Course Project Report

Detection of Autism using Machine Learning

Submitted By

Rajesh Kumar (2420682)
Prajwal Gabhane (2420685)
Abhishek Bhamare (2420693)
Pranav Pujara (2420060)

as part of the requirements of the course

IT for Healthcare [Jul-Nov 2024]

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

Master of Technology in Information Technology

under the guidance of

Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

undergone at

DEPARTMENT OF INFORMATION TECHNOLOGY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

AUG-DEC 2024

DEPARTMENT OF INFORMATION TECHNOLOGY

National Institute of Technology Karnataka, Surathkal

C E R T I F I C A T E

This is to certify that the Course project Work Report entitled **Detection of Autism using Machine Learning** is submitted by the group mentioned below -

Details of Project Group

Name of the Student	Register No.	Signature with Date
Rajesh Kumar	2420682	
Prajwal Gabhane	2420685	
Abhishek Bhamare	2420693	
Pranav Pujara	2420060	

this report is a record of the work carried out by them as part of the course **IT for Healthcare (IT820)** during the semester **Aug-Dec 2024**. It is accepted as the Course Project Report submission in the partial fulfillment of the requirements for the award of the degree of **Master of Technology in Information Technology.**

(Name and Signature of Course Instructor)
Dr. Sowmya Kamath S.
Associate Professor
Dept. of IT, NITK Surathkal

D E C L A R A T I O N

We hereby declare that the project report entitled **Detection of Autism using Machine Learning** submitted by us for the course **IT for Healthcare (IT820)** during the semester **Aug-Dec 2024**, as part of the partial course requirements for the award of the degree of Master of Technology in Information Technology at NITK Surathkal is our original work. We declare that the project has not formed the basis for the award of any degree, associateship, fellowship or any other similar titles elsewhere.

Details of Project Group

Name of the Student	Register No.	Signature with Date
1. Rajesh Kumar	2420682	
2. Prajwal Gabhane	2420685	
3. Abhishek Bhamare	2420693	
4. Pranav Pujara	2420060	

Place: NITK, Surathkal
Date: December 5, 2024

Detection of Autism using Machine Learning

Rajesh Kumar[1], Prajwal Gabhane[2], Abhishek Bhamare[3], Pranav Pujara[4]

***Abstract*— Autism Spectrum Disorder (ASD) is a neuro-developmental condition that affects communication, behavior, and social interactions. Early and accurate prognosis plays an important function in providing well timed intervention and support, significantly improving life of a person. This venture explores a data-driven method to detect autism, with tailored methodologies designed to address the diagnostic needs of two categories: adults and kids. For adults, the detection procedure includes studying responses to a structured set of behavioral questions, together with demographic info, family records, and health-associated records. Several machine learning algorithms, which include Random Forest, XGBoost, and k-Nearest neighbours (KNN) and AdaBoost Classifier are applied to classify the if the person has autism or not. Complete data preprocessing techniques, consisting of managing missing values, encoding specific data which are as strings, and feature engineering, are employed to enhance model accuracy and computational efficiency. For children, the analysis focuses on facial image analysis, wherein advanced deep learning techniques like MobileNet and InceptionV3 are used to detect autism-associated traits through facial features. Preprocessing steps including picture resizing, normalization, and data augmentation are incorporated to address variability inside the dataset and make certain robustness in the model's predictions. The findings underline the importance of leveraging era to help in the early diagnosis of neuro-developmental problems, paving the manner for broader applications in healthcare.**

Keywords: **Autism, Machine Learning, Facial Image Analysis, Feature Engineering, Classification, Neuro-developmental disorder Diagnosis**

***EDIT LINK** to your Overleaf report:* `https://www.overleaf.com/3258573467ytpcrptzbrsq#0bf355`

I. INTRODUCTION

Autism Spectrum disease (ASD) is a lifelong neuro-developmental disorder that affects people throughout numerous domains, which includes social interactions, communication, and behavior. The spectrum of autism is wide, ranging from moderate demanding situations that may not interfere appreciably with each day functioning to intense impairments that require huge assist. At the same time, as the exact causes of ASD stay uncertain, early detection is essential for improving outcomes through early intervention. Timely diagnosis permits individuals to acquire suitable healing procedures, that could drastically enhance their developmental trajectory, social abilties, and standard first-rate of lifestyles.

However, diagnosing autism remains a tough undertaking. The diagnostic procedure traditionally involves a comprehensive scientific assessment, which can also encompass behavioral observations, developmental history, and standardized diagnostic equipment. Whilst precious, this method may be time-consuming, subjective, and calls for get entry to the skilled professionals. As a result, there is a developing interest in exploring automatic and scalable answers that could assist healthcare companies in diagnosing ASD extra efficaciously and as it should be. Current advancements in Machine learning(ML) and deep learning(DL) strategies offer promising possibilities to develop computerized structures which can useful resource for the analysis of autism. These methods can include large volumes of information and extract complicated features that won't be much apparent to human evaluators. In particular, this project focuses on two different methods to autism detection, one for adults and one for children, leveraging the strength of machine learning and deep learning in methods to detect autism.

This project makes use of established facts derived from responses to behavioral questionnaires and demographic, family history, and health-related attributes. Machine learning techniques consisting of Random forest, XGBoost, k-Nearest neighbour(KNN) and AdaBoost Classifier are used to classify the probability of autism primarily based on these attributes. Feature engineering, along with techniques to handle missing statistics and normalization functions, are integrated to improve the performance and make correct predictions.

For kids, the detection process includes using their facial photograph. The facial pictures of children come across signs and symptoms of autism, the usage of superior deep learning models like MobileNet and InceptionV3. These architectures are well-desirable for learning image features, specially in the context of detecting subtle facial features associated with autism. Picture preprocessing steps consisting of resizing, normalization, and data augmentation are used to optimize the version's capability to generalize throughout distinctive pictures of children.

Both the methodologies approaches show the potential of machine learning and deep learning in presenting early and accurate autism detection. While the methodologies for adults and kids are distinct, both contribute to the general purpose of growing scalable, automated solutions which can aid with the early identification of ASD, improving accessibility to the people to the necessary sources and aid to detect this condition at an earlier stage of development.

This project explores the applications of machine learning and deep learning in the analysis of autism, contributing to the developing a path of research in healthcare era. With the aid of supplying separate methodologies tailor-made to unique age groups, the project aims to strengthen the abilities of automated structures in autism detection, in the end enhancing the accessibility and effectiveness of early diagnostic equipment.

II. Literature Review

The study by Chowdhury and Iraj (2020) was titled *"Predicting Autism Spectrum Disorder Using Machine Learning Classifiers"*; it was presented at the International Conference on Recent Trends on Electronics, Information, Communication & Technology. They applied machine learning classifiers to predict Autism Spectrum Disorder. They looked for patterns related to ASD using data-driven methods, exploiting the power of classifiers, such as Support Vector Machines, Decision Trees, and Random Forests for predictive analysis. From a dataset comprising behavioral and diagnostic features, they have shown that machine learning can attain rather significant accuracy in identifying cases of ASD. Such a study speaks software in the form of more advanced algorithms for the automation of ASD detection, and hence, its promise for facilitating near-patient early diagnosis and intervention. (Chowdhury and Iraj, 2020).

Europe PMC (2016) is the repository for all life science and biomedical literature, facilitating access to peer-reviewed research articles and preprints. The above study can be found in Europe PMC, focusing on improvements in perceiving and diagnosing ASD by using diverse approaches. It stresses the contribution that cutting-edge technologies such as machine learning and neuroimaging make to the discovery of crucial behavioral and biological markers for ASD. The repository's open-access framework enables researchers worldwide to access crucial data, fostering collaboration and innovation in ASD diagnosis and treatment. This is a prime example of how research and technology can work together towards a better understanding of complex medical conditions. (Europe PMC, 2016).

In the paper *"Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques"* by Vakadkar et al. (2021), published in SN Computer Science, explored machine learning methods like Support Vector Machines (SVM), Random Forest and K-Nearest Neighbors (KNN) for ASD detection in children. The study emphasized the significance of feature selection and model optimization towards diagnostic accuracy enhancement. Results obtained from this work serve as evidence of the possibility of machine learning to be used in assisting early diagnosis of ASD and intervention on time (Vakadkar et al., 2021).

The study by Farooq et al. (2023) was titled *"Detection of Autism Spectrum Disorder (ASD) in Children and Adults Using Machine Learning"*, published in Scientific Reports, investigated the use of machine learning algorithms for ASD detection across age groups. The research utilized diverse models to analyze behavioral and diagnostic data, demonstrating high accuracy in distinguishing autistic individuals from non-autistic ones. This study highlights the effectiveness of machine learning in enhancing ASD diagnosis, offering a scalable solution for early and accurate identification in both children and adults (Farooq et al., 2023).

The review by Hyde et al. (2019) was titled *"Applications of Supervised Machine Learning in Autism Spectrum Disorder Research"*, published in the Review Journal of Autism and Developmental Disorders, examined the application of supervised machine learning techniques in ASD research. The study highlighted various algorithms, including Support Vector Machines and Random Forest, used for analyzing behavioral, genetic, and neuroimaging data. It emphasized the potential of machine learning to improve ASD diagnosis, understand its underlying mechanisms, and support personalized interventions. This review underscores the growing role of artificial intelligence in advancing ASD research and clinical practices (Hyde et al., 2019).

Reddy and Andrew (2024), in their study *"Diagnosis of Autism in Children Using Deep Learning Techniques by Analyzing Facial Features"* published in Engineering Proceedings, explored the use of deep learning for ASD diagnosis through facial analysis. By employing advanced neural networks, the study focused on identifying distinct facial characteristics associated with autism. The results demonstrated the effectiveness of deep learning in achieving high accuracy, highlighting its potential as a non-invasive diagnostic tool for early detection of ASD in children (Reddy and J., 2024).

III. METHODOLOGY

This project is using two separate methodologies for detecting autism. The first methodology uses machine learning techniques, with focus on feature engineering and ML based classification models. The second methodology is involved in using transfer learning approach to detect autism from the facial images of children. Shown below are both methodologies, elaborated in detail.

A. Autism Prediction based on behavioural and health related questions

1) Dataset Overview: The dataset used in this approach includes questions and responses around:

- **Behavioral Questions:** A set of 10 questions used to evaluate behavioral traits typically associated with autism.
- **Demographic Details:** Data such as age, ethnicity, and country of residence, of the person.
- **Health and Family Background:** Data like a if person has history of jaundice or someone in family has autism.

The target column includes a binary value whether an individual is having autism or not.

2) Data Preprocessing: Preprocessing steps are applied to the dataset to make it ready for machine learning algorithms to work on. These steps include:

1) **Handling Missing values:** Missing or incomplete entries were handled through either removal or averaging (if there is scope to do so).
2) **Feature encoding:**
   - Categorical variables (e.g., ethnicity, country) were converted into numerical formats using methods like one-hot encoding or label encoding.
   - Binary answers (like yes/no questions) were transformed into numerical values of 0 and 1.

3) **Scaling and Normalization:** Continuous attributes (e.g., age) are normalized to ensure stability and efficient training of the models.
4) **Feature Selection:** Unnecessary or redundant features, or the features that are derived are eliminated, resulting in a feature matrix comprising 37 attributes (including the one hot encoded features).

3) Model Training: Several machine learning algorithms are used to build predictive ML models, including:

- **Random Forest Classifier:** A method utilizing decision trees to deliver reliable predictions and also some insights into feature importance.
- **K-Neighbors:** An distance based algorithm that classifies data points based on proximity measures.
- **XgBoost Classifier:** A gradient boosting algorithm that models complex relationships and handles class imbalances effectively.
- **AdaBoost Classifier:** An ensemble learning algorithm that combines multiple weak learning algorithms to create a strong classifier, improving accuracy while minimizing overfitting.

4) Training Process: The training process have the following steps:

1) **Splitting the Dataset:** The data is divided into training and testing subsets (in this case 80%-20% split) for performance evaluation.
2) **Hyperparameter Optimization:** Grid search and random search were used to identify optimal parameters for each model starting for some standard values.
3) **Cross-Validation:** K-fold cross-validation ensured the models' generalizability and reduced the overfitting risks.

5) Evaluation Metrics: The performance of the models is assessed using the following metrics:

- **Accuracy:** The proportion of correctly predicted outputs.
- **Precision:** The ratio of true positives to all predicted positives, highlighting the model's ability to reduce false positives.
- **Recall (Sensitivity):** The ratio of true positives to actual positives, showing the model's effectiveness in identifying autism correctly.
- **F1 Score:** The harmonic mean of precision and recall, providing a balanced evaluation of false positives and negatives.

6) Model Deployment: The selected model is saved using serialization methods like `joblib` or `pickle`, for use in our application that we designed for demonstrating this work.

7) Application Development: A user-friendly application is created to facilitate autism prediction based on user input:

1) **Input Page:** The application displays 10 behavioral questions and some relevant demographic, health-related and family history questions.
2) **Prediction Workflow:**
   - User responses are preprocessed using the same pre processing steps as applied on the training dataset.
   - The preprocessed data is fed into the saved model to generate predictions.
3) **Output Presentation:** The application provides the classification result (autism or not) based on the result given by the trained model (used the best performing model).

B. Transfer Learning for Autism Detection Using Facial pictures of children

1) Dataset Overview: The dataset used in this methodology comprises 3014 facial pictures of children, having two classes: children with autism and children without autism. The dataset is preprocessed so as to input into the deep learning models and is split into training, validation, and test sets.

2) Data Pre-processing and data Augmentation: To increase the quality and diversity of the training data, the following pre-processing steps are applied:

1) **Image Resizing:** All images are resized to match the input dimensions as accepted by the MobileNet and InceptionV3 models.
2) **Normalization:** Pixel values were normalized to the range [0, 1] to have faster training convergence.
3) **Data Augmentation:** The training set is augmented with additional pictures to improve model generalization and reduce overfitting. Approaches used for augmentation includes:
   - Random rotations
   - Horizontal flips
   - Cropping
   - Adjustments to brightness and contrast
   - Adding noise

3) Transfer Learning Approach: Transfer learning is employed to utilise the pretrained MobileNet and InceptionV3 models capabilities, these models are fine-tuned for our use case of autism detection by feeding in our dataset.

1) **Pre-trained Model Selection:**
   - MobileNet and InceptionV3 are selected as base models since they are efficient and provide robust performance in image classification tasks.
   - Pretrained weights from the ImageNet dataset are used to initialize these models, allowing them to extract meaningful features from the input images.
2) **Model Modification:**
   - The last few layers of each base model are removed.
   - Custom classification layers are added, including:
     - A global average pooling layer.
     - Dense layers with ReLU activation function.
     - Dropout layers to reduce the overfitting risks.
     - A final dense layer with sigmoid activation(for binary classification).
3) **Fine-Tuning:**

- The last 30 layers of the base models are unfreezed to enable fine-tuning.
- This allows the models to adapt pretrained features to the features of the autism detection task while preserving the general features learned during pre-training.

4) Training Process:

1) **Optimization:**
   - The models are trained using the Adam optimizer with a experimentally chosen learning rate.
   - Early stopping is used to halt training when validation performance plateaus.
2) **Batch Processing:** Mini-batch gradient descent is used, with a batch size optimized for performance and computational efficieny.
3) **Loss Function:** Binary cross-entropy is used as the loss function to measure classification error.

5) Model Evaluation: The performance of both MobileNet and InceptionV3 models is evaluated on the test set using the following metrics:

- **Accuracy:** Overall proportion of correctly classified images.
- **Precision:** Ratio of true positive to all positive predictions.
- **Recall (Sensitivity):** Ratio of true positive predictions to all actual positive cases.
- **F1 Score:** Harmonic mean of precision and recall, providing a balance between the two(Precision and recall).
- **Training Time:** The total time taken to train the model on the training dataset.
- **Evaluation Time:** The time taken to evaluate the model on the test dataset.

6) Model Selection: After evaluation, the model with the best performance on the test set was selected. The criteria for selection were based on the F1 score, overall accuracy and computational time. The chosen model demonstrated superior capability in detecting autism from facial images with less computational time.

7) Application Development: A user-friendly application is developed to deploy the best-performing model. The application workflow includes:

1) **Input:** Uploading a facial picture of a child.
2) **Pre-processing:** The image is resized, normalized, and transformed same as done with the training dataset.
3) **Prediction Pipeline:**
   - The preprocessed image is passed to the trained model.
   - The model outputs a prediction telling whether the child is classified as having autism or not.
4) **Output:** The application provides the prediction result (i.e. if the child is having autism or not).

IV. Results and Analysis

A. Behavioral and Demographic Data

In this aspect of our project, we analyzed behavioral questionnaire responses and demographic details using multiple machine learning models.The Confusion matrix of all models that is Random Forest, K-neighbors, XgBoost, AdaBoost is shown in Fig. 1, Fig. 2, Fig. 3, Fig. 4 respectively. The Receivr Operating Characteristic(ROC) Curve of all models that is Random Forest, K-neighbors, XgBoost, AdaBoost is shown in Fig. 5, Fig. 6, Fig. 7, Fig. 8respectively.

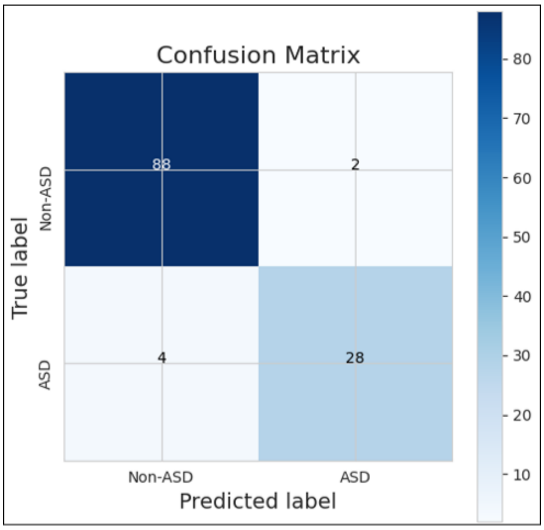


Fig. 1: Confusion matrix for Random Forest Classifier

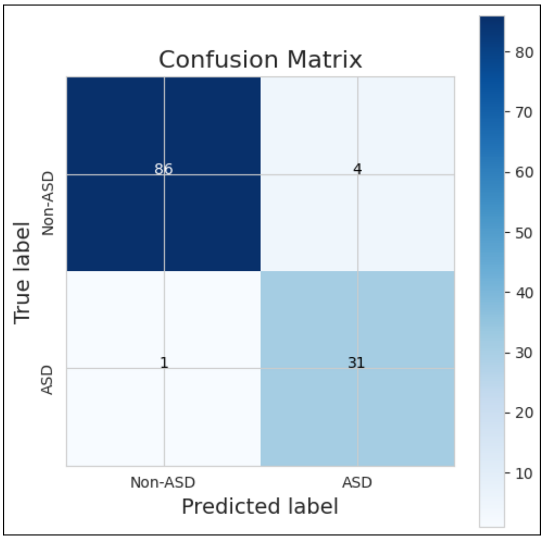


Fig. 2: Confusion matrix for KNN Classifier

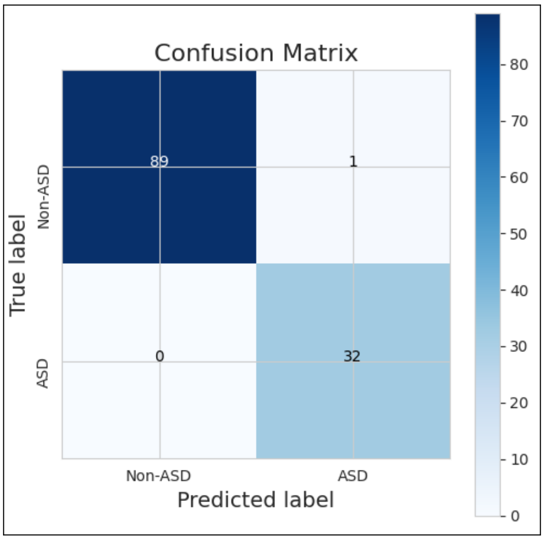


Fig. 3: Confusion matrix for XGBoost Classifier

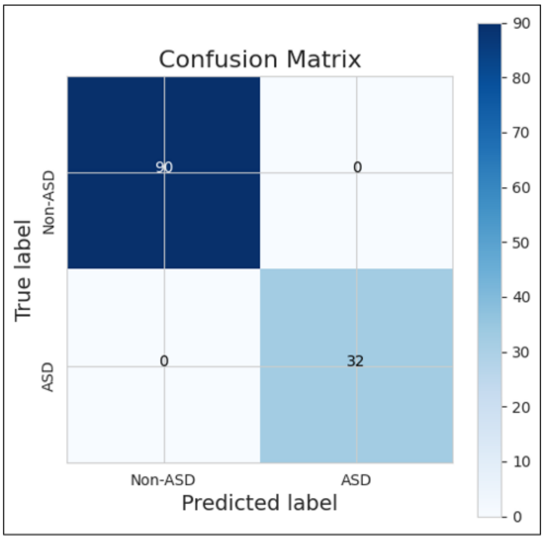


Fig. 4: Confusion Matrix for the AdaBoost Classifier

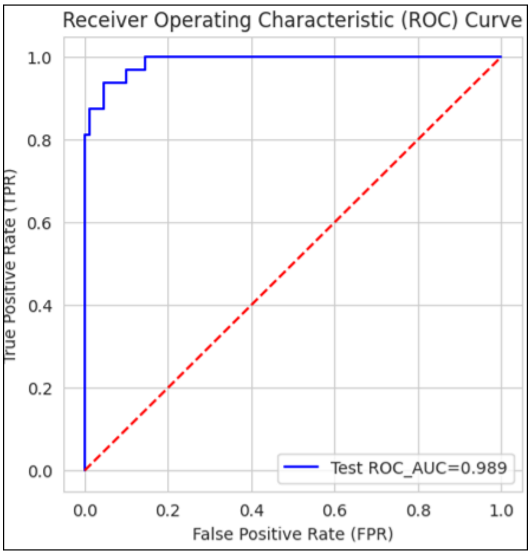


Fig. 5: ROC Curve for Random Forest Classifier

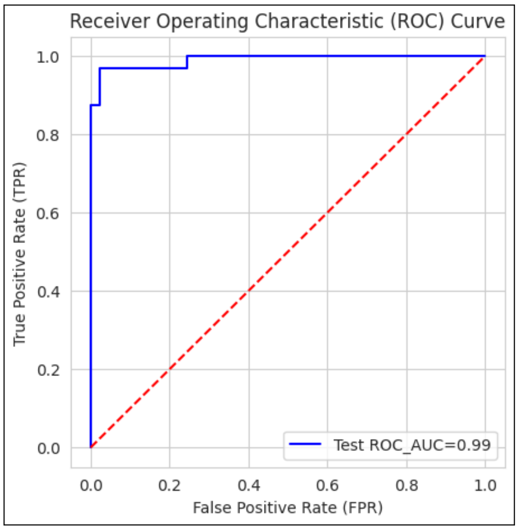


Fig. 6: ROC Curve for KNN Classifier

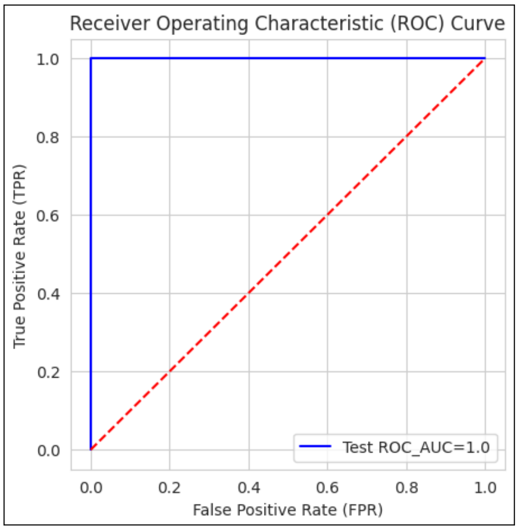


Fig. 7: ROC Curve for XGBoost Classifier

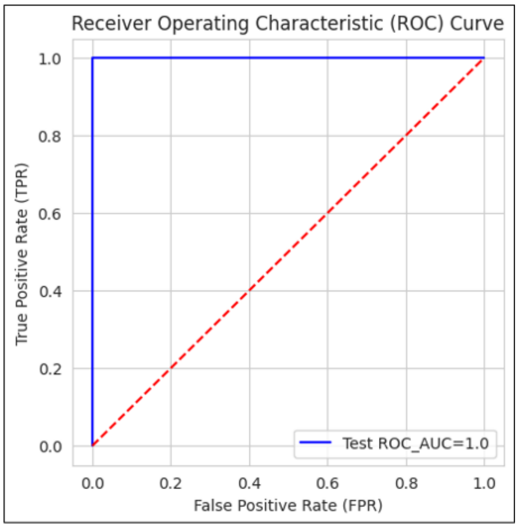


Fig. 8: ROC Curve for AdaBoost Classifier

In Table I, the performance metrics for different classifiers are shown. Among four classifier, The AdaBoost Classifier is best performing model with accuracy of 100%.

TABLE I: Performance metrics for questionnaire dataset

Model	Precision	Recall	F1 Score	Accuracy	AUC
Random Forest	0.952	0.954	0.952	0.95	0.98
K Neighbors	0.964	0.963	0.959	0.95	0.99
XgBoost	0.992	0.993	0.987	0.99	1
AdaBoost	1	1	1	1	1

B. Image-Based Detection

In this aspect, we used Transfer Learning approach with InceptionV3 and MobileNet as pre-trained model, to analyze children's images for autism detection. Confusion matrix for Inceptionv3 based model and for MobileNet based model is shown in Fig.9 and Fig.10 respectively. The performance metrics are shown in Table II.The model build using MobileNet is found to be best performing with accuracy of 92% .

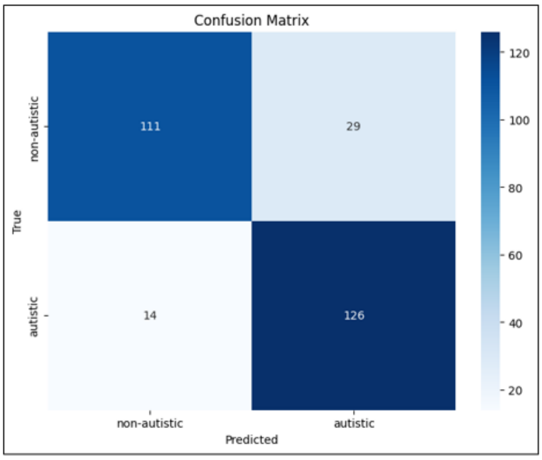


Fig. 9: Confusion Matrix for the Inceptionv3 based model

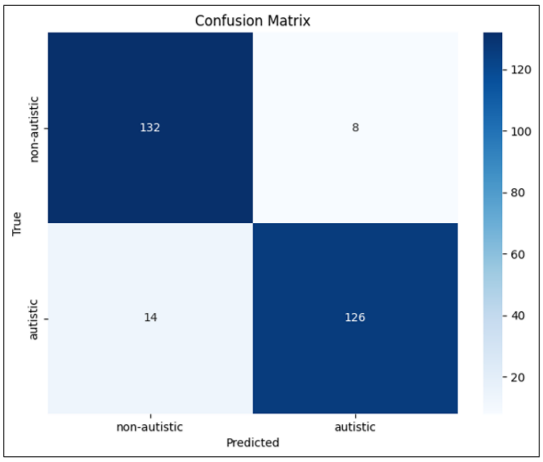


Fig. 10: Confusion Matrix for the MobileNet based model

TABLE II: Performance metrics for image dataset

	Precision	Recall	F1-score	Accuracy	AUC	Training time (seconds)	Evaluation time (seconds)
InceptionV3	0.85	0.85	0.85	0.85	0.95	3089.17	251.35
MobileNet	0.92	0.92	0.92	0.92	0.97	1559.50	102.12

V. CONCLUSIONS

In this project, we have efficiently developed a machine learning based system to detect autism leveraging two complementary processes: behavioral analysis and image-based analysis.

The primary technique applied a questionnaire together with demographic details to predict the presence of autism condition in a person. With the aid of using advanced data preprocessing techniques and feature engineering, we decreased dataset complexity and increased the model performance. Four machine learning classifier, which includes Random Forest,K-neighbors,XgBoost,AdaBoost were used to analyzed the predictions.The AdaBoost classifier is found to be best performing model among other four.

The second method centered on children, wherein image-based analysis was conducted by using Transfer Learning approach with InceptionV3 and MobileNet as pretrained model. The model build using MobileNet is found to be best performing model. These models validated their functionality to pick out autism-associated attributes in kid's facial features, offering an additional diagnostic angle.

Eventually, we incorporated each techniques into a user-friendly application, showcasing the real-world applicability of our models. This comprehensive solution not only increases the accuracy and reliability of autism detection, but also emphasizes the capability of AI in helping early diagnosis and intervention.

This project highlights the significance of interdisciplinary techniques in addressing medical situations and creates a basis for similar research and development in automatic autism detection systems.

REFERENCES

Chowdhury, K. and Iraj, M. A. (2020). Predicting autism spectrum disorder using machine learning classifiers. In *2020 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTEICT)*.

Europe PMC (2016). Europe pmc.

Farooq, M. S. et al. (2023). Detection of autism spectrum disorder (asd) in children and adults using machine learning. *Scientific Reports*, 13(1):9605.

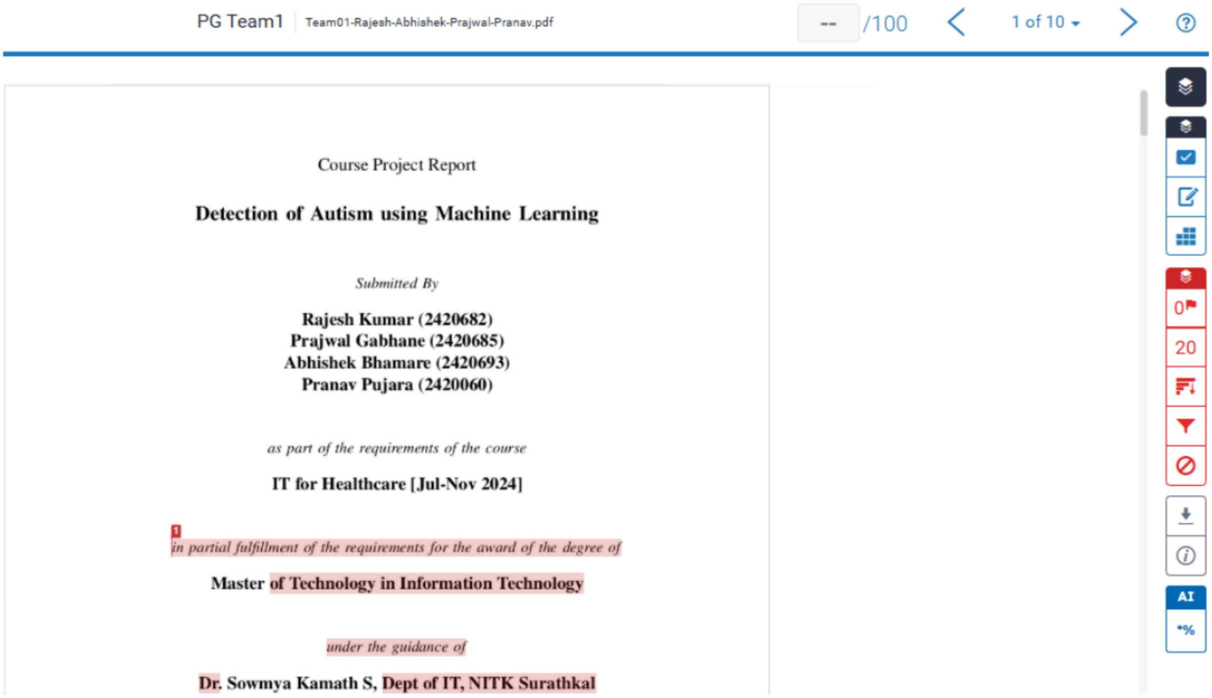
Hyde, K. K. et al. (2019). Applications of supervised machine learning in autism spectrum disorder research: A review. *Review Journal of Autism and Developmental Disorders*, 6(2):128–146.

Reddy, P. and J., A. (2024). Diagnosis of autism in children using deep learning techniques by analyzing facial features. *Engineering Proceedings*, 59(1):198.

Vakadkar, K. et al. (2021). Detection of autism spectrum disorder in children using machine learning techniques. *SN Computer Science*, 2(5).

APPENDIX

ACCEPTED FOR EVALUATION


PG Team1 Team01-Rajesh-Abhishek-Prajwal-Pranav.pdf -- /100 1 of 10

Course Project Report

Detection of Autism using Machine Learning

Submitted By

Rajesh Kumar (2420682)
Prajwal Gabhane (2420685)
Abhishek Bhamare (2420693)
Pranav Pujara (2420060)

as part of the requirements of the course

IT for Healthcare [Jul-Nov 2024]

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

Master of Technology in Information Technology

under the guidance of

Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal