VISVESVARAYA TECHNOLOGICAL UNIVERSITY

BELAGAVI - 590 018, KARNATAKA.

Major Project

ON

"IDENTIFICATION AND PREDICTION OF AUTISM DISORDER"

Submitted in the partial fulfillment of requirements for the

MAJOR PROJECT(BCS685)

Submitted by

Tanushree S C	**4BD22CD047**
Varshini N H	**4BD22CD050**
Sneha S C	**4BD22CD043**
Navya B R	**4BD22CD021**

PROJECT GUIDE:
Prof. Usha C M.Tech
Assistant Professor
CS&E (Data Science)

PROGRAM COORDINATOR:
Dr. Pradeep N Ph. D
Professor and Head
CS&E (Data Science)

Bapuji Institute of Engineering & Technology
Department of Computer Science and Engineering
(Data Science)
Davanagere - 577004

2024-2025

Bapuji Institute of Engineering and Technology

Davanagere -577004

Department of Computer Science and Engineering (Data Science)

CERTIFICATE

This is to certify that **TANUSHREE S CHAKRASALI, SNEHA S CHIKKARADDI, VARSHINI N H** and **NAVYA B R** bearing **USN 4BD22CD047, 4BD22CD043, 4BD22CD050** and **4BD22CD021** respectively of **Computer Science and Engineering (Data Science)** department have satisfactorily submitted the major project Synopsis entitled **"IDENTIFICATION AND PREDICTION OF AUTISM DISORDERS"**. The project synopsis has been approved as it satisfies the academic requirements for the academic year 2024-2025.

14/05/2025

PROJECT GUIDE
Prof. Usha C M.Tech
Assistant Professor
CS&E (Data Science)

PROJECT COORDINATOR
Prof. Gowramma B H M.Tech (ph. D)
Assistant Professor
CS&E (Data Science)

PROGRAM COORDINATOR
Dr. Pradeep N ph. D
Professor and Head
CS&E (Data Science)

Date: 16-05-2025

Place: Davangere

Bapuji Educational Association (Regd.)

Bapuji Institute of Engineering and Technology, Davangere-577004

Vision and Mission of the Computer Science and Engineering (Data Science) Department

VISION:

"To provide a quality and holistic education in data science, data analytics, data visualization, industry collaborations and research for empowering individuals to derive knowledge, thereby transform the potentials in data for the betterment of society."

MISSION:

1.	Educate and prepare students with a strong foundation in data science, equipping them with the skills, knowledge, and ethical principles needed to excel in data-driven fields.
2.	Foster collaborations with industries to adopt modern data science and visualization tools which solves the real-world problems that have societal benefits.
3.	Cultivate a culture of life-long learning with intellectual curiosity in data science and nurturing individuals who are passionate about data driven decision making.

Program Educational Objectives (PEOs):

PEO1	Graduates will work in the area of applications of software development, Artificial Intelligence, Machine Learning, Data Analytics and Data Visualization.
PEO2	Graduate will exhibit professional ethics and moral value with capabilities of working as an individual and as a team member in the corporate world to contribute toward the need of the dynamic requirements of industry and society.
PEO3	Graduates will become responsible successful software professionals with leadership and managerial quality in the modern software industries based on their strong skills on theoretical and practical foundation.

Program Specific Outcomes (PSOs):

PSO1	Students will be able to develop sustainable and efficient algorithm solutions for the real time problems by applying their problems solving skills.
PSO2	Students will be able to develop a solution for the given problem in the area of Artificial Intelligence, Data Analytics and any other societal problems through conducive environment and infrastructure.

ABSTRACT

Transformer models, renowned in natural language processing and computer vision, are increasingly applied in healthcare, particularly for the detection and prediction of autism spectrum disorder (ASD). ASD is a complex neurodevelopmental condition marked by challenges in social interaction, communication, and behavior. Transformers excel at capturing long-range dependencies, rendering them effective for analyzing diverse ASD-related data, such as neuroimaging, genetic sequences, and speech. Vision Transformers (ViTs) can evaluate MRI/fMRI scans to detect atypical brain patterns, while sequence transformers focus on communication and behavior.

ACKNOWLEDGEMENT

With profound respect and gratitude, we extend our heartfelt thanks to our esteemed institution, **"BAPUJI INSTITUTE OF ENGINEERING AND TECHNOLOGY",** for its unwavering support, well-qualified faculty, and state-of-the-art laboratories equipped with all the necessary resources.

We express our sincere thanks to our guide, **Prof. Usha C**, for her constant encouragement, unwavering support, and invaluable guidance throughout this project. Her expertise and mentorship have been pivotal to our success.

Our heartfelt gratitude goes to **Dr. Pradeep N**, the Head of the Department of Computer Science & Engineering (Data Science), for his consistent support, valuable insights, and encouragement, which have greatly facilitated the progress of our project.

We also extend our sincere appreciation to our principal, **Dr. Aravind H B**, for his continuous moral support and motivation, which have inspired us throughout our journey.

A special thanks to all the faculty members of the **Department of Computer Science and Engineering (Data Science)** for their constructive feedback, suggestions, and guidance, which have significantly contributed to our learning and growth.

We are deeply grateful to our family members and friends for their unwavering support, advice, and encouragement, which have been a source of strength during this endeavor. Finally, we thank everyone who has contributed to the successful completion of this project.

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Description of the Project

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects social interaction, communication, and behavior. The early identification and accurate prediction of ASD are critical, as early intervention has been shown to significantly improve developmental outcomes for children. Traditional methods of diagnosing ASD, which often involve clinical interviews, behavioral observations, and standardized diagnostic assessments, can be time-consuming and subjective, leaving room for misdiagnosis or delayed diagnosis [1].

The etiology of the disease remains unclear; nonetheless, it is believed to be associated with biological factors including genetic anomalies, neuroinflammation, and adverse prenatal conditions. The significant increase in the number of children diagnosed with Autism Spectrum Disorder (ASD) highlights the necessity for further research on this population. In particular, appropriate clinical practices are crucial [2]. In recent years, machine learning (ML) and deep learning (DL) models, particularly transformer models, have shown great potential in various fields, including healthcare. Transformers are a class of deep learning models that excel at handling sequential data and capturing long-range dependencies within datasets. Originally developed for natural language processing (NLP), transformer models have expanded into other domains, such as image recognition, genetics, and medical data analysis, due to their flexibility and power [4][5].

The issue of autism spectrum disorder (ASD) has been rapidly increasing across all age groups within the human population. Timely identification of this neurological condition can significantly contribute to the preservation of the individual's mental and physical well-being. Autism spectrum disorder pertains to challenges associated with the development of the human brain. An individual affected by autism spectrum disorder typically faces difficulties in social interactions and communication with others[1].

In the context of ASD, transformers can be applied to a wide variety of data types, including behavioral patterns, neuroimaging data, genetic information, and even speech or audio recordings. These models can identify subtle patterns, extract meaningful features, and make accurate predictions that might not be immediately apparent to human clinicians. As such, transformer-based systems are emerging as promising tools for the identification,

classification, and prediction of ASD.

The application of transformer models in ASD presents a transformative opportunity to enhance early diagnosis, predict developmental trajectories, and personalize intervention strategies. This introduction will explore how transformer models are being used to improve the identification and prediction of autism spectrum disorder, highlighting their potential benefits and challenges in the healthcare field [6].

1.2 Problem Statement

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects communication, behavior, and social interactions. Prompt and accurate diagnosis of ASD is crucial for effective intervention and support. However, traditional diagnostic methods are often protracted, subjective, and require specialized knowledge. This project aims to address the challenges related to the identification and prediction of ASD by developing a Transformer-based machine learning model. By leveraging behavioral screening data in conjunction with demographic information, the model will uncover patterns that assist in determining the presence of ASD in individuals. The goal is to create a reliable, automated system that can aid healthcare professionals in the early identification of ASD and improve decision-making processes.

1.3 Proposed Solution

This project proposes the development of a Transformer-based deep learning model designed to detect and forecast Autism Spectrum Disorder (ASD). The model will be trained using a dataset that includes behavioral traits, demographic details, and screening answers. Prior to training, the data will be preprocessed to ensure its cleanliness and readiness. Following this, the Transformer model will examine the data to identify patterns that suggest the probability of an individual being diagnosed with autism. The model's effectiveness will be measured using metrics like accuracy and precision, and it will be benchmarked against other machine learning models to evaluate its performance.

CHAPTER 2

LITERATURE SURVEY

ML approach has become one of the most common areas in finding functional patterns detect many diseases and find appropriate solutions. Furthermore, treating autism patients uses different methods to detect and identify patients as affected or not. **Azian A. et al**. [1] have suggested three ways: Selection Operator (LASSO), Least Absolute Shrinkage, and Chi-square, to test ML approaches that could be used for regression and classification. These techniques are K Nearest Neighbors (K-NN), Random Forest (RF), and Logistic Regression (LR). The experimental results showed the highest accuracy of the LR model, with 97.541%, among other techniques utilized. Their approach depends on 13 features selected based on the Chi-square selection model.

Suman R., and Sarfaraz M. [2], made try to find if there are specific ways to predict the ASD risk. They used Logistic Regression, Nave Bayes, K-NN, Support Vector Machines, Convolutional Neural Networks, and Neural Networks to test the necessary criteria and to analyze and classify ASD. Three ASD datasets were used (children, adults, and adolescents) to test the proposed methods. There are 292 samples for 21 attributes of ASD patients to examine children, 740 samples for 21 attributes of adults, and 104 for 21 attributes of adolescents. After applying the ML methods and extracting the results, it was found that the proposed model obtained the highest accuracy rate in the CNN classifier with 98.30%, 99.53%, and 96.88% for testing and detecting ASD for children, adults, and adolescents, respectively.

Nishat MM et al. [3] proposed a model based on ML algorithms to predict ASD and its psychological disorders that significantly affect the Individual's behavior in social life. They used the quadratic discriminant algorithm and linear analysis to detect and analyze the disease and find the appropriate treatment plan to reduce its severity. To build the ML model, they used the data of the University of California, Irvine (UCI) reservoir for analysis and discovery. They comprehensively evaluated the sensitivity, F1 score, accuracy, and Youden index. The experimental results showed the efficiency and effectiveness of the proposed model, as the Quadruple Analysis Algorithm (QDA) showed its high accuracy of 99.77% after adjusting the hyperparameters.

Md. Mokhlesur R. et al. [4], the authors made use of ML methods and highlighted essential topics related to autism. In addition, they confirmed the need to identify the best traits of autism, enhance categorization, and maintain top precision. The researchers can create an ML

system to produce promising results in detecting ASD by selecting the suitable important autistic traits and minimizing data dimensionality. Depending on the authors, several factors that affect accuracy must be addressed.

For example, include an unbalanced and insignificant data set, faulty sampling procedures, and feature redundancy.

Nurul A. et al. [5] proposed a model based on different ML techniques to identify and analyze the ASD classification. The dataset to be analyzed consists of 16 characteristics, including 703 autistic and non- autistic patients. ML methods were used to predict the state of ASD: support vector machine, K-NN, naïve Bayes (NB), J48, Bagging and Stacking.

These techniques were conducted in a simulated environment using the Waikato environment for knowledge analysis (WEKA) platform. The obtained results showed the accuracy, sensitivity, and superiority of some of the ML classifiers over the other studied classifiers, Stacking, J48, SVM, and Bagging, with an accuracy of 100% and a lower error rate.

Md Delowar H. and Muhammad Ak. et al. [6], tried to identify the most critical characteristics and automate the diagnostic procedure for ASD by utilizing current classification algorithms for better diagnosis. They tested ASD datasets from toddlers, children, teenagers, and adults. They evaluated and identified the best performing classifier among the latest classification techniques to discover the ASD of the above dataset. The results showed the efficiency and superiority of the multilayer perceptron (MLP) classifier, which achieved the highest accuracy of 100% for the four datasets. They also found that the best technique for the four ASD datasets was the “Relief F″ feature selection technique to rank the most important traits.

CHAPTER 3

OBJECTIVES

- To collect a high-quality and relevant dataset that includes behavioral, demographic, and screening information related to Autism Spectrum Disorder (ASD).
- To perform preprocessing on the collected data, including handling missing values, encoding categorical variables, normalization, and data splitting to prepare it for training the Transformer model.
- To design and implement a Transformer-based deep learning model that can learn from tabular autism related data. The model will be trained to identify patterns that distinguish individuals with ASD from those without.
- To evaluate the model using key metrics like accuracy, precision, recall, and F1- score to determine its effectiveness in identifying ASD.
- To explore the possibility of early autism prediction through feature importance analysis

CHAPTER 4

METHODOLOGY

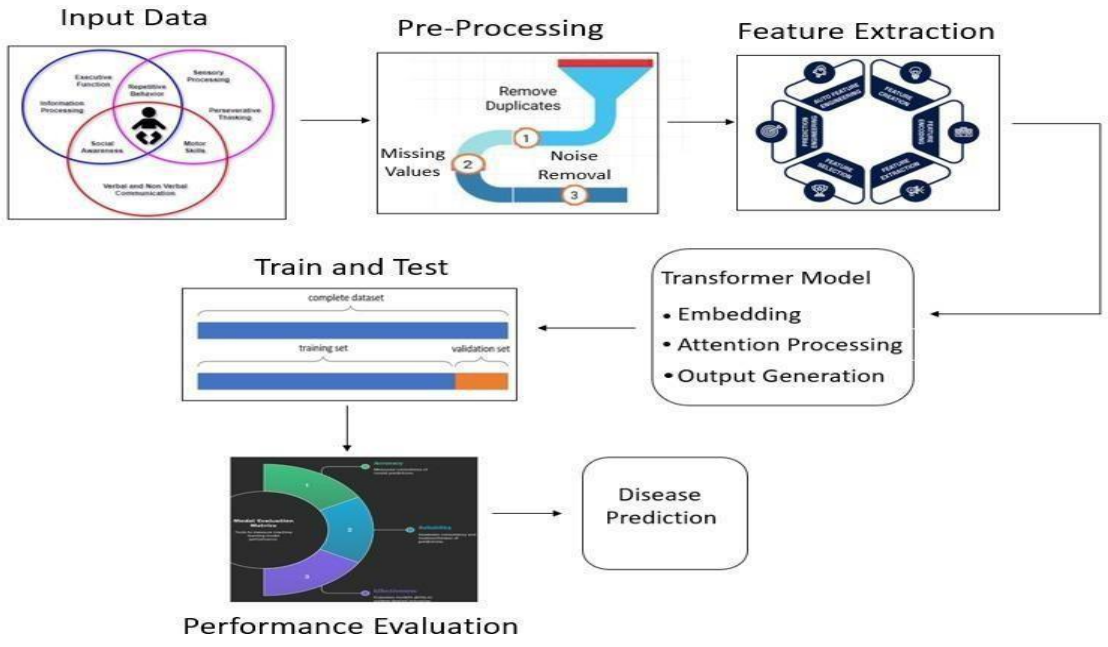

Fig. 4.1 Methodology

- Dataset

  Input: Raw dataset that contains the information needed to train a model.

- Preprocessing

  This step prepares the data for modeling:

  Missing Values & Outliers: Fill in or remove missing data and outliers.

  Noise Removal: Clean unwanted variations in data.

  Normalize: Scale features to a common range to improve model performance.

- Feature Extraction:

  Converts cleaned data into meaningful numerical features. These features help the model understand important patterns.

- Transformer Model:

  Uses embedding's and attention to learn complex relationships in the data. It generates outputs that represent the predicted condition.

- Train and Test:

  Splits the dataset into training and validation parts. The model learns on the training set and is tuned using the validation set.

- Disease Prediction:

  The trained model produces a final prediction about the disease or risk. This output is used for clinical interpretation.

- Performance Evaluation:

  Measures accuracy, precision and other metrics. This checks how well the model performs in real scenarios

CHAPTER 5

SOFTWARE REQUIREMENTS SPECIFICATIONS

5.1 Software Requirements

- Python - Due to extensive ML and DL libraries
- Numpy - Data manipulation
- Jupyter Notebook

5.2 Hardware Requirement

- Minimum 4GB RAM
- Ryzen 5 or intel-core i5
- Operating System

5.3 Possible Outcomes

1. Enhanced data quality through effective preprocessing techniques.
2. Created a precise machine learning model to differentiate between ASD and non-ASD cases.
3. Recognized essential characteristics that enhance the accuracy of ASD prediction.
4. Improved the efficiency of ASD diagnosis using AI in mental health analysis.
5. Achieved high performance metrics validating model reliability.

REFERENCES

CITED SOURCES:

[1] A.A. Abdullah, S. Rijal, S.R. Dash

Evaluation on machine learning algorithms for classification of autism spectrum disorder (ASD)

J. Phys. Conf. Ser., 1372 (1) (2019), Article 012052

[2] S. Raj, S. Masood

Analysis and detection of autism spectrum disorder using machine learning techniques Procedia Computer.

Sci., 167 (2020), pp. 994-1004, 10.1016/j.procs.2020.03.399

[3] M.M. Nishat, *et al.*

Detection of autism spectrum disorder by discriminant analysis algorithm (2022), pp. 473-482

Cited by (43)

[4] Md. Mokhlesur R. et al.

M.M. Rahman, O.L. Usman, R.C. Muniyandi, S. Sahran, S. Mohamed, R.A. Razak

A review of machine learning methods of feature selection and classification for autism

Spectrum Disorder Brain Sci., 10 (12) (2020), p. 949

[5] Nurul A. et al

N.A. Mashudi, N. Ahmad, N.M. Noor

Classification of adult autistic spectrumdisorder using machine learning approach IAES Int. J.

Artif. Intell., 10 (3) (September 2021), pp. 743-751.

[6] Md Delowar H., and Muhammad A. k. et al.

M.D. Hossain, M.A. Kabir, A. Anwar, M.Z. Islam

Detecting autism spectrum disorder using machine learning technique Health Inf. Sci. Syst., 9(1) (2021)