixels). These patches are then processed via a CNN to get significant features specific to every patch, enabling optimized representation and understanding of the input image. The subsequent transformation incorporates passing the patch vectors via a stack of transformer encoder layers.

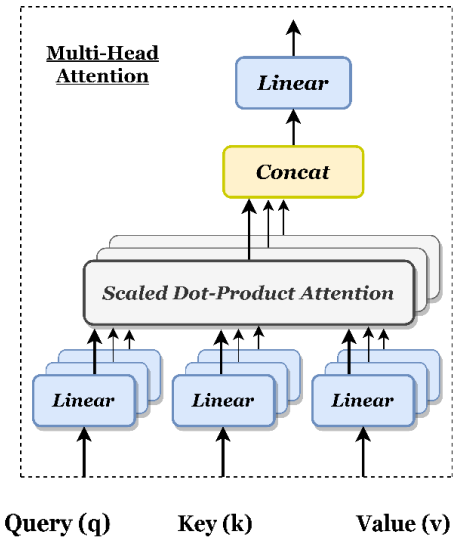


Figure 5.4 Diagrammatic representation of the workflow of multi-head attention segment in transformers

Within this architecture, the attention mechanism conducts computation repeatedly through instances called attention heads. Subsequently, the outcome of these parallel attention calculations is aggregated to generate a consolidated attention score. This is termed as multi-head attention. Figure 5.4 represents the internal structure of multi-head attention in transformers which enables the model to learn and focus on different sections of the input sequence simultaneously. In other words, each head conducts the self-attention mechanism independently, empowering the network to learn and grasp long-term dependencies between the individual patches of the image. This capability is specifically pertinent for image classification, as it fosters an understanding of how different portions within the image collectively contribute towards the overall classification. Further, the outcome of this transformation is a sequence of vectors that holistically combines the essential features of the input image. These vectors act as potent input for image categorization tasks and other computer vision endeavors, showcasing the ViT power in this domain [164].

Attention: Attention is an essential mechanism within deep learning networks that facilitates the focus on specific portions of input during processing. This concept takes inspiration from the human cognitive process, by how humans concentrate on different elements of an image when assimilating visual

information. By adopting attention, deep learning networks can efficiently prioritize relevant information, enabling them to perform more precise tasks across domains.

5.3. Proposed Architecture

The proposed ASD-CEVT model is relatively inspired by the architecture of ResNet. The ResNet incorporates the concept of skip connection (refers to the inclusion of the actual input into the outcome of every convolutional block). Building upon the original vision transformer architecture, ASD-CEVT introduces a novel enhancement where the original image (actual input) is repeatedly given to the output of every encoder layer. This iterative process is achieved by integrating a parallel CNN block alongside the transformer network. Specifically, the whole facial image of autistic individuals undergoes processing via CNN block, which creates embedding of the image as output, and this output is iteratively concatenated with the output of every encoder layer. The continual integration of the initial image representation throughout the encoding process ensures the preservation of significant details, making the architecture remember the actual image at every end of the encoder output as shown in Figure 5.5. The ASD-CEVT architecture incorporates both global feature information (via vision transformer) and local detailed features via CNN. This fusion approach ensures robust feature extraction, enabling the model to perform better across the dataset.

The CNN block proposed in the architecture incorporates a 2D conv layer in a stacked manner and a 1D average pooling layer. Table 5.1 represents the size of the kernel and the filters at various levels/layers of our architecture.

Table 5.1 Filters and kernel size used at various layers of the ASD-CEVT architecture

Layer 1	16 filters	Kernel size = 3
Layer 2	25 filters	Kernel size = 5
Layer 3	D filters	Kernel size = 5

The average pooling layer was used to calculate the output vector V_{img} with size D, which refers to the mapping of input to the D dimension. In the transformer section, the input was partitioned in $N = \frac{HW}{P^2}$ patches, having (P, P) as the resolution of every patch. These patches were then flattened among all the D dimensions to make a sequence of flattened patches V_P, with a total size of $N \times (P^2 \cdot C)$.

Further the patch embeddings were made by mapping all the patches to dimensions D via trainable linear projection:

$$E_P=[V_P^1 S, V_P^2 S, \dots\dots, V_P^N S\ ] \quad (1)$$

$S \in \mathbb{R}^{(P^2 \cdot C) \times D}$ (2)

A learnable 1-D position embedding with size D was incorporated for all the N+1 embeddings, where the position of additional V_{class} embedding was set to 0. The path embedding E_0 was calculated as:

$$E_0 = [\, V_{class}, E_P \,] + S_{Pos} \quad (3)$$

$$S_{Pos} \in \mathbb{R}^{(N+1) \times D} \quad (4)$$

Further, encoder layers L were stacked, with E_0 as the input of layer one, then E_l was calculated by the concatenation of V_{img} (image embedding) to the output E_l of every encoder layer *l*.

Where, l = 1,2,3……, L (5)

$$\hat{E}_l = [E_l, V_{img}] \quad (6)$$

$$\hat{E}_l \in \mathbb{R}^{(N+1+1) \times D} \quad (7)$$

$\hat{E}_l$ is given as input to the next encoder layer L. Here the encoder layer of ASD-CEVT inputs representation of the complete input facial image including the output of the previous encoder layer.

Training and Evaluation

For the training of ASD-CEVT, cross-entropy as loss function, Adam optimizer, batch size of 20 with 50 epochs, with a learning rate of 0.0001 was used. Table 5.2 presents the values of the hyper-parameters adopted while implementing the architecture.

Table 5.2 Hyper-parameters adopted for the development of ASD-CEVT architecture

Parameters	**Size**
Embed_dimension	*256*
Hidden_dimension	*512*
Learning_rate	*0.0001*
Loss_function	*Cross_entropy*
Optimizer	*Adam*
Batch_size	*20*
Num_epoch	*50*
Patch_size	*4*
Num_classes	*2*

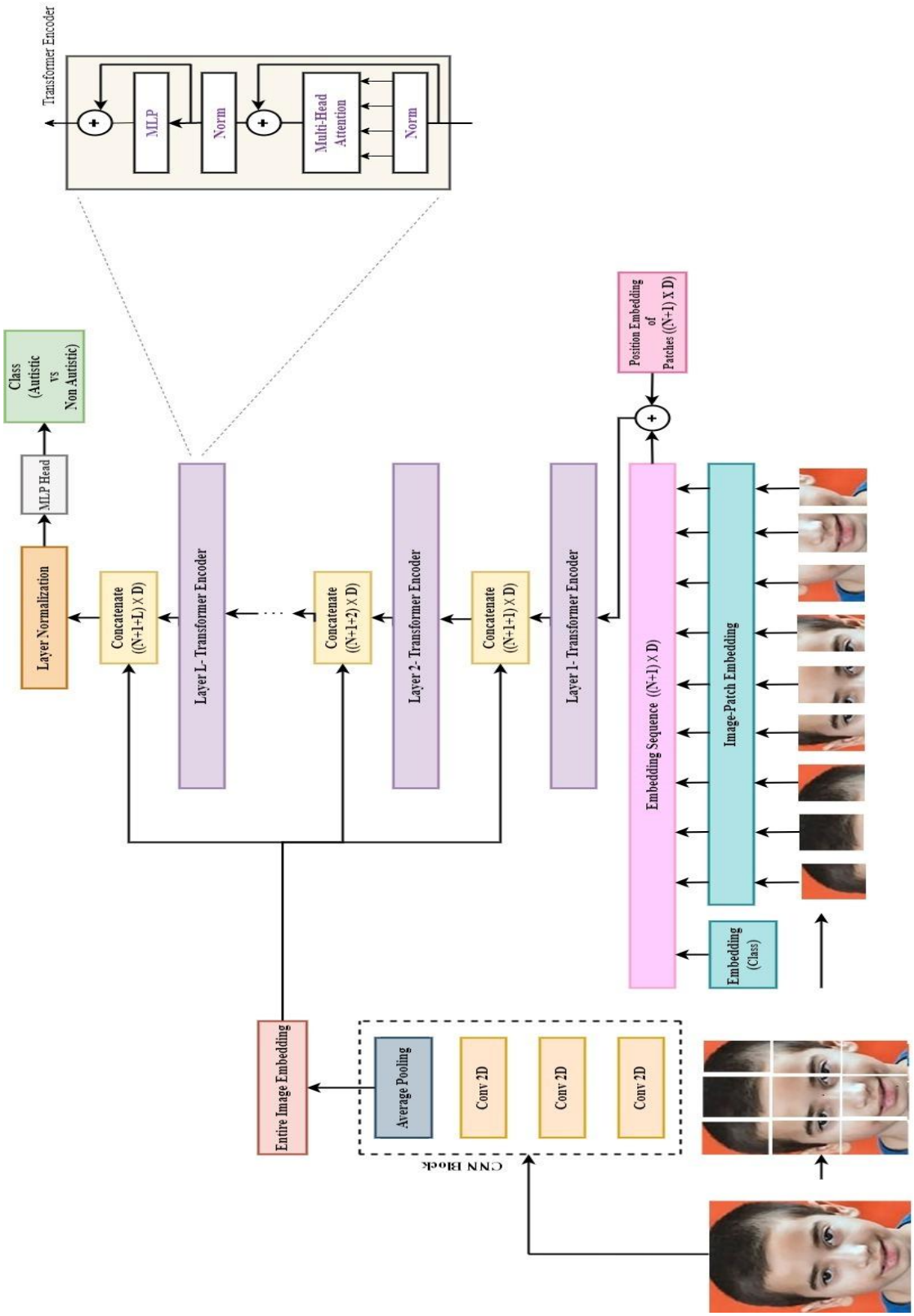


Figure 5.5 Overall structure of the proposed ASD-CEVT architecture comprising facial images

In order to adapt the existing weights on the facial data, a low learning rate was chosen. Label smoothing was performed while training to prevent overfitting, thus making the model to generalize well.

Algorithm: ASD-CEVT framework for ASD classification

Input: Set of facial images of autistic individuals

Configuration parameters: D, N, C, P, L, learning rate, batch size, epochs

Output: Trained ASD-CEVT model

1. Initialize CNN block:

   Initialize a CNN block with a stacked 2D convolution layer and 1D average pooling layer

   Configure convolutional layers with specific filter size and kernel size

   Computer V_{img} with dimension D

2. Initialize Transformer:

   Partition image into N= patches, each with size (P, P)

   Create a sequence of flattened patches V_p with size $N \times (P^2 \cdot C)$

   Introduce learnable V_{class} embedding of size D and 1D position embedding

   Compute E_0 using Equation 3

3. Encoder Layer:

   Repeat L times

   For i = 1 to L

   Concatenate V_{img} + output

   Compute $\hat{E}_l$ by applying the encoder layer to concatenated embedding using Equation 6

4. Training

   Cross entropy loss function for training

   Apply adam optimizer with learning rate

   Set batch size = 20, epoch = 50

   Implement label smoothing

5. Output: Obtained a trained ASD-CEVT model capable of accurately classifying ASD from

facial images

The algorithm shown above represents the generic view of the overall methodology adopted in our work.

5.4. Experimental Setup and Results Analysis

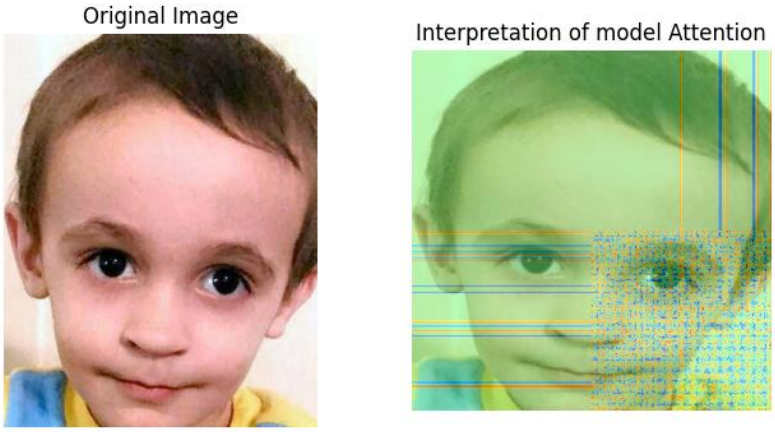
In this segment, we discussed the experimental outcomes obtained from the developed autism spectrum disorder (ASD) diagnosis architecture to showcase the efficacy of the proposed architecture. Furthermore, we conducted a comparative analysis to contrast the findings of our work with various existing state-of-the-art approaches.

5.4.1. Performance Evaluation Parameters

Performance measures are employed to measure how appropriately a model is detecting/predicting the outcomes. These measures are also crucial while comparing models. In this work, we employed accuracy, precision, and recall as these measures are most crucial when working in the healthcare domain.

5.4.2. Result Analysis

In this section, we analyze and describe the outcomes of our experimental evaluation on the classification performance of four distinct deep learning models, namely VGG16, Resnet, AlexNet, and the developed ASD-CEVT architecture. These models were trained using the Adam optimizer and evaluated via three key parameters.

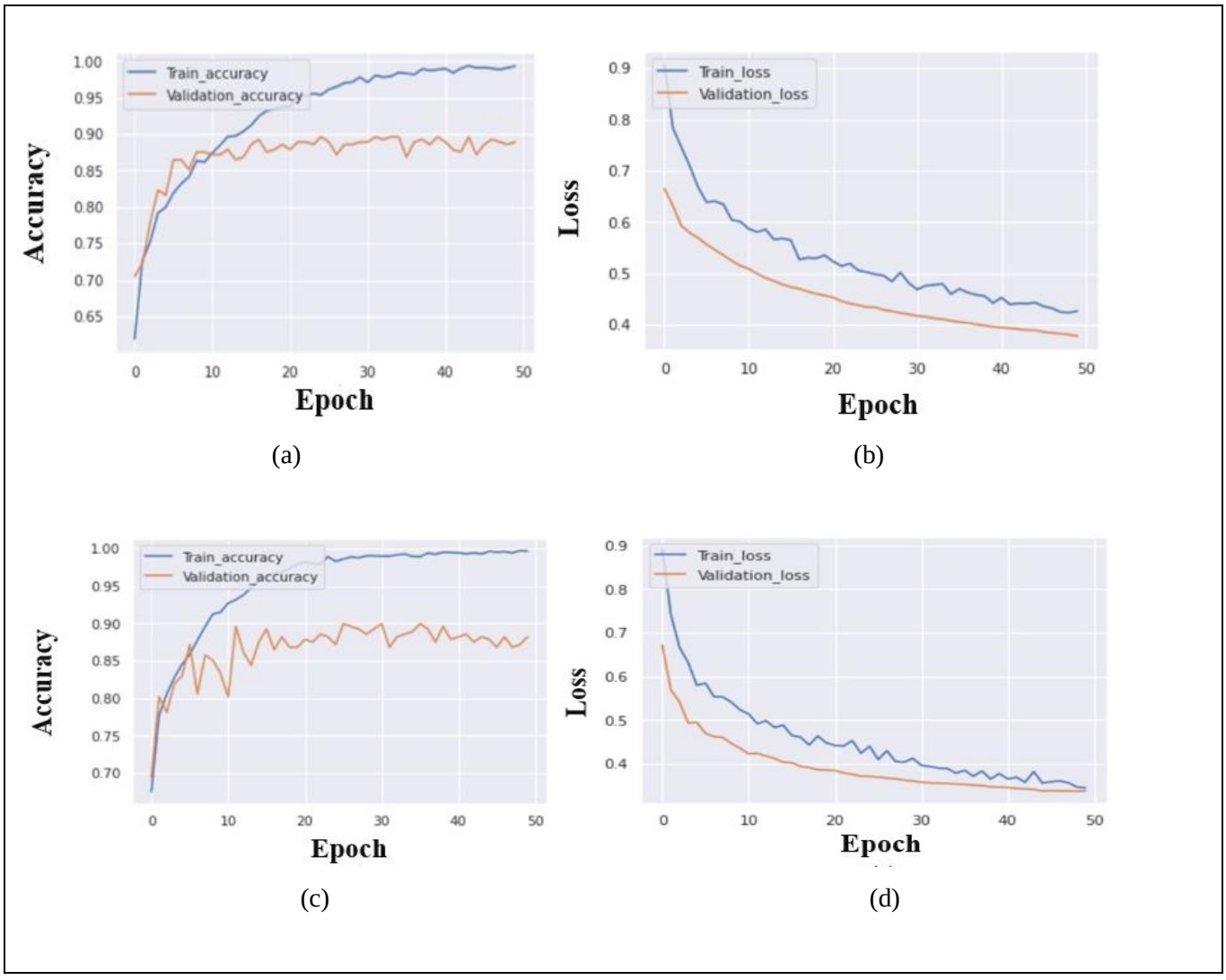


(a) (b)

Figure 5.6 Visual illustration of (a) original input (autistic child) image vs the (b) interpretation of the model attention

Figure 5.6 showcases the original/actual input image of an autistic child vs the interpretation done by the model attention component. The area highlighted with a colored pattern in (b) illustrates the region of high importance. This region incorporates the most crucial points identified by the architecture. As per the existing research and the studies reviewed in this work, the main facial biomarkers are the portion from eyes to lips. Proving that our ASD-CEVT architecture has also assigned importance to those features.

The obtained outcomes along with accompanying graphs, highlight the effectiveness and comparison of each architecture in addressing the task of classifying ASD.



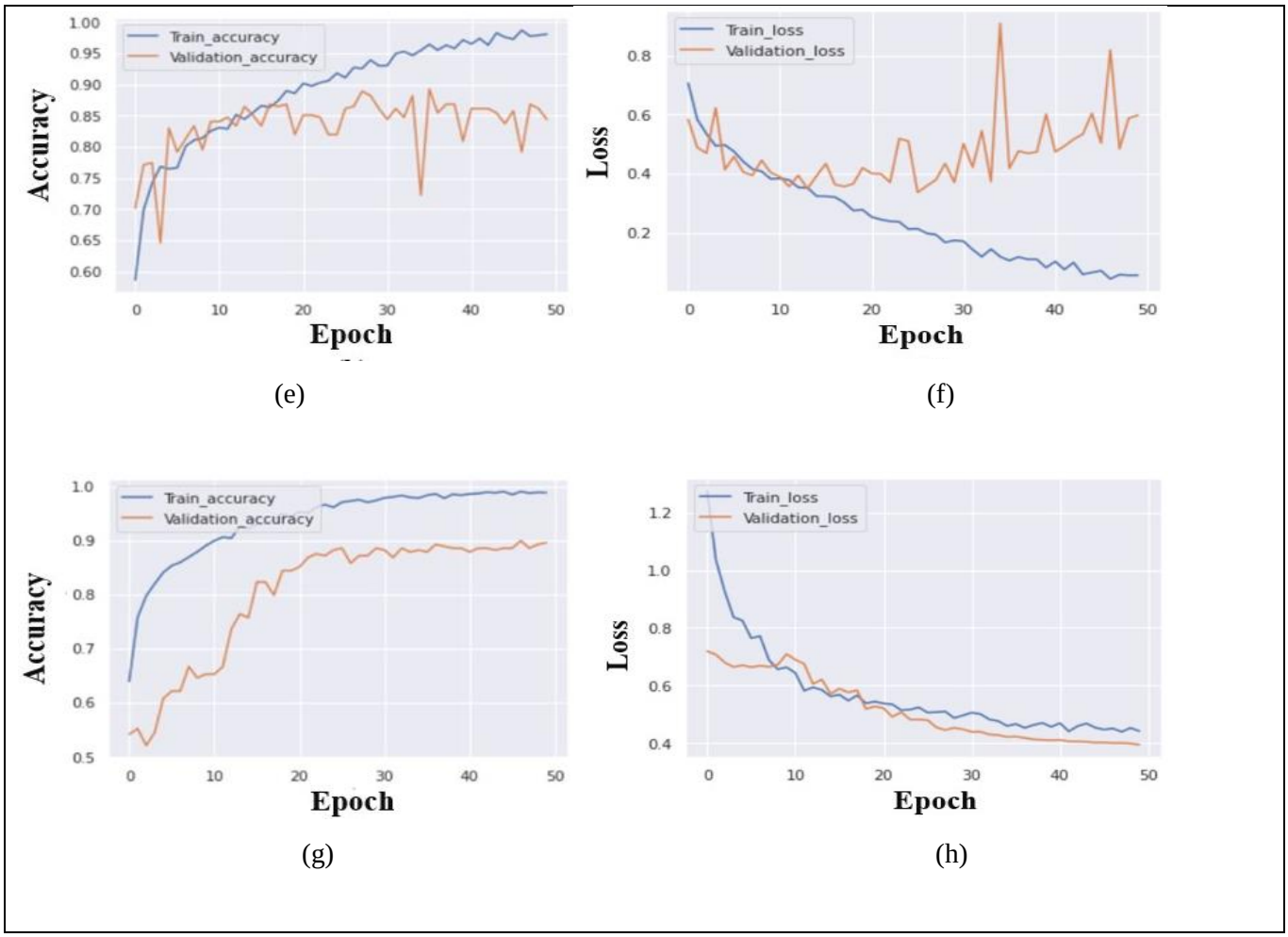


Figure 5.7 Graphical Plots of accuracy and model loss, where (a), (b) corresponds to ASD-CEVT; (c), (d) corresponds to AlexNet; (e), (f) corresponds to ResNet; and (g), (h) corresponds to VGG16 respectively

Figures 5.7 (c) and (d) portray that Alexnet achieved an accuracy of 0.908 demonstrating good classification performance on our dataset. The Resnet architecture showcases strong performance across all parameters (Figure 5.7 (e) and (f)) with an accuracy of 0.866% it maintains a balance in the values between precision and recall indicating its efficiency in achieving decent accuracy.

Table 5.3 Model performance achieved on various parameters

Model	Accuracy	Precision	Recall
ResNet	0.866	0.866	0.866
AlexNet	0.908	0.0908	0.0908
VGG16	0.835	0.835	0.835
ASD-CEVT	**0.924**	**0.909**	**0.909**

Figures 5.7 (g) and (h) demonstrate the results of the VGG16 architecture showcasing notable performance. It achieved an accuracy of 0.835, having lower precision and recall values. Our developed

ASD-CEVT architecture outperforms all parameters (Figure 5.7 (a) and (b)) with a value of 0.924. The results suggest that ASD-CEVT architecture is well suited to our dataset, showing robustness in making accurate classifications on facial images. Table 5.3 shows the results on various metrics for each of the four models. The analysis relies on the 300 test samples, where the highest accuracy achieved is 92.4% (on our proposed ASD-CEVT model).

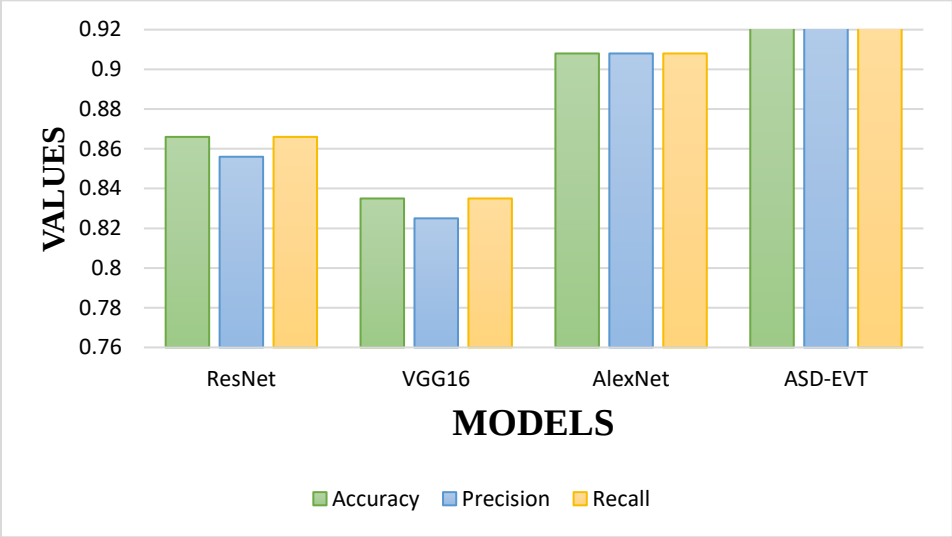
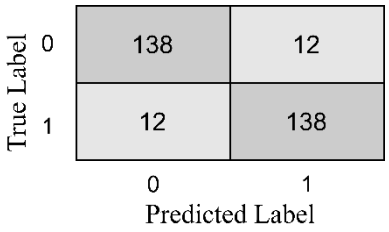
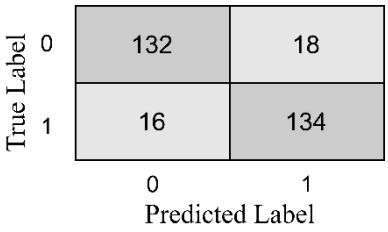


Figure 5.8 Comparison graph of classification performance of the developed model with baseline models

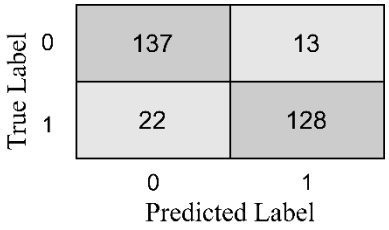
Figure 5.8 illustrates a visual comparison of the performance achieved by each model on the autism facial dataset allowing for an easy comparison of the true model's performance.



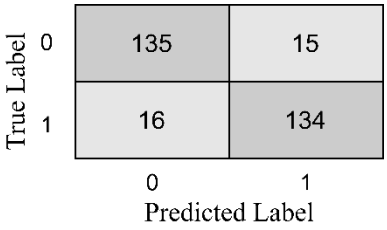
(a)



(b)



(c)



(d)

Figure 5.9 Confusion matrix of four models where (a) ASD-CEVT, (b) ResNet, (c) VGG16, (d) AlexNet respectively

Confusion matrix in Figure 5.9 visually illustrates the architecture's performance. Each architecture is evaluated on a 300-image set.

As seen from the confusion matrix, ASD-CEVT performed the best with only 12 input images incorrectly predicted as compared to the other employed models.

5.5. Discussion

Our research investigates the diagnosis of autism spectrum disorder (ASD) utilizing an enhanced convolution network-based transformer architecture applied to facial images of autistic and healthy (typically developing) children. While there are several methods for identifying autism within the current diagnostic framework, the most prominent approach remains brain MRIs and interview-based evaluation, despite its costly and lengthy diagnosis timeline. The significance of early detection offers the best prospects for ASD individuals to lead normal lives. Consequently, the impetus behind our work is self-evident i.e., to formulate a straightforward and precise identification strategy applicable from an early age. Recent trends in healthcare research underscore the utilization and efficacy of pattern recognition, image processing, and facial identification in this domain. Following the research literature on ASD and the fact that facial patterns (biomarkers) reflect underlying psychological functioning, facial identification emerges as a promising avenue for ASD. Studies developing the ASD identification architecture are still limited, as the limitations inherent in the training dataset pose several challenges for researchers. The dataset employed lacks data diversity as it is biased toward white ethnicity, posing a challenge in accurately identifying and generalizing individuals from other ethnicities. Furthermore, the dataset size and suboptimal visual quality are constraints. Additionally, relying solely on facial biomarkers is deemed inadequate, integrating and developing multi-modal architecture can notably augment accuracy. Table 5.4 showcases a comparative analysis of recent research findings providing their performance on various metrics. Some of the performance values mentioned represent the values near our developed architecture but most of them lack real-world applicability.

Table 5.4 Comparison with recent state-of-the-art strategies with their performance on various parameters

Reference	Purpose	Techniques	Highest Test Accuracy	Precision	Recall	F1	AUC
[165]	ASD	MobileNet	87%	87%	87%	87%	-

	recognition via facial expressions						
[166]	Detection of ASD via deep image learning strategies	VGG19, NasNet large, Inception V3, ResNet50	87.50%	-	-	-	-
[167]	Classification of ASD via facial images	(VGG16, MobileNet, VGG19), + LSTM	75.85%	76.56%	76.85%	75.69%	-
[168]	ASD diagnosis via facial landmarks	Hybrid VGG19, MobileNet V2	92%	92%	92%	92%	-
[169]	Detecting autistic individuals using face image	MobileNet, Inception V3, InceptionResNet V2	87%	-	-	-	-
[170]	ASD detection system for face images	CNN	91%	-	-	-	-
[171]	ASD diagnosis using deep learning	VGG16, EfficientnetB0, VGG19	87.9%	-	-	-	93.06%
[172]	Detection of ASD at early stage	MobileNet V1	92.1%	-	-	-	-
[173]	Face images based ASD diagnosis	VGG16, KNN, Random forest, Gradient boost, VGG19	0.88	0.87	0.88	0.88	-
[174]	ASD diagnosis	ResNet34,	92%	-	-	-	-

	using transfer learning	VGG19, AlexNet, ResNet50, VGG16, MobileNet V2					
Our Proposed	**Convolutional enhanced vision transformers for diagnosing ASD**	**CNN incorporated vision transformer**	**92.4%**	**90.9%**	**90.9%**	**-**	**96.3%**

5.6.Chapter Summary

This chapter presents a comprehensive exploration of our major contributions toward advancing Autism Spectrum Disorder (ASD) detection using facial image analysis. We proposed an optimized transfer learning-based face identification architecture capable of capturing autism with high accuracy, accompanied by visualizations of key facial features influencing the model's decisions. The novel enhanced vision transformer architecture integrates an attention mechanism with CNN blocks in parallel, supported by skip connections, to facilitate effective feature extraction and smooth information flow across layers. The proposed architecture addresses critical challenges, including overfitting, thereby improving generalization capabilities. Furthermore, this work pioneers a relatively underexplored area of image processing by focusing on ASD classification using facial images, diverging from conventional approaches that rely on brain imaging or EEG signals. The results underscore the potential of facial image-based diagnosis as an innovative and effective approach for ASD detection.

Chapter 6 SELF-SUPERVISED AND SELF-DISTILLATION APPROACH FOR ASD

The early and accurate diagnosis of Autism Spectrum Disorder (ASD) remains a complex challenge due to the diverse presentation of symptoms and the need for large annotated datasets. To address these limitations, this chapter explores the integration of self-supervised learning (SSL) and self-distillation techniques as innovative approaches for ASD diagnosis. By leveraging unlabeled data through SSL, the framework aims to extract robust feature representations, while self-distillation enhances the generalization and efficiency of the predictive model. This dual approach not only reduces dependency on annotated datasets but also improves the model's interpretability and diagnostic accuracy. The chapter presents a comprehensive analysis of these techniques, including their implementation, evaluation, and contribution toward advancing ASD detection methodologies. Section 6.1 provides an overview of the chapter, focusing on self-supervised and self-distillation approaches for ASD diagnosis. Section 6.2 outlines the background, emphasizing Self-Supervised Learning (SSL) and Knowledge Distillation (KD). Section 6.3 introduces the Autism Facial Image Dataset used in the study. Sections 6.4 and 6.5 describe the integration of Transformers and Masked Autoencoders, while Section 6.6 explores the self-distillation process. Section 6.7 presents the experiments and results, detailing the experimental setup and analysis. Section 6.8 discusses the findings and their implications. Finally, Section 6.9 summarizes the chapter's key contributions and insights.

6.1. Overview

This work delves into the development of an advanced deep-learning framework designed to improve Autism Spectrum Disorder (ASD) classification using facial images. The work focuses on addressing the challenges associated with early ASD detection, which is often hindered by the limitations of conventional diagnostic methods. The proposed framework, S/SD-ASD (Self-Supervised and Self-Distillation Learning for ASD), aims to provide a more accurate and efficient alternative for diagnosing ASD by utilizing cutting-edge techniques in deep learning and self-supervised learning. The core of this approach lies in the use of a Masked Auto-Encoder (MAE), which is an unsupervised model that learns to reconstruct masked regions of input data in this case, facial images from the known portions. This mechanism allows the model to focus on the underlying patterns within the image data and enhance feature extraction. By incorporating self-supervision, the model is trained to predict and infer missing parts of the image, improving its ability to learn from limited data. The MAE framework is further augmented by Self-Distillation, where a student model learns from the teacher model’s global features through logits-based knowledge distillation. This helps in transferring valuable knowledge from the more

complex teacher model to the simpler student model, thereby improving the latter's efficiency in representing important features from the data. A key innovation of the S/SD-ASD framework is its ability to work effectively with limited data, a common issue in ASD diagnosis where large labeled datasets are often unavailable. By combining self-supervised pretext tasks and self-distillation, the model is able to train on the target dataset alone without the need for additional external knowledge, addressing the scarcity of labeled data and reducing the risk of overfitting.

6.2. Background

6.2.1. Self-supervised learning (SSL)

A lot of deep learning architectures are trained through supervised learning, which needs a substantial amount of labeled data. However, some field lack datasets as extensive as those used in general-purpose applications. To address this issue, pre-trained models can be an effective solution [175]. These models are initially trained on large datasets and then fine-tuned for specific tasks. This approach has two main advantages: first, the parameters of a model trained on a large dataset provide a strong starting point for further training, facilitating faster convergence; second, such a model can effectively extract hierarchical semantic information, reducing the risk of overfitting on smaller datasets. Consequently, the performance of these models heavily depends on the size of the labeled dataset. In the medical field, gathering and labeling data in this domain is both costly and time-consuming, largely due to patient privacy concerns and the need for high-quality annotations. To address this challenge, self-supervised learning presents an excellent alternative, as it can be trained on unlabelled datasets.

Self-supervised learning/SSL leverages innovative pretext tasks for various applications, such as data augmentation, active learning, alignment, and anomaly detection. The general approach involves proposing a pretext task allowing the network to generate pseudo-labels from the data's attributes. These pseudo-labels then serve as supervisory signals during training. The resulting model can be transferred to the target data domain or task. In terms of representation learning, SSL can match the performance of supervised learning. Given the nature of human learning, which often does not rely on large labeled datasets, humans can learn effectively from unlabeled samples, demonstrating the potential of self-supervised learning in scenarios with limited labeled data.

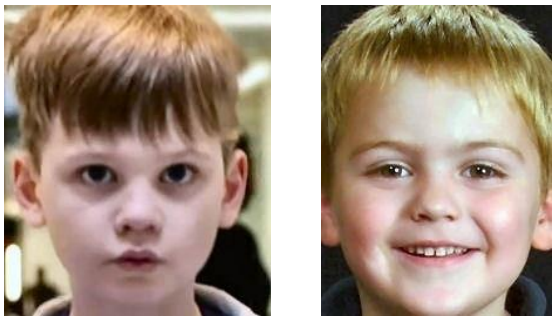
6.2.2. Knowledge Distillation/KD

With the evolution of neural networks, Knowledge distillation/KD has emerged as a significant model compression and transfer learning technique, gaining prominence in domains of artificial intelligence due to its effectiveness and simplicity [176]. KD addresses many practical problems by employing a "teacher-

student-network" training approach, wherein the knowledge from a trained model (the teacher network) is distilled and transferred to another model (the student network). This enhances the student model's learning capability by leveraging the teacher model's knowledge. Typically, the teacher module is a high-capacity network, while the student module is a smaller, more efficient network. Thus, knowledge distillation facilitates the training of a smaller student network under the supervision of a larger teacher module. Unlike the other compression algorithms, KD can effectively work across networks with different structures. Deep learning has achieved remarkable success across numerous domains, primarily due to the large, complex models capable of learning intricate patterns and features from data. However, these large models pose deployment challenges, especially on mobile devices. Knowledge distillation offers a solution by producing smaller models that retain the functionality of their larger counterparts. Consequently, an increasing number of researchers are focusing on knowledge distillation to overcome these deployment challenges. Table 2 summarizes the list of publicly accessible autism datasets.

6.3. Autism facial image dataset

The primary challenge encountered in this research was the lack of substantial and publicly available autism image datasets, which are essential for developing an image classification model. To address this, we employed an image dataset of autistic children from the Kaggle repository, as it is freely accessible [177]. Fig 6.1 provides a diagrammatic representation of the dataset, consisting of images of 2,940 subjects. This dataset includes 2D-RGB face images of both autistic and typically developing children aged 2-14 years, with the majority aged between 2-8 years.



(a) (b)

Fig 6.1 (a) Autistic, and (b) Non-Autistic sample from the image dataset

The gender ratio (male to female) in the dataset is approximately 3:1, and the ratio of autistic to non-autistic subjects is 1:1. It is crucial to note that the dataset lacks meta-information such as the clinical history of participants, the severity level of ASD among individuals, socio-economic background, and

ethnicity. Many of the facial images were of suboptimal quality concerning brightness, image size, and face alignment. Table 6.2 shows the data distribution of the employed image dataset.

Table 6.1 Distribution of Autism Facial Image Dataset

Type	Class		Total
	Autistic	**Non-Autistic**	
Train	1176	1176	2352
Test	294	294	588
Subjects (Total)	1470	1470	2940

6.4. Transformers

Convolutional Neural Networks (CNNs) exhibit limited capability in capturing localized relationships of foci in medical domain. In this paper, we employ a transformer-based attention strategy to identify long-term relationships. When applied directly to patch sequences, transformers prove to be highly effective in image classification tasks. Compared to the most advanced CNNs, the ViT achieves enhanced diagnostic performance with significantly fewer computational resources during training.

Fig 6.2 illustrates the ViT architecture, incorporating three primary components: (a) patch embedding; (b) positional embedding; and (c) transformer module.

Patch embedding: Within this, data is handled sequentially. Initially, a high-dimensional image is converted into a sequence format. The facial images are first uniformly resized to 224×224, passed through a convolutional layer, and then flattened into $x_P \in \mathbb{R}^{N\times(P^2.C)}$, where P signifies the resolution of each image patch. The number of patches is determined by $N = HW/P^2$. Each patch is linearly projected into a D-dimensional vector space, with $D = P^2C$.

Position embedding: In our work, this layer enhances patch embedding with positional information. A learnable $1D$ -positional embedding $LP_E \in R^{(N+1)\times D}$ is typically used in traditional ViT's.

Transformer module: This block consists of layer normalization ($layer_{norm}$), multi-head self-attention (MSA), and a multi-layer perceptron (MLP). Input images $X \in R^{H\times W\times C}$ out of the patch embedding layer are converted into a sequence of patches, concatenated with a class token for classification. This process parallels the transformer architecture used in natural language processing (NLP). The resultant patches are concatenated to the LP_E positional embedding and fed into the transformer encoder for computation. The class token outputted from this process is then inputted into the MLP module for classification. Formulas for these computations are detailed as follows:

$$A_0 = [x_{cls} : x_p^1 LP;\ x_p^2 LP;, \dots,\ x_p^N LP] + LP_E\ , LP\ \in \mathbb{R}^{(P^2.C)\times D} \quad (1)$$

$$A_l = MSA\big(\ layer_{norm}(A_{l-1})\big) +\ A_{l-1}\ , l = 1, 2, \dots, L \quad (2)$$

$$A_l = MLP\left(layer_{norm}(\dot{A}_l)\right) +\ \dot{A}_l \quad (3)$$

$$y = MLP(A_L^0) \quad (4)$$

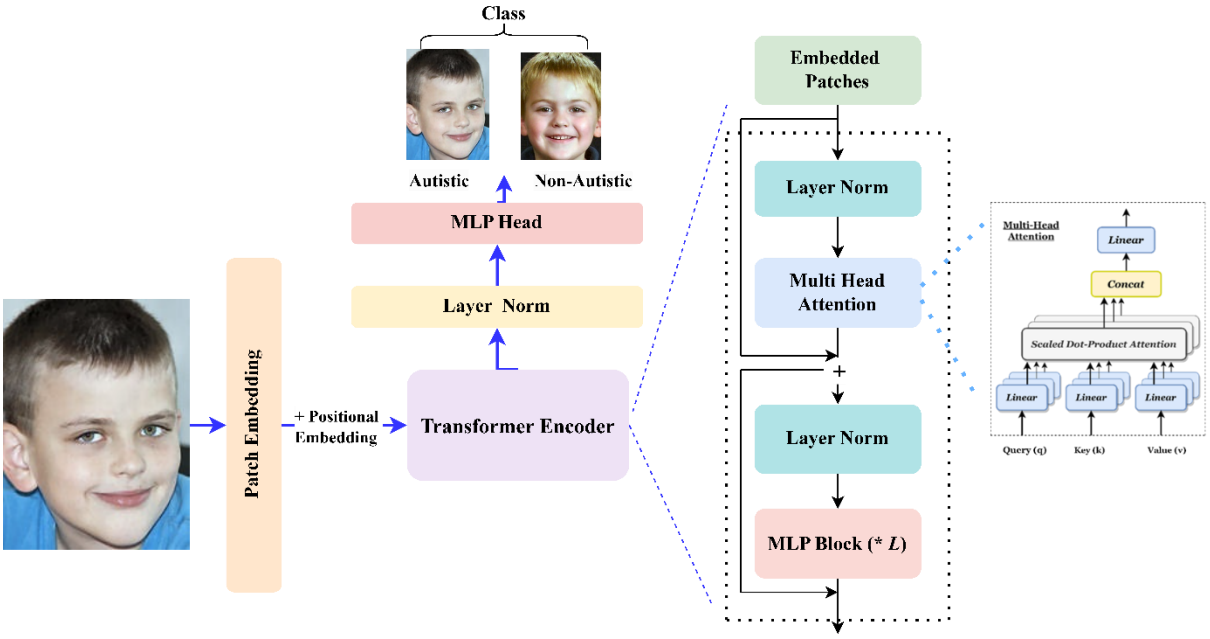


Fig 6.2 Visual representation of the ViT module

6.5. Image modeling using masks/Masked Autoencoder

In this segment, we unraveled the masked auto-encoder (MAE) components utilized in the S/SD-ASD method: decoder, encoder, and loss function. Fig 6.4 illustrates the structure of MAE. The encoder branch of MAE utilizes the transformer block from the vision transformer (ViT). Following procedures akin to ViT, the input image is initially divided into disjoint/non-overlapping image patches $X \in R^{H\times W\times C}$, which are then linearly mapped to obtain LP_E patch embeddings. Subsequently, a random masking process occurs at a specified ratio. Visible patches are denoted as Vis_P, while invisible patches are $Invis_P$. Upon receiving positional information from P_{info}, the encoder generates corresponding latent representations essential for subsequent image reconstruction. The encoder outputs an encoded vector along with a mask token, constituting the complete token set fed into the MAE decoder. Each token receives relevant positional embedding, whereas the mask token represents a shared, learnable vector symbolizing the anticipated missing pixel. Failure to include positional embedding can lead to suboptimal

image reconstruction, particularly from the mask token without associated patch position information. To optimize computational efficiency, the decoder is designed to be lightweight.

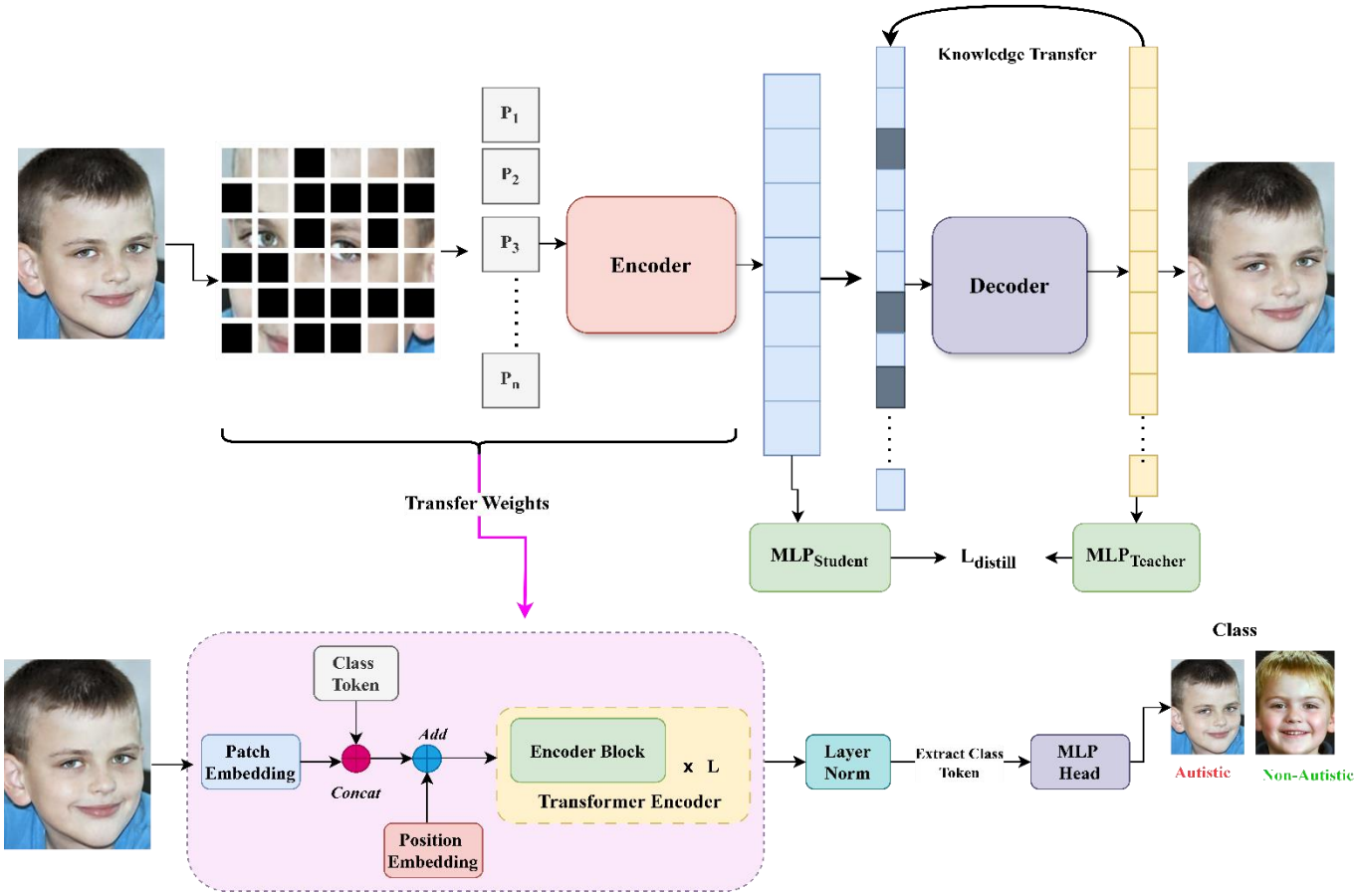


Fig 6.2 Diagrammatic illustration of the proposed S/SD-ASD architecture

The reconstruction entropy loss $\boldsymbol{L}$ in conjunction with MAE is computed by comparing the original region value y with the masked image patch's predicted value y_P, using MSE.

$$\boldsymbol{L_{MSE}} = \text{MSE}(y_P, y) \quad (5)$$

y represents the pixel value of the original facial image and y_P represents the pixel value of the predicted facial image patch.

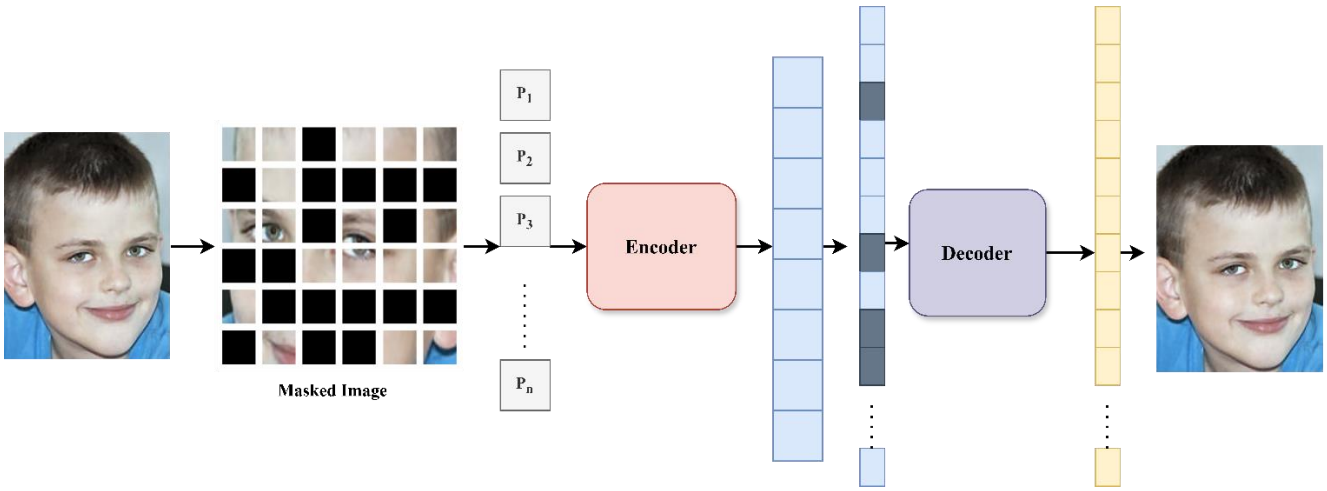


Fig 6.4 Visual representation of the workflow of MAE

6.6. Self-distillation

Fig 6.5 illustrates a diagrammatic representation of the self-distillation module. The image is noted for containing a significant amount of spatial redundancy compared to the semantic framework of language. It emphasizes that even if a portion of the image is obscured, sufficient identification can often still be derived from the remaining visible areas. However, this principle should not be misconstrued to imply that high-definition results can be achieved from an imperfect image. In contrast to this perspective, Mean Absolute Error (MAE) training focuses on the visible but insufficiently global features rather than the entire image. This approach can be enhanced by incorporating supervised labels into the network, a straightforward and effective method ensuring the network comprehends the necessary concepts for reconstruction. Fig 6.3 demonstrates the integration of two separate supervised branches into the system: one following the encoder and another after the decoder, facilitating comprehensive processing of all patches (40% local) during training. A student mapping vector distills. The encoder outputs global information. The goal is to augment the encoder's capacity for feature extraction through the decoder based on the teacher mapping vector.

$$M_a = MLP(Vis_P);\ M_b = MLP(Vis_d) \tag{6}$$

$$L_{distill} = 0.5L_{ce}(M_a;\ y) + 0.5L_{ce}(M_a;\ M_b) \tag{7}$$

$$\boldsymbol{L} = \boldsymbol{L_{MSE}} + L_{distill} \tag{8}$$

The proposed distillation method employs a single MLP layer in each referenced student and teacher branch within the encoder and decoder. To minimize $\boldsymbol{L}$ in the MAE process, vector Vis_P from the encoder

and vector Vis_d from the decoder are branched from vectors M_a and M_b, ensuring alignment between the two vector distributions and total loss formulation.

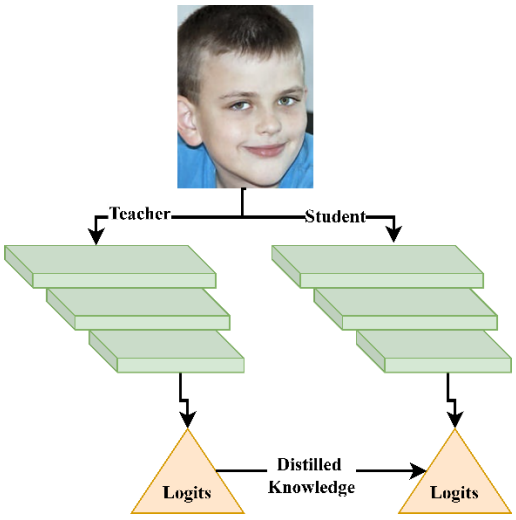


Fig 6.5. Schematic representation of self-distillation

6.7. Result Analysis and Discussion

This section sub-divides into various sections explaining the experimental setup, and visual and statistical comparison of the proposed work with SOTA techniques. To validate the effectiveness of S/SD-ASD, an ablation study was also conducted.

6.7.1. Experimental setup

The picture samples used in the renal datasets were resized to 224 × 224 in size. The training was finished after 200 epochs of pre-training and 60 epochs of fine-tuning. Every patch was split into sixteen-by-sixteen-inch pieces. The parameters used in this paper were changed to match those in the MAE.

6.7.2. Result Discussion and Visualization

The section provides a comprehensive evaluation of the results obtained by using modern computational approaches to autism spectrum disorder (ASD) diagnosis. This section digs into the insights acquired from applying S/SD-ASD to facial imaging data, intending to improve diagnostic accuracy and better understand the underlying visual indicators of ASD. This article explains how AI-driven methodologies contribute to interpreting and diagnosing ASD from medical pictures by analyzing model-generated visualizations in detail, such as patch segmentation, masked regions, token extraction, saliency maps, and essential feature identification.

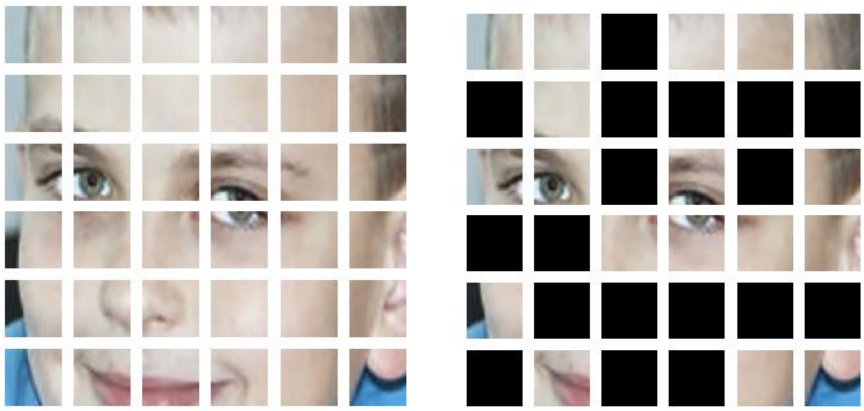


Fig 6.6 Visualization of the original patched image and the masked patch image obtained after inputting the original image

Upon inputting an image into the vision transformer (ViT) model for autism spectrum disorder (ASD) diagnosis, a series of visualizations are generated to elucidate the model's processing and interpretation: The initial step involves dividing the original medical image into smaller patches. This segmentation allows the ViT model to efficiently process spatially localized information, enhancing its ability to capture fine-grained details crucial for diagnosis. As part of the self-supervised learning (SSL) strategy, specific patches of the image are masked during training and replaced with visible patches (as shown in Fig 6.6). This technique encourages the model to learn robust features invariant to variations in localized image regions, thus improving its generalization capability.

After processing each patch, the model takes out tokens corresponding to abstract features. These tokens all capture essential visual details including texture, color, and spatial connections within the image, mentioned in Fig 6.7. These characteristics are essential for later classification steps or diagnostic decision-making.

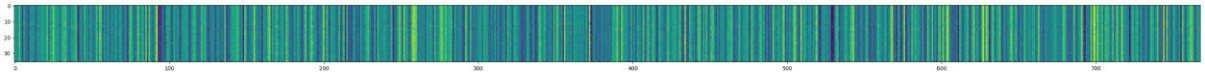


Fig 6.7 Embedding vectors obtained after masking the image

The portions of the input image that contribute most significantly to the model's decision-making process are highlighted in a saliency map that is created. This map provides insights into the particular visual cues suggestive of ASD by graphically indicating the areas that are critical for the model's categorization output. Certain areas or elements inside the image are recognized as crucial for the diagnosis of ASD based on the saliency map. The areas that are highlighted indicate the locations where the model has identified relevant patterns or irregularities in visual cues related to autism spectrum disorder. For the

facial features, eyes and lower part of the face i.e. lips and jawline are the main biomarkers which categorizes autistic and non-autistic individuals. Same phenomenon is learned by our model and it highlighted the significant areas of the face as seen in Fig 6.8.

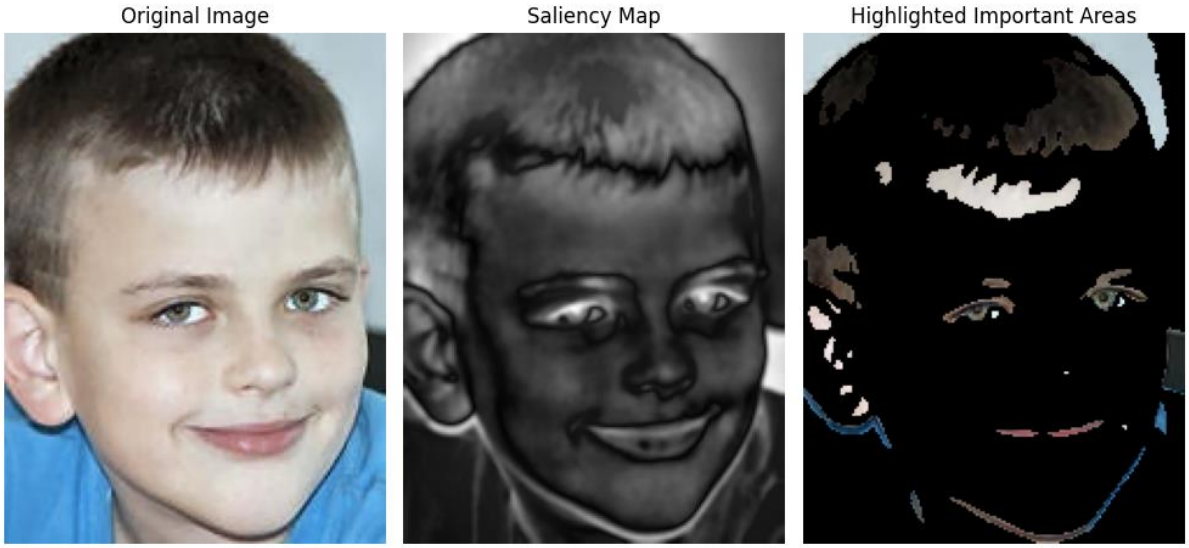


Fig 6.8 Visual comparison of the original image, saliency map, and the important regions highlighted by the model

These visualizations not only validate the model's decision-making processes but also provide vital insights into the complex visual signals associated with ASD, allowing for more informed clinical decisions and furthering the field of computer-assisted autism diagnosis.

Table 6.2 Fine-tuning (%) accuracy outcomes with various mask ratios for S/SD-ASD

Ratio$_{mask}$	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.75**	**0.8**	**0.9**
Fine-tune accuracy (%)	96.32	96.98	97.21	97.43	97.5	96.32	97.42	94.06	91.08

Table 6.3 shows the performance metrics attained by the model during the downstream fine-tuning task at various mask ratios (ranging from 0.1 to 0.9). The results show that our model performs optimally at a mask ratio of 50% i.e., 0.5, exceeding the MAE ideal mask ratio of 75%. This highlights the different qualities of ASD facial images as opposed to natural images. ASD facial images have richer features and better information density, requiring more advanced reconstruction guiding algorithms. Distillation processes can further improve the encoder's feature extraction capabilities, allowing for the creation of more informative features in later decoding phases. Notably, our model performs well even at a high

mask ratio of 90% (0.90), demonstrating its capacity to extract useful insights from the few visible patches used during pre-training, which are critical for downstream fine-tuning tasks.

6.7.2.1. Comparison with classification methods

This section discusses the outcomes obtained on MAE, BEiT, ViT, VGG16, VGG19, MobileNetV2, and ResNet50 with respect to our proposed S/SD-ASD. As seen from Table 6.4, we compare several model's performance on the ASD diagnosis problem by employing supervised and self-supervised learning strategies. In comparison to other models, the suggested approach, S/SD-ASD, exhibits substantial superiority across several criteria. In particular, S/SD-ASD exhibits a 97.5% accuracy rate, surpassing both MAE and BEiT by 6.54% and 6.66%, respectively. S/SD-ASD outperforms supervised models ViT by 9.84%, VGG19 by 27.35%, MobileNetV2 by 6.40%, and ResNet50 by 9.31% in accuracy. Furthermore, S/SD-ASD consistently outperforms other models in measures such as precision, recall, and F1 score, demonstrating its effectiveness in identifying distinguishing characteristics that are pertinent to the diagnosis of ASD. Compared to current state-of-the-art techniques, these results demonstrate the promise of S/SD-ASD as a reliable technique for improving diagnostic accuracy in autism.

Table 6.3 Comparison of models performance on various measures

Model	Accuracy	Precision	Recall	F1
Self-Supervised				
MAE	91.53	93.10	88.80	90.31
BEiT	91.42	95.49	88.76	91.04
Supervised				
ViT	88.77	96.53	80.68	87.98
VGG19	70.15	68.54	66.55	67.53
MobileNetV2	91.10	93.27	88.79	90.97
ResNet50	88.19	93.13	82.79	87.60
Proposed: S/SD-ASD	97.5	97.31	97.89	97.60

During the fine-tuning stage, we finished all of the verifications for the epochs used into the S/SD-ASD technique. To ensure a fair comparison, the S/SD-ASD and MAE algorithms were pre-trained for 200 epochs and both use the ViT-B/16 backbone network structure. Evaluations were carried out across 60 epochs after the pre-trained weights were transferred to the fine-tuning system. Fig 6.9, illustrates the fine-tuning performance of these strategies graphically. It can be seen from Fig 6.9., that the S/SD-ASD

approach continuously produced great accuracy right from the beginning of the epochs. MAE, on the other hand, started out with inferior accuracy. This discrepancy can be explained by the S/SD-ASD method's pre-training reconstruction, which enables it to learn features with more detailed information on facial features.

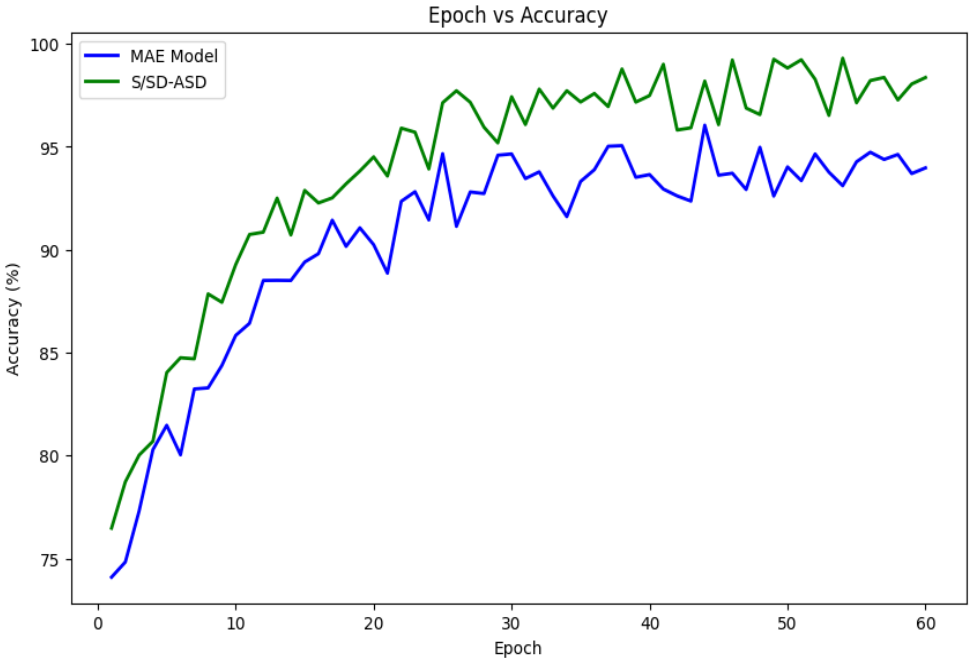


Fig 6.9 Fine-tuning (%) accuracy of MAE with the proposed S/SD-ASD

6.7.2.2. Ablation Study

In this segment, we conducted ablation experiments of our S/SD-ASD model on autism facial image dataset to validate the effectiveness of the proposed approach. We examine the effects of various pre-training objectives on the experimental outcomes in Table 6.5(a). Here, **FT** denotes fine-tuning in the downstream task, **KD** stands for self-distillation, and $\mathbf{I_{REC}}$ stands for masked image reconstruction. As a single pre-training target, the results suggest that masked picture reconstruction has a modest advantage over self-distillation. This benefit is probably caused by the encoder only trains 50% of the visible picture patches, which helps the model reach its full potential by enabling it to extract as many useful features as possible. Furthermore, this procedure might lessen superfluous information in the facial images, which would minimize interference and improve the functionality of the model. The advantages of each strategy are combined when $\mathbf{I_{REC}}$ and **KD** are used, leading to a more successful overall performance. We present experimental investigations of the count of MLP layers on the decoder and encoder sides in Table 6.5(b). Using two linear layers reduces the accuracy of the model for the encoder. Our hypothesis is that the encoded vector will include comparatively more important information if just 50% (0.5) of the viewable

(unmasked) image is fed into the encoder. As a result, adding two linear layers to this vector causes more information to be lost, which reduces the encoder's capacity to extract features. By processing the entire image encoding, on the other hand, the decoder preserves more global information. With just one linear layer, it can successfully generate a vector that is linearly separable. However, some information is lost during transmission and transformation when two linear layers are used.

Table 6.4 S/SD-ASD ablation cases on the facial image data

(a) Pre-training				**(b) MLP layers**				
I_{REC}	✓		✓	**Encoder**	1	1	2	2
KD		✓	✓	**Decoder**	1	2	1	2
FT (%)	*91.37*	*90.56*	*97.5*	***FT (%)***	*97.5*	*92.13*	*95.81*	*95.46*

6.7.2.3. Comparison with existing ASD architectures

Table 6.6 showcases the comparison of various previous architectures with the proposed S/SD-ASD. As seen from the table below, our proposed model, S/SD-ASD, demonstrates the effectiveness of integrating knowledge distillation (KD), vision transformer (ViT), and masked autoencoder (MAE), achieving an unprecedented accuracy of 97.5%. This achievement marks a significant advancement over previous models. The highest prior accuracy reported in the literature was 96%, achieved by Khan et al. (2024). Our model surpasses this by 1.56%, indicating a notable improvement. Hosseini et al. (2022) achieved an accuracy of 94.6%, and our model improves upon this by 2.9%. Lu and Perkowski (2021) attained an accuracy of 95%, which our model exceeds by 2.5%. Additionally, Ahmad et al. (2023) reported an accuracy of 92%, and our model demonstrates a substantial enhancement of 5.5% over this result. These improvements underscore the superior performance and effectiveness of our proposed approach. By integrating KD, ViT, and MAE, we have developed a model that not only outperforms existing methodologies but also sets a new benchmark for accuracy in this domain. The incremental improvements over the best prior results highlight the robustness and innovation inherent in our approach, making a significant contribution to advancing the state of the art.

Table 6.5 Comparative evaluation of ASD classification methods using facial image dataset

Reference	**Year**	**Methods**	**Maximum Accuracy**
Dodia et al. [178]	2024	MobileNet	89.58%
Anjum et al. [179]	2024	MobileNet; EfficientNet; Xception; VGG19; VGG16	88.33%
Kurniawan and	2024	VGG19	75.85%

Gunawan [180]			
Khan et al. [181]	2024	MobileNetV2; ResNet50; Xception; DenseNet121; VGG16	96%
Ahmad et al. [182]	2023	ResNet50; ResNet34; MobileNetV2; AlexNet; VGG19; VGG16	92%
Gaddala et al. [183]	2023	VGG19; VGG16	84%
Hosseini et al. [184]	2022	MobileNet	94.6%
Lu and Perkoswki [185]	2021	VGG16	95%
Proposed: S/SD-ASD	2024	KD; ViT; MAE	97.5%

6.8. Discussion

The neurological illness known as autism spectrum disorder (ASD) is typified by difficulties with communication, social interaction, and repetitive activities. It is essential to recognize and diagnose ASD since early intervention can greatly enhance the lives of those who are impacted. ASD symptoms are influenced by several factors, including brain development, environmental circumstances, and genetic susceptibility. The frequency of ASD has increased noticeably in recent decades, underscoring the critical need for precise and effective diagnostic instruments. In contrast to medical illnesses, the diagnosis of ASD is largely based on clinical evaluations and behavioral observations, which leaves it vulnerable to subjectivity and variation amongst physicians. ASD is a challenge in terms of early detection because of its diverse presentation and lack of conclusive biological indicators. Unfortunately, the lack of complete and labeled datasets required for training robust models causes many current techniques to suffer. This restriction highlights the significance of creating techniques that can use self-supervised learning (SSL) to utilize tiny datasets efficiently. To improve the classification accuracy of ASD, we present in this paper a strategy we call S/SD-ASD, which combines SSL and self-distillation (SD) techniques. To maximize feature extraction capabilities essential for precise classification, S/SD-ASD first masks portions of the input data and then reconstructs the masked regions using visible patches. The approach successfully directs the model toward learning discriminative features pertinent to ASD diagnosis by integrating label information into the reconstruction job. A significant obstacle in ASD research is addressed by the S/SD-ASD methodology, which shows encouraging results in obtaining high classification accuracy even with little labeled data. It reduces the effect of data shortage and improves diagnostic accuracy by enhancing the encoder's feature extraction capabilities through SD loss. Furthermore, the method can be applied practically in clinical situations where data availability is sometimes limited due to its flexibility to small

training datasets. Although S/SD-ASD works well, it has several drawbacks, especially during the reconstruction and labeling stages. Reconstruction with random masking may limit the quality of feature extraction, requiring attention techniques to be used for refining. Furthermore, the reliance on labeled datasets presents difficulties in situations where there are few complete labeled datasets. Further investigations could focus on ways to reduce labeling dependencies by employing contrastive learning and enhancing feature quality by using unlabeled datasets.

6.9. Chapter Summary

This chapter presents a novel, innovative methodology for ASD classification using facial images, combining self-supervised learning with self-distillation. The S/SD-ASD framework effectively tackles challenges such as data scarcity, enhances feature extraction, and achieves superior diagnostic accuracy. It represents a valuable contribution to the fields of autism research and healthcare technology, advancing the development of automated systems for early ASD detection. To ensure the robustness of the proposed approach, an ablation study is conducted, where different components of the model are individually analyzed to confirm their contribution to the overall performance. Additionally, the chapter highlights the significance of this research in the broader context of computer-aided diagnostic systems, particularly in the field of autism. By leveraging deep learning and self-supervised learning, the proposed method provides a promising direction for developing automated, scalable tools to assist healthcare professionals in the timely diagnosis of ASD. This work not only contributes to the academic literature on ASD diagnosis but also has potential implications for clinical applications, offering a reliable and efficient tool to support clinicians in their decision-making processes.

Chapter 7 CONCLUSION AND FUTURE SCOPE

This chapter provides a comprehensive overview of the various approaches employed for Autism Spectrum Disorder detection using artificial intelligence-based frameworks. Section 7.1 outlines the key research contributions of this study. Section 7.2 highlights the limitations of the proposed framework, and Section 7.3 explores potential directions for future work.

7.1. Research Summary

This thesis tackles the pressing challenge of Autism Spectrum Disorder (ASD) diagnosis by introducing an artificial intelligence-based framework that combines computational intelligence and multimodal data integration. ASD diagnosis remains hindered by subjective assessment techniques, limited data interpretability, and insufficient generalizability. The work spans a systematic literature review, the development of advanced diagnosis models, multimodal framework design, and detailed comparative analysis to advance the field of ASD detection and diagnosis.

To achieve Research Objective 1 (RO1), a comprehensive systematic literature review was conducted to explore the landscape of ASD diagnostic methods. The review synthesizes information on traditional diagnostic techniques, machine learning (ML), and deep learning (DL) applications in ASD research, highlighting their limitations such as the lack of multimodal data utilization, overfitting issues, and restricted scalability. By identifying these gaps, this objective lays the groundwork for the development of more robust, accurate, and interpretable models.

For Research Objective 2 (RO2), two advanced ASD diagnosis models were developed to address key challenges in computational intelligence. The first model, AFF-BPL (Adaptive Feature Fusion Technique for the Diagnosis of Autism Spectrum Disorder using Bat-PSO-LSTM-based Framework), introduces a novel feature fusion method. This model employs a hybrid optimization approach combining Bat Algorithm and Particle Swarm Optimization (PSO) to select optimal features, followed by an LSTM-based classifier for precise diagnosis. The adaptive feature fusion enhances the model's ability to process diverse data types while improving classification accuracy and computational efficiency.

The second model, WS-BiTM (Integrating White Shark Optimization with Bi-LSTM for Enhanced Autism Spectrum Disorder Diagnosis), focuses on leveraging White Shark Optimization (WSO) for optimal feature selection. This model integrates WSO with a Bidirectional Long Short-Term Memory (Bi-LSTM) network to improve the model's generalization and mitigate overfitting. Experimental results demonstrate that WS-BiTM outperforms conventional models, particularly in datasets with high-

dimensional and complex features. Both models are validated using paired t-tests, ablation studies, and extensive performance metrics, including accuracy, precision, sensitivity, and specificity.

To address Research Objective 3 (RO3), a multimodal ASD detection framework was designed, incorporating diverse data modalities such as clinical records, brain MRI images, and facial features. This framework employs Channel-Based Attention Combination (CBAC) for effective feature fusion and integrates Vision Transformers with LSTM networks for robust analysis of multimodal inputs. The proposed framework bridges the gaps of single-modality models by leveraging the complementary strengths of different data types, leading to significant improvements in diagnostic accuracy. Evaluations on the ABIDE dataset using leave-one-site-out cross-validation further affirm the efficacy of the proposed framework.

Finally, to fulfill Research Objective 4 (RO4), an extensive comparative analysis was conducted between the proposed models and state-of-the-art ASD diagnosis techniques. The analysis underscores the advantages of the proposed models in terms of accuracy, computational efficiency, and robustness. The study also provides critical insights into the limitations of existing approaches, reinforcing the significance of multimodal data and hybrid optimization techniques for advancing ASD diagnosis research.

The contributions of this thesis lie at the intersection of artificial intelligence, computational intelligence, and healthcare, offering new possibilities for the timely and accurate diagnosis of ASD. The proposed frameworks demonstrate substantial potential for integration into clinical practice, providing valuable tools for early detection and personalized intervention strategies. This research not only enhances diagnostic accuracy but also establishes a foundation for future advancements in artificial intelligence applications for healthcare challenges.

7.2. Limitations of the Work

No one is perfect in the world, and every study has certain limits and constraints. This work is also subject to the following limitations:

- **Dataset Diversity:** The datasets used in this study are limited in terms of demographic and geographical diversity, which may impact the generalizability of the proposed framework across different populations.

- **Sample Size:** Despite employing multiple datasets, the relatively small sample size, especially in specific age groups or demographic categories, may influence the robustness and statistical reliability of the findings.

- **Feature Scope:** The study primarily focuses on features derived from clinical data, MRI scans, and facial images. However, it does not incorporate other potentially impactful modalities such as genetic data or behavioral assessments. The data for non-medical health determinants risk factors for some regions of the study area are not accessible.

- **Focus on Binary Classification:** The framework is tailored for binary classification (ASD vs. non-ASD) and does not explore multi-class problems, such as predicting ASD severity or comorbidities.

- **Dependency on High-Quality Data**: The framework's performance depends on the availability of high-quality and preprocessed data, which may not always be accessible in practical scenarios.

7.3. Future Aspects

Following are the future perspectives of the work:

- *Hybrid Networks for MRI Analysis:* Develop hybrid networks combining deep learning and traditional techniques to analyze MRI scans, enhancing early ASD detection through neuroimaging.
- Integration of more nature-inspired algorithms (e.g., Grey Wolf Optimizer, Ant Colony Optimization) with deep learning models to improve feature selection and model efficiency.
- Analyze functional MRI data, incorporating multimodal neuroimaging techniques (e.g., fMRI, DTI) to study neural mechanisms underlying ASD.
- *Facial Image Data Across Age Groups:* Extend analysis to datasets containing facial images from diverse age groups to identify phenotypic markers of ASD for early detection.
- *Scalable Algorithms for Diverse Datasets:* Develop scalable models to handle large, diverse datasets, ensuring better generalization and clinical applicability.
- *Longitudinal ASD Studies:* Conduct longitudinal studies to understand ASD progression and predict outcomes using time-series analysis and recurrent neural networks.

- *Global Dataset Creation:* Develop globally diverse datasets with demographic variations to enhance model generalization and inclusivity.
- *Integration of Genomic Data:* Explore the inclusion of genomic data to identify potential genetic markers and their correlation with neuroimaging findings.
- *Clinical Validation and Trials:* Collaborate with clinicians to validate AI-based frameworks in real-world settings through clinical trials.

- *Personalized Treatment Insights:* Use AI models to predict individual responses to therapies, aiding in the development of precision medicine strategies for ASD treatment.
- *Cross-Cultural ASD Studies:* Conduct cross-cultural analyses to identify region-specific ASD markers, improving diagnostic models for diverse populations.
- *Synthetic Data Generation:* Employ generative models, such as GANs, to synthesize realistic neuroimaging or facial datasets for augmenting training data and overcoming data scarcity.
- *Collaborative Learning Models:* Explore federated or collaborative learning approaches to train models across decentralized, privacy-preserving datasets from multiple institutions.
- *Multi-View Learning:* Implement multi-view learning techniques to combine information from different sources, such as 2D images, 3D MRI scans, and clinical data, for comprehensive analysis.
- *Ethical and Societal Implications:* Address ethical considerations in ASD diagnosis using AI, focusing on bias mitigation, transparency, and the societal impact of automated decision-making.

References

[1] N. S. Ajmi, D. A. George, M. B. Megha, and J. Mohan, "A Review of Machine Learning Techniques for Detecting Autism Spectrum Disorders," *Int. Conf. Sustain. Comput. Data Commun. Syst. ICSCDS 2022 - Proc.*, pp. 148–155, 2022, doi: 10.1109/ICSCDS53736.2022.9760909.

[2] "abide_I @ fcon_1000.projects.nitrc.org." [Online]. Available: https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html.

[3] "abide_II @ fcon_1000.projects.nitrc.org." [Online]. Available: https://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html.

[4] "AU @ www.nimhgenetics.org." [Online]. Available: https://www.nimhgenetics.org/download-tool/AU.

[5] "national-database-for-autism-research-ndar @ catalog.data.gov." [Online]. Available: https://catalog.data.gov/dataset/national-database-for-autism-research-ndar.

[6] "autism-screening @ www.kaggle.com." [Online]. Available: https://www.kaggle.com/datasets/faizunnabi/autism-screening.

[7] "index @ www.nimhgenetics.org." [Online]. Available: https://www.nimhgenetics.org/.

[8] V. Prasad, G. V. Sriramakrishnan, and I. Diana Jeba Jingle, "Autism spectrum disorder detection using brain MRI image enabled deep learning with hybrid sewing training optimization," *Signal, Image Video Process.*, vol. 17, no. 8, pp. 4001–4008, 2023, doi: 10.1007/s11760-023-02630-y.

[9] T. Thanarajan, Y. Alotaibi, S. Rajendran, and K. Nagappan, "Eye-Tracking Based Autism Spectrum Disorder Diagnosis Using Chaotic Butterfly Optimization with Deep Learning Model," *Comput. Mater. Contin.*, vol. 76, no. 2, pp. 1995–2013, 2023, doi: 10.32604/CMC.2023.039644.

[10] S. Loganathan, C. Geetha, A. R. Nazaren, and M. Harin Fernandez Fernandez, "Autism spectrum disorder detection and classification using chaotic optimization based Bi-GRU network: An weighted average ensemble model," *Expert Syst. Appl.*, vol. 230, no. June, p. 120613, 2023, doi: 10.1016/j.eswa.2023.120613.

[11] C. Vidyadhari, A. Karrothu, P. Manickavasagam, and S. Anjali Devi, "Autism Spectrum Disorder Detection Using Fractional Social Driving Training-Based Optimization Enabled Deep Learning," *Multimed. Tools Appl.*, no. 0123456789, 2023, doi: 10.1007/s11042-023-16784-x.

[12] B. S. Kumar and D. Jayaraj, "Resilient Artificial Fish Swarm Optimization-Based Enhanced Convolutional Neural Network for Autism Spectrum Disorder Classification," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 4, pp. 1297–1310, 2023.

[13] V. Bhandage, M. R. K, S. Muppidi, and B. Maram, "Autism spectrum disorder classification using Adam war strategy optimization enabled deep belief network," *Biomed. Signal Process. Control*, vol. 86, no. PA, p. 104914, 2023, doi: 10.1016/j.bspc.2023.104914.

[14] G. Anurekha and P. Geetha, "An Intelligent Hybrid Ensemble Gene Selection Model for Autism Using DNN," *Intell. Autom. Soft Comput.*, vol. 35, no. 3, pp. 3049–3064, 2023, doi: 10.32604/iasc.2023.029127.

[15] B. Suresh Kumar and D. Jayaraj, "Zealous Particle Swarm Optimization Based Reliable Multi-Layer Perceptron Neural Networks for Autism Spectrum Disorder Classification," *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 1, pp. 301–314, 2023.

[16] A. M. Almars, M. Badawy, and M. A. Elhosseini, "ASD2-TL∗ GTO: Autism spectrum disorders detection via transfer learning with gorilla troops optimizer framework," *Heliyon*, vol. 9, no. 11, p. e21530, 2023, doi: 10.1016/j.heliyon.2023.e21530.

[17] S. Rama Sree *et al.*, "Jellyfish Search Optimization with Deep Learning Driven Autism Spectrum Disorder Classification," *Comput. Mater. Contin.*, vol. 74, no. 1, pp. 2195–2209, 2023, doi: 10.32604/cmc.2023.032586.

[18] S. Kadry, V. E. Jessy, V. Rajinikanth, and R. G. Crespo, "Automatic classification of normal/AD brain MRI slices using whale-algorithm optimized hybrid image features," *J. Ambient Intell. Humaniz. Comput.*, vol. 14, no. 10, pp. 14237–14248, 2023, doi: 10.1007/s12652-023-04662-1.

[19] G. V. Sriramakrishnan, V. V. Rani, S. Thatavarti, and B. Maram, "Chronological pelican remora optimization-enabled deep learning for detection of autism spectrum disorder," *Signal, Image Video Process.*, 2023, doi: 10.1007/s11760-023-02741-6.

[20] R. V. Arumugam and S. Saravanan, “Automated multi-class skin cancer classification using white shark optimizer with ensemble learning classifier on dermoscopy images,” *Multimed. Tools Appl.*, no. 0123456789, 2024, doi: 10.1007/s11042-024-18973-8.

[21] J. Singh, R. Ramya, and V. M., “Dense net with shark mud ring optimization for severity detection of tuberculosis using sputum image,” *Biomed. Signal Process. Control*, vol. 91, no. February, p. 105929, 2024, doi: 10.1016/j.bspc.2023.105929.

[22] A. I. Hammouri, M. S. Braik, H. H. Al-hiary, and R. A. Abdeen, *A binary hybrid sine cosine white shark optimizer for feature selection*, vol. 2. Springer US, 2024.

[23] N. A. Alawad, B. H. Abed-alguni, M. A. Al-Betar, and A. Jaradat, “Binary improved white shark algorithm for intrusion detection systems,” *Neural Comput. Appl.*, vol. 35, no. 26, pp. 19427–19451, 2023, doi: 10.1007/s00521-023-08772-x.

[24] J. Talukdar, D. K. Gogoi, and T. P. Singh, “A comparative assessment of most widely used machine learning classifiers for analysing and classifying autism spectrum disorder in toddlers and adolescents,” *Healthc. Anal.*, vol. 3, no. February, p. 100178, 2023, doi: 10.1016/j.health.2023.100178.

[25] L. Kang, J. Chen, J. Huang, and J. Jiang, “Autism spectrum disorder recognition based on multi-view ensemble learning with multi-site fMRI,” *Cogn. Neurodyn.*, vol. 17, no. 2, pp. 345–355, 2022, doi: 10.1007/s11571-022-09828-9.

[26] Y. Tang, G. Tong, X. Xiong, C. Zhang, H. Zhang, and Y. Yang, “Multi-site diagnostic classification of Autism spectrum disorder using adversarial deep learning on resting-state fMRI,” *Biomed. Signal Process. Control*, vol. 85, no. February, p. 104892, 2023, doi: 10.1016/j.bspc.2023.104892.

[27] S. Loganathan, C. Geetha, A. Rosy, M. Harin, and F. Fernandez, “Autism spectrum disorder detection and classification using chaotic optimization based Bi-GRU network : An weighted average ensemble model,” *Expert Syst. Appl.*, vol. 230, no. February, p. 120613, 2023, doi: 10.1016/j.eswa.2023.120613.

[28] Z. A. T. Ahmed *et al.*, “Facial Features Detection System to Identify Children with Autism Spectrum Disorder: Deep Learning Models,” *Comput. Math. Methods Med.*, vol. 2022, 2022, doi: 10.1155/2022/3941049.

[29] H. Kwon *et al.*, “Sparse Hierarchical Representation Learning on Functional Brain Networks for Prediction of Autism Severity Levels,” *Front. Neurosci.*, vol. 16, no. July, 2022, doi: 10.3389/fnins.2022.935431.

[30] A. Gaspar, D. Oliva, S. Hinojosa, I. Aranguren, and D. Zaldivar, “An optimized Kernel Extreme Learning Machine for the classification of the autism spectrum disorder by using gaze tracking

images," *Appl. Soft Comput.*, vol. 120, p. 108654, 2022, doi: 10.1016/j.asoc.2022.108654.

[31] J. Han, G. Jiang, G. Ouyang, and X. Li, "A Multimodal Approach for Identifying Autism Spectrum Disorders in Children," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2003–2011, 2022, doi: 10.1109/TNSRE.2022.3192431.

[32] F. Sadeghian, H. Hasani, and M. Jafari, "Feature Selection Based on Genetic Algorithm in the Diagnosis of Autism Disorder by fMRI," *Casp. J. Neurol. Sci.*, vol. 7, no. 2, pp. 74–83, 2021, doi: 10.32598/cjns.7.25.5.

[33] N. A. Ali, A. R. Syafeeza, A. S. Jaafar, S. Shamsuddin, and N. K. Nor, "LSTM-based Electroencephalogram Classification on Autism Spectrum Disorder," *Int. J. Integr. Eng.*, vol. 13, no. 6, pp. 321–329, 2021, doi: 10.30880/ijie.13.06.028.

[34] D. Pavithra and A. N. Jayanthi, "An improved adaptive neuro fuzzy interference system for the detection of autism spectrum disorder," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 7, pp. 6885–6897, 2021, doi: 10.1007/s12652-020-02332-0.

[35] Z. Li, Y. Xia, and H. Sahli, "CSA-DE/EDA: a Novel Bio-inspired Algorithm for Function Optimization and Segmentation of Brain MR Images," *Cognit. Comput.*, vol. 11, no. 6, pp. 855–868, 2019, doi: 10.1007/s12559-019-09663-x.

[36] A. T. Umrani and P. Harshavardhanan, "Efficient classification model for anxiety detection in autism using intelligent search optimization based on Deep CNN," *Multimed. Tools Appl.*, no. 0123456789, 2024, doi: 10.1007/s11042-023-17911-4.

[37] P. V. K. Sandeep and N. S. Kumar, "Pain detection through facial expressions in children with autism using deep learning," *Soft Comput.*, vol. 28, no. 5, pp. 4621–4630, 2024, doi: 10.1007/s00500-024-09696-x.

[38] A. El Mouatasim and M. Ikermane, "Control learning rate for autism facial detection via deep transfer learning," *Signal, Image Video Process.*, vol. 17, no. 7, pp. 3713–3720, 2023, doi: 10.1007/s11760-023-02598-9.

[39] P. Moridian *et al.*, "Automatic Autism Spectrum Disorder Detection Using Artificial Intelligence Methods with MRI Neuroimaging: A Review."

[40] G. Wan *et al.*, "FECTS: A Facial Emotion Cognition and Training System for Chinese Children with Autism Spectrum Disorder," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/9213526.

[41] F. Zhang, Y. Wei, J. Liu, Y. Wang, W. Xi, and Y. Pan, "Identification of Autism spectrum disorder based on a novel feature selection method and Variational Autoencoder," *Comput. Biol. Med.*, vol. 148, 2022, doi: 10.1016/j.compbiomed.2022.105854.

[42] Y. Chen *et al.*, "Adversarial Learning Based Node-Edge Graph Attention Networks for Autism

Spectrum Disorder Identification," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–12, 2022, doi: 10.1109/TNNLS.2022.3154755.

[43] C. Pang *et al.*, "Improving model robustness via enhanced feature representation and sample distribution based on cascaded classifiers for computer-aided diagnosis of brain disease," *Biomed. Signal Process. Control*, vol. 79, no. P1, p. 104047, 2023, doi: 10.1016/j.bspc.2022.104047.

[44] N. Wang, D. Yao, L. Ma, and M. Liu, "Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with resting-state fMRI," *Med. Image Anal.*, vol. 75, p. 102279, 2022, doi: 10.1016/j.media.2021.102279.

[45] D. Mason and F. Happé, "The role of alexithymia and autistic traits in predicting quality of life in an online sample," *Res. Autism Spectr. Disord.*, vol. 90, no. June 2021, 2022, doi: 10.1016/j.rasd.2021.101887.

[46] H. Sharif and R. A. Khan, *A Novel Machine Learning Based Framework for Detection of Autism Spectrum Disorder (ASD)*, vol. 36, no. 1. Taylor & Francis, 2022.

[47] G. Jee, S. Chouhan, M. K. Gourisaria, and R. K. Tiwari, "Detection of Autism Spectrum Disorder Through Orthogonal Decomposition and Pearson Correlation for Feature Selection," *4th Int. Conf. Recent Trends Comput. Sci. Technol. ICRTCST 2021 - Proc.*, pp. 103–109, 2022, doi: 10.1109/ICRTCST54752.2022.9781971.

[48] Z. Sherkatghanad *et al.*, "Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network," *Front. Neurosci.*, vol. 13, no. January, pp. 1–12, 2020, doi: 10.3389/fnins.2019.01325.

[49] N. Chaitra, P. A. Vijaya, and G. Deshpande, "Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework," *Biomed. Signal Process. Control*, vol. 62, no. July, p. 102099, 2020, doi: 10.1016/j.bspc.2020.102099.

[50] A. Wawer, I. Chojnicka, L. Okruszek, and J. Sarzynska-Wawer, "Single and Cross-Disorder Detection for Autism and Schizophrenia," *Cognit. Comput.*, vol. 14, no. 1, pp. 461–473, 2022, doi: 10.1007/s12559-021-09834-9.

[51] C. M. Parlett-Pelleriti, E. Stevens, D. Dixon, and E. J. Linstead, "Applications of Unsupervised Machine Learning in Autism Spectrum Disorder Research: a Review," *Rev. J. Autism Dev. Disord.*, no. 0123456789, 2022, doi: 10.1007/s40489-021-00299-y.

[52] P. Lu, X. Li, L. Hu, and L. Lu, "Integrating genomic and resting State fMRI for efficient autism spectrum disorder classification," *Multimed. Tools Appl.*, vol. 81, no. 14, pp. 19183–19194, 2022, doi: 10.1007/s11042-020-10473-9.

[53] M. Wang, J. Guo, Y. Wang, M. Yu, and J. Guo, "Multimodal Autism Spectrum Disorder Diagnosis Method Based on DeepGCN," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp.

3664–3674, 2023, doi: 10.1109/TNSRE.2023.3314516.

[54] J. Li *et al.*, "MMASD: A Multimodal Dataset for Autism Intervention Analysis," *ACM Int. Conf. Proceeding Ser.*, no. 2, pp. 397–405, 2023, doi: 10.1145/3577190.3614117.

[55] C. Song, S. Wang, M. Chen, H. Li, F. Jia, and Y. Zhao, "A multimodal discrimination method for the response to name behavior of autistic children based on human pose tracking and head pose estimation," *Displays*, vol. 76, no. November 2022, p. 102360, 2023, doi: 10.1016/j.displa.2022.102360.

[56] M. M. Hasan, C. N. Watling, and G. S. Larue, "Validation and interpretation of a multimodal drowsiness detection system using explainable machine learning," *Comput. Methods Programs Biomed.*, vol. 243, no. October 2023, 2024, doi: 10.1016/j.cmpb.2023.107925.

[57] Y. Du *et al.*, "A new multimodality fusion classification approach to explore the uniqueness of schizophrenia and autism spectrum disorder," *Hum. Brain Mapp.*, vol. 43, no. 12, pp. 3887–3903, 2022, doi: 10.1002/hbm.25890.

[58] R. Y. Y. Chan, C. M. V. Wong, and Y. N. Yum, "Predicting Behavior Change in Students With Special Education Needs Using Multimodal Learning Analytics," *IEEE Access*, vol. 11, no. June, pp. 63238–63251, 2023, doi: 10.1109/ACCESS.2023.3288695.

[59] X. Fu, E. Patrick, J. Y. H. Yang, D. D. Feng, and J. Kim, "Deep multimodal graph-based network for survival prediction from highly multiplexed images and patient variables," *Comput. Biol. Med.*, vol. 154, no. February, 2023, doi: 10.1016/j.compbiomed.2023.106576.

[60] L. A. Passos, J. P. Papa, J. Del Ser, A. Hussain, and A. Adeel, "Multimodal audio-visual information fusion using canonical-correlated Graph Neural Network for energy-efficient speech enhancement," *Inf. Fusion*, vol. 90, no. September 2022, pp. 1–11, 2023, doi: 10.1016/j.inffus.2022.09.006.

[61] H. D. Le, G. S. Lee, S. H. Kim, S. Kim, and H. J. Yang, "Multi-Label Multimodal Emotion Recognition With Transformer-Based Fusion and Emotion-Level Representation Learning," *IEEE Access*, vol. 11, no. February, pp. 14742–14751, 2023, doi: 10.1109/ACCESS.2023.3244390.

[62] N. Jaafar and Z. Lachiri, "Multimodal fusion methods with deep neural networks and meta-information for aggression detection in surveillance," *Expert Syst. Appl.*, vol. 211, no. July 2022, p. 118523, 2023, doi: 10.1016/j.eswa.2022.118523.

[63] L. Herath, D. Meedeniya, J. Marasinghe, V. Weerasinghe, and T. Tan, "Autism spectrum disorder identification using multi-model deep ensemble classifier with transfer learning," *Expert Syst.*, no. December 2022, pp. 1–23, 2024, doi: 10.1111/exsy.13623.

[64] S. De Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna, and A. M. P. Michalek, "Computational Decision Support System for ADHD Identification," *Int. J. Autom. Comput.*, vol.

18, no. 2, pp. 233–255, 2021, doi: 10.1007/s11633-020-1252-1.

[65] D. Haputhanthri *et al.*, “Integration of Facial Thermography in EEG-based Classification of ASD,” *Int. J. Autom. Comput.*, vol. 17, no. 6, pp. 837–854, 2020, doi: 10.1007/s11633-020-1231-6.

[66] M. Zhang, Q. Cui, Y. Lü, W. Yu, and W. Li, “A multimodal learning machine framework for Alzheimer’s disease diagnosis based on neuropsychological and neuroimaging data,” *Comput. Ind. Eng.*, vol. 197, no. 6, p. 110625, 2024, doi: 10.1016/j.cie.2024.110625.

[67] Q. Yu *et al.*, “A transformer-based unified multimodal framework for Alzheimer’s disease assessment,” *Comput. Biol. Med.*, vol. 180, no. August, 2024, doi: 10.1016/j.compbiomed.2024.108979.

[68] J. Sheng *et al.*, “A hybrid multimodal machine learning model for Detecting Alzheimer’s disease,” *Comput. Biol. Med.*, vol. 170, no. February, p. 108035, 2024, doi: 10.1016/j.compbiomed.2024.108035.

[69] Z. Li *et al.*, “Joint Self-Supervised and Supervised Contrastive Learning for Multimodal MRI Data: Towards Predicting Abnormal Neurodevelopment,” vol. 157, no. August, 2023, [Online]. Available: http://arxiv.org/abs/2312.15064.

[70] M. Zhang, L. Sun, Z. Kong, W. Zhu, Y. Yi, and F. Yan, “Pyramid-attentive GAN for multimodal brain image complementation in Alzheimer’s disease classification,” *Biomed. Signal Process. Control*, vol. 89, no. October 2023, 2024, doi: 10.1016/j.bspc.2023.105652.

[71] J. H. Moon, H. Lee, W. Shin, Y. H. Kim, and E. Choi, “Multi-Modal Understanding and Generation for Medical Images and Text via Vision-Language Pre-Training,” *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 12, pp. 6070–6080, 2022, doi: 10.1109/JBHI.2022.3207502.

[72] O. Alpar, “A mathematical fuzzy fusion framework for whole tumor segmentation in multimodal MRI using Nakagami imaging,” *Expert Syst. Appl.*, vol. 216, no. July 2022, p. 119462, 2023, doi: 10.1016/j.eswa.2022.119462.

[73] M. Tang, P. Kumar, H. Chen, and A. Shrivastava, “Deep multimodal learning for the diagnosis of autism spectrum disorder,” *J. Imaging*, vol. 6, no. 6, 2020, doi: 10.3390/jimaging6060047.

[74] M. K. Elakkiya and Dejey, “Novel deep learning models with novel integrated activation functions for autism screening: AutiNet and MinAutiNet,” *Expert Syst. Appl.*, vol. 238, no. PD, p. 122102, 2024, doi: 10.1016/j.eswa.2023.122102.

[75] H. S. Nogay and H. Adeli, “Multiple Classification of Brain MRI Autism Spectrum Disorder by Age and Gender Using Deep Learning,” *J. Med. Syst.*, vol. 48, no. 1, 2024, doi: 10.1007/s10916-023-02032-0.

[76] N. Li *et al.*, “Joint learning of multi-level dynamic brain networks for autism spectrum disorder

diagnosis," *Comput. Biol. Med.*, vol. 171, no. January, 2024, doi: 10.1016/j.compbiomed.2024.108054.

[77] S. Parui, D. Samanta, N. Chakravorty, U. Ghosh, and J. J. P. C. Rodrigues, "Artificial intelligence and sensor-based autism spectrum disorder diagnosis using brain connectivity analysis ☆ , ☆☆," *Comput. Electr. Eng.*, vol. 108, no. April, p. 108720, 2023, doi: 10.1016/j.compeleceng.2023.108720.

[78] M. Mishra and U. C. Pati, "A classification framework for Autism Spectrum Disorder detection using sMRI: Optimizer based ensemble of deep convolution neural network with on-the-fly data augmentation," *Biomed. Signal Process. Control*, vol. 84, no. February, p. 104686, 2023, doi: 10.1016/j.bspc.2023.104686.

[79] S. Dc, B. Gadgay, S. Farheen, and M. A. Waheed, "A Machine Learning Approach for Early Detection and Diagnosis of Autism and Normal Controls and Estimating Severity Levels Based on Face Recognition," *2022 Int. Conf. Emerg. Trends Eng. Med. Sci. ICETEMS 2022*, pp. 35–40, 2022, doi: 10.1109/ICETEMS56252.2022.10093412.

[80] B. A. Kumar and N. K. Misra, "Masked face age and gender identification using CAFFE-modified MobileNetV2 on photo and real-time video images by transfer learning and deep learning techniques," *Expert Syst. Appl.*, vol. 246, no. October 2023, p. 123179, 2024, doi: 10.1016/j.eswa.2024.123179.

[81] E. Özbay, F. A. Özbay, and F. S. Gharehchopogh, "Kidney Tumor Classification on CT images using Self-supervised Learning," *Comput. Biol. Med.*, vol. 176, no. May, 2024, doi: 10.1016/j.compbiomed.2024.108554.

[82] Z. Tan, Y. Yu, J. Meng, S. Liu, and W. Li, "Self-supervised learning with self-distillation on COVID-19 medical image classification," *Comput. Methods Programs Biomed.*, vol. 243, no. October 2023, p. 107876, 2024, doi: 10.1016/j.cmpb.2023.107876.

[83] S. Yang *et al.*, "A self-supervised image aesthetic assessment combining masked image modeling and contrastive learning," *J. Vis. Commun. Image Represent.*, vol. 101, no. May, p. 104184, 2024, doi: 10.1016/j.jvcir.2024.104184.

[84] Y. Bai *et al.*, "Masked autoencoders with handcrafted feature predictions: Transformer for weakly supervised esophageal cancer classification," *Comput. Methods Programs Biomed.*, vol. 244, no. October 2023, p. 107936, 2024, doi: 10.1016/j.cmpb.2023.107936.

[85] X. Ma, C. Liu, C. Xie, L. Ye, Y. Deng, and X. Ji, "Disjoint Masking With Joint Distillation for Efficient Masked Image Modeling," *IEEE Trans. Multimed.*, vol. 26, pp. 3077–3087, 2024, doi: 10.1109/TMM.2023.3306840.

[86] Z. Chen, D. Agarwal, K. Aggarwal, W. Safta, M. M. Balan, and K. Brown, "Masked Image

Modeling Advances 3D Medical Image Analysis," *Proc. - 2023 IEEE Winter Conf. Appl. Comput. Vision, WACV 2023*, pp. 1969–1979, 2023, doi: 10.1109/WACV56688.2023.00201.

[87] L. Qi, Z. Jiang, W. Shi, F. Qu, and G. Feng, "GMIM: Self-supervised pre-training for 3D medical image segmentation with adaptive and hierarchical masked image modeling," *Comput. Biol. Med.*, vol. 176, no. September 2023, p. 108547, 2024, doi: 10.1016/j.compbiomed.2024.108547.

[88] C. Liu, Y. Cheng, and S. Tamura, "Masked image modeling-based boundary reconstruction for 3D medical image segmentation," *Comput. Biol. Med.*, vol. 166, no. September, p. 107526, 2023, doi: 10.1016/j.compbiomed.2023.107526.

[89] L. Qi, W. Shi, Y. Miao, Y. Li, G. Feng, and Z. Jiang, "Intra-modality masked image modeling: A self-supervised pre-training method for brain tumor segmentation," *Biomed. Signal Process. Control*, vol. 95, no. PA, p. 106343, 2024, doi: 10.1016/j.bspc.2024.106343.

[90] C. Jayakumaran and J. D. Sweetlin, "An optimized neural network with inertia weight variation of PSO for the detection of autism," *Proc. Int. Conf. Smart Technol. Comput. Electr. Electron. ICSTCEE 2020*, pp. 439–445, 2020, doi: 10.1109/ICSTCEE49637.2020.9277247.

[91] Z. Ma, G. Wu, P. N. Suganthan, A. Song, and Q. Luo, "Performance assessment and exhaustive listing of 500+ nature-inspired metaheuristic algorithms," *Swarm Evol. Comput.*, vol. 77, no. February 2022, p. 101248, 2023, doi: 10.1016/j.swevo.2023.101248.

[92] F. R. Llorella, J. M. Azorín, and G. Patow, "Black hole algorithm with convolutional neural networks for the creation of brain-computer interface based in visual perception and visual imagery," *Neural Comput. Appl.*, vol. 35, no. 8, pp. 5631–5641, 2023, doi: 10.1007/s00521-022-07542-5.

[93] H. M. Balaha and A. E. S. Hassan, *Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm*, vol. 35, no. 1. Springer London, 2023.

[94] Y. Liang, Y. Lin, and Q. Lu, "Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM," *Expert Syst. Appl.*, vol. 206, no. June, p. 117847, 2022, doi: 10.1016/j.eswa.2022.117847.

[95] M. M. Eid *et al.*, "Meta-Heuristic Optimization of LSTM-Based Deep Network for Boosting the Prediction of Monkeypox Cases," *Mathematics*, vol. 10, no. 20, pp. 1–20, 2022, doi: 10.3390/math10203845.

[96] M. Shehab *et al.*, *A Comprehensive Review of Bat Inspired Algorithm: Variants, Applications, and Hybridization*, vol. 30, no. 2. Springer Netherlands, 2023.

[97] A. S. Alphonse, S. Abinaya, and K. S. Arikumar, "A Novel Monogenic Sobel Directional Pattern (MSDP) and Enhanced Bat Algorithm-Based Optimization (BAO) with Pearson Mutation (PM) for Facial Emotion Recognition," *Electron.*, vol. 12, no. 4, 2023, doi:

10.3390/electronics12040836.

[98] K. Balasubramanian, K. Ramya, and K. Gayathri Devi, “Optimized adaptive neuro-fuzzy inference system based on hybrid grey wolf-bat algorithm for schizophrenia recognition from EEG signals,” *Cogn. Neurodyn.*, vol. 0123456789, 2022, doi: 10.1007/s11571-022-09817-y.

[99] S. Pulipati, R. Somula, and B. R. Parvathala, “Nature inspired link prediction and community detection algorithms for social networks: a survey,” *Int. J. Syst. Assur. Eng. Manag.*, 2021, doi: 10.1007/s13198-021-01125-8.

[100] N. A. Ali, A. R. Syafeeza, A. S. Jaafar, and M. K. M. F. Alif, “Autism spectrum disorder classification on electroencephalogram signal using deep learning algorithm,” *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, pp. 91–99, 2020, doi: 10.11591/ijai.v9.i1.pp91-99.

[101] H. Li *et al.*, “Newly Emerging Nature-Inspired Optimization-Algorithm Review, Unified Framework, Evaluation, and Behavioural Parameter Optimization,” *IEEE Access*, vol. 8, pp. 72620–72649, 2020, doi: 10.1109/ACCESS.2020.2987689.

[102] L. Jain, R. Katarya, and S. Sachdeva, “Opinion leader detection using whale optimization algorithm in online social network,” *Expert Syst. Appl.*, vol. 142, p. 113016, 2020, doi: 10.1016/j.eswa.2019.113016.

[103] B. Khorram and M. Yazdi, “A New Optimized Thresholding Method Using Ant Colony Algorithm for MR Brain Image Segmentation,” *J. Digit. Imaging*, vol. 32, no. 1, pp. 162–174, 2019, doi: 10.1007/s10278-018-0111-x.

[104] R. Abitha and S. Mary Vennila, “A Swarm Based Symmetrical Uncertainty Feature Selection Method for Autism Spectrum Disorders,” *Proc. 3rd Int. Conf. Inven. Syst. Control. ICISC 2019*, no. Icisc, pp. 665–669, 2019, doi: 10.1109/ICISC44355.2019.9036454.

[105] R. P. Cherian, N. Thomas, and S. Venkitachalam, “Weight optimized neural network for heart disease prediction using hybrid lion plus particle swarm algorithm,” *J. Biomed. Inform.*, vol. 110, no. August 2019, p. 103543, 2020, doi: 10.1016/j.jbi.2020.103543.

[106] N. S. Kumar, J. Mahil, A. S. Shiji, and K. P. Joshua, “Detection of Autism in Children by the EEG Behavior Using Hybrid Bat Algorithm-Based ANFIS Classifier,” *Circuits, Syst. Signal Process.*, vol. 39, no. 2, pp. 674–697, 2020, doi: 10.1007/s00034-019-01197-9.

[107] K. Choudhary and N. Goel, “A review on face recognition techniques,” *Int. Conf. Commun. Electron. Syst. Des.*, vol. 8760, no. 4, p. 87601E, 2013, doi: 10.1117/12.2012238.

[108] L. Zhang *et al.*, “Cognitive Load Measurement in a Virtual Reality-Based Driving System for Autism Intervention,” *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 176–189, 2017, doi: 10.1109/TAFFC.2016.2582490.

[109] X. Guo, K. C. Dominick, A. A. Minai, H. Li, C. A. Erickson, and L. J. Lu, “Diagnosing autism

spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method," *Front. Neurosci.*, vol. 11, no. AUG, pp. 1–19, 2017, doi: 10.3389/fnins.2017.00460.

[110] J. Dhalia Sweetlin, H. K. Nehemiah, and A. Kannan, "Computer aided diagnosis of pulmonary hamartoma from CT scan images using ant colony optimization based feature selection," *Alexandria Eng. J.*, vol. 57, no. 3, pp. 1557–1567, 2018, doi: 10.1016/j.aej.2017.04.014.

[111] X.-S. Yang, S. Deb, S. Fong, X. He, and Y.-X. Zhao, "COVER FEATURE EMERGING COMPUTING PARADIGMS COVER FEATURE EMERGING COMPUTING PARADIGMS and Educational Consultant," 2016.

[112] Y. D. Zhang *et al.*, "Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm," *Multimed. Tools Appl.*, vol. 77, no. 17, pp. 22629–22648, 2018, doi: 10.1007/s11042-017-5023-0.

[113] N. S. D'Souza *et al.*, "Deep sr-DDL: Deep structurally regularized dynamic dictionary learning to integrate multimodal and dynamic functional connectomics data for multidimensional clinical characterizations," *Neuroimage*, vol. 241, no. July, p. 118388, 2021, doi: 10.1016/j.neuroimage.2021.118388.

[114] A. M. E. Khalil, S. E. K. Fateen, and A. Bonilla-Petriciolet, "MAKHA-A new hybrid swarm intelligence global optimization algorithm," *Algorithms*, vol. 8, no. 2, pp. 336–365, 2015, doi: 10.3390/a8020336.

[115] S. M. Kesavan, S. Tanavade, M. Al Balushi, S. S. Al Araimi, and A. Al Khazraji, "Development of heuristic algorithm based tool to extract and evaluate tumour section from brain MRI and CT image," pp. 375–379, 2021, doi: 10.1049/icp.2021.0878.

[116] L. Herath, D. Meedeniya, J. Marasingha, and V. Weerasinghe, "Optimize Transfer Learning for Autism Spectrum Disorder Classification with Neuroimaging: A Comparative Study," *ICARC 2022 - 2nd Int. Conf. Adv. Res. Comput. Towar. a Digit. Empower. Soc.*, pp. 171–176, 2022, doi: 10.1109/ICARC54489.2022.9753949.

[117] F. G. Mohammadi, M. H. Amini, and H. R. Arabnia, "Applications of nature-inspired algorithms for dimension reduction: Enabling efficient data analytics," *Adv. Intell. Syst. Comput.*, vol. 1123, no. April, pp. 67–84, 2020, doi: 10.1007/978-3-030-34094-0_4.

[118] S. K. Suguna, R. Ranganathan, J. Sangeetha, S. Shandilya, and S. K. Shandilya, *Application of nature—inspired algorithms in medical image processing*. Springer International Publishing, 2019.

[119] D. E. Gbenga and E. I. Ramlan, "Understanding the limitations of particle swarm algorithm for dynamic optimization tasks: A survey towards the singularity of PSO for swarm robotic

applications," *ACM Comput. Surv.*, vol. 49, no. 1, 2016, doi: 10.1145/2906150.

[120] "autism-screening-on-adults @ www.kaggle.com." [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/autism-screening-on-adults.

[121] Y. S. Lin, S. S. F. Gau, and C. C. Lee, "A Multimodal Interlocutor-Modulated Attentional BLSTM for Classifying Autism Subgroups during Clinical Interviews," *IEEE J. Sel. Top. Signal Process.*, vol. 14, no. 2, pp. 299–311, 2020, doi: 10.1109/JSTSP.2020.2970578.

[122] K. Khan and R. Katarya, "Machine Learning Techniques for Autism Spectrum Disorder: Current trends and future directions," *2023 Int. Conf. Innov. Trends Inf. Technol. ICITIIT 2023*, no. 1, pp. 1–7, 2023, doi: 10.1109/ICITIIT57246.2023.10068658.

[123] E. Waizbard-Bartov, D. Fein, C. Lord, and D. G. Amaral, "Autism severity and its relationship to disability," *Autism Res.*, vol. 16, no. 4, pp. 685–696, 2023, doi: 10.1002/aur.2898.

[124] M. Khodatars *et al.*, "Deep learning for neuroimaging-based diagnosis and rehabilitation of Autism Spectrum Disorder: A review," *Comput. Biol. Med.*, vol. 139, 2021, doi: 10.1016/j.compbiomed.2021.104949.

[125] D. Lasantha, S. Vidanagamachchi, and S. Nallaperuma, "CRIECNN: Ensemble convolutional neural network and advanced feature extraction methods for the precise forecasting of circRNA-RBP binding sites," *Comput. Biol. Med.*, vol. 174, no. March, p. 108466, 2024, doi: 10.1016/j.compbiomed.2024.108466.

[126] L. O. Lyra, A. E. Fabris, and J. B. Florindo, "A multilevel pooling scheme in convolutional neural networks for texture image recognition," *Appl. Soft Comput.*, vol. 152, no. December 2023, 2024, doi: 10.1016/j.asoc.2024.111282.

[127] Z. Tan, H. Madzin, B. Norafida, R. W. O. Rahmat, F. Khalid, and P. S. Sulaiman, "SwinUNeLCsT: Global–local spatial representation learning with hybrid CNN–transformer for efficient tuberculosis lung cavity weakly supervised semantic segmentation," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 4, p. 102012, 2024, doi: 10.1016/j.jksuci.2024.102012.

[128] R. K. Vasanthakumari, R. V. Nair, and V. G. Krishnappa, "Improved learning by using a modified activation function of a Convolutional Neural Network in multi-spectral image classification," *Mach. Learn. with Appl.*, vol. 14, no. October, p. 100502, 2023, doi: 10.1016/j.mlwa.2023.100502.

[129] E. Özbay and F. Altunbey Özbay, "Interpretable features fusion with precision MRI images deep hashing for brain tumor detection," *Comput. Methods Programs Biomed.*, vol. 231, 2023, doi: 10.1016/j.cmpb.2023.107387.

[130] Y. C. Tseng, C. W. Kuo, W. C. Peng, and C. C. Hung, "al-BERT: a semi-supervised denoising technique for disease prediction," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, pp. 1–19, 2024, doi: 10.1186/s12911-024-02528-w.

[131] L. Min, Z. Fan, F. Dou, J. Sun, C. Luo, and Q. Lv, “Adaption BERT for Medical Information Processing with ChatGPT and Contrastive Learning,” *Electron.*, vol. 13, no. 13, 2024, doi: 10.3390/electronics13132431.

[132] H. Muizelaar, M. Haas, K. van Dortmont, P. van der Putten, and M. Spruit, “Extracting patient lifestyle characteristics from Dutch clinical text with BERT models,” *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, pp. 1–15, 2024, doi: 10.1186/s12911-024-02557-5.

[133] “64dece7304f468474ed45342880865a4e7eb119d @ fcon_1000.projects.nitrc.org.” [Online]. Available: https://fcon_1000.projects.nitrc.org/indi/abide/.

[134] X. Yang, N. Zhang, and P. Schrader, “Machine Learning with Applications A study of brain networks for autism spectrum disorder classification using resting-state functional connectivity,” *Mach. Learn. with Appl.*, vol. 8, no. March, p. 100290, 2022, doi: 10.1016/j.mlwa.2022.100290.

[135] L. Herath, D. Meedeniya, M. A. J. C. Marasingha, and V. Weerasinghe, “Autism spectrum disorder diagnosis support model using Inception V3,” *Proc. - Int. Res. Conf. Smart Comput. Syst. Eng. SCSE 2021*, vol. 4, pp. 1–7, 2021, doi: 10.1109/SCSE53661.2021.9568314.

[136] M. Rakić, M. Cabezas, K. Kushibar, A. Oliver, and X. Lladó, “Improving the detection of autism spectrum disorder by combining structural and functional MRI information,” *NeuroImage Clin.*, vol. 25, no. January, p. 102181, 2020, doi: 10.1016/j.nicl.2020.102181.

[137] Z. Rakhimberdina, X. Liu, and T. Murata, “Population graph-based multi-model ensemble method for diagnosing autism spectrum disorder,” *Sensors (Switzerland)*, vol. 20, no. 21, pp. 1–18, 2020, doi: 10.3390/s20216001.

[138] S. W. H. Khor, A. Q. Md Sabri, and A. Othmani, “Autism classification and monitoring from predicted categorical and dimensional emotions of video features,” *Signal, Image Video Process.*, 2023, doi: 10.1007/s11760-023-02699-5.

[139] A. Qayyum, I. Razzak, M. Tanveer, and M. Mazher, “Spontaneous Facial Behavior Analysis using Deep Transformer Based Framework for Child–Computer Interaction,” *ACM Trans. Multimed. Comput. Commun. Appl.*, 2022, doi: 10.1145/3539577.

[140] R. Jayasudha, C. Suragali, J. T. Thirukrishna, and B. Santhosh Kumar, “Hybrid optimization enabled deep learning-based ensemble classification for heart disease detection,” *Signal, Image Video Process.*, 2023, doi: 10.1007/s11760-023-02656-2.

[141] M. Wang, D. Zhang, J. Huang, P. T. Yap, D. Shen, and M. Liu, “Identifying Autism Spectrum Disorder with Multi-Site fMRI via Low-Rank Domain Adaptation,” *IEEE Trans. Med. Imaging*, vol. 39, no. 3, pp. 644–655, 2020, doi: 10.1109/TMI.2019.2933160.

[142] M. C. Caschera, P. Grifoni, and F. Ferri, “Emotion Classification from Speech and Text in Videos Using a Multimodal Approach,” *Multimodal Technol. Interact.*, vol. 6, no. 4, 2022, doi:

10.3390/mti6040028.

[143] S. M. M. Hasan, P. Uddin, A. Ulhaq, and G. Krishnamoorthy, “A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders,” *IEEE Access*, vol. 11, no. December 2022, pp. 15038–15057, 2023, doi: 10.1109/ACCESS.2022.3232490.

[144] T. Zhang, S. Li, B. Chen, H. Yuan, and C. L. P. Chen, “AIA-Net: Adaptive Interactive Attention Network for Text–Audio Emotion Recognition,” *IEEE Trans. Cybern.*, pp. 1–13, 2022, doi: 10.1109/TCYB.2022.3195739.

[145] S. Timms, S. Lodhi, J. Bruce, and E. Stapleton, “Auditory symptoms and autistic spectrum disorder: A scoping review and recommendations for future research,” *J. Otol.*, no. xxxx, 2022, doi: 10.1016/j.joto.2022.08.004.

[146] B. Elshoky, O. Ibrahim, and A. Ali, “Machine Learning Techniques Based on Feature Selection for Improving Autism Disease Classification,” *Int. J. Intell. Comput. Inf. Sci.*, vol. 21, no. 2, pp. 65–81, 2021, doi: 10.21608/ijicis.2021.61582.1058.

[147] M. Hosseinzadeh *et al.*, “A review on diagnostic autism spectrum disorder approaches based on the Internet of Things and Machine Learning,” *J. Supercomput.*, vol. 77, no. 3, pp. 2590–2608, 2021, doi: 10.1007/s11227-020-03357-0.

[148] F. Thabtah, “Machine learning in autistic spectrum disorder behavioral research: A review and ways forward,” *Informatics Heal. Soc. Care*, vol. 44, no. 3, pp. 278–297, 2019, doi: 10.1080/17538157.2017.1399132.

[149] M. S. Jaliaawala and R. A. Khan, “Can autism be catered with artificial intelligence-assisted intervention technology? A comprehensive survey,” *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 1039–1069, 2020, doi: 10.1007/s10462-019-09686-8.

[150] S. Raj and S. Masood, “Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques,” *Procedia Comput. Sci.*, vol. 167, no. 2019, pp. 994–1004, 2020, doi: 10.1016/j.procs.2020.03.399.

[151] G. A. G. Y. M. Y. S. Dhanyatha Sriram, “IRJET- Prediction of Autism Spectrum Disorder based on Machine Learning Approach,” *Irjet*, vol. 8, no. 7, 2021.

[152] T. Ciceri, L. Squarcina, A. Giubergia, A. Bertoldo, P. Brambilla, and D. Peruzzo, “Review on deep learning fetal brain segmentation from Magnetic Resonance images,” *Artif. Intell. Med.*, vol. 143, no. December 2022, p. 102608, 2023, doi: 10.1016/j.artmed.2023.102608.

[153] H. Selcuk Nogay and H. Adeli, “Diagnostic of autism spectrum disorder based on structural brain MRI images using, grid search optimization, and convolutional neural networks,” *Biomed. Signal Process. Control*, vol. 79, no. P2, p. 104234, 2023, doi: 10.1016/j.bspc.2022.104234.

[154] B. Pandey, D. Kumar Pandey, B. Pratap Mishra, and W. Rhmann, “A comprehensive survey of

deep learning in the field of medical imaging and medical natural language processing: Challenges and research directions," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5083–5099, 2022, doi: 10.1016/j.jksuci.2021.01.007.

[155] W. Yin, S. Mostafa, and F. X. Wu, "Diagnosis of Autism Spectrum Disorder Based on Functional Brain Networks with Deep Learning," *J. Comput. Biol.*, vol. 28, no. 2, pp. 146–165, 2021, doi: 10.1089/cmb.2020.0252.

[156] M. U. Emon, M. S. Keya, A. R. Sozib, S. Islam, F. A. Imran, and R. Zannat, "A Comparative Analysis of Autistic Spectrum Disorder (ASD) Disease for Children using ML Approaches," no. 03, pp. 121–126, 2021.

[157] M. Radhakrishnan, K. Ramamurthy, K. K. Choudhury, D. Won, and T. A. Manoharan, "Performance analysis of deep learning models for detection of Autism Spectrum Disorder from EEG signals," *Trait. du Signal*, vol. 38, no. 3, pp. 853–863, 2021, doi: 10.18280/ts.380332.

[158] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage Clin.*, vol. 17, no. November 2016, pp. 16–23, 2018, doi: 10.1016/j.nicl.2017.08.017.

[159] B. Banire, D. Al Thani, M. Qaraqe, and B. Mansoor, "Face-Based Attention Recognition Model for Children with Autism Spectrum Disorder," *J. Healthc. Informatics Res.*, vol. 5, no. 4, pp. 420–445, 2021, doi: 10.1007/s41666-021-00101-y.

[160] M. A. Rahaman *et al.*, "Deep multimodal predictome for studying mental disorders," *Hum. Brain Mapp.*, no. July, pp. 1–14, 2022, doi: 10.1002/hbm.26077.

[161] M. F. H. Siddiqui, P. Dhakal, X. Yang, and A. Y. Javaid, "A Survey on Databases for Multimodal Emotion Recognition and an Introduction to the VIRI (Visible and InfraRed Image) Database," *Multimodal Technol. Interact.*, vol. 6, no. 6, 2022, doi: 10.3390/mti6060047.

[162] J. Liu, "Image Classification Algorithm Based on Deep Learning-Kernel Function," vol. 2020, no. 1, 2020.

[163] S. S. Joudar, A. S. Albahri, and R. A. Hamid, "Intelligent triage method for early diagnosis autism spectrum disorder (ASD) based on integrated fuzzy multi-criteria decision-making methods," *Informatics Med. Unlocked*, vol. 36, no. November 2022, p. 101131, 2023, doi: 10.1016/j.imu.2022.101131.

[164] A. Z. Guo, "Automated Autism Detection based on Characterizing Observable Patterns from Photos," *IEEE Trans. Affect. Comput.*, vol. X, no. X, pp. 1–6, 2020, doi: 10.1109/TAFFC.2020.3035088.

[165] R. Sadik, S. Anwar, and L. Reza, "AutismNet: Recognition of Autism Spectrum Disorder from Facial Expressions using MobileNet Architecture," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 10,

no. 1, pp. 327–334, 2021, doi: 10.30534/ijatcse/2021/471012021.

[166] J. S. Roy and P. G. Datta, “Autism spectrum disorder and detection of autism.,” *Mymensingh Med. J.*, vol. 21, no. 1, pp. 188–189, 2012.

[167] W. Y. Kurniawan, P. H. Gunawan, and N. Aquarini, “Comparison of the Keras-LSTM Algorithms for Classifying Autism Spectrum Disorder Using Facial Images,” *2023 Int. Conf. Data Sci. Its Appl. ICoDSA 2023*, pp. 7–12, 2023, doi: 10.1109/ICoDSA58501.2023.10276630.

[168] H. Alkahtani, T. H. H. Aldhyani, and M. Y. Alzahrani, “Deep Learning Algorithms to Identify Autism Spectrum Disorder in Children-Based Facial Landmarks,” *Appl. Sci.*, vol. 13, no. 8, 2023, doi: 10.3390/app13084855.

[169] Y. Khosla, P. Ramachandra, and N. Chaitra, “Detection of autistic individuals using facial images and deep learning,” *CSITSS 2021 - 2021 5th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solut. Proc.*, pp. 1–5, 2021, doi: 10.1109/CSITSS54238.2021.9683205.

[170] S. R. Arumugam, S. G. Karuppasamy, S. Gowr, O. Manoj, and K. Kalaivani, “A Deep Convolutional Neural Network based Detection System for Autism Spectrum Disorder in Facial images,” *Proc. 5th Int. Conf. I-SMAC (IoT Soc. Mobile, Anal. Cloud), I-SMAC 2021*, pp. 1255–1259, 2021, doi: 10.1109/I-SMAC52330.2021.9641046.

[171] P. Reddy and J. Andrew, “Diagnosis of Autism in Children Using Deep Learning Techniques by Analyzing Facial Features †,” 2024.

[172] T. Akter *et al.*, “Improved transfer-learning-based facial recognition framework to detect autistic children at an early stage,” *Brain Sci.*, vol. 11, no. 6, 2021, doi: 10.3390/brainsci11060734.

[173] T. Farhat, S. Akram, H. S. AlSagri, Z. Ali, A. Ahmad, and A. Jaffar, “Facial Image-Based Autism Detection: A Comparative Study of Deep Neural Network Classifiers,” *Comput. Mater. Contin.*, vol. 0, no. 0, pp. 1–10, 2023, doi: 10.32604/cmc.2023.045022.

[174] I. Ahmad, J. Rashid, M. Faheem, A. Akram, N. A. Khan, and R. ul Amin, “Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks,” *Healthc. Technol. Lett.*, no. August 2023, pp. 1–13, 2024, doi: 10.1049/htl2.12073.

[175] S. C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, “Self-supervised learning for medical image classification: a systematic review and implementation guidelines,” *npj Digit. Med.*, vol. 6, no. 1, 2023, doi: 10.1038/s41746-023-00811-0.

[176] L. Huang, Y. Chen, and X. He, “Spectral-Spatial Masked Transformer With Supervised and Contrastive Learning for Hyperspectral Image Classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, 2023, doi: 10.1109/TGRS.2023.3264235.

[177] “autism-image-data @ www.kaggle.com.” [Online]. Available:

https://www.kaggle.com/datasets/cihan063/autism-image-data?resource=download.

[178] S. Dodia, V. Meshram, J. Kasle, S. Gomase, H. Amrit, and R. Sarse, “Autism Spectrum Disorder (ASD) Detection from Facial Images using MobileNet,” pp. 1–7, 2024, doi: 10.1109/i2ct61223.2024.10543439.

[179] J. Anjum, N. A. Hia, A. Waziha, and K. A. Kalpoma, “Deep Learning-Based Feature Extraction from Children’s Facial Images for Autism Spectrum Disorder Detection,” *ACM Int. Conf. Proceeding Ser.*, pp. 155–159, 2024, doi: 10.1145/3660853.3660888.

[180] W. Y. Kurniawan and P. H. Gunawan, “Classification of Autism Spectrum Disorder Based on Facial Images Using the VGG19 Algorithm,” *J. Comput. Sci. Eng.*, vol. 18, no. 1, pp. 1–9, 2024, doi: 10.5626/JCSE.2024.18.1.1.

[181] B. Khan, S. M. Bhatti, and A. Akram, “Autism Spectrum Disorder Detection in Children Via Deep Learning Models Based on Facial Images,” *Bull. Bus. Econ.*, vol. 13, no. 1, 2024, doi: 10.61506/01.00241.

[182] I. Ahmad, J. Rashid, M. Faheem, A. Akram, N. A. Khan, and R. ul Amin, “Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks,” *Healthc. Technol. Lett.*, no. December 2023, pp. 1–13, 2024, doi: 10.1049/htl2.12073.

[183] L. K. Gaddala *et al.*, “Autism Spectrum Disorder Detection Using Facial Images and Deep Convolutional Neural Networks,” *Rev. d’Intelligence Artif.*, vol. 37, no. 3, pp. 801–806, 2023, doi: 10.18280/ria.370329.

[184] M. P. Hosseini, M. Beary, A. Hadsell, R. Messersmith, and H. Soltanian-Zadeh, “Deep Learning for Autism Diagnosis and Facial Analysis in Children,” *Front. Comput. Neurosci.*, vol. 15, no. January, pp. 1–7, 2022, doi: 10.3389/fncom.2021.789998.

[185] A. Lu and M. Perkowski, “Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application,” *Brain Sci.*, vol. 11, no. 11, 2021, doi: 10.3390/brainsci11111446.

List of Publications

Journal Publication

1. Khan, Kainat, and Rahul Katarya. "AFF-BPL: An adaptive feature fusion technique for the diagnosis of autism spectrum disorder using Bat-PSO-LSTM based framework." Journal of Computational Science 83 (2024): 102447. https://doi.org/10.1016/j.jocs.2024.102447, (Impact Factor: 3.1, Publisher: Elsevier), ***(SCIE Indexed-Published)***

2. Khan, Kainat, and Rahul Katarya. " WS-BiTM: Integrating White Shark Optimization with Bi-LSTM for Enhanced Autism Spectrum Disorder Diagnosis" Journal of Neuroscience Methods (2024): 110319. https://doi.org/10.1016/j.jneumeth.2024.110319, (Impact Factor: 2.7, Publisher: Elsevier), ***(SCIE Indexed-Published)***

3. K. Khan and R. Katarya, "MCBERT: A Multi-Modal Framework for the Diagnosis of Autism Spectrum Disorder", in Biological Psychology, vol. 194, 108976, 2024, doi.org/10.1016/j.biopsycho.2024.108976, (Impact Factor: 2.8, Publisher: Elsevier), ***(SCIE Indexed-Published)***

Conference Publication

1. K. Khan and R. Katarya, "Machine Learning Techniques for Autism Spectrum Disorder: current trends and future directions," 2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT), Kottayam, India, 2023, pp. 1-7, doi: 10.1109/ICITIIT57246.2023.10068658. **(*Published) (Scopus-Indexed)***

2. K. Khan and R. Katarya, "LSTVision: A Multi-Modal Framework for the Diagnosis of Autism Spectrum Disorder utilizing LSTM and Vision Transformer", 5th International Conference on Smart Sensors and Applications (ICSSA) 2024, Malaysia. **(*Published) (Scopus-Indexed)***

Communicated Journal

1. S/SD-ASD: Self-Supervised and Self-Distillation Learning Approach for Autism Spectrum Disorder Classification using Facial Images [*Engineering Analysis with Boundary Elements, Elsevier, Impact Factor: 4.2*] (***Under Review)***

2. ASD-CEVT: Convolutional Enhanced Vision Transformer Architecture for the Diagnosis of Autism Spectrum Disorder [*Brain Research, Elsevier, Impact Factor: 2.7*] (***Under Review)***

3. Bio-Inspired Techniques in Autism Spectrum Disorder: Comprehensive Survey and Future Trajectories [*Genetic Programming and Evolvable Machines, Springer, Impact Factor: 1.7*] (***Submitted)***

4. Computer-Aided Diagnosis of Autism Spectrum Disorder via Facial Imaging [*IEEE Transactions on Neural Systems and Rehabilitation Engineering, Impact Factor: 4.8]* ***(In Drafting mode)***

AWARD

Won the Best Paper award for my research paper titled " LSTVision: A Multi-Modal Framework for Autism Spectrum Diagnosis Utilizing LSTM and Vision Transformer" at ICSSA 2024, Penang, Malaysia

BIOGRAPHY

(Research Scholar 2022-2024)

Ms. Kainat Khan is currently designated as a Ph.D research scholar in the Department of Computer Science, at Delhi Technological University, Delhi, India. She has completed her M.tech from the University School of Information, Communication, and Technology, Delhi. She has completed an undergraduate degree (B.Tech, Computer Science) from the Noida International University. She has published various research papers in SCIE/SCOPUS/IEEE/ELSEVIER-indexed International Conferences/Journals. Her research areas of interest include Artificial Intelligence, Healthcare, Data mining, Machine learning, Deep learning, and Natural Language Processing. She is currently doing her research on public health surveillance using machine learning techniques. She was also awarded the eminent “Best Paper Award” in 2024 at ICSSA, Penang, Malaysia.