IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⧓ Peer-reviewed & Refereed journal ⧓ Vol. 11, Issue 4, April 2024

DOI: 10.17148/IARJSET.2024.11451

AUTISM SPECTRUM DETECTION

C M Rithika[1], Disha Gupta[2], Priyanka A H[3], Rakshita P Kulkarni[4], Roopa K Murthy[5]

Student, Artificial Intelligence and Machine Learning, K S Institute of Technology, Bengaluru, India[1-4]

Professor, Artificial Intelligence and Machine Learning, K S Institute of Technology, Bengaluru, India[5]

Abstract: The increasing prevalence of Autism Spectrum Disorder (ASD) underscores the need for accurate early detection methods to facilitate timely intervention. This study investigates the efficacy of computational models in ASD detection by leveraging both numerical data and image datasets. Employing Support Vector Machine, Logistic Regression, Random Forest, and Neural Network algorithms for numerical data analysis, and utilizing an EfficientNet model for image data analysis, a comprehensive approach is adopted. The numerical dataset, consisting of 2000 samples, yields an accuracy rate of up to 98% with grid search cross-validation using a Decision Tree classifier. Meanwhile, the image dataset, comprising 2500 images, achieves a 94% accuracy rate with the EfficientNet model. By integrating findings from both numerical and image analyses, this study provides a comprehensive report comparing the results and demonstrating the potential of combined approaches in enhancing ASD detection accuracy.

Keywords: CNN (Convolution Neural Networks), Deep learning, SVM (Support Vector Machine), Pre-processing, Feature Extraction, Segmentation.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a rapidly growing neurodevelopmental disorder characterized by significant impacts on information processing and social interaction skills. The exact etiology of ASD remains unclear, but both genetic and environmental factors are believed to be contributing influences. The Centers for Disease Control and Prevention (CDC) reports a concerning rise in ASD prevalence, with estimates suggesting a staggering 1 in 44 children in the US diagnosed with the condition. Early diagnosis is paramount for implementing effective interventions and improving the quality of life for individuals with ASD. However, there is currently no cure for this disorder. Traditional methods for diagnosing ASD heavily rely on questionnaires, clinical observations, and potentially subjective evaluations.

This approach can present significant challenges, particularly for young children who may exhibit subtle symptoms that are easily missed. Additionally, resource limitations and inconsistencies in healthcare professional expertise can further complicate the diagnostic process, potentially leading to delayed diagnoses or misdiagnoses. In response to these challenges, researchers are actively exploring the potential of Machine Learning (ML) and Deep Learning (DL) for earlier, more objective ASD diagnosis. These techniques offer exciting possibilities for analyzing various data points that can be particularly difficult to assess in young children, including behavioral patterns, facial features, and speech. ML algorithms have the potential to analyze large datasets with greater accuracy and consistency, potentially overcoming limitations faced by traditional diagnostic methods.

This paper details our implementation of a deep learning model for ASD diagnosis using facial features extracted from photographs. We leverage a publicly available dataset from Kaggle to train and evaluate our model. Our ultimate aim is to contribute to the development of automated and objective tools for earlier ASD identification. This not only has the potential to improve the lives of individuals with ASD but also holds the promise of reducing the burden on healthcare systems by facilitating earlier interventions and improved resource allocation.

II. LITERATURE SURVEY

1. A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders [1]: The paper proposes a comprehensive framework for evaluating ML techniques in the early detection of autism. The methodology involves the approach of four different Feature Scaling (FS) strategies and eight ML algorithms applied to four standard ASD datasets representing different age groups. Various statistical evaluation measures are employed to identify the best-performing classification methods and effective feature scaling techniques for each age group. The study includes risk factor calculation and attribute importance ranking using different Feature Selection Techniques (FSTs). The proposed framework achieves promising results, outperforming existing approaches, and provides detailed insights for healthcare.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 4, April 2024

DOI: 10.17148/IARJSET.2024.11451

2. A Multimodal Approach for Identifying Autism Spectrum Disorders in Children [3]: Addressing the challenge of identifying developmental diversity in children, a novel multimodal diagnosis framework is proposed. This methodology combines Electroencephalogram (EEG) and Eye-Tracking (ET) data, capturing both internal neurophysiological and external behavioral perspectives. The approach entails a two-step process utilizing Stacked Denoising Autoencoders (SDAE): initially, separate SDAE models learn features from EEG and ET modalities, followed by a third SDAE model that fuses these learned features. By automatically capturing correlations and complementarity between modalities in a latent feature space, this multimodal identification model generates informative representations with enhanced discriminability. Evaluation on a dataset comprising 40 ASD children and 50 Typically Developing (TD) children demonstrates superior performance compared to unimodal and simple fusion methods. This proposed framework holds promising potential for providing objective and accurate ASD diagnosis to assist clinicians.

3 Wavelet Analysis on Thermal Images for Affective States Recognition of Children with ASD [7]: challenges in socio-emotional engagement strategies for children on the Autism Spectrum by introducing a non-invasive method using thermal imaging to identify affective states. The hypothesis was based on cutaneous temperature changes associated with pulsating blood flow in the frontal face area. The methodology included a structured experimental setup inducing affective state expressions, a wavelet-based pattern detection technique for thermal imaging data analysis, and evaluation against a baseline model for typically developing children aged 5 to 9. The results demonstrated an 88% classification accuracy in identifying affective states in autistic children, suggesting the potential for more effective responses to improve social-emotional interaction in this population.

III. DATASETS

```
Data columns (total 28 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   CASE_NO_PATIENT'S                               1985 non-null   int64
 1   A1                                              1985 non-null   int64
 2   A2                                              1985 non-null   int64
 3   A3                                              1985 non-null   int64
 4   A4                                              1985 non-null   int64
 5   A5                                              1985 non-null   int64
 6   A6                                              1985 non-null   int64
 7   A7                                              1985 non-null   int64
 8   A8                                              1985 non-null   int64
 9   A9                                              1985 non-null   int64
 10  A10_Autism_Spectrum_Quotient                    1985 non-null   int64
 11  Social_Responsiveness_Scale                     1976 non-null   float64
 12  Age_Years                                       1985 non-null   int64
 13  Qchat_10_Score                                  1946 non-null   float64
 14  Speech Delay/Language Disorder                  1985 non-null   object
 15  Learning disorder                               1985 non-null   object
 16  Genetic_Disorders                               1985 non-null   object
 17  Depression                                      1984 non-null   object
 18  Global developmental delay/intellectual disability 1985 non-null   object
 19  Social/Behavioural Issues                       1971 non-null   object
 20  Childhood Autism Rating Scale                   1985 non-null   int64
 21  Anxiety_disorder                                1985 non-null   object
 22  Sex                                             1985 non-null   object
 23  Ethnicity                                       1985 non-null   object
 24  Jaundice                                        1985 non-null   object
 25  Family_mem_with_ASD                             1985 non-null   object
 26  Who_completed_the_test                          1985 non-null   object
 27  ASD traits                                      1985 non-null   object
```

Fig1: Numerical Dataset

The Datasets used are both numerical and image datasets to enhance ASD detection and screening. The numerical ASD screening dataset focuses on gathering personal and medical information, including age, gender, ethnicity, and health data such as whether the individual was born with jaundice or has immediate family members with pervasive developmental disorders. It includes answers to standardized questions (AQ-10-Adult) related to behavioural traits associated with ASD and tracks app usage and the screening method chosen based on age.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 4, April 2024
DOI: 10.17148/IARJSET.2024.11451

Responses to ten screening questions are recorded as binary values (0 or 1), representing ASD traits, and a final score is calculated using the scoring algorithm based on the screening method. This dataset contains 28 columns of patient information and supports classification tasks within medical, health, and social science domains. By combining image and numerical data, the implementation offers a comprehensive approach for effective early diagnosis of ASD, potentially leading to better healthcare outcomes and resource allocation.

The ASD image dataset consists of 5,880 images organized into two settings with equal number of images for autistic and non-autistic classes. Setting 1 includes 2,940 images, with each class containing 1,470 images at a resolution of 124 × 124 pixels. Setting 2 includes 2,940 images at a resolution of 248 × 248 pixels. Each image is a 3-channel image capturing a straight face posed in front of a camera, providing balanced representation across classes and supporting training, validation, and testing tasks.

Fig 2: Autistic Images from Image Dataset

Fig 3: Non-Autistic Images from Image Dataset

IV. METHODOLOGY

The proposed approach for ASD detection integrates two main tasks:

1. Facial Image-based ASD Detection: Utilizes neural network models to classify facial images as indicative of ASD or not using Deep Learning Models.

2. Numerical ASD Assessment: Evaluates ASD based on patient questionnaire responses and other numerical data using Machine learning Models.

The approach includes a Flask-based website that combines these methods and generates a report on whether a person has ASD or not.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ∺ Peer-reviewed & Refereed journal ∺ Vol. 11, Issue 4, April 2024

DOI: 10.17148/IARJSET.2024.11451

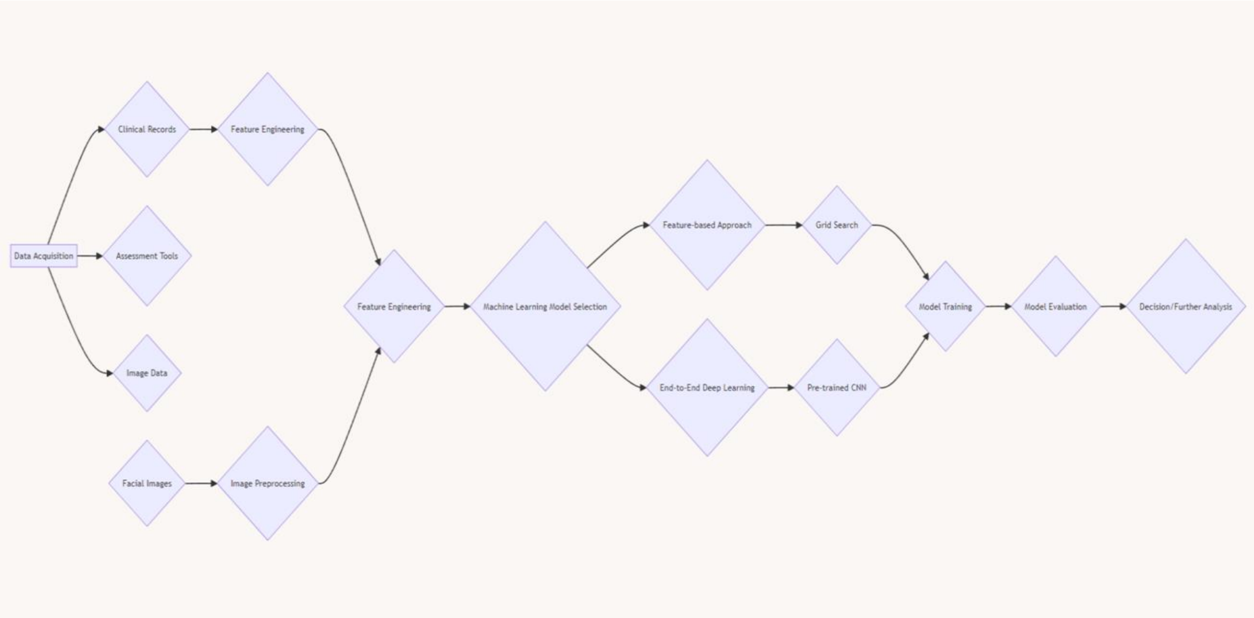

Fig 4: Overview of the proposed approach

1.Data pre-processing, Augmentation and Annotation:

The preliminary stage of data preparation for the ASD implementation involves cleaning and preprocessing the collected data, including handling missing values, normalization, and feature engineering to extract meaningful information. For the image dataset, data augmentation techniques such as resizing, recycling, and rotating are utilized to enhance training efficiency and generate an augmented set of images for training and validation sets. The images are resized to specific dimensions to ensure compatibility with the required architecture, and the dataset consists of training, testing, and validation sets. For the numerical dataset, label encoding converts special categories into numerical format, which is necessary for the algorithm, and the encoded dataframe is printed for reference. Data visualization involves histogram plotting with the `hist()` function to generate visualizations of the numerical data, as well as box plots based on selected numerical features for deeper analysis.

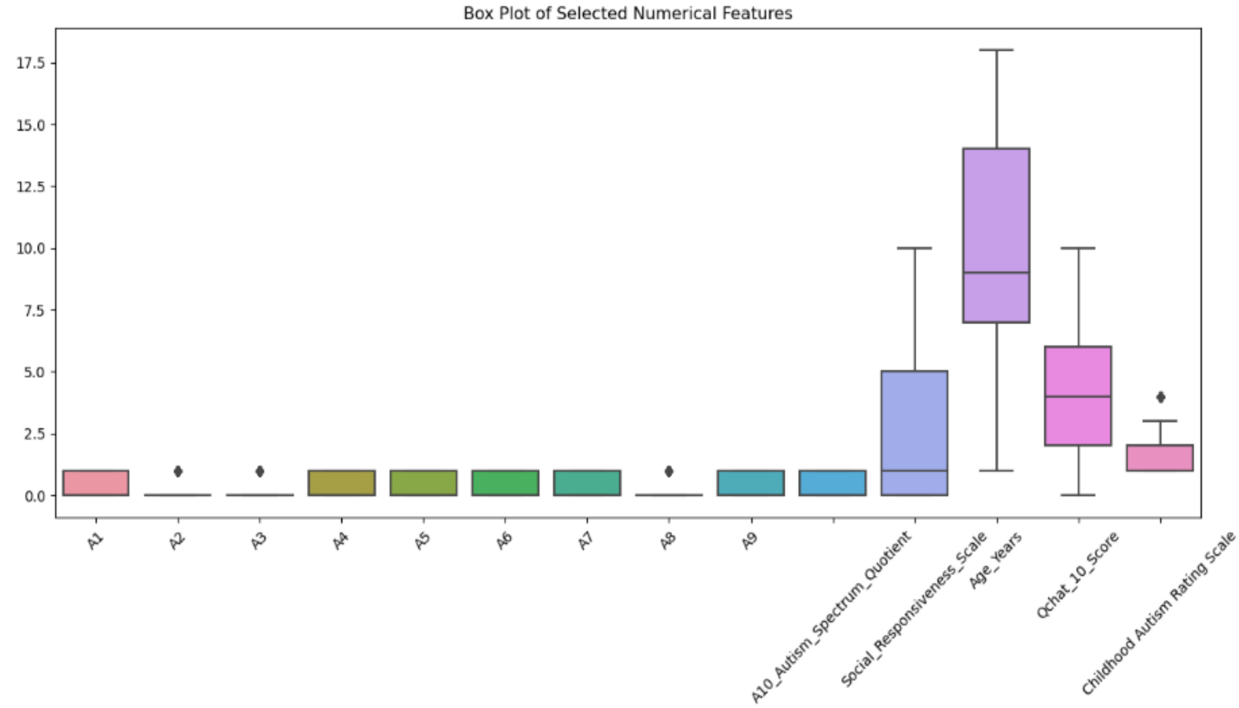

Fig 5: Data visualization of Numerical Dataset\

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 4, April 2024

DOI: 10.17148/IARJSET.2024.11451

In the dataset splitting phase, the ASD traits target variable, which indicates whether a person has ASD or not, is removed, and the dataset is divided into a 70-30 ratio for training and testing. Preprocessing techniques such as QuantilTransformer, Normalizer, and MaxAbsScaler enhance data quality and model performance, making the implementation more effective in the early diagnosis of ASD and potentially leading to better healthcare outcomes.

2. Model Development:

The image dataset uses deep learning models such as VGG16, EfficientNet-B0, EfficientNet-B7, and InceptionV3. VGG16 is a deep neural network for complex feature extraction. EfficientNet-B0 and B7 offer efficient scaling and a balance of model size and accuracy. InceptionV3 uses inception modules for high performance in image recognition tasks. For the numerical dataset, machine learning models like AdaBoost, RandomForest, Logistic Regression, Support Vector Machine (SVM), and GridSearch Decision Tree are implemented. AdaBoost combines weak classifiers to create a strong one, while RandomForest uses multiple decision trees for high accuracy and robustness. Logistic Regression is a simple model for binary classification, and SVM effectively separates classes. GridSearch Decision Tree optimizes decision trees using grid search for the best hyperparameters, leading to accurate data classification.

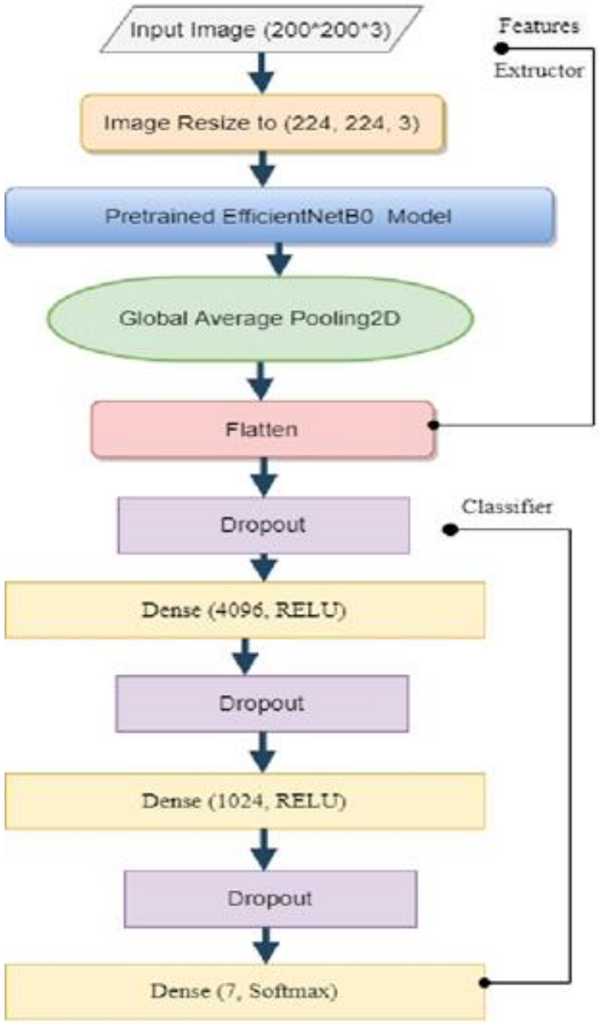

Fig 6: EfficientNet-B0 model Architecture

For the image dataset, the EfficientNet-B0 model provided the highest accuracy in ASD detection tasks. EfficientNet-B0 is a pre-trained deep learning model known for its efficient scaling and balance of model size and performance. The model incorporates several inverted residual blocks and uses a compound scaling method to adjust the network's depth, width, and resolution, resulting in an optimized architecture. The EfficientNet-B0 model was fine-tuned for ASD detection using the image dataset, leading to high accuracy in classifying images as either autistic or non-autistic. In the numerical dataset, GridSearch Decision Tree yielded the highest accuracy for ASD detection tasks.

DOI: 10.17148/IARJSET.2024.11451

This model employs decision trees, which make decisions based on various features, and optimizes them using a grid search technique to identify the best hyperparameters. By systematically exploring combinations of tree depth, splitting criteria, and other parameters, GridSearch Decision Tree finds the optimal configuration for the dataset. This approach ensures robust classification and prediction performance, effectively identifying patterns in the numerical data.

3. Model training and Evaluation

For the image dataset, various deep learning models were explored, including VGG16, EfficientNet-B0, EfficientNet-B7, and InceptionV3, each known for their architecture and performance in image classification tasks. VGG16 is a deep Convolutional Neural Network (CNN) architecture with a series of convolutional layers that extract complex features from images.

EfficientNet models, such as B0 and B7, are known for their scalability and efficiency, providing a balance between model size and accuracy. InceptionV3 is another CNN architecture that leverages inception modules for feature extraction and is known for its high performance in image recognition tasks. In this implementation, the final model chosen was EfficientNet-B0, which achieved an accuracy of 0.8942. This model was selected for its high accuracy and efficient resource usage.

Models	Accuracy
VGG16	0.8533
EfficientNet-B0	0.8942
EfficientNet-B7	0.8100
InceptionV3	0.8367

Table 1: Accuracy of the five DL models on the Image Dataset

For the numerical dataset, machine learning models such as AdaBoost, RandomForest, Logistic Regression, Support Vector Machine (SVM), and GridSearch Decision Tree were evaluated. AdaBoost is an ensemble learning method that combines multiple weak classifiers to form a strong classifier, known for its robustness against noise.

RandomForest is an ensemble learning method that uses multiple decision trees to create a strong model, providing high accuracy and resistance to overfitting. Logistic Regression is a statistical model for binary classification, useful for predicting the probability of an event based on input features. SVM is a supervised learning model that works well for classification tasks by finding the optimal hyperplane that separates different classes.

Models	Accuracy
AdaBoost	0.96
RandomForest	0.96
Logistic Regression	0.90
Support Vector Machine (SVM)	0.86
GridSearch Decision Tree	0.99

Table 2: Accuracy of the five ML models on Numerical Dataset

The final model chosen for the numerical dataset was the GridSearch Decision Tree, which achieved an accuracy of 0.99. This approach optimizes a decision tree by searching for the best hyperparameters using grid search, resulting in a model that effectively captures the underlying patterns in the data. This combination of deep learning and machine learning models allows for comprehensive analysis and improved accuracy in ASD detection and screening.

DOI: 10.17148/IARJSET.2024.11451

V. RESULTS

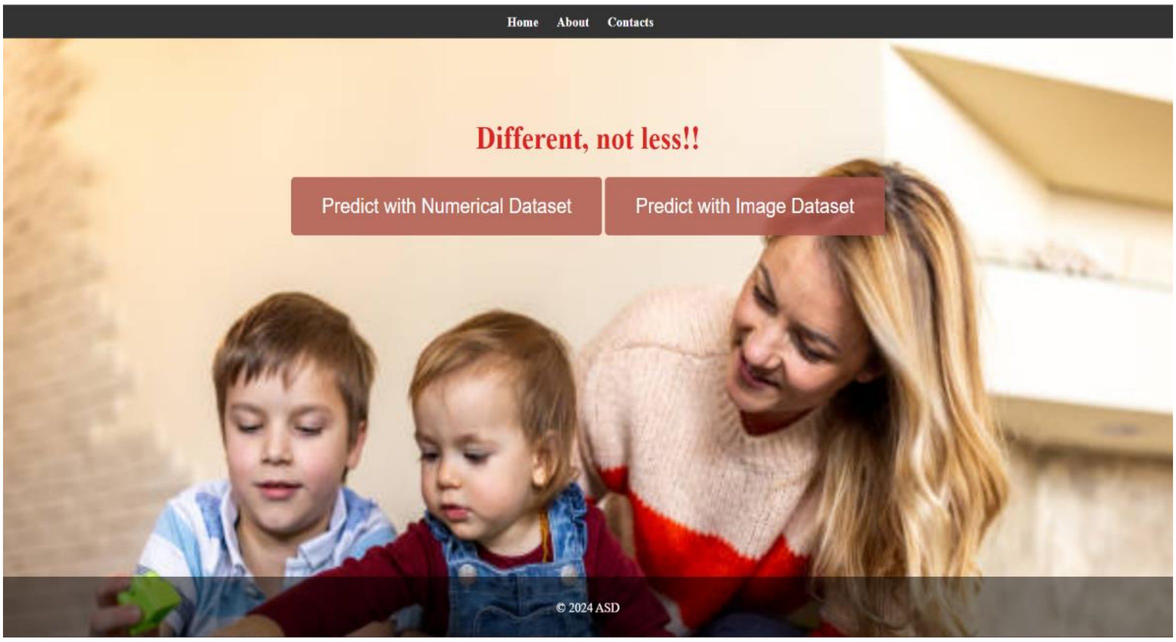

Fig 7: Home page

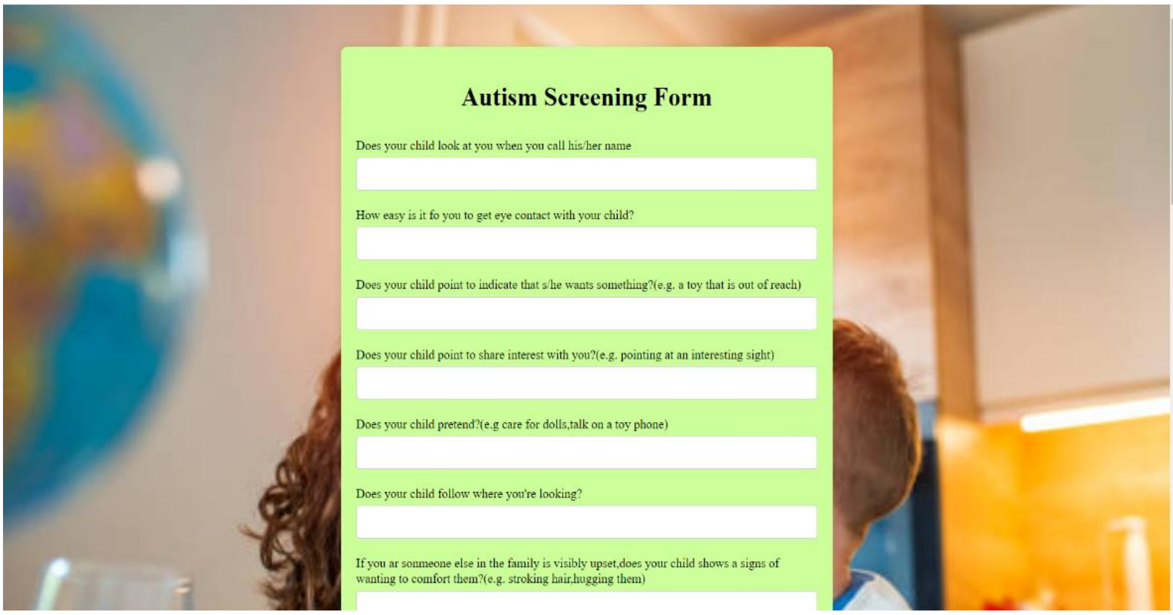

Fig 8: Autism Screening Test page

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁝ Peer-reviewed & Refereed journal ⁝ Vol. 11, Issue 4, April 2024
DOI: 10.17148/IARJSET.2024.11451

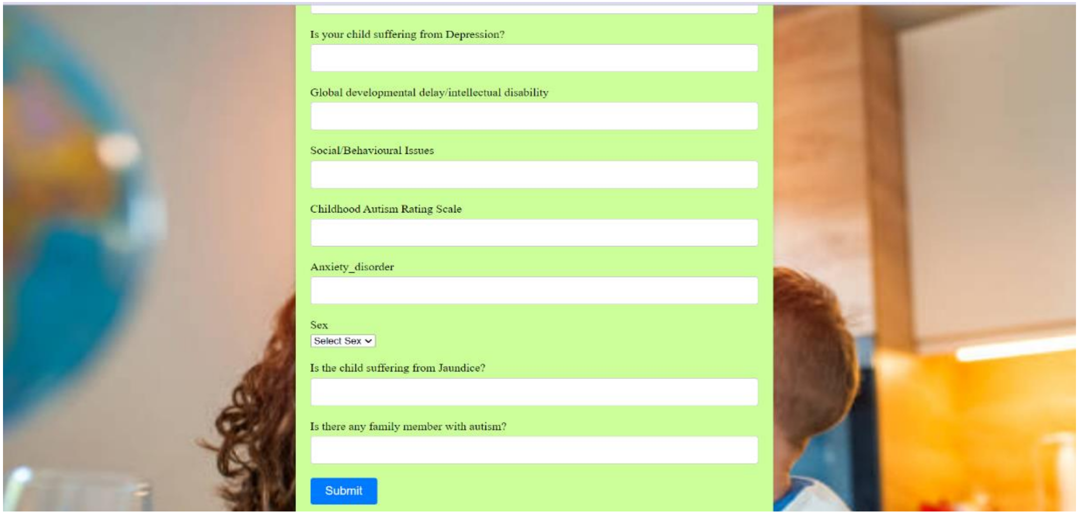

Fig 9: Autism Screening Test page

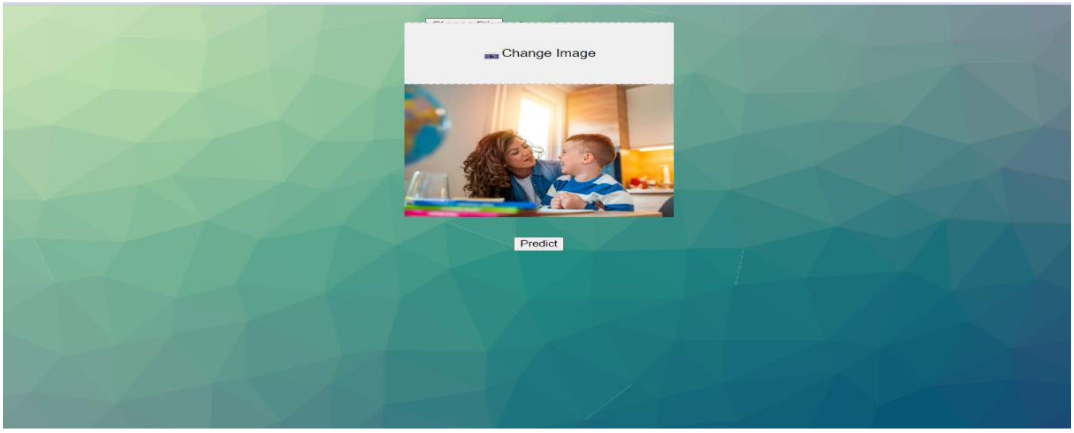

Fig 10: Image Test page

Fig 11: Report page

DOI: 10.17148/IARJSET.2024.11451

VI. CONCLUSION

Early diagnosis of Autism Spectrum Disorder (ASD) is crucial for ensuring individuals receive appropriate interventions and support to maximize their potential. While there's no single definitive test for ASD, advancements in Machine Learning (ML) show promise as supplementary tools for early detection in both children and adults. Studies indicate that ML models can achieve accuracy rates exceeding 80% in ASD detection using various data points. This offers exciting possibilities for improving screening and aiding professionals in the diagnosis process.

Research into ML-based ASD detection is ongoing, with efforts focused on:

- Refining algorithms and incorporating more comprehensive datasets to enhance the precision of ASD detection.
- Developing tools that can identify potential ASD cases at even younger ages, allowing earlier intervention.
- Exploring the potential of ML for assessing the severity of ASD symptoms

REFERENCES

[1]. S. M. Mahedy Hasan, M. P. Uddin, M. A. Mamun, M. I. Sharif, A. Ulhaq and G. Krishnamoorthy, "A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders," in IEEE Access, vol. 11, pp. 15038-15057, 2023, doi: 10.1109/ACCESS.2022.3232490.
[2]. M. Kohli, A. K. Kar and S. Sinha, "The Role of Intelligent Technologies in Early Detection of Autism Spectrum Disorder (ASD): A Scoping Review," in IEEE Access, vol. 10, pp. 104887-104913, 2022, doi: 10.1109/ACCESS.2022.3208587.
[3]. J. Han, G. Jiang, G. Ouyang, and X. Li, "A Multimodal Approach for Identifying Autism Spectrum Disorders in Children," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 30, pp. 2003-2011, 2022, doi: 10.1109/TNSRE.2022.3192431.
[4]. M. A. Rashidan et al., "Technology-Assisted Emotion Recognition for Autism Spectrum Disorder (ASD) Children: A Systematic Literature Review," in IEEE Access, vol. 9, pp. 33638-33653, 2021, doi: 10.1109/ACCESS.2021.3060753.
[5]. K. D. Bartl-Pokorny et al., "Robot-Based Intervention for Children with Autism Spectrum Disorder: A Systematic Literature Review," in IEEE Access, vol. 9, pp. 165433-165450, 2021, doi: 10.1109/ACCESS.2021.3132785.
[6]. F. Ke, S. Choi, Y. H. Kang, K. -A. Cheon and S. W. Lee, "Exploring the Structural and Strategic Bases of Autism Spectrum Disorders with Deep Learning," in IEEE Access, vol. 8, pp. 153341-153352, 2020, doi: 10.1109/ACCESS.2020.3016734.
[7]. N. Rusli, S. N. Sidek, H. M. Yusof, N. I. Ishak, M. Khalid, and A. A. A. Dzulkarnain, "Implementation of Wavelet Analysis on Thermal Images for Affective States Recognition of Children with Autism Spectrum Disorder," in IEEE Access, vol. 8, pp. 120818-120834, 2020, doi: 10.1109/ACCESS.2020.3006004.
[8]. K. Khowaja et al., "Augmented Reality for Learning of Children and Adolescents with Autism Spectrum Disorder (ASD): A Systematic Review," in IEEE Access, vol. 8, pp. 78779-78807, 2020, doi: 10.1109/ACCESS.2020.2986608.
[9]. R. Munoz, C. Morales, R. Villarroel, Á. Quezada and V. H. C. De Albuquerque, "Developing a Software That Supports the Improvement of the Theory of Mind in Children with Autism Spectrum Disorder," in IEEE Access, vol. 7, pp. 7948-7956, 2019, doi: 10.1109/ACCESS.2018.2890220.