People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

IBN KHALDOUN UNIVERSITY OF TIARET

Dissertation

Presented to:

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
DEPARTEMENT OF COMPUTER SCIENCE

in order to obtain the degree of:

MASTER

Specialty: Computer science engineering

Presented by:
BOUKERMA Malak
BELAID Zohra Inasse

On the theme:

Optimization and Comparison of Deep Learning Models for Early Detection of Autism Spectrum Disorders Using the TASD Dataset

Defended publicly on / /2025 in Tiaret in front the jury composed of:

Mr. BENGHENI Abdelmalek	MCA	Tiaret University	Chairman
Mr. DAOUD Mohamed Amine	MCB	Tiaret University	Supervisor
Mr. MEZZOUG Karim	MAA	Tiaret University	Examiner

2024-2025

Dedication

To my precious parents and beloved siblings,

To my grandparents and all my family,

To all my dear friends,

And to Inasse, my friend and teammate,

This work is dedicated to you.

Malak

I'm so grateful to my parents for their unwavering support, through thick and thin.

And to my sister, brothers, grandparents, thank you for always encouraging me and believing in my potential.

Also my incredible friends especially my teammate Malak.

I also want to extend a huge thank you to the BEKKAR family, my second family, for all their help and invaluable advice throughout my life.

Inasse

Acknowledgement

First and foremost, all praise and gratitude belong to Allah (SWT), whose blessings have guided us through every step of this journey. Without His grace, this accomplishment would not have been possible.

We would like to express our deepest gratitude to our supervisor, Mr. **Daoud Mohamed Amine**, for being not only an exceptional mentor but also a great teacher. His support, guidance, and profound expertise have been a cornerstone of this work. We are truly honored to have had the opportunity to learn from him.

To the members of the jury Mr. **Bengheni Abdelmalek** and Mr**. Mezzoug Karim**, we extend our heartfelt thanks for dedicating your time and effort to evaluate this work. Your thoughtful comments and constructive feedback have not only enriched this thesis but have also contributed significantly to our academic development.

To our beloved families —our parents, siblings, and extended relatives—your unconditional love, sacrifices, and constant encouragement have been our anchor throughout this journey. Your belief in us has been a source of strength and motivation, and we are forever grateful for your support.

To our dear friends, thank you for standing by our side during the challenging moments and celebrating the small victories along the way.

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social communication, restricted interests, and repetitive behaviors. Early detection of ASD is critical, as it significantly enhances the effectiveness of intervention and long-term developmental outcomes. However, traditional diagnostic methods—based on clinical observation, parental interviews, and psychological testing—are often subjective, time-consuming, and inaccessible, particularly in under-resourced regions.

This research explores the use of Artificial Intelligence (AI), specifically machine learning (ML) and deep learning (DL) models, as innovative solutions to enhance the accuracy, speed, and accessibility of ASD diagnosis. Using the Autism Spectrum Disorder Screening Dataset for Toddlers, the study compares the performance of various ML algorithms such as K-Nearest Neighbors (KNN), Decision Tree (DT), and Random Forest (RF), alongside DL architectures like Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM).

Experimental results demonstrate that all models achieved high classification performance, with deep learning models—especially RNN and LSTM—delivering near-perfect accuracy (99.52%).

The study culminated in the development of **Spectria**, a multilingual web-based application that integrates the best-performing models into an accessible, user-friendly diagnostic tool. **Spectria** empowers caregivers and health professionals by offering a fast, free, and culturally adaptable solution for early ASD screening.

Keywords: Autism Spectrum Disorder, Early Detection, Machine Learning, Deep Learning, Artificial Intelligence.

ملخص

اضطراب طيف التوحد (ASD) هو اضطراب عصبي معقد يتميز بصعوبات في التواصل الاجتماعي، والاهتمامات المقيدة، والسلوكيات التكرارية. يُعد الكشف المبكر عن هذا الاضطراب أمرًا بالغ الأهمية، حيث يُسهم بشكل كبير في تحسين فعالية التدخل العلاجي وتطور الحالة على المدى البعيد. ومع ذلك، فإن الطرق التشخيصية التقليدية، والتي تعتمد على الملاحظة السريرية والمقابلات مع أولياء الأمور والاختبارات النفسية، غالبًا ما تكون ذاتية، وتستغرق وقتًا طويلًا، ويصعب الوصول إليها في المناطق ذات الموارد المحدودة.

تتناول هذه الدراسة استخدام تقنيات الذكاء الاصطناعي، وبشكل خاص التعلم الآلي والتعلم العميق، بهدف تطوير أدوات تشخيصية دقيقة وسريعة وسهلة الاستخدام للكشف المبكر عن اضطراب طيف التوحد. تم استخدام قاعدة بيانات "فحص طيف التوحد للأطفال الصغار" لتدريب واختبار عدة نماذج من بينها: أقرب الجيران (KNN)، شجرة القرار (DT)، الغابة العشوائية (RF)، الشبكات العصبية الاصطناعية (ANN)، الشبكات العصبية المتكررة (RNN)، وذاكرة المدى الطويل (LSTM).

أظهرت نتائج التجارب أن جميع النماذج قدمت أداءً عالياً، وكانت النماذج العميقة، وخاصة RNN وLSTM، الأفضل من حيث الدقة حيث وصلت إلى 99.52 %. تؤكد هذه النتائج قدرة النماذج على تحليل الأنماط السلوكية وتحديد مؤشرات التوحد من خلال بيانات محدودة.

تم تتويج هذا العمل بتطوير تطبيق إلكتروني يحمل اسم **Spectria**، وهو منصة متعددة اللغات تدمج النماذج الأكثر دقة في واجهة سهلة الاستخدام. يهدف **Spectria** إلى تسهيل الفحص المبكر لأطفال التوحد وتمكين أولياء الأمور والمهنيين في المجال الصحي من خلال أداة تقنية دقيقة ومجانية.

الكلمات المفتاحية: اضطراب طيف التوحد، الكشف المبكر، التعلم الآلي، التعلم العميق، الذكاء الاصطناعي، أدوات التشخيص، فحص التوحد.

Résumé

Le trouble du spectre de l'autisme (TSA) est un trouble neurodéveloppemental complexe, caractérisé par des difficultés dans la communication sociale, des intérêts restreints et des comportements répétitifs. La détection précoce du TSA est cruciale, car elle améliore considérablement l'efficacité des interventions thérapeutiques et les perspectives de développement à long terme. Cependant, les méthodes de diagnostic traditionnelles, reposant sur l'observation clinique, les entretiens parentaux et les tests psychologiques, présentent souvent des limites en termes de subjectivité, de durée et d'accessibilité, notamment dans les régions sous-équipées.

Cette recherche propose l'utilisation de l'intelligence artificielle, notamment des modèles d'apprentissage automatique (ML) et d'apprentissage profond (DL), pour développer des outils de diagnostic plus précis, rapides et accessibles pour le dépistage du TSA. En utilisant la base de données de dépistage de l'autisme chez les tout-petits, plusieurs algorithmes ont été comparés, parmi lesquels : K-Nearest Neighbors (KNN), Arbre de Décision (DT), Forêt Aléatoire (RF), Réseaux de Neurones Artificiels (ANN), Réseaux Neuronaux Récurrents (RNN) et Mémoire à Long Terme (LSTM).

Les résultats expérimentaux montrent que tous les modèles ont obtenu d'excellentes performances, les modèles profonds—en particulier RNN et LSTM—atteignant une précision presque parfaite de 99,52 %.

Cette étude a abouti au développement de **Spectria**, une application web multilingue intégrant les meilleurs modèles identifiés. **Spectria** permet aux parents et professionnels de santé de réaliser un dépistage rapide, accessible et adapté culturellement, contribuant ainsi à une prise en charge plus précoce et plus efficace du TSA.

Mots-clés : Trouble du spectre de l'autisme, Détection précoce, Apprentissage automatique, Apprentissage profond, Intelligence artificielle, Outils de diagnostic, Dépistage TSA.

Table of content

List of Figures

List of tables

List of abbreviations

ASD - Autism Spectrum Disorder

AI - Artificial Intelligence

ML - Machine Learning

DL - Deep Learning

CNN - Convolutional Neural Network

RNN - Recurrent Neural Network

LSTM - Long Short-Term Memory

NLP - Natural Language Processing

ABIDE - Autism Brain Imaging Data Exchange

fMRI - functional Magnetic Resonance Imaging

sMRI - structural Magnetic Resonance Imaging

ADI-R - Autism Diagnostic Interview-Revised

PDD-NOS - Pervasive Developmental Disorder-Not Otherwise Specified

CDD - Childhood Disintegrative Disorder

DSM-5 - Diagnostic and Statistical Manual of Mental Disorders, 5th Edition

M-CHAT - Modified Checklist for Autism in Toddlers

CHAT - Checklist for Autism in Toddlers

EEG - Electroencephalography

BCI - Brain-Computer Interface

NIFTI - Neuroimaging Informatics Technology Initiative

CSV - Comma-Separated Values

KNN - K-Nearest Neighbors

ANN - Artificial Neural Network

ADOS - Autism Diagnostic Observation Schedule

AQ-10 - Autism Spectrum Quotient-10

PDD - Pervasive Developmental Disorder

WHO - World Health Organization

EI - Early Intervention

DNA - Deoxyribonucleic Acid

RNA - Ribonucleic Acid

MECP2 - Methyl-CpG Binding Protein 2 (gene associated with Rett Syndrome)

B12 - Vitamin B12

D - Vitamin D

BERT - Bidirectional Encoder Representations from Transformers

GPT - Generative Pre-Trained Transformer

T5 - Text-to-Text Transfer Transformer

TP - True Positives

TN - True Negatives

FP -False Positives

FPR- False Positive Rate

ROC -Receiver Operating Characteristic curve

TPR- True Positive Rate

DT- Decision Tree

RF- Random Forest

API- Application Programming Interface

ReLU -Rectified Linear Unit

tanh- Hyperbolic tangent

HTTP-Hypertext Transfer Protocol

General introduction

General introduction

In a world full of children with unique personalities and characteristics, some face challenges that set them apart in significant ways. Among these atypical cases are children whose behaviors, communication styles, and ways of interacting with the world differ from the norm. These differences can often be observed in how they talk, move, and engage with their surroundings. One of the most prominent and widely discussed conditions within this spectrum is Autism Spectrum Disorder (ASD).

The term “autism” was first introduced in the early 20th century, with groundbreaking studies conducted in the 1940s by researchers such as Leo Kanner and Hans Asperger [1]. Over the years, the understanding of autism has evolved significantly, shifting from a narrow perspective to a broader acknowledgment of its complexity and diversity. Today, autism is recognized as a spectrum, encompassing a wide range of abilities and challenges.

Children with autism often experience difficulties in social interaction and communication, as well as repetitive behaviors or restricted interests. These characteristics may manifest in different ways, such as difficulty maintaining eye contact, sensitivity to sensory stimuli, or an intense focus on specific topics. While these traits can present challenges, they also highlight the unique ways individuals with autism experience and interpret the world.

The growing awareness of ASD has shed light on the importance of understanding and supporting children with autism. Research continues to uncover the biological, genetic, and environmental factors that contribute to its development, as well as the interventions and strategies that can improve the quality of life for those affected. By fostering greater acceptance and inclusivity, society can create environments where children with autism can thrive and reach their full potential.

Artificial Intelligence (AI) has revolutionized countless fields and has proven to be a powerful tool for understanding complex challenges and delivering innovative solutions. One area where AI has been a game-changer is in addressing the needs of individuals with Autism Spectrum Disorder (ASD). By leveraging advanced machine learning and deep learning models and sophisticated algorithms, researchers and developers have created tools that not only aid in early diagnosis but also provide tailored interventions and support for children with autism.

1.Research problematic and motivation

1 .1 Problematic

Despite increasing awareness of Autism Spectrum Disorder (ASD), the early and accurate diagnosis of the condition remains a significant challenge due to the wide variability of symptoms across individuals, limited access to specialized diagnostic services—particularly in low-resource settings—and the emotional and practical difficulties faced by families in recognizing and responding to their children's behaviors. Traditional diagnostic tools, such as observational checklists, are often time-consuming and may lack the sensitivity required for early detection, frequently necessitating follow-up assessments. These limitations highlight a critical gap in current diagnostic practices and underscore the urgent need for more efficient, accessible, and sensitive diagnostic approaches to facilitate timely intervention and improve outcomes for individuals with ASD.

1.2 Motivation

Studies have demonstrated that children who receive interventions before age 3 often experience significant improvements in developmental outcomes, including cognitive function, social interaction, and adaptive behaviors [2] .By using machine learning and deep learning to automate parts of the diagnostic process, early ASD identification could become more accessible and efficient, particularly in underserved areas. The goal of this research is to explore and compare machine learning and deep learning models for ASD detection to support global efforts for early and accurate diagnoses.

2.Research Questions

2.1 Primary Research Question

How can different machine learning and deep learning models be utilized to develop accurate, efficient, and accessible diagnostic tools for early detection of Autism Spectrum Disorder (ASD)?

2.2 Sub-questions

- How can artificial intelligence techniques be leveraged to accurately detect autism spectrum disorder (ASD)?
- Which machine learning or deep learning models demonstrate the highest effectiveness in identifying autism-related patterns?

3.Research Aims and Objectives

3.1 Aim

The objective of this study is to conduct a comprehensive comparative analysis of various machine and deep learning models to evaluate their effectiveness in the detection of Autism Spectrum Disorder (ASD). This analysis will not only focus on the accuracy and performance of each model but also consider their adaptability, scalability, and practical applicability across diverse populations and global healthcare settings. By examining these models in varied contexts, the study aims to identify the most suitable approaches for supporting early, reliable, and accessible ASD diagnosis.

3.2 Objectives

- Analyze and summarize current literature on ASD detection, with a focus on needs and challenges.
- Test and propose machine learning models (e.g.RFs, KNNs) and deep learning models (e.g ANNs , LSTMs) using Autism screening data for toddlers dataset.
- Compare model performance, focusing on metrics like **accuracy, recall and precision**.
- Developing effective early detection tool also contributes to a deeper understanding of the early signs of autism, which can lead to advancements in treatment methods.

4.Thesis Structure

Chapter 1: Background

This chapter provides an overview of Autism Spectrum Disorder (ASD), including its definition, symptoms, causes, and types. It highlights the challenges of traditional diagnostic methods, which rely heavily on behavioral observation and clinical judgment, often resulting in delays and inconsistent outcomes. Then introduces Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) as transformative tools capable of improving diagnostic accuracy and efficiency. It explores the growing application of AI in psychology, particularly in areas like behavioral analysis and cognitive assessment. Finally, a comprehensive literature review outlines recent advancements in AI-driven ASD detection, summarizing key findings, datasets used, and limitations faced in prior research.

Chapter 2: Proposed Approaches and methodology

This chapter details the methodology developed for the research, starting with an overview of the proposed approach for early ASD detection using AI techniques. It describes the dataset used, including its origin, features, and labeling. The data preprocessing steps—such as data cleaning, categorical encoding, and feature scaling—are then explained to ensure the data is prepared for model training. The chapter also defines the evaluation metrics used to assess model performance, including accuracy, precision, recall, F1-score, and the ROC curve, ensuring a comprehensive understanding of model effectiveness.

Chapter 3: Results and discussion

This part presents the experimental results obtained from the implemented AI models. It includes a detailed analysis of performance outcomes based on the chosen evaluation metrics and compares various machine learning and deep learning algorithms to determine the most effective approach. The comparative analysis highlights strengths and weaknesses of each model, drawing insights from their behavior on the dataset and suggesting which configurations yield the most accurate and reliable ASD predictions.

Chapter 4: Our Solution

This chapter introduces the practical solution developed through the research, beginning with the tools and platforms used to build and test the system. It presents the project's branding, including the design and rationale behind the logo, and showcases the user interface developed for interacting with the model. The interface is designed for accessibility and ease of use, targeting healthcare professionals and caregivers. The chapter demonstrates how the final system functions in a real-world context and how it streamlines the ASD detection process.

General Conclusion

Chapter 1: Background

Chapter 1: Background

1.Introduction

In this first chapter, we introduce the basic concepts of Autism Spectrum Disorder, its fundamental characteristics, and the range of symptoms and signs commonly found and the basic concepts of artificial intelligence, its primary branches with a focus on machine learning and deep learning techniques. The chapter presents a comprehensive overview of the fundamentals and literature regarding early detection of Autism Spectrum Disorder (ASD) using artificial intelligence. Furthermore, the chapter explores potential theories of autism and identifies issues that arise when diagnosing the condition with typical approaches such as interviews and observation tools. These limitations underscore the need for more effective and scalable solutions. And the architectures commonly applied for ASD detection, such as Artificial Neural Networks (ANNs) and Long Short-Term Memory networks (LSTMs). By integrating psychological insights into action with advances in AI, this chapter attempts to establish a solid theoretical foundation for describing and justifying the methodological choices of this research.

2.Definitions

2.1 Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD) is a multifaceted neurodevelopmental condition characterized by atypical brain development that influences various aspects of an individual's behavior, communication, and social interaction. It encompasses a broad range of symptoms and challenges that differ widely from person to person. The word "spectrum" reflects this diversity, indicating that no two individuals with ASD will present the exact same traits or experiences. [3]

ASD impacts how individuals perceive and interpret the world around them, leading to unique behavioral patterns and responses to stimuli. These challenges are generally categorized into three core domains: communication, social interaction, and restrictive or repetitive behaviors. However, the degree to which these areas are affected varies greatly. For example, some individuals may exhibit pronounced difficulties in verbal communication, relying entirely on nonverbal methods such as gestures and assistive devices, while others may have advanced vocabularies but struggle to engage in reciprocal conversations.

The severity of ASD symptoms ranges across a spectrum, from mild to profound:

1. **Mild ASD**: Individuals may have strong intellectual and verbal abilities, enabling them to live relatively independent lives. They might require minimal support; such as help in navigating social situations or managing sensory sensitivities.
2. **Moderate ASD**: People in this category might experience more noticeable challenges with social interaction, communication, and behavioral flexibility. They often need additional support in educational or occupational settings and may rely on structured routines to maintain daily functioning.
3. **Severe ASD**: Individuals with severe ASD often face significant impairments in all core domains, requiring comprehensive, lifelong support for daily activities, safety, and overall well-being. Intellectual disabilities and co-occurring medical conditions such as epilepsy or gastrointestinal issues are more common in this group.

Figure 1 Levels of autism

2.1.1 Symptoms and Signs of Autism

The symptoms of ASD emerge early in childhood, often noticeable before the age of 2, though some cases become evident later. [4]

Symptoms can be broadly categorized as follows:

Social and Communication Challenges

- Limited or atypical verbal and nonverbal communication skills.
- Difficulty interpreting facial expressions, tone of voice, and gestures.
- A preference for solitary activities and a lack of interest in group play.

- Delayed speech development, or in some cases, complete lack of speech.

Behavioral and Sensory Patterns

- Restricted and repetitive behaviors such as hand-flapping, rocking, or spinning objects.
- Insistence on sameness, with resistance to changes in routines or environments.
- Intense focus on specific interests, sometimes accompanied by encyclopedic knowledge of the topic.
- Hypersensitivity or hyposensitivity to sensory input, such as loud noises, bright lights, or textures.

2.1.2 Causes of autism

Autism spectrum disorder has no single known cause, and scientists are still uncertain about its exact origins. The complexity of the disorder, along with the wide variation in symptoms, suggests that multiple contributing factors may be involved.

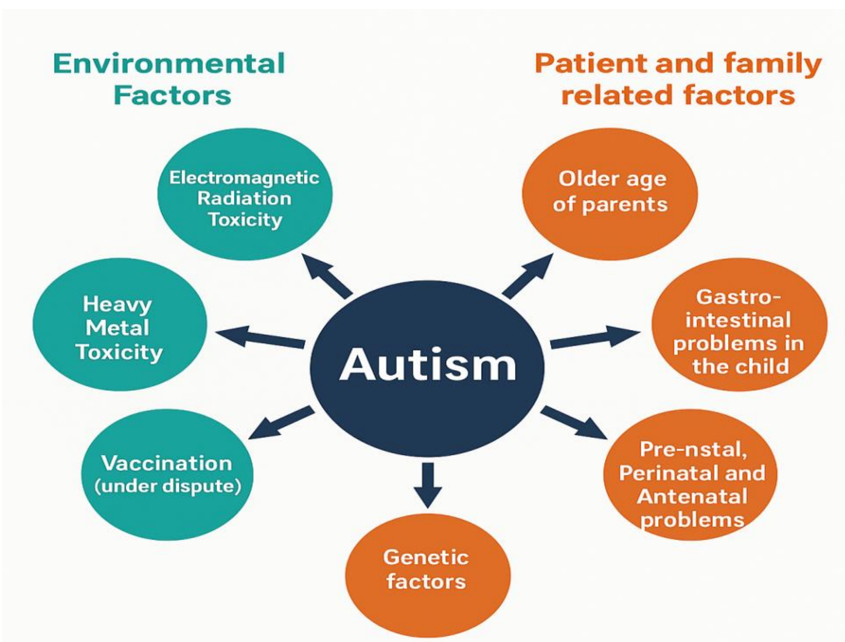

Figure 2 Probable causes of autism.

a) Genetic causes: Autism is highly heritable: It is estimated at least 50% of genetic risk is predicted by common genetic variation, and another 15-20% is due to spontaneous mutations or predictable inheritance patterns. The remaining genetic risk is yet to be determined.

For this study, researchers performed whole genome sequencing in 4,551 individuals from 1,004 families with at least two children diagnosed with autism. This group included 1,836 children with autism and 418 children without an autism diagnosis. [5]

Many researchers found seven potential genes that are predicted to increase the risk of autism: PLEKHA8, PRR25, FBXL13, VPS54, SLFN5, SNCAIP, and TGM1. [6]

b) Perinatal factors (before birth) : Physical health of the mother, such as the presence of metabolic syndrome or viral infections in the first trimester or mental health issues in her, such as depression and anxiety, particularly during weeks 21 to 32 of pregnancy ;also the parental age especially paternal age, above 34 years.Certain prenatal medications, such as antiepileptic drugs and antidepressant medications and a lower socioeconomic status. [6]

c) Antenatal factors (during birth) : Fetal complications, which may lead to a lack of oxygen ,or bleeding during pregnancy, also premature or late birth. [6]

d) Postnatal environmental factors (after birth) : Postnatal infections, such as meningitis or encephalitis, can impact brain development and potentially increase the risk of autism. Vitamin deficiencies, particularly in essential nutrients like vitamin D and B12, may also play a role in cognitive and neurological development. Additionally, inadequate levels of care, including neglect or lack of proper stimulation, can affect a child's early development. Finally, exposure to harmful chemicals, such as heavy metals or pollutants, during infancy may contribute to neurological alterations. While these factors alone may not directly cause autism. [6]

e) Neurobiological imbalances: Brain imaging studies have shown that people on the autism spectrum may experience differences in brain development early in life, such as changes to the formation of the amygdala and corpus callosum. [7]

2.1.3 Types of Autism Spectrum Disorders

ASD, as defined by the DSM-5, includes all autism-related diagnoses under one umbrella. Previously, these were categorized into distinct subtypes, each reflecting different manifestations of autism:

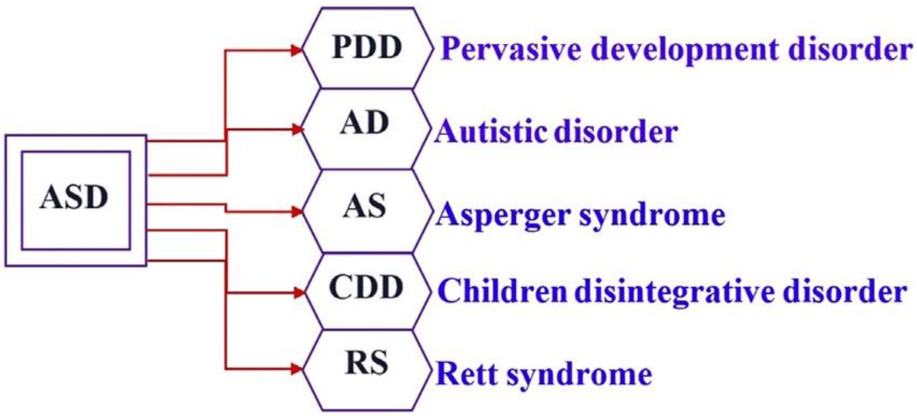

Figure 3 The different types of Autism.

a) Asperger's Syndrome: now considered part of the broader Autism Spectrum Disorder (ASD), is typically associated with average to above-average intelligence and strong verbal skills. Individuals with Asperger's often display a deep and focused interest in specific topics, sometimes developing an exceptional level of expertise in their chosen area. While their cognitive and language abilities may be well-developed, they often face significant challenges in social communication and interaction. These difficulties can include trouble understanding nonverbal cues such as facial expressions, tone of voice, and body language. They may also struggle with forming and maintaining relationships, interpreting social norms, or engaging in reciprocal conversations. Despite these challenges, many people with Asperger's have unique strengths, including attention to detail, persistence, and a strong sense of honesty and loyalty.

b)Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS): often referred to as "atypical autism," was a diagnosis used for individuals who demonstrated clear characteristics of autism but did not fully meet the criteria for other specific autism spectrum subtypes such as classic autism or Asperger's Syndrome. This category served as a kind of catch-all for those with significant social, communication, and behavioral challenges that were consistent with autism spectrum disorders, but varied in intensity or presentation. Symptoms of PDD-NOS were typically less severe than those seen in classic autism but more noticeable or disabling than those associated with Asperger's. Individuals with this diagnosis might have had delays in language development, difficulty with social interactions, repetitive behaviors, and unusual sensitivities to sensory input, though the specific combination and severity of symptoms varied widely. Because of this variability, PDD-NOS was often considered a "milder" or "atypical" form of autism, though it still posed meaningful challenges in daily functioning. Since the release of the DSM-5 in 2013, PDD-NOS has been subsumed under the umbrella term Autism Spectrum Disorder (ASD). [8]

c)Childhood Disintegrative Disorder (CDD): Childhood Disintegrative Disorder (CDD), also known as Heller's syndrome, is a rare and severe form of Autism Spectrum Disorder (ASD) that is marked by a period of normal development followed by a dramatic and unexpected regression. Typically, children with CDD develop language, motor skills, social engagement, and other age-appropriate behaviors up until around the age of 2 to 4. However, during this critical period, they begin to lose previously acquired skills in multiple areas, including speech and language, social interaction, play, toileting, and motor coordination. This regression is often sudden and can be deeply distressing for families. In many cases, the onset of regression coincides with the appearance of seizures or other neurological symptoms. Unlike other forms of ASD, the abrupt and widespread loss of abilities in CDD is what distinguishes it clinically. Though exceedingly rare, CDD reflects one of the most profound manifestations of autism-related disorders and typically results in significant lifelong challenges. Like other subtypes, CDD is now categorized under Autism Spectrum Disorder in the DSM-5.

d)Autistic Disorder: Autistic Disorder, often referred to as "classic autism," represents the more pronounced and traditionally recognized form of autism on the spectrum. Individuals with this diagnosis typically exhibit significant delays or abnormalities in speech and language development, pronounced difficulties in social interaction, and a range of repetitive or restrictive behaviors. These may include hand-flapping, insistence on routines, limited interests, or sensory sensitivities. In many cases, intellectual disabilities are also present, affecting learning and adaptive functioning. However, it's important to note that some individuals with classic autism may possess extraordinary abilities in specific areas, such as mathematics, art, memory, or music—a phenomenon sometimes referred to as "splinter skills" or savant abilities. The challenges associated with Autistic Disorder can be extensive, often requiring specialized support in communication, education, and daily living skills. Since the introduction of the DSM-5, Autistic Disorder is no longer a separate diagnosis and has been incorporated into the broader category of Autism Spectrum Disorder (ASD), which reflects the wide variability in symptoms and functioning across individuals.

e) Rett Syndrome: Rett Syndrome is a rare neurodevelopmental disorder that was once included within the autism spectrum due to some overlapping features but is now recognized as a distinct genetic condition. It is primarily caused by mutations in the *MECP2* gene, which plays a critical role in brain development and function. Rett Syndrome almost exclusively affects females, as the mutation is typically fatal in males during early development. Children with Rett Syndrome usually experience normal growth and development for the first 6 to 18 months of life, followed by a period of rapid and severe regression. During this phase, they progressively lose acquired motor skills, purposeful hand use, speech, and social engagement. One of the most characteristic signs is repetitive hand-wringing or hand-washing movements, often paired with slowed growth, breathing irregularities, and seizures. The condition leads to profound intellectual disability and severe physical impairments, often requiring lifelong care. While Rett Syndrome shares some surface-level similarities with autism, its known genetic cause and distinct developmental trajectory have led to its classification as a separate disorder.

Understanding these categories highlights the diversity of autism presentations and emphasizes the need for individualized diagnostic and therapeutic approaches.

Type of Autism	**Characteristics**
Autistic Disorder	Delays in language, communication, and social skills
Asperger Syndrome	Mild symptoms, no significant language delays
PDD-NOS	Atypical autism with fewer symptoms

Childhood Disintegrative Disorder	Severe developmental regression
Rett syndrome	Progressive loss of motor skills and language

Table 1 types of autism

2.1.4 Traditional methods of autism diagnosis

The comprehensive evaluation for autism spectrum disorder for young children includes several visits with the clinical and developmental child psychologists. These experts specialize in evaluating toddlers and young children suspected of having autism spectrum disorder or other communication disorders. Here are some methods used:

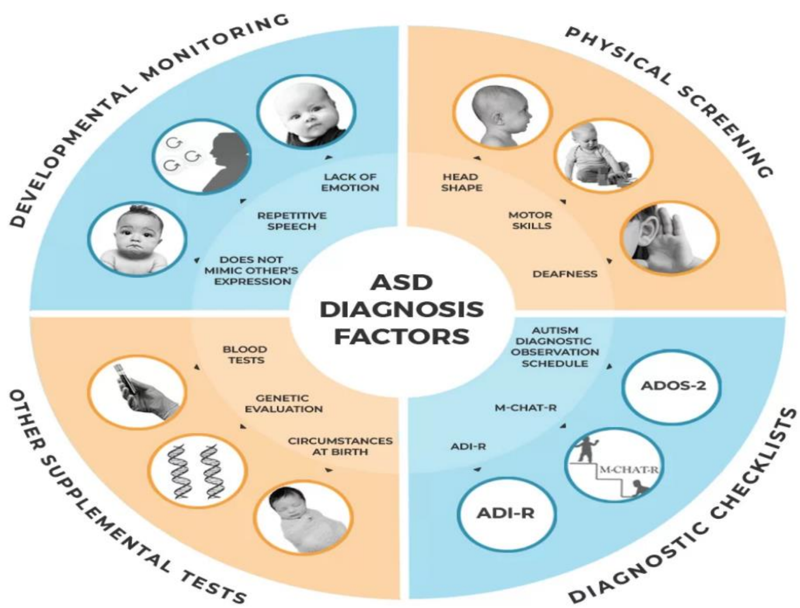

Figure 4 ASD diagnosis factors [9]

a) **Diagnostic Interviews**: The experts use the Autism Diagnostic Interview™, Revised (ADI™-R), a structured interview conducted with parents that focuses on a child's behavior in three main areas: qualities of reciprocal social interaction, communication and language, and restricted and repetitive interests and behaviors.The ADI™-R asks general and specific questions about a child's current and past behavior. Verbal and nonverbal communication, play and social skills, and restricted interests and repetitive behaviors are scored. [10]

b) **Neuropsychological Testing**: A neuropsychological evaluation involves cognitive and achievement testing, as well as further specialized testing of memory, attention, and executive function, to pinpoint a child's abilities and deficits in learning and communicating. During this evaluation, the child's parents provide the specialists with a comprehensive history of the child's behavior and symptoms since birth. The experts may also gather information from teachers or directly observe the child in the classroom. [10]

c**)** **Hearing and Vision Screening**: Doctors may also seek to rule out hearing and vision problems that could be causing symptoms that are similar to those of autism spectrum disorder. A doctor may test a child's hearing and vision with simple auditory and visual exams. [10]

d)**Genetic Testing**: Chromosome analysis and microarray are usually performed initially. These look for additional or missing DNA material and can reveal the genetic cause of autism spectrum disorder in a small percentage of children. These tests involve taking a blood sample from the child and having geneticists and genetic counselors interpret the results. [10]

1.1.5 Inconveniences of Autism Diagnosis Methods

Current autism diagnosis methods often present significant challenges, including lengthy assessment processes, subjective interpretations, and limited accessibility, which can delay early intervention and create barriers for individuals and families seeking timely support.

Method	**Inconveniences**
Diagnostic Interviews	- Relies heavily on parental recall, which may be inaccurate or incomplete, especially for early childhood behaviors. - Interpretation of responses can vary depending on the clinician, affecting consistency. - The process is lengthy (1.5–2 hours) and demands significant parental time. - Parents may unconsciously underreport or overreport behaviors due to emotional or social concerns.
Neuropsychological Testing	- Evaluations are expensive, limiting accessibility. - The testing environment may overwhelm children with autism due to unfamiliarity, lengthy sessions, or difficult tasks. - Requires specialized professionals, which may not be available in all areas, especially underserved regions.
Hearing and Vision Screening	- Detection of hearing or vision problems might divert attention away from considering autism as a potential underlying cause. - Children with autism may struggle to engage in the testing due to anxiety, communication barriers, or difficulty understanding instructions, leading to inconclusive or invalid results.

Genetic Testing	- Can be costly and results often take weeks or months to return. - Even when genetic anomalies are found, they usually do not directly guide treatment or interventions. - The complexity and length of the evaluation can delay diagnosis, missing early intervention opportunities. - Access to qualified professionals and tools is limited in rural or low-income areas.

Table 2 Inconveniences of traditional methods

Artificial Intelligence is revolutionizing a wide range of industries, particularly healthcare, by streamlining complex processes, enhancing diagnostic accuracy, and enabling the early detection of conditions such as Autism Spectrum Disorder (ASD).

2.2 Artificial Intelligence (AI)

Artificial Intelligence (AI) is the simulation of human intelligence processes by machines, enabling them to perform tasks such as reasoning, learning, and problem-solving. [11]

The primary branches of AI include:

1. **Machine Learning (ML)**: Algorithms trained on data to identify patterns and make predictions.
2. **Deep Learning (DL)**: A subset of ML using artificial neural networks to analyze large datasets.
3. **Natural Language Processing (NLP)**: AI systems designed to process, understand, and generate human language.

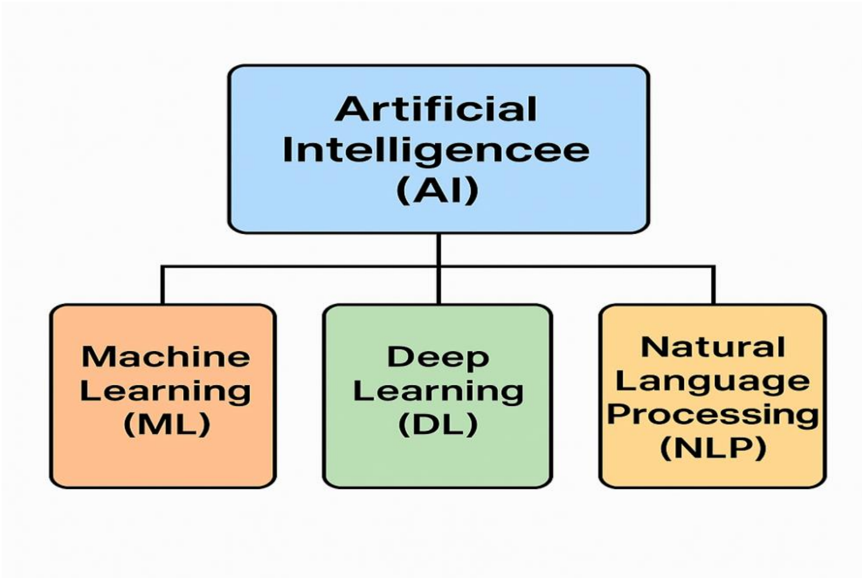

Figure 5 Artificial Intelligence (AI) and its branches.

2.2.1 Machine learning

Machine learning is a subset of artificial intelligence that enables a system to autonomously learn and improve without being explicitly programmed. Machine learning algorithms work by recognizing patterns and data and making predictions when new data is inputted into the system.

In broad strokes, three kinds of models are often used in machine learning: supervised, unsupervised, and reinforcement.

a) Supervised learning

Supervised learning is a machine learning model that uses labeled training data (structured data) to map a specific input to an output. In supervised learning, the output is known (such as recognizing a picture of an apple) and the model is trained on data of the known output. In simple terms, to train the algorithm to recognize pictures of apples, feed it pictures labeled as apples. [12]

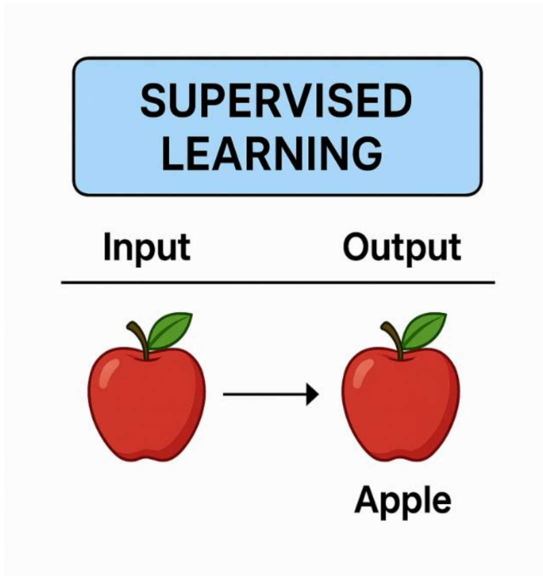

Figure 6 Supervised learning explained.

b) Unsupervised learning

Unsupervised learning is a machine learning model that uses unlabeled data (unstructured data) to learn patterns. Unlike supervised learning, the output is not known ahead of time.Rather, the algorithm learns from the data without human input (thus, unsupervised) and categorizes it into groups based on attributes.

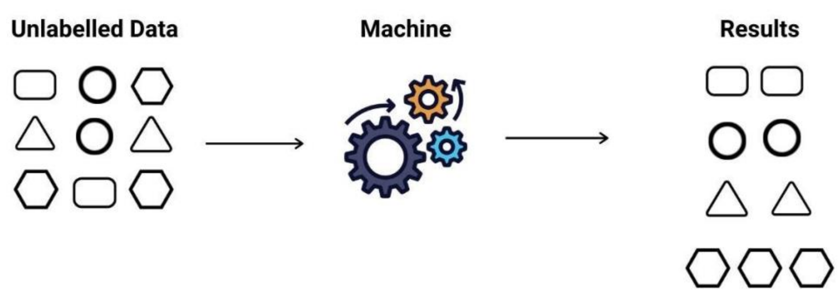

Figure 7 unsupervised learning explained.

Note: A mixed approach machine learning called semi-supervised learning is also often employed, where only some of the data is labeled. In semi-supervised learning, the algorithm must figure out how to organize and structure the data to achieve a known result. For instance, the machine learning model is told that the end result is an apple, but only some of the training data is labeled as an apple.

c)Reinforcement learning

Reinforcement learning is a machine learning model that can be described as "learn by doing" through a series of trial and error experiments. An "agent" learns to perform a defined task through a feedback loop until its performance is within a desirable range. The agent receives positive reinforcement when it performs the task well and negative reinforcement when it performs poorly.

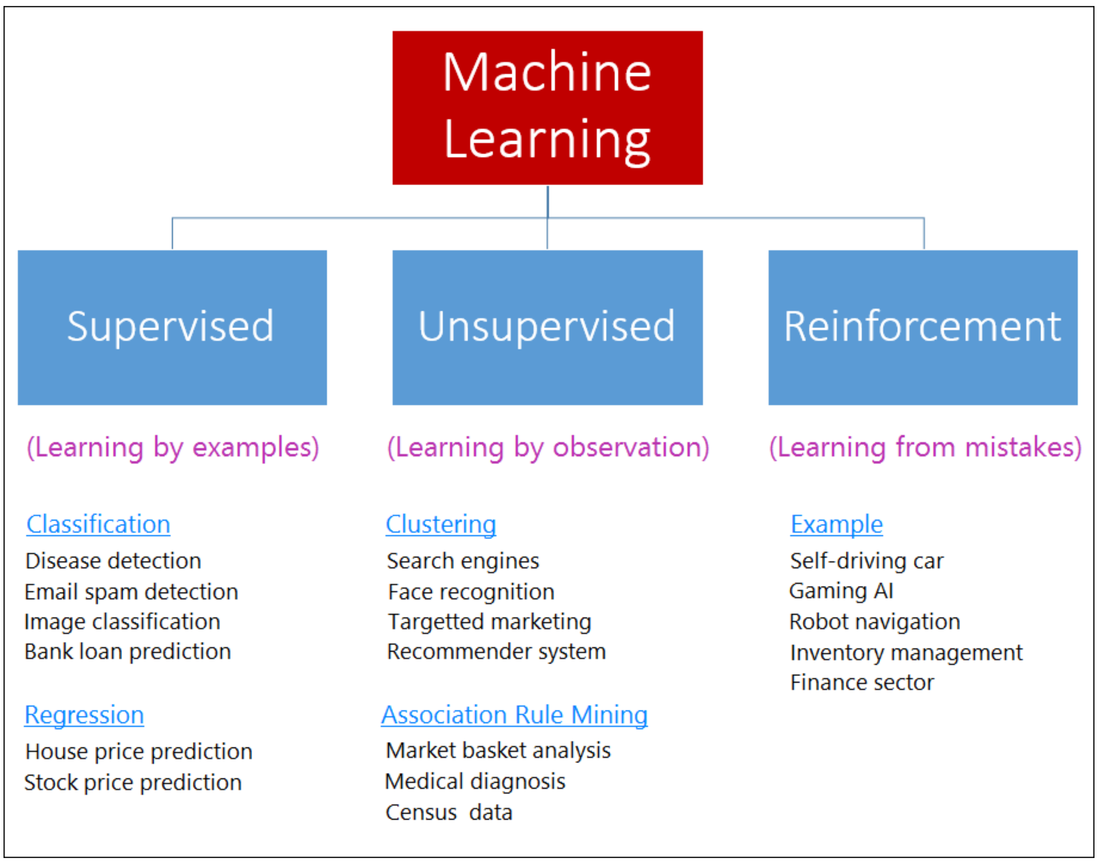

Figure 8 Machine learning types.

2.2.1.1 Machine learning techniques:

a. K-Nearest Neighbors (KNN)

Definition

KNN is a simple, instance-based machine learning algorithm that classifies a data point based on how its neighbors are classified. It finds the k closest points to a query point and assigns the most common class among them. [13]

Key Features

- Very simple to understand and implement.
- No explicit training phase (lazy learning).
- Sensitive to the choice of k and the distance metric.
- Can be computationally expensive for large datasets.
- Works well with smaller datasets.

b. Decision Tree (DT)

Definition

A Decision Tree is a supervised machine learning algorithm that splits data into branches based on feature values, forming a tree-like structure to make predictions. Each internal node represents a decision based on a feature, each branch represents an outcome of that decision, and each leaf node represents a final prediction. [14]

Key Features

- Easy to understand and interpret.
- Handles both classification and regression tasks.
- Requires little data preprocessing.
- Prone to overfitting on training data.
- Performs poorly with small variations in data (high variance).

c. Random Forest (RF)

Definition

Random Forest is an ensemble learning method that builds many decision trees and combines their outputs to make more accurate and stable predictions. [15]

Key Features

- Reduces overfitting compared to a single decision tree.
- Works well with both classification and regression tasks.
- Can handle large datasets and high-dimensional spaces.
- Provides feature importance measures.
- Requires more memory and computational power.

2.2.2 Deep Learning

Deep learning is a subset of machine learning that uses artificial neural networks to process and analyze information. Neural networks are composed of computational nodes that are layered within deep learning algorithms. Each layer contains an input layer, an output layer, and a hidden layer. The neural network is fed training data which helps the algorithm learn and improve accuracy. When a neural network is composed of three or more layers it is said to be "deep," thus deep learning. [16]

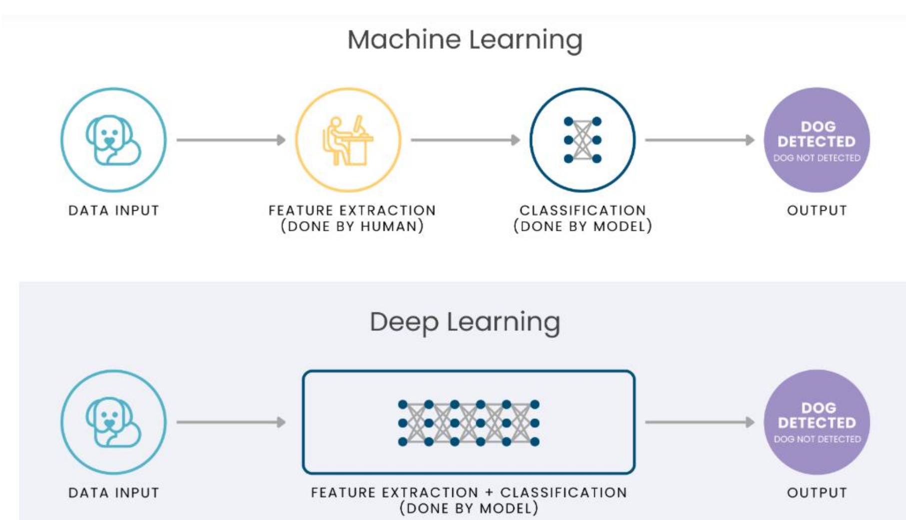

Figure 9 Machine learning vs deep learning.

2.2.2.1 Deep Learning techniques

a. Artificial Neural Network (ANN)

Definition

ANN is a computing system inspired by the biological neural networks in human brains. It consists of layers of nodes (neurons) that can learn complex patterns in data.

Key Features

- Capable of learning non-linear relationships.
- Can approximate any continuous function (universal approximators).
- Requires large datasets and computational power.
- Needs careful tuning of hyperparameters (like learning rate, number of layers).
- Prone to overfitting without regularization techniques.

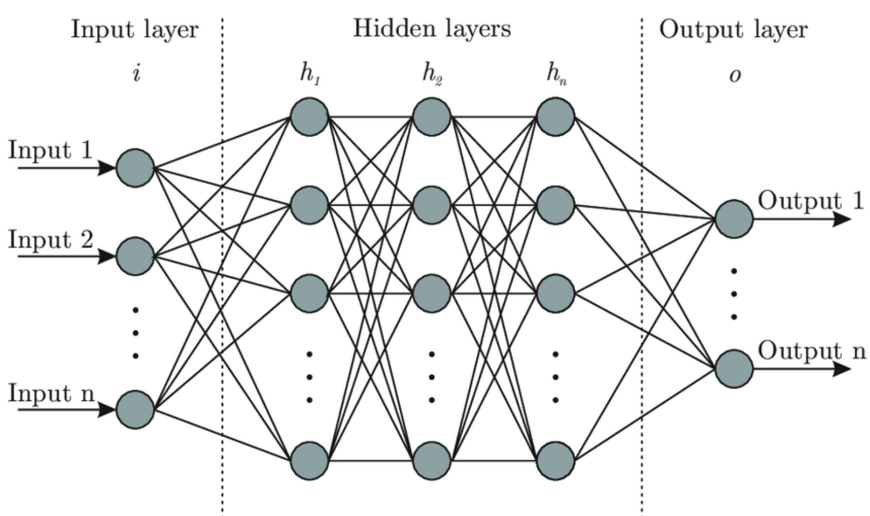

Figure 10 ANN Architecture

b. Recurrent Neural Networks (RNN)

Definition

RNNs are a type of neural network designed for sequential data, such as time series, text, or speech. They have a "memory" mechanism that allows them to retain information from previous steps in the sequence, making them suitable for tasks like language modeling, machine translation, and speech recognition [17]

Key Features

- Recurrent connections for handling sequential data.
- Can suffer from vanishing/exploding gradients in long sequences.
- Variant like LSTM (Long Short-Term Memory) address these issues.

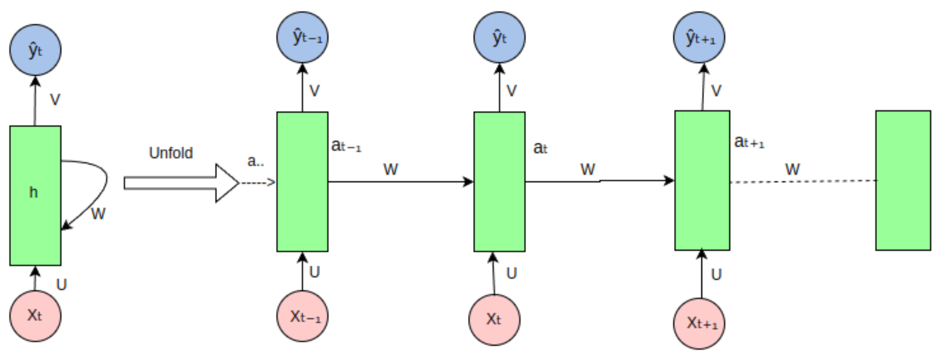

Figure 11 RNN architecture

c. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is an enhanced version of the Recurrent Neural Network (RNN) designed by Hochreiter & Schmidhuber. LSTMs can capture long-term dependencies in sequential data making them ideal for tasks like language translation, speech recognition and time series forecasting. [18]

Key features

- LSTM can store information for a long time.
- It uses gates to control what information to keep or remove.
- LSTM remembers important information even over many steps.
- It can choose what to remember and what to forget.
- LSTM can learn from both past and future data.
- It trains better and is more stable than basic RNNs.
- LSTM works well for tasks like language, time series, and speech.
- It keeps a cell state for memory and a hidden state for output.

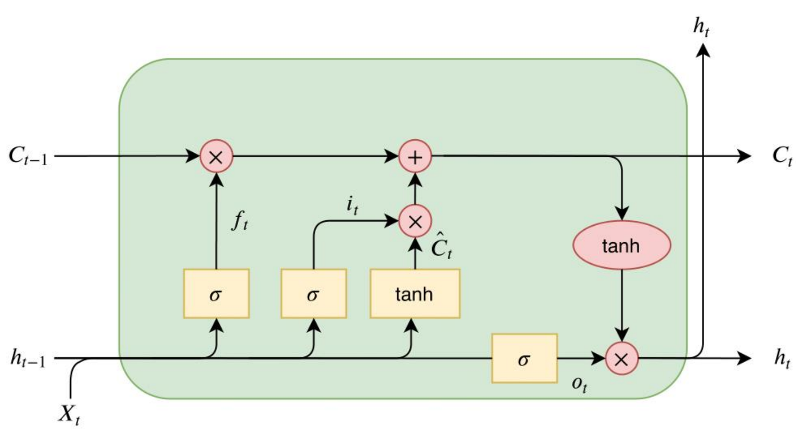

Figure 12 LSTM architecture

2.2.3 Natural language processing

Definition

Natural Language Processing (NLP) is a subfield of artificial intelligence (AI) and computational linguistics that focuses on enabling computers to understand, interpret, and generate human language. It combines techniques from machine learning, deep learning, and linguistics to process textual or spoken data. [19]

Key Features

Text Processing: Tokenization, stemming, lemmatization.

Syntax & Semantics: Parsing, part-of-speech tagging, named entity recognition (NER).

Contextual Understanding: Word embeddings (e.g., Word2Vec, BERT), transformers.

Generation & Translation: Machine translation, text summarization, chatbots.

3.Artificial Intelligence in psychology

Artificial Intelligence (AI) is transforming psychological research and practice by providing advanced tools for understanding human behavior, cognition, and mental health. Its ability to process complex, large-scale data enables psychologists to gain insights that were previously difficult or impossible to obtain. AI techniques, including machine learning (ML), natural language processing (NLP), and deep learning (DL), are increasingly used across various domains of psychology.

1. Understanding Thought Processes and Behavior

AI systems can analyze text, speech, facial expressions, and even body language to infer psychological states. For example:

- Sentiment analysis detects emotional tone in written or spoken language.
- Facial expression recognition helps in identifying emotions or mood changes.

- Personality detection algorithms estimate personality traits from online behavior or speech patterns.

2. Diagnosing and Treating Mental Illness

AI tools are increasingly used to support mental health care:

- Machine learning models can identify early signs of depression, anxiety, schizophrenia, and other disorders by analyzing data from social media posts, voice recordings, questionnaires, or EEG signals.
- AI systems can detect biomarkers in brain signals (e.g., EEG, fMRI) to support early and more accurate diagnosis.
- Chatbots and virtual therapists, powered by NLP, deliver basic cognitive behavioral therapy (CBT) or supportive conversations, increasing access to mental health resources.

3. Neuroimaging and Brain Research

AI, particularly deep learning, enhances the analysis of brain imaging data:

- It can process fMRI and EEG data to uncover patterns of brain activation linked to specific psychological states.
- AI helps map regions of the brain responsible for emotions, memory, language, and cognitive control, aiding in the study of neurodevelopmental and neurodegenerative disorders.

4. Educational Psychology and Development

AI applications support psychological insights into learning and development:

- Systems can monitor and evaluate learner behavior, providing feedback and predicting academic performance.
- Personalized learning platforms adapt content based on a student's strengths, weaknesses, and preferences.
- Video-based AI tools track children's developmental milestones, aiding in early detection of developmental disorders like autism.

5. Social and Group Psychology

AI also contributes to understanding broader societal and group behavior:

- It analyzes social media content, images, and interaction patterns to detect trends in public sentiment, group dynamics, or collective behavior.
- AI can predict social influence, conformity, or the spread of ideas within networks, which is valuable for both research and policy-making.

4.Literature review

In recent years, machine learning (ML) and deep learning (DL) techniques have shown great promise in the early detection and diagnosis of Autism Spectrum Disorder (ASD). These approaches are capable of identifying patterns in complex datasets with high accuracy, making them valuable tools for supplementing traditional diagnostic methods. Several notable studies have explored the use of various ML models—including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and hybrid models like CNN-LSTM—on a wide range of datasets, including behavioral questionnaires, neuroimaging data, and genetic information.

In the study titled "Analysis and Detection of Autism Spectrum Disorder using Machine Learning Techniques", Suman Raj and Sarfaraz Masood (2020) [20] applied CNN, SVM, ANN, and KNN techniques to datasets from children, adolescents, and adults, achieving accuracies as high as 99.53% with CNN on adult data. Similarly, Nawshin Haque et al. (2024), in their work "A Comparative Study of Early Autism Spectrum Disorder Detection Using Deep Learning-Based Models," [21] conducted a comparative analysis using CNN, RNN, and ANN models on children's and toddlers' datasets, reporting 100% accuracy with ANN on the children's dataset and CNN on the toddlers' dataset. In the domain of neuroimaging, Sherkatghanad et al. (2020), in their study "Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network" [22], utilized resting-state fMRI data from the ABIDE I dataset and implemented a CNN model, which achieved a classification accuracy of 70.22%, reflecting both the potential and limitations of using brain imaging data for ASD detection. Additionally, Fan et al. (2023) introduced a novel method titled "DeepASDPred: A CNN-LSTM-Based Deep Learning Method for Autism Spectrum Disorders Risk RNA Identification," [23] where they trained a CNN-LSTM model on RNA transcript sequences and gene expression data, achieving 93.8% accuracy, 88.7% sensitivity, and 97.1% specificity, thus demonstrating the effectiveness of transcriptomic biomarkers in ASD risk prediction.

The table below provides a detailed summary of these studies, including dataset characteristics, machine learning techniques used, and their corresponding results:

<table>
<tr><th>Paper</th><th>Year</th><th>Dataset</th><th>Machine learning techniques</th><th>Results</th></tr>
<tr><td rowspan="9">Analysis and detection of Autism Spectrum Disorder using Machine Learning techniques.</td><td rowspan="9">2020</td><td rowspan="3">Children's Dataset: 292 instances with 21 attributes.</td><td>CNN</td><td>98.30%</td></tr>
<tr><td>SVM</td><td>98.30%</td></tr>
<tr><td>ANN</td><td>98.30%</td></tr>
<tr><td rowspan="3">Adult Dataset: 704 instances with 21 attributes.</td><td>CNN</td><td>99.53%</td></tr>
<tr><td>KNN</td><td>95.75%</td></tr>
<tr><td>ANN</td><td>97.64%</td></tr>
<tr><td rowspan="3">Adolescents' Dataset: 104 instances with 21 attributes.</td><td>CNN</td><td>96.88%</td></tr>
<tr><td>SVM</td><td>95.23%</td></tr>
<tr><td>KNN</td><td>80.95%</td></tr>
<tr><td>Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network</td><td>2020</td><td>Resting-state fMRI data from the ABIDE I dataset</td><td>CNN</td><td>70.22%</td></tr>
</table>

DeepASDPred: a CNN-LSTM-based deep learning method for Autism spectrum disorders risk RNA identification	2023	ASD risk gene dataset integrating RNA transcript sequences (from GENCODE) and gene expression values (from BrainSpan Atlas); 1005 positive samples and 1590 negative samples	CNN-LSTM	**Accuracy 93.8%, Sensitivity 88.7%, Specificity 97.1%**
A comparative study of early autism spectrum disorder detection using deep learning based models	2024	**Children's Dataset**: 292 instances with 22 attributes.	CNN	**98%**
			RNN	**96%**
			ANN	**100%**
		Toddlers' Dataset: 1,054 instances with 19 attributes.	CNN	**100%**
			RNN	**99.52%**
			ANN	**99.05%**

Table 3 Related works results

5.Conclusion

The first chapter highlights the complexity of Autism Spectrum Disorder (ASD) and the critical need for early, accurate diagnosis to enable timely interventions. Traditional diagnostic methods, though valuable, face challenges such as subjectivity, time constraints, and limited accessibility. These limitations signal the necessity for innovative solutions.

Artificial Intelligence (AI), particularly machine learning (ML) and deep learning (DL), has emerged as a transformative tool in ASD research. Advanced models like Convolutional Neural Networks (CNNs), Artificial Neural Networks (ANNs) and Recurrent Neural Networks (RNNs) have demonstrated exceptional accuracy in analyzing neuroimaging, genetic, and behavioral data, these technologies can pave the way for more inclusive, accurate, and accessible support systems, ultimately improving outcomes for individuals with ASD and their families.

Chapter 2: Approaches and methodology

Chapter 2: Approaches and methodology

1.Introduction

Nowtimes, the importance of datasets in machine learning and data analysis has grown immensely. A dataset is a structured collection of related data points, organized to enable meaningful analysis and insights. Particularly in fields like healthcare and social sciences, datasets serve as the foundation for predictive modeling, decision-making, and diagnostic support.

This chapter focuses on the Autistic Spectrum Disorder Screening Data for Toddlers. This dataset, created by Dr. Fadi Thabtah, is a valuable resource designed to facilitate early and efficient detection of ASD through behavioral and demographic assessments. The dataset contains features carefully selected to distinguish ASD traits.

To prepare the dataset for effective machine learning and deep learning application, some data pre-processing techniques are employed. Pre-processing involves cleaning the raw data, handling missing values, encoding categorical attributes, removing duplicates, splitting the data into training and testing sets, and scaling feature values. Each of these steps ensures the quality and reliability of the data, which is essential for building accurate and robust models.

It also explores the evaluation of model performance using metrics such as accuracy, precision, recall and F1 score and tools such as the Receiver Operating Characteristic (ROC) curve, which measures the ability of the classifier to distinguish between classes.

Through a systematic exploration of our dataset, preprocessing techniques, machine learning and deep learning models, and evaluation metrics, this chapter lays the groundwork for developing predictive systems that can assist in the early diagnosis of Autism Spectrum Disorder.

2.The dataset

Definition

A dataset is a structured collection of data organized and stored together for analysis or processing. The data within a dataset is typically related in some way and taken from a single source or intended for a single project. A dataset can include many different types of data, from numerical values to text, images or audio recordings. The data within a dataset can typically be accessed individually, in combination or managed as a whole entity. [24]

A.Data Set Name: Autistic Spectrum Disorder Screening Data for Toddlers . [25]

Date: July, 22, 2018.

Author: Dr Fadi Thabtah

Definition: a time-efficient and accessible ASD screening is imminent to help health professionals and inform individuals whether they should pursue formal clinical diagnosis. The rapid growth in the number of ASD cases worldwide necessitates datasets related to behaviour traits. Presently, very limited autism datasets associated with clinical or screening are available and most of them are genetic in nature. Hence, dr. Thabtah proposes a new dataset related to autism screening of toddlers that contained influential features to be utilised for further analysis especially in determining autistic traits and improving the classification of ASD cases. In this dataset, they record ten behavioural features (Q-Chat-10) plus other individuals characteristics that have proved to be effective in detecting the ASD cases from controls in behaviour.

Data Type	Task	Area	Format Type	Number of Instances	Number of Attributes	Does the data set contain missing values
Predictive and Descriptive: Nominal/ categorical, binary and continuous	Classification	Medical, health and social science	Non-Matrix	1054	18	No

Table 4 The toddlers database characteristics.

Variable in Dataset	**Corresponding Q-chat-10-Toddler Features**
A1	Does your child look at you when you call his/her name?
A2	Is it easy for you to get eye contact with your child?
A3	Does your child point to indicate that s/he wants something? (e.g. a toy that is out of reach)
A4	Does your child point to share interest with you? (e.g. pointing at an interesting sight)
A5	Does your child pretend? (e.g. care for dolls, talk on a toy phone)
A6	Does your child follow where you’re looking?
A7	If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them)
A8	Would you describe your child’s first words as: common ?

A9	Does your child use simple gestures? (e.g. wave goodbye)
A10	Does your child stare at nothing with no apparent purpose?

Table 5 Details of variables mapping to the Q-Chat-10 screening methods.

Feature	**Type**	**Description**
A1: Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A2: Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A3: Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A4: Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A5: Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A6: A6: Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A7: Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used

A8: Question 8 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A9: Question 9 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A:10 Question 10 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Age	Number	Toddlers (months)
Score by Q-chat-10	Number	1-10 (Less than or equal 3 no ASD traits; > 3 ASD traits
Sex	Character	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean (yes or no)	Whether the case was born with jaundice
Family member with ASD history	Boolean (yes or no)	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician ,etc.
Class variable	String	ASD traits or No ASD traits (automatically assigned by the ASDTests app). (Yes / No)

Table 6 Features collected and their descriptions.

3. Proposed approach

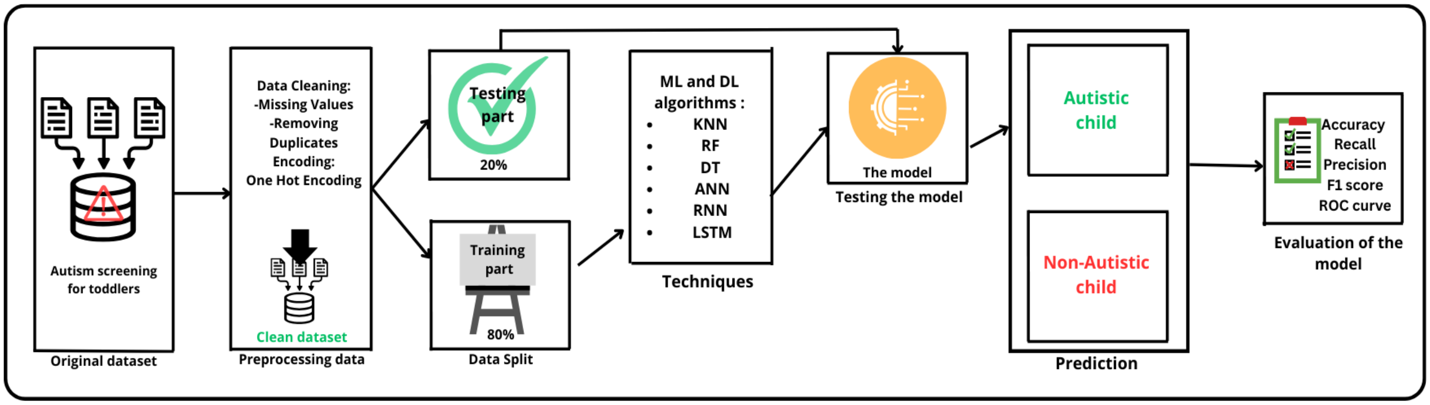

Figure 13 Proposed approach of Autism Prediction

The Autistic Spectrum Disorder Screening Data for Toddlers dataset required preprocessing to address missing values and format issues before modeling. We employed mode imputation to handle missing data and implemented One-Hot Encoding to transform categorical variables into numerical format. The dataset was then divided into training and testing subsets to ensure proper model evaluation. Both machine learning and deep learning models were developed by selecting appropriate architectures, compilers, optimizers, loss functions, and performance metrics. After training while monitoring validation performance to prevent overfitting, we evaluated the models using multiple metrics including accuracy, precision, recall, and F1 score on the test set, ultimately preparing the validated model for deployment and prediction tasks.

4.Data Pre-processing

Data Pre-processing refers to the transformations applied to the data before feeding it to our algorithms. The data gathered from different sources is collected in raw format which is not feasible for the analysis. Data Preprocessing technique is used to convert the raw data into a clean data set.

Why preprocessing?

Real world data are usually:

Incomplete: lacking attribute values, lacking certain attributes of interest, or containing only aggregate data.

Noisy: containing errors or outliers.

Inconsistent: containing discrepancies in codes or names.

4.1Preprocessing techniques

A. Data Cleaning: It is the process of identifying and correcting errors or inconsistencies in the dataset. It involves handling missing values, removing duplicates, and correcting incorrect or outlier data to ensure the dataset is accurate and reliable. Clean data is essential for effective analysis, as it improves the quality of results and enhances the performance of data models. [26]

- **Missing Values:** This occurs when data is absent from a dataset. You can either ignore the rows with missing data or fill the gaps manually, with the attribute mean, or by using the most probable value. This ensures the dataset remains accurate and complete for analysis [26]
- **Removing Duplicates:** It involves identifying and eliminating repeated data entries to ensure accuracy and consistency in the dataset. This process prevents errors and ensures reliable analysis by keeping only unique records [26]

B. Encoding:

Encoding is the process of converting categorical or text-based data into numerical form so that machine learning algorithms can understand and process it. Most algorithms (like neural networks, support vector machines, etc.) cannot work directly with raw text or categories; they require numerical inputs. Encoding preserves the original information while making it usable for computation, ensuring that models can interpret features correctly during training and prediction. Different encoding methods exist depending on the data and the algorithm used — such as label encoding, one-hot encoding, ordinal encoding, and others.

Choosing the right encoding method is important because it can impact the performance and accuracy of the model.

- **One Hot Encoding:**

One Hot Encoding is a method for converting categorical variables into a binary format. It creates new columns for each category where 1 means the category is present and 0 means it is not. The primary purpose of One Hot Encoding is to ensure that categorical data can be effectively used in machine learning models. [27]

5.Data Splitting

Data Splitting is the process of dividing a dataset into training, validation, and testing sets in preparation for the training and testing of a machine learning model.

DATA SPLITTING

```
[ ]  X = df1.drop(columns=["Class/ASD Traits "])
     y = df1["Class/ASD Traits "]

     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Figure 14 Data splitting code

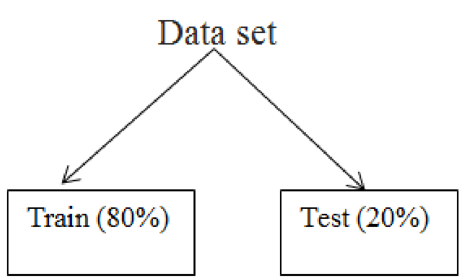

Figure 15 Data splitting

6.Feature Scaling

Scaling is a technique used in machine learning to normalize or standardize the range of independent variables (features). This is crucial because many algorithms (like KNN, SVM, logistic regression, neural networks) perform better when features are on the same scale.

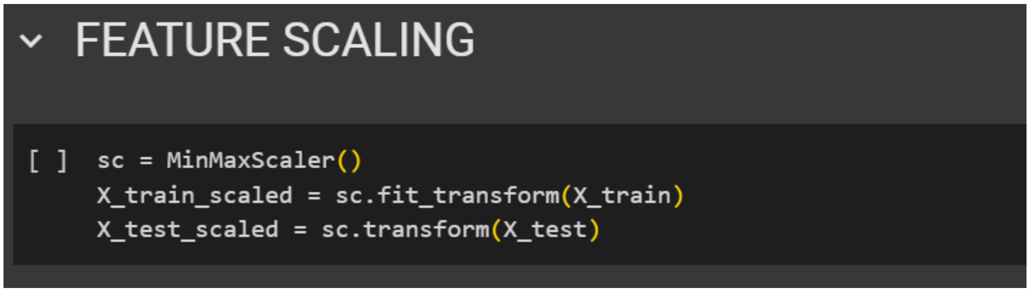

Figure 16 Feauture scaling

MinMax Scaler: It is a way of data scaling, where the minimum of feature is made equal to zero and the maximum of feature equal to one. MinMax Scaler shrinks the data within the given range, usually of 0 to 1. It transforms data by scaling features to a given range. It scales the values to a specific value range without changing the shape of the original distribution. [28]

The formula is:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where:

- X is the original feature value.
- X_{min} and X_{max} are the minimum and maximum values of the feature, respectively.

7.Classifiers

7.1 Machine learning classifiers

a. Decision Tree Classifier

Key Hyperparameters

- max_depth=2.
- min_samples_split=10
- min_samples_leaf=5
- max_features='sqrt':
- random_state=42
- **Training**
- Trained on the scaled training data (X_train_scaled, y_train) and tested on scaled test data.

b. Random Forest Classifier

Key Hyperparameters

- n_estimators=10: Uses 10 decision trees.
- max_depth=5
- min_samples_split=10.
- min_samples_leaf=5
- random_state=42

Training

- Trained on scaled training data; predictions via majority voting.

c. KNN Classifier

The KNeighborsClassifier was used with **default hyperparameters**, which includes:

- n_neighbors = 5 (the 5 nearest neighbors are considered).
- Distance metric is **Euclidean distance**.

- Uniform weights (all neighbors contribute equally to the vote).
- No dimensionality reduction or weighting was applied.

7.2 Deep learning algorithms

a. Artificial Neural Network (ANN)

A **feedforward neural network** was implemented using Keras' Sequential API.

Architecture

- **Input Layer**:
  - Accepts input_dim=29.
- **Hidden Layer 1**:
  - Dense(32) neurons.
  - Activation: **ReLU**.
- **Hidden Layer 2**:
  - Dense(16) neurons.
  - Activation: **ReLU**.
- **Output Layer**:
  - Dense(1) neuron.
  - Activation: **Sigmoid** (used for binary classification).

b. Recurrent Neural Network (RNN)

Although our dataset is not sequential, the data was reshaped to simulate time-series input by converting it to 3D.

Architecture

- **Input Shape**: (1, 29).
- **SimpleRNN Layer**:
    - 64 units.
    - Activation: **tanh**.
- **Dropout Layer**:
    - Dropout rate: 0.3 (to reduce overfitting).
- **Dense Layer**:
    - Dense(32) neurons.
    - Activation: **ReLU**.
- **Output Layer**:
    - Dense(1) neuron.
    - Activation: **sigmoid**.

c. Long Short-Term Memory Network (LSTM)

Similar to the RNN, the input was reshaped to 3D for LSTM input.

Architecture

- **Input Shape**: (1, 29) — features treated as a single time step.
- **LSTM Layer**:
    - 64 units.

- Activation: **ReLU**.

- **Output Layer**:
    - Dense(1) neuron.
    - Activation: **sigmoid**.

8.Evaluation metrics

8.1. Confusion matrix

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm. [29]

The confusion matrix consists of four basic characteristics (numbers) that are used to define the measurement metrics of the classifier. These four numbers are:

- **TP** = True Positives
- **TN** = True Negatives
- **FP** = False Positives
- **FN** = False Negatives

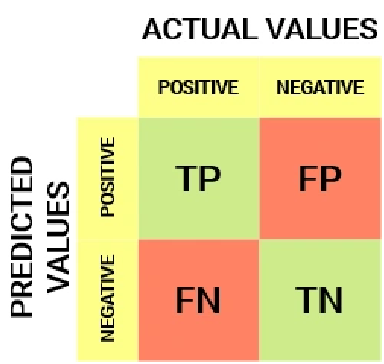

Figure 17 Confusion matrix

Performance metrics of an algorithm are accuracy, precision, recall, and F1 score, which are calculated on the basis of the above-stated TP, TN, FP, and FN.

a. Accuracy

Definition: The ratio of correctly predicted observations to the total observations. [30]

$$\text{Accuracy}= \frac{TP+TN}{TP+TN+FP+FN}$$

b. Precision

Definition: The ratio of correctly predicted positive observations to the total predicted positives. [30]

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

c. Recall (Sensitivity or True Positive Rate)

Definition: The ratio of correctly predicted positive observations to all observations in the actual positive class. [30]

$$\text{Recall}=\frac{TP}{TP+FN}$$

d. F1 Score

Definition: The harmonic mean of precision and recall. It balances the two by giving a single score that accounts for both false positives and false negatives. [30]

$$F1=2\times \frac{Precision \times Recall}{Precision+ Recall}$$

8.2. Receiver Operating Characteristic curve

The diagnostic performance of a test, or the accuracy of a test to discriminate diseased cases from normal cases is evaluated using Receiver Operating Characteristic (ROC) curve analysis. ROC curves can also be used to compare the diagnostic performance of two or more raters. [30]

Interpretation of ROC curves

Interpreting ROC (Receiver Operating Characteristic) curves involves analyzing the trade-off between a model's true positive rate (TPR) and false positive rate (FPR) across different classification thresholds. Here's a breakdown of how to interpret them effectively:

1. Axes Meaning

- **X-axis**: False Positive Rate (FPR) = FP / (FP + TN)
- **Y-axis**: True Positive Rate (TPR), also known as **Sensitivity or Recall** = TP / (TP + FN)

Each point on the ROC curve represents a TPR/FPR pair corresponding to a specific decision threshold.

2. Ideal ROC Curve

- The closer the curve follows the **top-left corner** of the plot, the **better the model**.
- An ideal model would have:
  - **TPR = 1**
  - **FPR = 0**
  - That point (0, 1) indicates **perfect classification** with no false positives and all true positives detected.

3. Diagonal Line (Random Guessing)

- A ROC curve that follows the **diagonal line from (0,0) to (1,1)** represents a model that is **no better than random guessing**.
- This is the **baseline**.

4. Area Under the Curve (AUC)

- **AUC-ROC** gives a **single value summary** of the ROC curve.
- **Range**: 0 to 1
  - **1.0** = perfect model
  - **0.5** = random model
  - < **0.5** = worse than random (the model might be misclassifying)
- **Interpretation of AUC scores**
  - **0.9 – 1.0**: Excellent
  - **0.8 – 0.9**: Good
  - **0.7 – 0.8**: Fair
  - **0.6 – 0.7**: Poor
  - **0.5 – 0.6**: Fail / No discrimination

5. Choosing a Threshold

- The ROC curve helps visualize **how different thresholds affect performance**.
- Depending on the application (e.g., in medical diagnostics),we use:
    - Higher **TPR** even with some increase in **FPR** (catch all positives).
    - Or lower **FPR** to reduce false alarms.

Chapter 3: Results and discussion

Chapter 3: Results and discussion

1.Introduction

This chapter presents the results of our comprehensive evaluation of machine learning and deep learning approaches for early detection of Autism Spectrum Disorder (ASD) in toddlers. Our study leverages the Autism Spectrum Disorder Screening Data for Toddlers to systematically compare the efficacy of various computational models in identifying ASD traits.

We structured our investigation into two distinct experiments. The first experiment evaluated traditional machine learning algorithms—specifically Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN)—to establish baseline performance metrics for ASD classification. The second experiment explored more sophisticated deep learning architectures, including Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks. For each model, we report key performance metrics including accuracy, precision, recall, and F1 score, supplemented by Receiver Operating Characteristic (ROC) curve analysis to visualize classification performance across different thresholds.

This systematic comparison provides valuable insights into the relative strengths and limitations of both traditional and advanced computational approaches for ASD screening. The findings presented here have significant implications for the development of reliable, automated screening tools that could assist healthcare professionals in early ASD detection.

2.Experiments

In our study, we conducted two experiments using the Autism Spectrum Disorder (ASD) Screening Data for Toddlers to evaluate the performance of different models in early detection of ASD. In the first experiment, we applied traditional machine learning models—Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN)—to assess their effectiveness in identifying ASD traits. In the second experiment, we focused on deep learning approaches, utilizing Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks. The comparison of results from both experiments provided valuable insights into the strengths and adaptability of machine learning versus deep learning techniques in the context of early ASD screening.

2.1 Experiment 01

1. KNN

Accuracy	**Precision**	**Recall**	**F1 Score**
97.16%	99.99%	95.77%	97.84%

Table 7 KNN Results

The KNN model achieved high overall performance with 97.16% accuracy. While its precision is excellent at 99.99% (indicating very few false positives), the recall is somewhat lower at 95.77%, suggesting that the model missed detecting some positive cases. The F1 score of 97.84% represents a good balance between precision and recall, but indicates there's room for improvement in identifying all positive cases.

2.Random Forest

Accuracy	**Precision**	**Recall**	**F1 Score**
98.57%	97.93%	99.99%	98.95%

Table 8 RF Results

The Random Forest classifier demonstrated excellent performance with 98.57% accuracy. Its precision of 97.93% shows it has relatively few false positives, while its outstanding recall of 99.99%

indicates it rarely misses positive cases. The F1 score of 98.95% reflects a strong balance between precision and recall, making it highly effective for this classification task.

3.Decision tree

Accuracy	**Precision**	**Recall**	**F1 Score**
98.57%	97.93%	99.99%	98.95%

Table 9 DT Results

The Decision Tree classifier shows identical performance metrics to the Random Forest model, with 98.57% accuracy, 97.93% precision, 99.99% recall, and 98.95% F1 score. This suggests both models are very effective at this classification task, with particularly strong ability to identify positive cases (high recall), though with a slightly higher rate of false positives compared to the KNN model.

Experiment 01: ROC curves

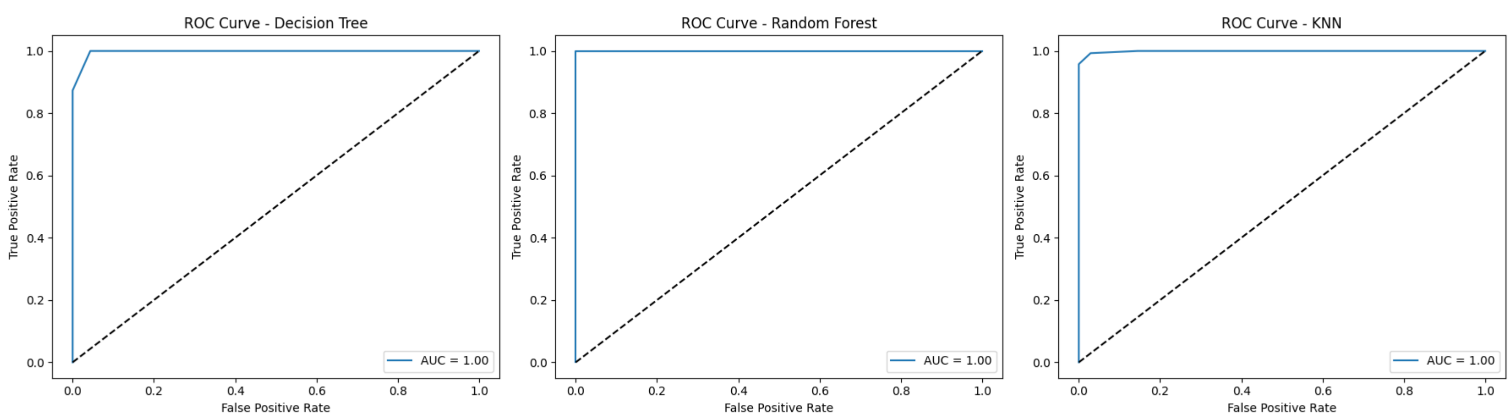

Figure 18 Experiment 01 ROC curves

The ROC curves for the Decision Tree, Random Forest, and KNN models illustrate differing levels of classification performance. The Decision Tree's curve closely approaches the top-left corner, indicating excellent discriminative ability with a high true positive rate and low false positive rate, consistent with its strong accuracy of 98.57%. The Random Forest curve hugs the top-left corner even more tightly, reflecting similar performance and a larger AUC, which confirms its enhanced discriminative power across classification thresholds and its position as the best-performing classical model. In comparison, the KNN model's ROC curve shows a small but noticeable deviation from the ideal curve, highlighting its relatively lower performance (97.16% accuracy), particularly in recall. This suggests that KNN struggles more to maintain high sensitivity without compromising specificity at certain thresholds.

4. Combined ML Roc curve

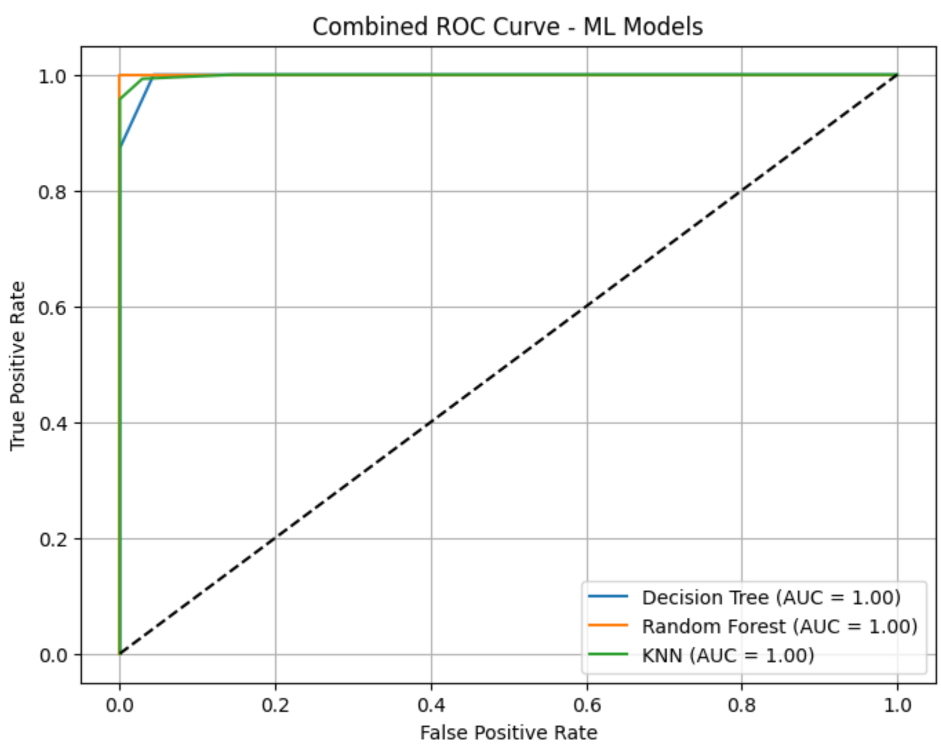

Figure 19 ML Combined curve

The combined plot clearly visualizes the relative performance of all three models. The Random Forest and Decision Tree curves appear nearly superimposed and closer to the perfect classifier point (top-left corner), while the KNN curve shows a visible gap. This confirms that the tree-based models outperform KNN for this classification task and explains the metrics differences observed in the tables.

Conclusion for experiment 01

Based on both metrics and visualizations, the Random Forest and Decision Tree classifiers demonstrated identical outstanding performance with 98.57% accuracy, 97.93% precision, 99.99% recall, and 98.95% F1 score. While KNN performed well with 97.16% accuracy and excellent precision (99.99%), it showed lower recall (95.77%), indicating it was less effective at identifying all positive cases.

The exceptional performance of both tree-based models suggests that the classification boundaries in your dataset can be effectively captured using hierarchical decision rules. The perfect recall scores (99.99%) of both Random Forest and Decision Tree models indicate they rarely miss positive cases, which could be particularly valuable in applications where false negatives must be minimized.

For applications where both precision and recall are important, either the Random Forest or Decision Tree would be recommended models from this experiment, with their identical performance metrics suggesting similar effectiveness for this particular classification task..

2.2 Experiment 02

1. Artificial Neural Network

Accuracy	**Precision**	**Recall**	**F1 Score**
98.10%	99.28%	97.88%	98.58%

Table 10 ANN Results

The ANN achieved solid performance with 98.10% accuracy. Its precision (99.28%) outperforms its recall (97.88%), indicating the model is better at avoiding false positives than missing positive cases. The F1 score of 98.58% shows good balance between precision and recall.

2. Recurrent Neural Network

Accuracy	Precision	Recall	F1 Score
99.52%	99.30%	99.99%	99.64%

Table 11 RNN Results

The RNN demonstrated exceptional performance with 99.52% accuracy. Its excellent precision (99.30%) and near-perfect recall (99.99%) result in a very high F1 score of 99.64%. This outstanding classification indicates the model excels at capturing both positive and negative cases with minimal errors, suggesting the sequential aspects of the data are valuable for classification.

3.Long short-term memory:

Accuracy	Precision	Recall	F1 Score
99.05%	98.61%	99.99%	99.30%

Table 12 LSTM Results

The LSTM network achieved excellent performance with 99.05% accuracy. While its precision (98.61%) is slightly lower than the RNN's, it maintains the same near-perfect recall (99.99%). The resulting F1 score of 99.30% is very high, though slightly below the RNN's performance, suggesting that for this particular task, the more sophisticated memory mechanisms in LSTM didn't provide additional benefits over the standard RNN.

Experiment 02: ROC curves

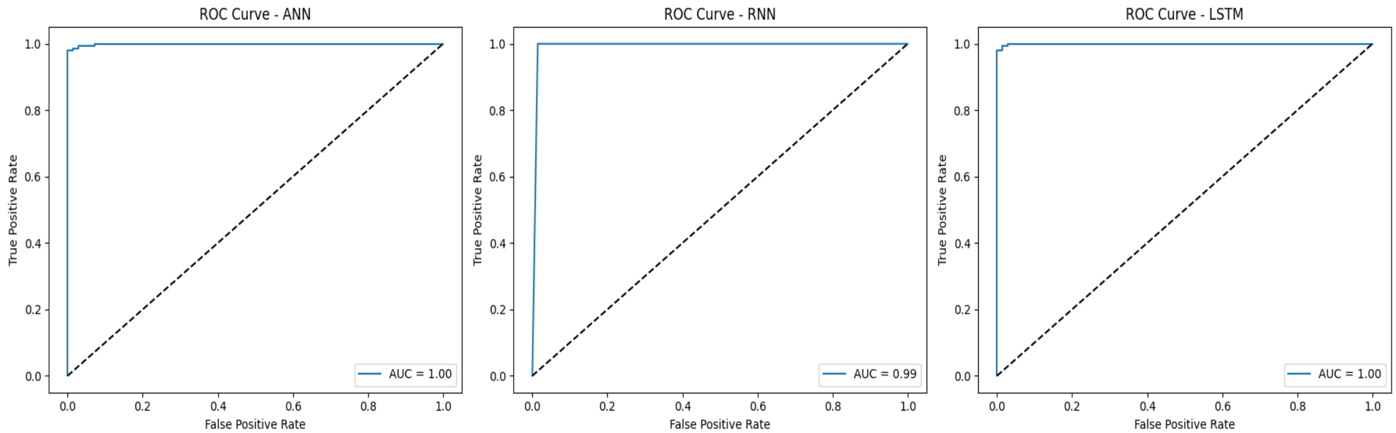

Figure 20 Experiment 02 ROC curves

The ROC curves for the ANN, RNN, and LSTM models highlight their varying classification performances. The ANN’s curve approaches the top-left corner, indicating strong but not perfect performance (98.10% accuracy), with a slight deviation suggesting some trade-offs between sensitivity and specificity at certain thresholds—trade-offs that the recurrent models manage more effectively. In contrast, the RNN’s ROC curve nearly hugs the top-left corner, illustrating near-perfect classification ability and aligning with its outstanding performance metrics , confirming its superior discrimination capability. Similarly, the LSTM's ROC curve mirrors that of the RNN, showing near-perfect adherence to the top-left edge and almost identical performance (99.05%), reinforcing that both recurrent architectures are equally proficient at optimizing the balance between sensitivity and specificity.

4.Combined Neural Models Roc Curve

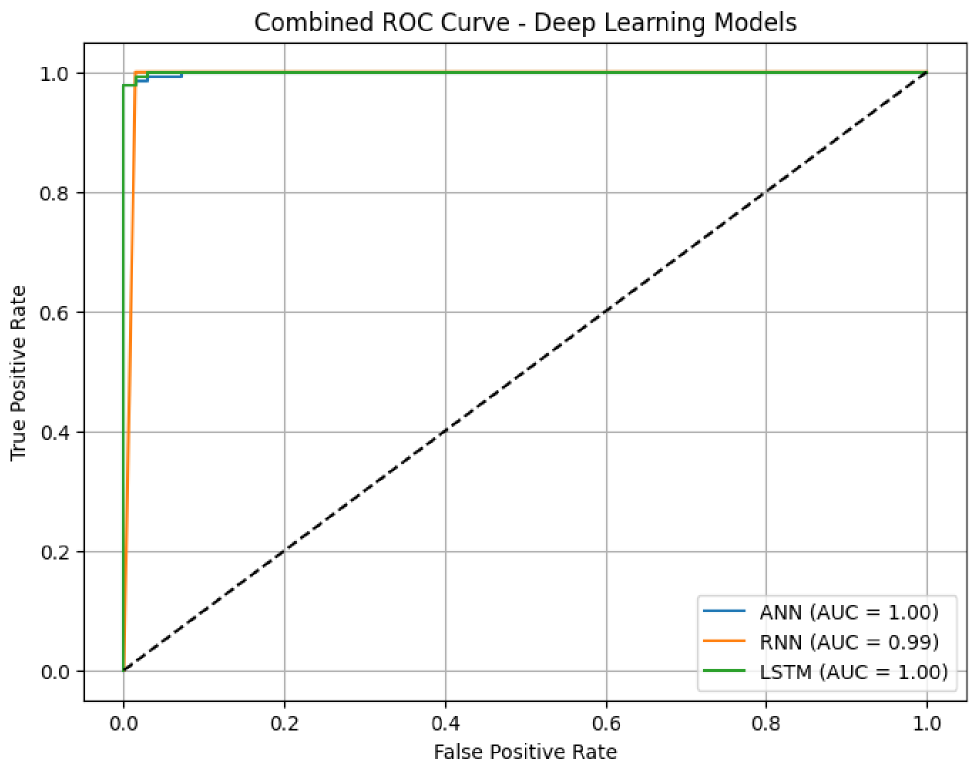

Figure 21 Combined DL ROC curves

The combined ROC curve for neural networks clearly illustrates that both RNN and LSTM significantly outperform the standard ANN. Their curves are practically indistinguishable from each other and hug the perfect classifier point more closely. The ANN curve, while still showing strong performance, has a visible gap compared to the recurrent models. This visualization perfectly reflects the numerical advantages seen in the metrics tables

Conclusion for Experiment 02

The recurrent models (RNN and LSTM) significantly outperformed the standard ANN in this classification task, achieving near-perfect results across all metrics. This suggests that capturing sequential patterns or dependencies in your data is crucial for optimal performance. Interestingly, the more complex LSTM architecture did not provide additional benefits over the standard RNN, indicating that the simple recurrent structure was sufficient to model the patterns in your data. For applications requiring the highest possible accuracy, either RNN or LSTM would be excellent choices, with RNN potentially preferred for its simpler architecture if computational efficiency is a concern.

3.Comparison of our experiments results with related works results

Study	**Dataset**	**Machine learning techniques**	**Results**
Analysis and detection of Autism Spectrum Disorder using Machine Learning techniques.	**Children's Dataset**	CNN	**98.30%**
		SVM	**98.30%**
		ANN	**98.30%**
	Adult Dataset	CNN	**99.53%**
		KNN	**95.75%**
		ANN	**97.64%**
	Adolescents' Dataset	CNN	**96.88%**
		SVM	**95.23%**
		KNN	**80.95%**
Automated Detection of Autism Spectrum Disorder Using a Convolutional Neural Network	**ABIDE I dataset**	CNN	**70.22%**

DeepASDPred: a CNN-LSTM-based deep learning method for Autism spectrum disorders risk RNA identification	**ASD risk gene dataset**	CNN-LSTM	**93.8%**
A comparative study of early autism spectrum disorder detection using deep learning based models	**Children's Dataset**	CNN	**98%**
		RNN	**96%**
		ANN	**100%**
	Toddlers' Dataset	CNN	**100%**
		RNN	**99.52%**
		ANN	**99.05%**
Our approach	**Toddlers' Dataset**	KNN	**97.16%**
		DT	**98.57%**
		RF	**98.57%**
		ANN	**98.10%**
		RNN	**99.52%**
		LSTM	**99.05%**

Table 13 Our Models vs related studies

Our KNN model achieved 97.16% accuracy, outperforming the KNN implementations in previous studies (95.75% on Adult Dataset and 80.95% on Adolescents' Dataset from the 2020 study, though the relatively lower recall (95.77%) shows room for improvement in detecting all positive cases. Both our Random Forest and Decision Tree models achieved identical performance metrics (98.57% accuracy, 97.93% precision, 99.99% recall), demonstrating the effectiveness of hierarchical decision rules for this classification task which was not used in related works, highlighting our contribution in evaluating these approaches for ASD detection. The near-perfect recall values are particularly significant for clinical applications where missing cases (false negatives) must be minimized.

Our ANN model (98.10% accuracy) performed comparably to ANNs in previous studies (98.30% on Children's Dataset, 97.64% on Adult Dataset from 2020), though slightly below the perfect accuracy reported in the 2024 study on the Children's Dataset. On the other hand, our RNN model achieved outstanding performance (99.52% accuracy), outperforming the RNN implementation in the 2024 study on the Children's Dataset (96%) and matching its performance on the Toddlers' Dataset (99.52%). As for, LSTM model (99.05% accuracy) significantly outperformed the CNN-LSTM approach in the 2023 DeepASDPred study (93.8%).

4.Conclusion

Our experimental results demonstrate the exceptional potential of both machine learning and deep learning approaches for early detection of Autism Spectrum Disorder in toddlers. Across both experiments, we observed consistently high performance metrics, with several models achieving near-perfect classification results.

This comparative study of multiple machine learning and deep learning approaches for ASD detection reveals important insights into model performance characteristics. The analysis demonstrates that various methodologies can achieve high accuracy rates ranging from 97.16% to 99.52%, with each offering distinct advantages. KNN provides strong precision, tree-based models offer excellent recall rates critical for clinical applications, and neural network architectures—particularly recurrent networks—capture complex patterns in the data. When compared with previous research, these models consistently match or exceed established benchmarks across different datasets. The identical performance of Random Forest and Decision Tree models (98.57% accuracy) suggests certain dataset characteristics respond well to hierarchical decision rules, while the particularly strong performance of the RNN model (99.52% accuracy) indicates potential benefits from sequential data processing. These findings contribute to the growing body of research on automated ASD detection by providing a comprehensive cross-model comparison and highlighting that implementation details, feature engineering, and appropriate model selection significantly impact detection accuracy. Future research should focus on model interpretability, generalizability across diverse populations, and integration into clinical workflows to maximize practical impact in ASD screening and diagnosis

Chapter 4: Our Solution

Chapter 4: Our Solution

1.Introduction

In this chapter, we introduce **Spectria**, our innovative web application to facilitate early diagnosis of Autism Spectrum Disorder (ASD) in toddlers through advanced deep learning techniques. Following the exceptional performance demonstrated by our deep learning experiment in the previous chapter, we have developed a solid, simple-to-use screening system that is accessible to parents, caregivers, and healthcare professionals alike.

Early identification of ASD is still an essential challenge within developmental medicine, and timely treatment makes a substantive difference in affected children's eventual outcomes. Routine screening has also typically required completion of long forms, specialized expertise to administer the screening, and lack of familiarity within underserved populations. **Spectria** overcomes these barriers with the use of advanced deep learning techniques to afford rapid, simple screening that is accessible and complements professional evaluation.

Spectria development required a cutting-edge technology backbone that combines front-end user experience with back-end analysis capabilities. In this chapter, we summarize the end-to-end technology used—Node.js, TypeScript, React, Supabase, TensorFlow, and Flask—and describe how these components play together to give an end-to-end screening solution. We also present the multilingual interface design facilitating accessibility within different linguistic and cultural settings as a demonstration of our commitment to making ASD screening devices available to as large an audience as possible.

By simplifying our high-performance deep learning techniques to an accessible, usable app, **Spectria** is a significant step towards putting sophisticated computational methods to work to solve real-world health issues. The ensuing technical deployment and user experience design provide a close-up accounting of how the groundbreaking screening device becomes a reality.

2.Tools

a. Node.js

Node.js is a powerful runtime environment that allows developers to execute JavaScript code outside of a web browser, usually on a server. It uses an event-driven, non-blocking I/O model that makes it efficient and ideal for building scalable web applications.

In our interface , we used Node.js to handle the backend operations of the website, such as managing server requests, connecting to databases, and providing the core logic that ties the front end and back end together, ensuring a fast and seamless experience for users. [31]

b. TypeScript

TypeScript is a strongly typed programming language that builds on JavaScript by adding static types, making it easier to catch bugs early in the development process and maintain large codebases. It enhances code quality, improves readability, and allows for better collaboration, especially on complex projects. We chose TypeScript for this project to make our codebase more robust and organized, reduce the chances of runtime errors, and ensure that as the project grows, it remains easy to manage and scale. [32]

c. React

React is a popular JavaScript library for building user interfaces, especially for creating dynamic and responsive single-page applications. It allows developers to build reusable UI components, manage application state efficiently, and deliver a smooth user experience without unnecessary page reloads. We used React to design and develop the front end of the website, creating an engaging and interactive environment where users could navigate easily, submit their stories, and view their results with a modern, clean interface.

d. Supabase

Supabase is an open-source backend-as-a-service platform that offers features like a hosted database (PostgreSQL), authentication, storage, and real-time data updates, making it an excellent alternative to services like Firebase. It provides an easy-to-use API and built-in security features that allow developers to manage data without worrying about building backend systems from scratch. [33] In our project, we used Supabase primarily to store the stories submitted by users and their analysis results, ensuring that all data is securely saved, easily retrieved, and efficiently managed without the need to build a full custom database system.

e. Anaconda and Jupyter Notebook

Anaconda is a distribution of Python and R for scientific computing and data science, bundling together essential libraries and tools, while Jupyter Notebook is an interactive coding environment that lets users write and execute code in a step-by-step, cell-based format. This setup is particularly useful for experimenting with data, visualizing results, and building machine learning models in a clear and iterative way. Anaconda and Jupyter Notebook helped us to develop the models needed for analyzing and processing the results, allowing us to test different approaches easily, and visually monitor the progress and performance of our experiments in an organized manner.

f. TensorFlow

TensorFlow is an open-source machine learning library developed by Google that provides a comprehensive ecosystem for building, training, and deploying machine learning and deep learning models. It supports complex numerical computations and offers a high level of flexibility, making it ideal for everything from simple experiments to production-ready AI systems. In our study, we utilised TensorFlow to build and train the machine learning and deep learning models that predict accurate results, taking advantage of TensorFlow's powerful tools for handling large datasets, fine-tuning model performance, and ultimately creating intelligent systems that can learn from and adapt to the data over time. [34]

g. Flask

Flask is a lightweight, open-source web framework written in Python that is designed for building web applications quickly and with minimal code. It provides the essential tools and libraries needed to create web servers, handle HTTP requests, and build APIs. Flask is highly flexible and easy to extend, making it ideal for deploying deep learning models and creating interactive, data-driven applications. In our context , Flask was used to link the frontend with the backend to interact with the trained models, input data, and receive real-time predictions. [35]

3.Spectria

In the pursuit of developing an effective and reliable solution for autism detection, we embarked on an extensive research journey that involved rigorous data preprocessing, exploratory data analysis, and the evaluation of multiple machine learning and deep learning models. After meticulous experimentation and validation, we are proud to introduce Spectria, a cutting-edge tool designed to deliver the most accurate and robust results in autism spectrum disorder (ASD) detection. Specteria represents the culmination of our efforts to harness the power of advanced deep learning techniques, including ANNs ,LSTMs and RNNs. to address the challenges associated with early and accurate autism diagnosis.

4.Our logo

‘**Spectria**’ is a name that encapsulates both the essence of autism as a spectrum and the integration of artificial intelligence (AI) in deep learning research. The term "**spectria**" is derived from "**spectrum**," which highlights the diverse range of experiences and characteristics within autism, acknowledging that no two individuals on the spectrum are alike. Additionally, "IA" within the name stands for "intelligence artificielle" (the French equivalent of AI), further reinforcing the connection to artificial intelligence. This dual significance reflects the project's mission to harness AI in understanding, analyzing, or assisting with autism-related challenges.

The logo visually represents this concept, featuring a brain to symbolize cognition and neural processes, while data elements signify the technological foundation of deep learning. The spectrum colors:red, orange, yellow, green, blue, indigo, and violet, incorporated into the design serve as a visual metaphor for the vast and varied nature of autism, reinforcing the idea that AI can be used to capture and interpret this diversity. Together, the name and logo work in harmony to express the innovative approach of Spectria in blending AI with autism research.

Figure 22 Spectria logo

5.Our interface

The **Spectria** interface has been carefully designed to be highly user-friendly and accessible to a wide audience, including parents, caregivers, and professionals. Recognizing the diverse backgrounds of potential users, the platform is available in three languages: English, Arabic, and French. This multilingual support ensures that users from different linguistic and cultural contexts can interact with the tool comfortably and without barriers.

Navigation through the interface is simple and intuitive. Users can move seamlessly between sections with minimal effort, guided by a clean layout and clear instructions at each step. The design prioritizes clarity and ease of use.

The screening process itself is streamlined for efficiency and clarity. Users are first presented with a form that requires some demographic information: age, gender, ethnicity, jaundice, asd history and who completed the test.

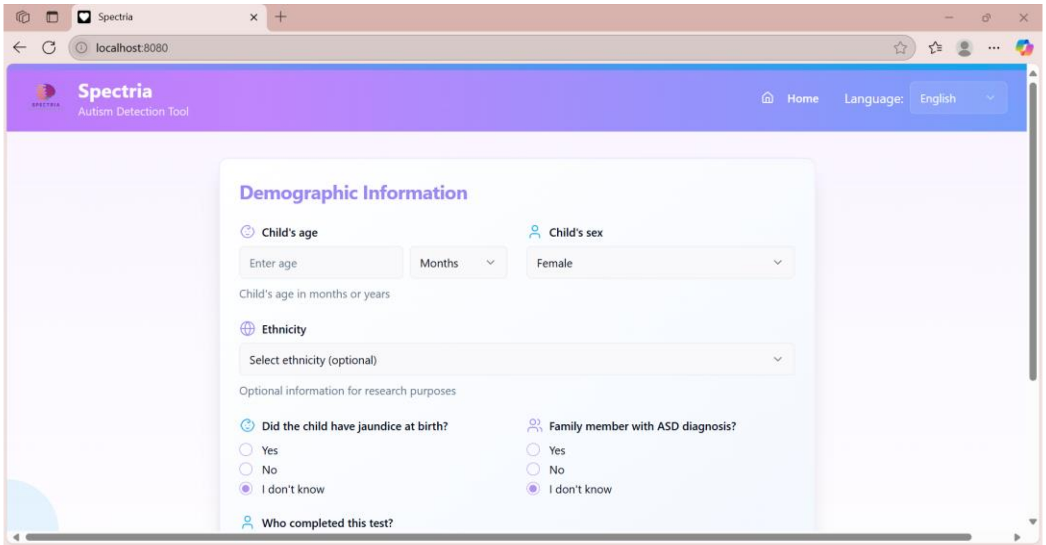

Figure 23 Demographic info page

Then faced with a series of simple, straightforward questions, each requiring only a "Yes" or "No" answer. These questions are specifically curated to cover behavioral patterns and developmental milestones relevant to autism spectrum assessment.

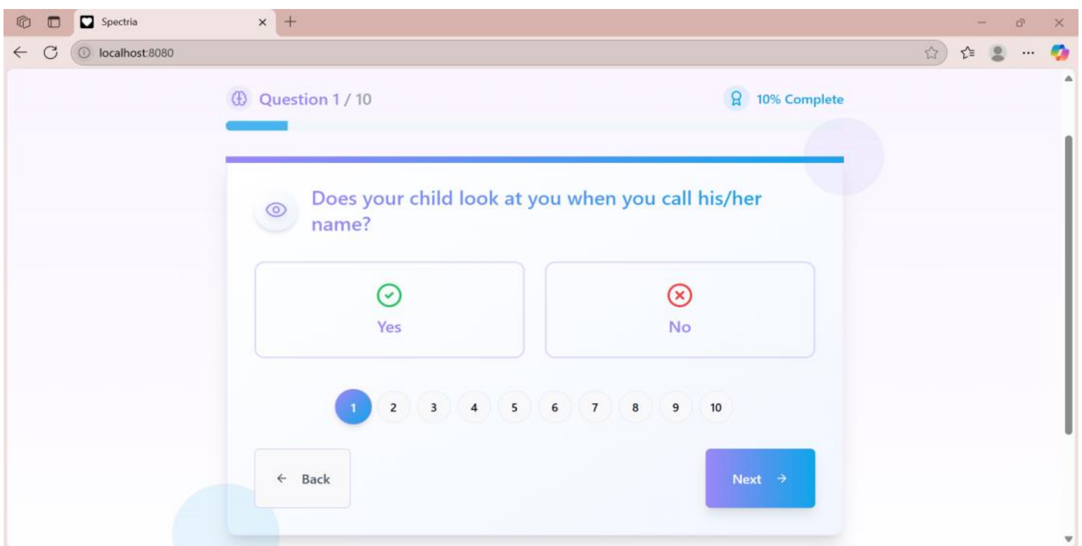

Figure 24 Questions Page

Once all responses are collected, **Spectria** automatically applies its trained deep learning model to the data. The model, developed and validated with rigorous testing, analyzes the inputs and provides an immediate assessment, indicating the potential level of autism using the score and the other factors.

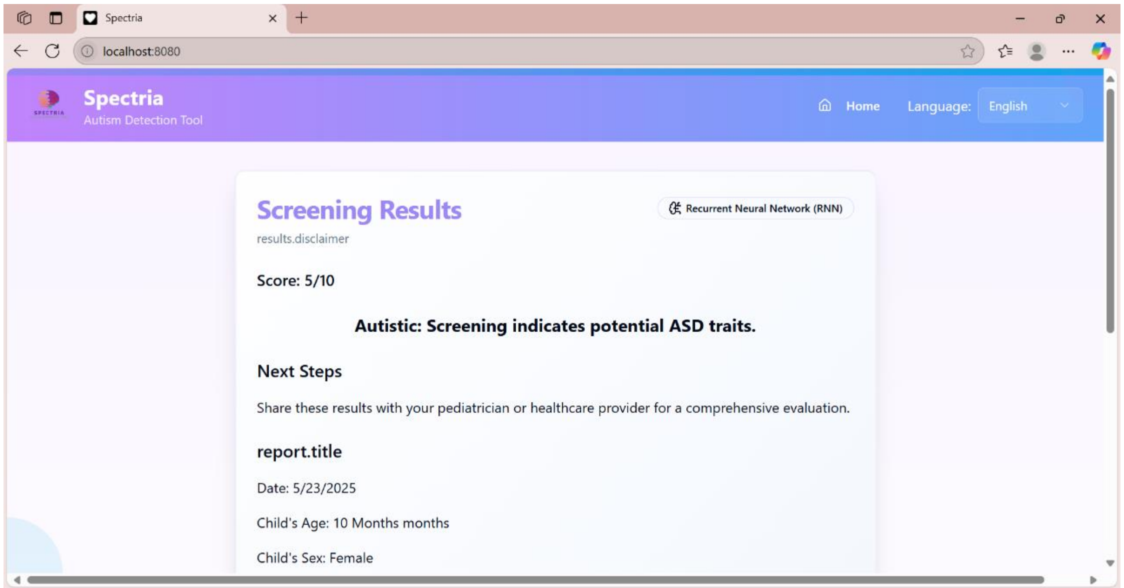

Figure 25 Results page

Users also have the option to print the full results, allowing them to easily share the screening outcome with healthcare providers, specialists, or educational institutions.

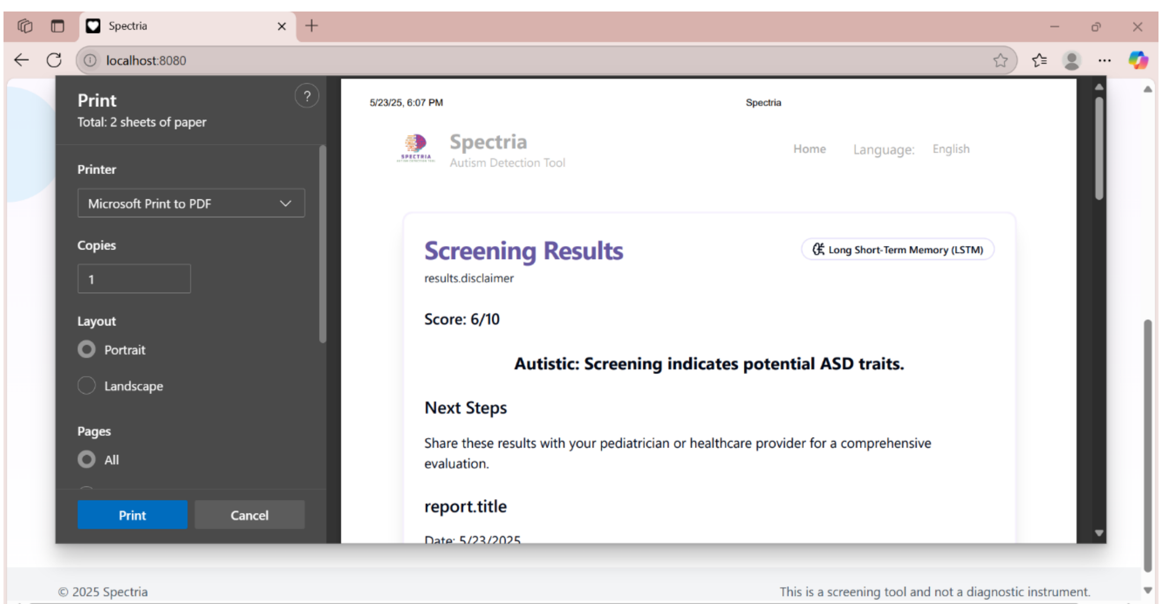

Figure 26 Print results

We have also put a section called **Add your story**, where users could share their stories and how early detection helped their child case.

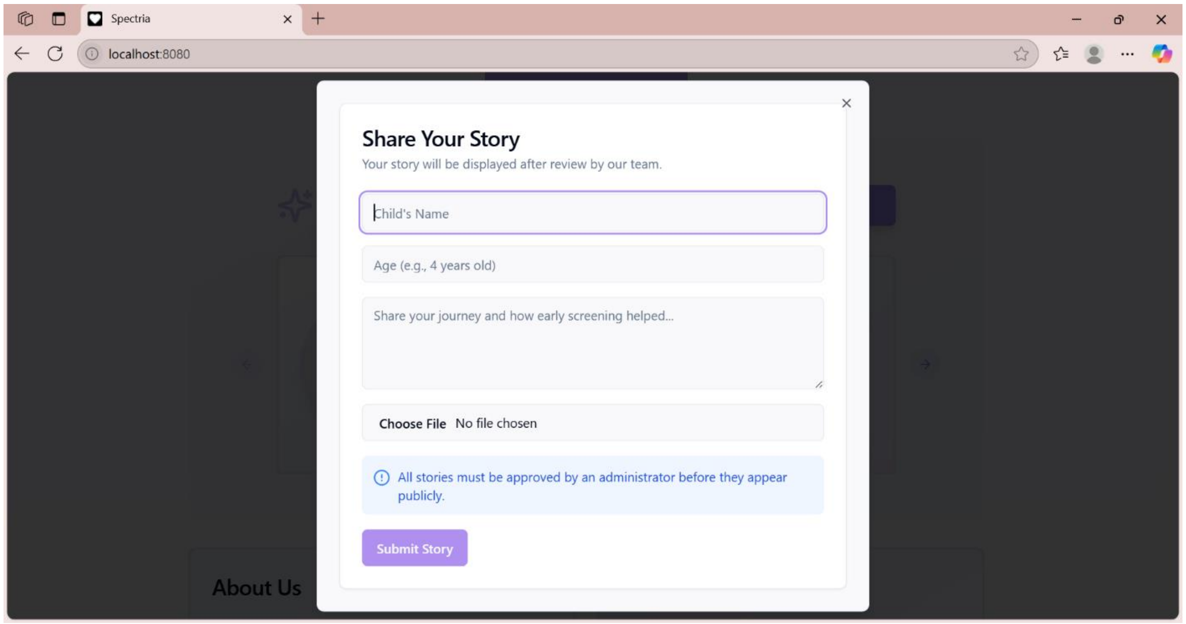

Figure 27 Share your story page

6.Conclusion

The creation and deployment of **Spectria** is an effective adaptation of cutting-edge deep learning research into a functional tool with the potential for real-world application to ASD screening. Through the utilization of the outstanding classification performance of our deep learning models and packaging them into an accessible, multilingual web-based application, we have developed a resource that stands to significantly reduce the barriers to early ASD screening.

Our end-to-end tech stack, illustrates how web technologies of today can seamlessly cross the gap between deep learning research and actual end-user use. The simple, easy-to-navigate interface design prioritizes user experience, making it so that even non-technical users can simply navigate through the screening process.

Add to this, Spectria's multilingual capabilities for English, Arabic, and French, reflecting the global reach of autism spectrum disorder and the importance of culturally sensitive measurement instruments. The straightforward yes/no question format, coupled with prompt feedback and printable report generation, gives a screening procedure that respects users' time pressures yet offers meaningful information to discuss with healthcare professionals.

General conclusion

General conclusion

Autism Spectrum Disorder (ASD) represents a significant challenge for early diagnosis and intervention, given its complex, diverse manifestations and the limitations of traditional diagnostic methods. Through this study, we have explored how modern artificial intelligence (AI) techniques, particularly deep learning, can provide innovative and effective solutions for enhancing early ASD detection.

We introduced **Spectria,** a user-friendly, multilingual tool that utilizes a combination of Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) to screen behavioral data and predict potential ASD traits with high accuracy. The results obtained through our experiments demonstrated that these models can achieve outstanding performance, offering a fast, accessible, and reliable initial screening method.

Beyond technical performance, **Spectria** has been designed with a deep commitment to user accessibility, cultural adaptability, and support for non-expert users, making it a practical tool for families, caregivers, and healthcare providers worldwide. Furthermore, the envisioned future enhancements, such as the integration of multimedia data analysis, continuous learning systems, and expanded language support, position **Spectria** as a project with significant potential for long-term evolution and impact.

In conclusion, this work highlights the powerful role that AI can play in early autism detection, making screening tools more available and effective for all communities. By combining technological innovation with a strong user-centered approach, Spectria contributes not only to the field of AI in healthcare but also to the broader mission of fostering early support, understanding, and inclusion for individuals on the autism spectrum.

1.Future plans

a. Dynamic and Adaptive Questionnaires

Future versions of **Spectria** could move beyond static question lists by adopting dynamic and adaptive questionnaires. Based on the user's previous responses, subsequent questions could be tailored to probe deeper into specific behavioral patterns. This approach would make the screening more personalized and increase the model's sensitivity to subtle manifestations of autism spectrum disorders.

b. Integration of Multimedia Inputs

To improve screening accuracy, **Spectria** could incorporate multimedia data collection. For example, users might upload short videos of behaviors, speech recordings, or even drawings. By analyzing visual and audio patterns through computer vision and speech recognition algorithms, the model could detect non-verbal cues that simple questionnaires might miss, resulting in more comprehensive and precise assessments.

c. Continuous Learning and Model Improvement

Implementing a continuous learning mechanism would allow **Spectria** to evolve over time. With proper user consent and anonymization, new screening data could be used to periodically retrain and fine-tune the model. This would help maintain high accuracy, adapt to new behavioral trends, and ensure that **Spectria** remains at the forefront of autism screening technologies.

d. Integration of Transformer-Based Architectures

A promising future enhancement is the adoption of transformer-based models, such as BERT, Vision Transformers (ViT), or custom transformer models adapted to behavioral data. Transformers have revolutionized natural language processing and computer vision through their ability to handle sequential and multi-modal data effectively. Integrating transformers could empower **Spectria** to manage more complex datasets, improve prediction robustness, and even allow multi-modal assessments combining text, audio, and video inputs.

e. Partnerships with Healthcare Providers

Establishing collaborations with hospitals, clinics, and autism centers would reinforce **Spectria**'s credibility and broaden its impact. By integrating **Spectria** as a preliminary screening tool in pediatric healthcare settings, early identification rates could be significantly improved, leading to earlier interventions, better long-term outcomes and extending or even building our own dataset.

Bibliography

[1] M. M. FITZGERALD, The history of autism in the first half century of the 20th century: new and revised.The history of autism in the first half century of the 20th century:p8.

[2] E. J. K. M. K. V. Dawson, "Social attention: a possible early indicator of efficacy in autism clinical trials," *Journal of neurodevelopmental disorders,* 2012.

[3] B. Z. S Dalili, Understanding Autism Spectrum Disorder (ASD) Unraveling the Mysteries.

[4] "When do children usually show symptoms of autism?," National Institute of Child Health and Human Development (NICHD),. Available: https://www.nichd.nih.gov/health/topics/autism/conditioninfo/symptoms-appear.

[5] P. L. R. K.-H. H. L. S Sandin, "The familial risk of autism," *JAMA,* 2014.

[6] " New Genetic Clues Uncovered in Largest Study of Families with Autism.," [Online]. Available: UCLA Health. New Genetic Clues Uncovered in Largest Study of Families with Autism. Available at: https://www.uclahealth.org/news/release/new-genetic-clues-uncovered-largest-study-families-with.

[7] "Autism," American Brain Foundation, Available: https://www.americanbrainfoundation.org/diseases/autism/.

[8] M. Mattila, " Autism spectrum disorders: An epidemiological and clinical study," 2014.

[9] " Complete Guide to Autism Tests," Available: https://autism.fratnow.com/blog/complete-guide-to-autism-tests/.

[10] N. Langone, "Autism Spectrum Disorder in Children: Diagnosis," Available: https://nyulangone.org/conditions/autism-spectrum-disorder-in-children/diagnosis.

[11] "What is Artificial Intelligence (AI)?," IBM, Available: https://www.ibm.com/topics/artificial-intelligence.

[12] "Deep learning vs Machine learning," Google Cloud, Available: https://cloud.google.com/discover/deep-learning-vs-machine-learning.

[13] T. &. H. P. Cover, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory,* p. 21.

[14] J. R. Quinlan, Induction of Decision Trees. Machine Learning.

[15] L. Breiman, Random Forests. Machine Learning.

[16] Y. B. Y. &. H. G. LeCun, "Deep learning," *Nature,* p. 521, 2015.

[17] I. B. Y. &. C. A. Goodfellow, "Deep Learning," *MIT Press,* 2016.

[18] J. G. F. &. E. D. Schmidhuber, " Learning nonregular languages: A comparison of simple recurrent networks and LSTM," *Neural computation.*

[19] O. T. M. Y. K. D. C. S. H. S. J. Baclic, " Natural language processing (NLP) a subfield of artificial intelligence," *CCDR,* 2020.

[20] S. &. M. S. Raj, "Analysis and detection of autism spectrum disorder using machine learning techniques," *Procedia Computer Science,* 2020.

[21] T. Rahman, "A comparative study of early autism spectrum disorder detection using deep learning based models.".

[22] Z. A. M. S. S. Z.-M. M. A. M. A. U. R. .. &. S. V. Sherkatghanad, "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in neuroscience,* 2020.

[23] Y. X. H. &. S. G. Fan, "DeepASDPred: a CNN-LSTM-based deep learning method for Autism spectrum disorders risk RNA identification.," *BMC bioinformatics,* 2023.

[24] "What is a Dataset?," Available: https://www.databricks.com/glossary/what-is-dataset.

[25] "Autism Screening for Toddlers Dataset," Kaggle,. Available: https://www.kaggle.com/datasets/fabdelja/autism-screening-for-toddlers.

[26] "Data Preprocessing in Data Mining," Available: https://www.geeksforgeeks.org/data-preprocessing-in-data-mining.

[27] "One hot encoding for machine learning," Available: https://www.geeksforgeeks.org/ml-one-hot-encoding.

[28] "Data Pre-processing with Sklearn using Standard and MinMax Scaler," Available: https://www.geeksforgeeks.org/data-pre-processing-wit-sklearn-using-standard-and-minmax-scaler/.

[29] N. S. K. K. S. A. S. Pushpa Singh, "Machine Learning and the Internet of Medical Things in Healthcare," *Academic Press,* pp. 89-111, 2021.

[30] https://scikit-learn.org/stable/api/sklearn.metrics.html.

[31] https://en.wikipedia.org/wiki/Node.js.

[32] https://gibbok.github.io/typescript-book/book/typescript-introduction

[33] https://supabase.com/.

[34] "Tensorflow,". Available: https://opensource.google/projects/tensorflow.

[35] https://flask.palletsprojects.com/.

[36] Q. Z. W. C. M. P. G. Y. X. &. Y. Q. Luo, "Self-Attention and Transformers: Driving the Evolution of Large Language Models," *EEE 6th International Conference on Electronic Information and Communication Technology,* p. 401, 2023.