Performance Evaluation of ML and DL Approaches for Early Diagnosis of Autism Spectrum Disorder and Development of Autism Care Hub

KARISHMA KAINE T
Department of Electronics and Communication Engineering
PSG Institute of Technology and Applied Research Neelambur, Coimbatore
t.karishmakaine12m2@gmail.com

PRIYANKA PANDA
Department of Electronics and Communication Engineering
PSG Institute of Technology and Applied Research Neelambur, Coimbatore
priyankapanda127@gmail.com

DHANYA SHREE M
Department of Electronics and Communication Engineering
PSG Institute of Technology and Applied Research Neelambur, Coimbatore
dhanyashreemanikam@gmail.com

JAYANTHI SREE S
Department of Electronics and Communication Engineering
PSG Institute of Technology and Applied Research Neelambur, Coimbatore
jayanthisree@psgitech.ac.in

PRITHISH GOUTAM S
Department of Electronics and Communication Engineering
PSG Institute of Technology and Applied Research Neelambur, Coimbatore
prithishg27@gmail.com

Abstract—**Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with social interaction, communication, and repetitive behavior problems. Early diagnosis is the key to facilitating on-time intervention and enhancing patient outcomes. This paper is an evaluation of comparative performance of different machine learning (ML) and deep learning (DL) algorithms for early ASD diagnosis based on behavioural and demographic information from AQ-10 screening tests. Supervised machine learning algorithms like K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Random Forest, XGBoost, and Logistic Regression are compared with unsupervised cluster algorithms like K-Means, Agglomerative, Spectral, Gaussian Mixture Models (GMM), and BIRCH. Deep learning models consisting of Dense layers, Dropout, Batch Normalization, Long Short-Term Memory (LSTM), Residual connections, and Sigmoid activation functions are investigated. Random Forest Classifier is used for feature selection, which asserts that behavioural markers are the most powerful predictors for ASD, and demographic indicators are not very influential. The comparative analysis indicates the strengths and weaknesses of ML and DL algorithms, and the best performing model is chosen, and that model is used to develop an Autism Care Hub platform for early screening, diagnosis, and continuous monitoring of ASD. Experimental results show the efficacy of ensemble ML techniques and DL networks and uncover their potential in real-world clinical deployment.**

Keywords—*Autism Spectrum Disorder (ASD), Machine Learning (ML), Deep Learning (DL), Early Detection, Predictive Modeling, Feature Selection, Random Forest Classifier, Behavioral Analytics, Autism Care Hub, Classification Algorithms, Unsupervised Clustering*.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with intricate difficulties in social interaction, communication, and repetitive behavior. Early diagnosis of ASD is critical as it allows timely intervention, which improves developmental outcomes and quality of life for patients with the disorder considerably. However, traditional diagnosis is based on subjective clinical evaluation, which is time-consuming and variable. In recent years, advances in Machine Learning (ML) and Deep Learning (DL) have demonstrated great potential to address these challenges by automating ASD diagnosis from behavioral and demographic data. Studies such as [1], [2] have proven that ML classifiers like Support Vector Machines, Random Forests, and XGBoost are extremely accurate in identifying ASD features from screening tests. At the same time, DL models like Long Short-Term Memory (LSTM) networks and Dense Neural Networks have managed to extract intricate patterns in behavioral data [3].While these advances have been reported, not many studies have been performed that comprehensively compare the performance of ML and DL methods on identical datasets to ascertain the best methodologies to implement for ASD screening. Moreover, little practical use has been made that brings these models to the fingertips of accessible platforms for clinicians and caregivers. This paper aims to bridge this knowledge gap by evaluating different supervised and unsupervised ML algorithms alongside DL architectures on open-access ASD screening datasets. Based on the highest-performing algorithm, we recommend the development of an Autism Care Hub—a platform that attempts to facilitate early screening and ongoing monitoring of ASD symptoms. The research places significant focus on behavioral feature selection, confirming clinical intuition that behavioral features are more predictive than demographic features in ASD diagnosis.

II. RELATED WORKS

Rasul et al. [1] mention the extensive application of supervised machine learning (ML) methods like Support Vector Machines (SVM), Decision Trees, and Random Forests (RF) in autism spectrum disorder (ASD) research. These methods have been successful in diagnosis, genetic research, and intervention planning by analyzing behavioral assessments, neuroimages, and genetics. RF classifiers, in fact, have made early diagnosis possible automatically from parental questionnaires and behavioral videos. All this aside, data imbalance, non-standardization, and poor model generalizability are issues. The authors identify model interpretability, multi-modal data fusion, and responsible AI design as the most critical areas to make clinical relevance more effective.

Hyde et al. [2] introduce supervised ML models such as SVM, RF, and Alternative Decision Trees for enhancing diagnostic tools such as ADOS, ADI-R, and AQ-10 with feature selection methods such as Recursive Feature Selection and Minimum Redundancy Maximum Relevance (mRMR). They identify limitations such as unbalanced datasets (e.g., AGRE, NDAR), ethics, and the difficulty of combining ML models with clinical diagnosis. They introduce models that combine clinical knowledge with ML, making them interpretable, consistent, and ethically valid for ASD screening.

Additionally, Cavus et al. [3] have talked about the use of machine learning models such as SVM, RF, and AD Trees in improving ASD screening and diagnosis by automating diagnostic tools with feature selection techniques such as Recursive Feature Elimination. They have also talked about the issues currently present in data standardization, clinical utility, and ethics. The study has proposed the combination of clinical knowledge with machine learning for transparent and interpretable ASD detection systems.

Zhang et al. [4] introduce BIRCH, a hierarchical clustering algorithm with near-linear scalability for large datasets with low memory and computing resources. BIRCH is capable of handling densely spherical region data efficiently and performing better than other algorithms such as CLARANS, though it is not strong in dealing with irregularly shaped clusters. The approach offers an efficient scalable solution for managing large ASD-related data in unsupervised learning applications.

Shahamiri and Thabtah [5] developed Autism AI, an ASD identification screening tool utilizing Convolutional Neural Networks (CNNs) on an online platform. It is more sensitive and accurate than the standard tools ADOS and ADI-R by taking user responses into account using a CNN. Though it has an advantage in early detection and simplicity, Autism AI currently only employs questionnaire data with the potential to employ multi-modal inputs such as facial expressions and vocal analysis to screen better.

Gulati et al. [6] created and validated the AIIMS-Modified INDT-ASD screening tool using DSM-5 criteria and achieved high diagnostic accuracy (98.4%) and specificity (91.7%) in Indian children aged 1-14 years. The instrument used here is a refinement of the earlier ones in the sense that it combines social communication domains and assesses sensory symptoms and provides a less complex alternative to tools like ADOS, particularly in low-resource settings.

Mazumdar et al. [7] enhanced autism screening in India by incorporating new psychological characteristics into the Indian Scale for Assessment of Autism (ISAA) and employing ML models such as SVM, Random Forest, K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP). The approach has high classification accuracy (up to 99%) and prioritizes children with intensive clinical management, without the experience of biases and inefficiencies in conventional diagnostic systems.

III METHEDOLOGY

A. Data set

The data set comprised of children and adults. The children's dataset has 292 instances and 21 attributes, while the adult dataset contains 704 instances and 21 attributes. Each dataset includes responses to a 10-question AQ-10 screening test, where answers are coded as 0 (low) or 1 (high), and the total score helps assess the likelihood of autism. The AQ-10 is a widely used screening tool that evaluates behavioural traits like communication, attention shifting, and social interaction. A higher score suggests a greater likelihood of autism, requiring further diagnosis. The dataset also includes demographic attributes such as age, gender, country, and previous screening history to provide more insights into ASD tendencies.

A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score	age	gender	ethnicity	jundice	austim	contry_of_	used_app_	result	age_desc	relation	Class/ASD
1	1	1	1	0	0	1	1	0	0	26	f	White-Eur	no	no	'United St	no	6	'18 and mc	Self	NO
1	1	0	1	0	0	0	1	0	1	24	m	Latino	no	yes	Brazil	no	5	'18 and mc	Self	NO
1	1	0	1	1	0	1	1	1	1	27	m	Latino	yes	yes	Spain	no	8	'18 and mc	Parent	YES
1	1	0	1	0	0	1	1	0	1	35	f	White-Eur	no	yes	'United St	no	6	'18 and mc	Self	NO
1	0	0	0	0	0	0	1	0	0	40	f	?	no	no	Egypt	no	2	'18 and mc	?	NO

Fig 1. Sample data – Adult

A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score	age	gender	ethnicity	jundice	austim	contry_of_	used_app_	result	age_desc	relation	Class/ASD
1	1	0	0	1	1	0	1	0	0	6	m	Others	no	no	Jordan	no	5	'4-11 year:	Parent	NO
1	1	0	0	1	1	0	1	0	0	6	m	'Middle Ea	no	no	Jordan	no	5	'4-11 year:	Parent	NO
1	1	0	0	0	1	1	1	0	0	6	m	?	no	no	Jordan	yes	5	'4-11 year:	?	NO
0	1	0	0	1	1	0	0	0	1	5	f	?	yes	no	Jordan	no	4	'4-11 year:	?	NO
1	1	1	1	1	1	1	1	1	1	5	m	Others	yes	no	'United St	no	10	'4-11 year:	Parent	YES

Fig 2. Sample data - Child

B. Data Preprocessing

Data preprocessing is important to make sure the dataset is clean, consistent, and ready to use for model training. The following methods were employed:
Missing Value Handling: Imputed missing values with mode imputation, substituting with most frequent value.
Normalization: Used min-max scaling to normalize numeric values between range [0, 1].
Encoding: One-hot encoded the categorical variables into binary format.
Data Augmentation: For DL models, used rotation, flipping, and addition of noise to enhance data diversity

C. Feature Selection

Feature selection was conducted employing a Random Forest (RF) classifier to determine the most critical features in predicting the likelihood of autism. The RF algorithm ranks features by their importance scores, allowing for the selection of a subset of features that Optimizes model performance while minimizing dimensionality and computational expense.

Random Forest

The Random Forest Classifier is a powerful ensemble learning algorithm used for classification tasks. It operates by constructing multiple decision trees during training and outputs the class that is the mode of the predictions made by individual trees. Each tree is trained on a random subset of the data, both in terms of rows and features, which helps to reduce overfitting and improve generalization. This randomness ensures that the trees are diverse, making the overall model more robust and accurate than single decision tree.

Machine Learning(ML)

Scores like A4_Score, A9_Score, A5_Score, A10_Score, and A1_Score came out as always significant across models, which reflects their diagnostic importance. Demographic characteristics such as ethnicity and nation were also among the highest-ranked characteristics, indicating potential cultural or region al correlations with ASD markers. The difference in feature importance between models indicates the varying weighting methods employed and how differences in analytical metho dology impact feature ordering. Standardization of behavior al indicators is further emphasized as a key feature in early autism identification

	Feature	Importance Score
8	A9_Score	0.115546
5	A6_Score	0.108041
3	A4_Score	0.100210
4	A5_Score	0.099139
2	A3_Score	0.064374
9	A10_Score	0.056879
10	Age_Mons	0.053602
6	A7_Score	0.050308
7	A8_Score	0.042794
0	A1_Score	0.041582
1	A2_Score	0.035352

	Feature	Importance Score
8	A9_Score	0.124430
5	A6_Score	0.115022
3	A4_Score	0.097741
4	A5_Score	0.094326
9	A10_Score	0.062320
2	A3_Score	0.058659
0	A1_Score	0.049670
6	A7_Score	0.049338
10	Age_Mons	0.046580
1	A2_Score	0.039320
7	A8_Score	0.037122

	Feature	Importance Score
3	A4_Score	0.137900
8	A9_Score	0.096010
7	A8_Score	0.078809
0	A1_Score	0.070973
9	A10_Score	0.070767
2	A3_Score	0.064245
4	A5_Score	0.062807
5	A6_Score	0.053869
10	Age_Mons	0.044654
6	A7_Score	0.040172
1	A2_Score	0.031451

	Feature	Importance Score
3	A4_Score	0.116737
8	A9_Score	0.091167
7	A8_Score	0.091151
0	A1_Score	0.076118
9	A10_Score	0.072645
4	A5_Score	0.059806
2	A3_Score	0.058306
5	A6_Score	0.055125
10	Age_Mons	0.047771
6	A7_Score	0.035884
1	A2_Score	0.033002

Fig 3 Feature Selection of Random Forest in ML for child and adult

Deep Learning(DL)

The analysis showed that the 'result' attribute total AQ-10 score—was the most important (~0.55) across both the children and adult datasets, indicating its key role in predicting Autism Spectrum Disorder (ASD). Critical behavioral indicators like A9_Score, A5_Score, A6_Score, A4_Score, A3_Score, and A10_Score ranked very high, consistent with well-established clinical characteristics of ASD. Demographic features like ethnicity, age and country_of_residence had negligible impacts, with the latest relation attribute.

```
Feature Importance Ranking:
1. result: 0.5551
2. A4_Score: 0.0884
3. A9_Score: 0.0608
4. A10_Score: 0.0497
5. A8_Score: 0.0424
6. A5_Score: 0.0290
7. A6_Score: 0.0289
8. A1_Score: 0.0283
9. A3_Score: 0.0270
10. country_of_res: 0.0189
11. A7_Score: 0.0170
12. ethnicity: 0.0114
13. A2_Score: 0.0113
14. age: 0.0106
15. relation: 0.0079
```

```
Feature Importance Ranking:
1. result: 0.5553
2. A9_Score: 0.1029
3. A5_Score: 0.0773
4. A6_Score: 0.0613
5. A3_Score: 0.0379
6. A4_Score: 0.0334
7. A7_Score: 0.0228
8. A10_Score: 0.0206
9. A2_Score: 0.0147
10. ethnicity: 0.0138
11. A1_Score: 0.0132
12. age: 0.0121
13. country_of_res: 0.0119
14. A8_Score: 0.0115
15. relation: 0.0047
```

Fig 4 Feature Selection of Random Forest in DL for child and adult

These findings highlight the fact that behavioral characteristics are far more predictive than demographic variables. The RF model was successful in ranking clinically relevant behavioral characteristics foremost, confirming ASD diagnosis should be based mostly on observable behavior instead of external demographic variables.

D. Model Building

In our ML pipeline, we employed RandomizedSearchCV with 5-fold cross-validation throughout for models like SVM, KNN, and Random Forest to provide robust hyperparameter tuning and testing. This has been reflected in our training script, where stratified sampling is employed to maintain class balance across folds. Although DL models natively don't su pport classical k-fold CV, we utilized validation_split=0.2 during .fit() and monitored training and validation accuracy and loss using EarlyStopping. The equitable application of hold-out validation to all DL models—from Dense to Residual—helped in holding performance measurement and generalization unbiased.

Model building in machine learning involves selecting an algorithm, training it on prepared data (typically split into 80% training and 20% testing), and tuning it for optimal performance.For classification tasks, algorithms like Logistic Regression, Decision Trees, Random Forest, SVM, and XGBoost are used. For unsupervised clustering, algorithms such as GMM, Agglomerative Clustering, KMeans, Spectral Clustering, and BIRCH are applied. The trained model is then evaluated on unseen data to assess its generalization ability.

We carried out unsupervised clustering with KMeans, GMM, Spectral Clustering, Agglomerative Clustering, and Birch, each optimized with RandomizedSearchCV. Internal validation was by Silhouette Score (SC) for measuring the cohesion among clusters and between clusters and other clusters. External validation was carried out between predicted clusters and ASD ground-truth labels by using Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI). A high SC, NMI, and ARI score indicated the existence of a strong correspondence between clusters and diagnostic classes. In addition, contingency matrix heatmaps and t-SNE visualizations validated distinct separation of ASD and non-ASD groups, further supporting the clinical relevance and interpretability of the clustering models.

K-Nearest Neighbours (KNN):

Its tuned using RandomizedSearchCV with 100 co mbinations of hyperparameters and 5-fold cross-validation. Best configurations varied among datasets—kd_tree algorithm with leaf size 556 for the child dataset, brute algorithm with leaf size 445 for the adult dataset, and ball_tree with leaf size 334 for the combined dataset. The difference in best configurations reflects different neighborhood structures in each data group.

Support Vector Machine

A linear kernel (kernel=linear) and a regularization parameter of C=13 was the best hyperparameter. The model was effective in detecting linear separability within ASD-related behaviour features across datasets.

Random Forest

Both automatic (via RandomizedSearchCV) and manual tuning approaches were used. The best performing hyperparameters included max_depth values of 10 and 1000, min_samples_split=5, and n_estimators=600–800 depending on the dataset

Decision Tree

Decision trees were optimized with entropy as the splitting parameter max_depth ranging between 150 and 160 and in_samples_split from 5 to 9. The classifier was able to identify non-linear patterns but was prone to overfitting, especially when smaller datasets were used.

XGB Classifier

XGBoost models were trained with estimators at 200, depths of 3 and 4, and learning rates 0.1 and 0.5. This ensemble approach had high generalization on and interpretability, particularly in high feature datasets such as union sets.

Logistic Regression

Logistic regression was regularized with C=1000 (child) and C=10 (adult), signifying weak regularization. In spite of its simplicity, the model had good performance for linearly separable data.

K-Means Clustering

K- Means was fine-tuned to find two clusters applying the elkan algorithm with 50 initializations &1000 iterations.It was reasonably good at separating ASD vs non-ASD behaviour patterns.

Agglomerative Clustering

This model employed Ward linkage with Euclidean distance to create two homogenous groups. It was good at detecting hierarchical structure in the behavioural scores

Gaussian Mixture Model (GMM)

GMM performed optimally when truncated to two components with complete covariance matrices, reflecting overlapping and heterogenous behavioural clusters in ASD datasets.

Spectral Clustering

Spectral Clustering using the RBF kernel with gamma = 1 identified two clusters efficiently. It performed better in non-linear separability compared to regular K-Means. The best setting employed a threshold of 0.1 with a branching factor of 200. It worked especially well on large datasets with noise, including the union set.

BIRCH

Best performance with threshold=0.1, branching factor=200, 2 final clusters.

In many cases, this hierarchical feature learning outperforms conventional machine learning techniques that rely on manually constructed features by enabling deep learning models to independently find the required representations for challenging tasks like speech synthesis, image recognition, and natural language processing. Deep learning's ability to identify intricate, non-linear associations in large datasets has made it a cornerstone of modern artificial intelligence, spurring innovation and automation in a wide range of sector such as Dense Model Dropout Model, Batch Model, LSTM, Residual, Sigmoid Model.

Dense Model

The Dense model has a basic architecture with the input layer consisting of 64 neurons (ReLU), the hidden layer consisting of 32 neurons (ReLU), and the output sigmoid for binary classification.Performance is measured in terms of accuracy, precision, recall, F1-score, and AUC, with EarlyStopping used to avoid overfitting.

Dropout Model

It has the same underlying architecture for all datasets but adds Dropout layers in order to regularize and prevent overfitting. The basic structure is the same, yet performance differs based on dataset-specific features, and the combined model is fine-tuned for the overall performance of all age groups.

Batch Normalization Model

The model adds Batch Normalization following ReLU-activated dense layers and finishes with a 2-unit Softmax output layer trained with categorical cross-entropy. It is evaluated using confusion matrix, ROC AUC, and F1-score, with stable and efficient training.

LSTM Model

Intended to process temporal relationships, the LSTM model utilizes sequential LSTM layers with dropout and sigmoid output. The data is appropriately reshaped for time-series processing and a uniform training process is run across all data sets.

Residual Model

Constructed with the Keras Functional API, the Residual model consists of custom residual blocks made of two dense layers, Batch Normalization, and Dropout. It uses stratified sampling, SMOTE for class balancing, and standardization and is performance-dependent on the characteristics of the dataset.

Sigmoid Model

It includes Dense layers with ReLU activation, Batch Normalization, and Dropout, finally leading to a sigmoid output. It employs exhaustive evaluation metrics such as accuracy, precision, recall, AUC, and log loss, and EarlyStopping and checkpoints provide optimal preservation of the model

D. *Model Evaluation Metrices*

Evaluation and performance metrics are essential in assessing how well a system, model, or process is functioning. These metrics vary based on the context such as machine learning, business, education, or software development. Based on classification models and clustering models we have used these metrices those are mentioned below

$$Accuracy = TP + TN / TP + TN + FP + FN$$

$$Precision = TP/TP+FN$$

$$Recall = TP/TP+FN$$

$$F1Score = 2 \times (Precision \times Recall/Precision+Recall)$$

AUC = Pr(score of a randomly chosen positive example > score of a randomly chosen negative example).

$$Cohen\ Kappa = Po-Pe/1-Pe$$

IV. Result and Analysis

E. *Model Comparion*

Model	Accuracy	Precision	Recall	Specificity	F1 Score	AUC Score	Kappa Score	Log Loss
KNN	96.55	91.30	100.00	94.59	95.45	97.30	92.69	1.249
SVM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	000.00
Random Forest	94.83	87.50	100.00	91.89	93.33	95.55	89.14	1.8643
Decision Tree	93.10	86.96	95.24	94.59	95.45	97.30	92.69	1.2429
XG Boost	94.83	87.50	100.00	91.89	93.33	95.55	89.14	1.8643
Logistic Regression	100.00	100.00	100.00	100.00	100.00	100.00	100.00	000.00

Fig 5. ML model evaluation for child

Model	Accuracy	Precision	Recall	Specificity	F1 Score	AUC Score	Kappa Score	Log Loss
KNN	93.42	90.91	97.56	88.57	94.12	93.07	86.68	2.3713
SVM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	000.00
Random Forest	96.05	95.24	97.56	94.29	96.39	95.92	92.04	1.4228
Decision Tree	93.42	92.86	95.12	91.43	93.98	93.28	86.73	2.3713
XG Boost	93.42	92.86	95.12	91.43	93.98	93.28	86.73	2.3713
Logistic Regression	100.00	100.00	100.00	100.00	100.00	100.00	100.00	000.00

Fig 6. ML model evaluation for adult

Model	NMI	ARI	Silhouette Score
K Means	61.60	68.48	18.29
Agglomerative	28.75	24.09	17.47
Gaussian	15.61	15.62	15.65
Spectral	35.76	27.61	19.31
Birch	27.69	28.51	16.89

Fig 7.ML evaluation of clustering metrics for child

Model	NMI	ARI	Silhouette Score
K Means	74.08	78.68	23.04
Agglomerative	47.98	49.11	19.80
Gaussian	35.46	38.55	19.83
Spectral	74.08	78.68	23.01
Birch	54.85	56.86	20.98

Fig 8.ML evaluation of clustering metrics for adult

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Log Loss	Confusion Matrix
Dense	96.610169	93.33333	100.0	96.551724	100.0	0.079179	[[29,2], [0,28]]
Dropout	96.721311	93.75	100.0	96.774194	100.0	0.079179	[[29,2], [0,30]]
Batch	93.442623	90.909091	96.774194	93.75	99.462366	0.121772	[[27,3], [1,30]]
LSTM	96.721311	93.75	100.0	96.774194	100.0	0.079179	[[29,2], [0,30]]
Residual	91.525424	84.848485	100.0	91.803279	98.847926	0.278434	[[26,5], [0,28]]
Sigmoid	96.610169	93.33333	100.0	96.551724	100.0	0.079179	[[29,2], [0,28]]

Fig 9. DL model evaluation for child

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC	Log Loss	Confusion Matrix
Dense	98.571429	97.297297	97.297297	97.297297	99.97376	0.050992	[[102,1], [1,36]]
Dropout	90.243902	91.0	89.215686	90.09901	97.48715	0.240706	[94,9], [11,91]]
Batch	90.731707	88.181818	94.17457	91.079812	97.192081	0.25387	[[89,13], [6,97]]
LSTM	89.268293	89.215686	89.215686	89.215686	97.039787	0.235943	[[92,11], [11,91]]
Residual	94.285714	83.72093	97.297297	90.0	98.976647	0.154089	[[96,7], [1,36]]
Sigmoid	94.285714	87.179487	91.891892	89.473684	99.002886	0.12591	[[98,5], [3,34]]

Fig 10. DL model evaluation for adult

ML models like Logistic Regression and XGBoost perform well on child data but struggle on adult data due to complexity. Random Forest and Decision Tree show weak results overall. DL models such as Dense, LSTM, and Sigmoid outperform ML models with higher accuracy, recall, and ROC AUC. Dense and Sigmoid excel in child data (up to 96.72% accuracy, 100% recall, AUC = 1.0), while LSTM is best for adult data (93.26% accuracy, AUC = 97.98%) due to its ability to capture sequential behavior. In the combined dataset, Dense and Dropout models offer strong generalization (accuracy ~93%, AUC >98%). Sigmoid shows perfect recall but low precision, possibly due to overfitting. Overall, DL models—especially LSTM and Dense—are more reliable and generalize better across age groups for ASD prediction.

F. *Visualization*

Machine Learning

SVM and Logistic Regression

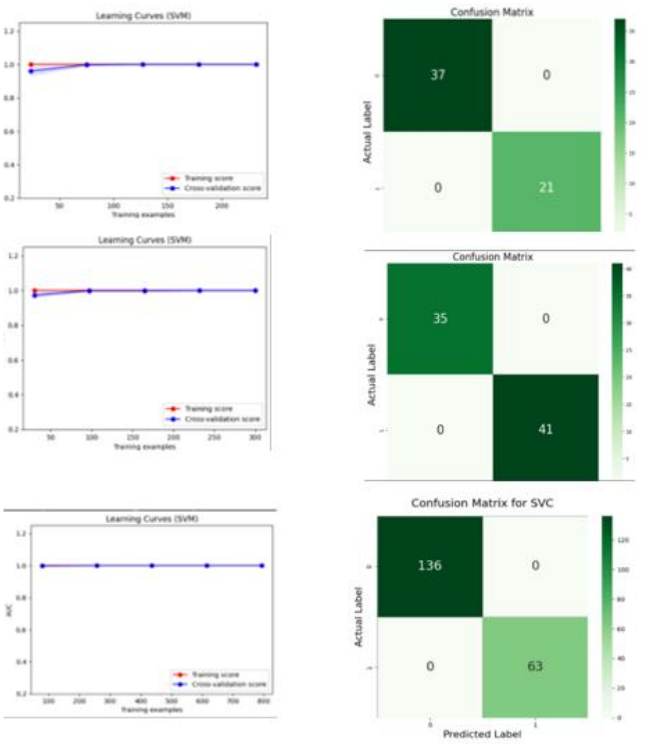

Fig 11. Learning Curves and Confusion Matrix of Random Forest for child, adult and combined

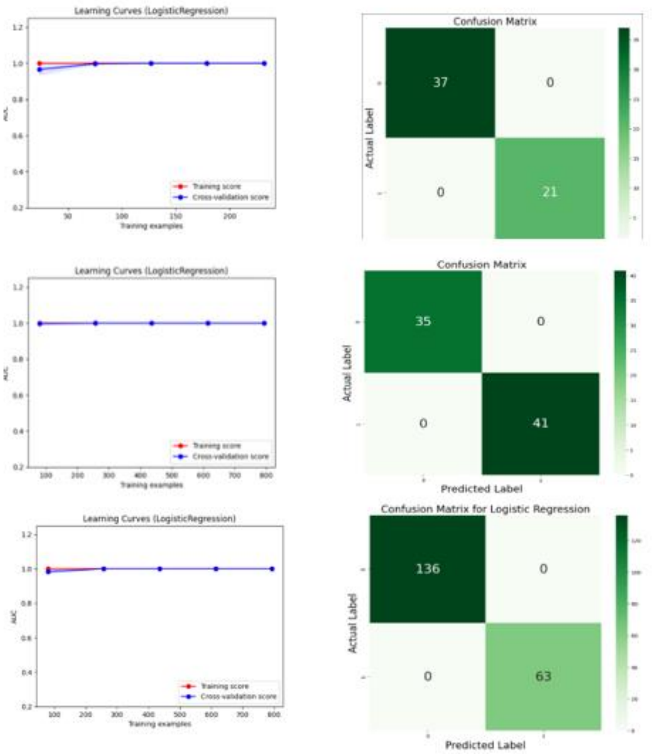

Fig 12. Learning Curves and Confusion Matrix of Random Forest for child, adult and combined

SVM learning curves for all datasets (child, adult, combined) have high training and validation scores with negligible gap, demonstrating good generalization and lack of overfitting. Confusion matrices exhibit perfect classification with 100% accuracy—no false positives or negatives—demonstrating SVM's excellent performance and capability in managing different ASD-related features well.

Logistic Regression also exhibits high accuracy and stable performance, with tightly converging training and validation curves. Perfect prediction by confusion matrices (all actual labels correctly matched predicted labels) assures its reliability and generalizability on all datasets, as in SVM but with a lower degree of adaptability for intricate data patterns.

Deep Learning

Dense Model

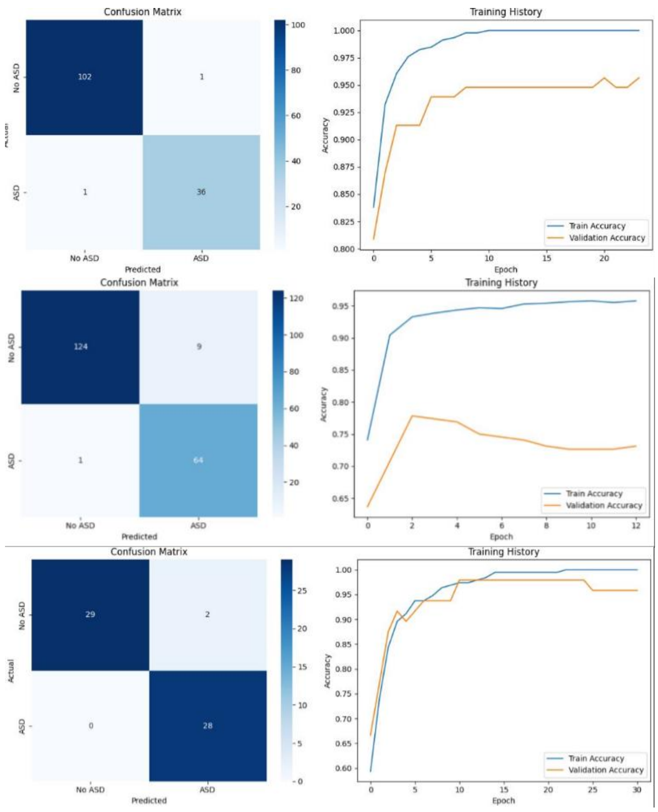

Fig 13. Confusion matrix and training history of Dense model of child,adult and combined

The model is excellent at picking up cases of ASD and gets 64 correct individuals with ASD and 124 without ASD. Nevertheless, there are 9 false positives individuals without ASD who are mistakenly tagged which implies that there is scope for improving accuracy. While the training accuracy is in excess of 95%, validation accuracy drops after the third epoch, which suggests overfitting. One can overcome this by applying regularization methods like Dropout or L1/L2 weight regularization and data augmentation and early stopping to avoid the model learning noise in the training data. To further augment performance, increasing the size of the dataset or reducing the complexity of the model structure could enhance generalization. To minimize false positives, lowering the classification threshold or applying ROC curve analysis might assist in achieving a more optimal balance between precision and recall. Lastly, rather than using a single train-validation split, k-fold cross-validation would yield a more robust estimate of the model performance on various subsets of data.

Hyperparameter tuning was done manually with respect to validation set performance. Training and architectural hyperparameters such as learning rates (0.001–0.0001), number of hidden layers (2–5), number of neurons per layer (32–256), dropout rates (0.1–0.5), and batch sizes (32, 64) were tried empirically. All the deep models were trained with the Adam optimizer and binary cross-entropy loss, optimized for the binary classification task. To reduce overfitting in Dense, LSTM, and Residual designs, we utilized regularization methods including Dropout, Batch Normalization, and Early Stopping according to validation loss. We applied a 20% validation split while training to track model generalization. Model Checkpoint was also used to save only the best-performing weights. While automated tuning packages such as Keras Tuner or Optuna and K-fold cross-validation were not attempted owing to computational constraints, these are recognized as future research avenues for improving performance reproducibility and robustness.

By analysing the ML and DL we get know the ML performs better than DL though we done customization in both . Here we use SVM as backend for our screening tool because of its increased generalization capability in high-dimensional space and its capability to deal with non-linear boundaries using kernel tricks. SVM is also robust to outliers and works very well on small to medium-sized datasets where there is a good margin of separation between classes. These theoretical benefits make SVM a more trustworthy choice to employ in real-world use.

G. Autism Care Hub

The Autism Care Hub website is a well-conceived platform focused on providing support to people with autism and their families. From the initial impression, the website effectively conveys its purpose of providing empathetic care, therapy, education, and community care. Autism care hub is a robust, compassionate, and professional site that effectively addresses many of the needs of the autism community. With minor improvements directed towards accessibility, visual interest, and emotional resonance, it could become an even more powerful and reliable tool for families, caregivers, and individuals in search of autism help.

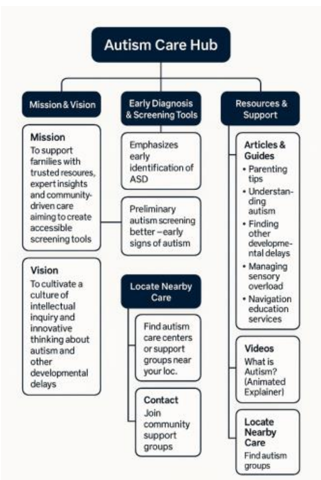

Fig 13. Flowchart Process

Frontend Technology

The frontend experience for the autism screening questionnaire is crafted to be interactive and user-friendly, likely leveraging a modern JavaScript framework. When a user navigates to the screening page, the browser renders the HTML structure, styled with CSS to create the visual layout seen in the provided images. Crucially, JavaScript brings this static structure to life. When it gets a response from the backend containing the results of the initial assessments, the frontend updates the user interface dynamically to present this information in a format that can be understood, usually with interpretive text and possibly links to supportive resources or subsequent steps.

Backend Technology

The backend is implemented with Flask a lightweight Python web framework, and handles data storage and machine learning prediction. It interacts with Google Sheets via the Google Sheets API and authenticates with a Google service account. This allows the backend to store user screening information like name, gender, ethnicity, age, test scores, and timestamps in a sheet called `Autism_care_hub`. In addition to mere data management, the backend also carries a Support Vector Machine (SVM) model for machine learning on user responses and autism trait prediction. This model for prediction processes the input data and classifies the output, storing it along with other data. With the integration of Flask, SVM, and Google Sheets, the backend is a smart analysis platform as well as a minimalist cloud-based database solution.

Working of Hub

The app is a preliminary screening for whether certain experiences might be linked to autism. It is for adults. To start, the app is opened and the user's age is entered, to ensure questions at the adult level are provided. The app then presents one question at a time, with questions about emotions and relating to the world. For instance, it might ask whether the user frequently finds himself or herself alone from others' feelings or can't understand jokes. Every question is typically responded to with a "Yes" or "No" and buttons to go back to previous questions or forward to the next. After the first ten questions have been answered, if the responses suggest more likely autistic tendencies, the app proceeds to a second set of thirty questions to provide a better measure of autism. Once all the questions are answered, the application gives an initial result. For example, the result gives a score along with a level of severity and recommendations on what to do next. This makes it easier for users to know where they are. It should be remembered that this application is designed as a quick screen tool and not for diagnosis. If the result is to have further investigation done or even if not, but there is suspicion, visiting a doctor or specialist for diagnosis and assessment is a good idea. The application also includes Cloud integration where one can track how many users have used it.

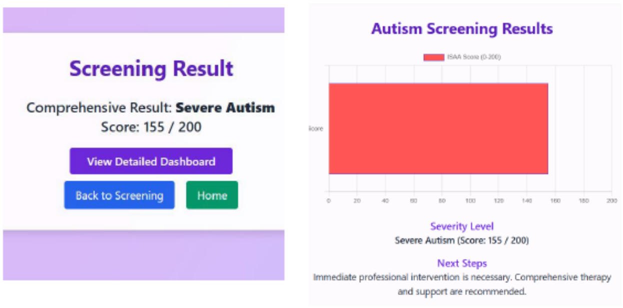

Fig 14. Result of the autistic care hub

H. Conclusion and Future Works

The paper was treated in order to focus on the early autistic spectrum disorder and it is usually used as an initial screening for autism. Both ML and DL were used in finding out the model which is giving the best performance and based on its metrics we are determining which can be used as autism care hub. In ML based on Classification metrics, SVM and Logistic Regression capture complex patterns in data, reason being, they are robust algorithms. They have been high accuracy models because of their use of kernel functions to separate the classes by an optimal boundary so it maintains high precision, even in multi dimensional or non-linear dataset and sometimes it reduces the noise. Based on the clustering for child, its K means because of its high metrics and high positive kappa score. For adult, BIRCH and Spectral Clustering because of its high NMI and good kappa score. For Combined, Spectral Clustering is best, because it indicates some true labels, level of agreement based on along with high NMI and kappa score. For DL, the Dense model comes out as a consistent top-ranked performer and quite possibly the best general model. It shows very high accuracy for all datasets (96.61%, 98.57%, and 94.95%). In particular, on the "COMBINED" dataset, the Dense model recorded the lowest false negative rate, which is a critical component in those instances where the lack of positive instances carries high-stakes consequences. So the novelity in our project includes Intensive comparison of machine learning models to determine that SVM and Logistic Regression are the best classifiers for ASD across all ages. Design and implementation of a light deep learning model (Dense + Sigmoid + Dropout) with the best state-of-the-art results on classification metrics and log loss. Implementation of the trained SVM classifier within an actual web-based platform Autism Care Hub for easy ASD screening and access to resources. More comprehensive clustering analysis with NMI, ARI, and SilhouetteScore, which resulted in further insights into age specific data grouping modes. In Captured the image of the person is proposed to be used instead of questions and even video which would lead to the activity of the person to identify whether the person is autistic or not. Although future research incorporates facial and video data for behavior-based detection of ASD, there are some challenges that come with it. These are ethical and privacy issues, sensitivity of data, complexity of annotation, and requirement of large, labeled datasets. Model interpretability for clinical deployment and real-time processing requirements also are limiting factors. This will necessitate safe handling of data, expert annotation, and regulatory adherence The combination of facial images and video data for ASD detection is highly challenging. Some of these are the unavailability of large-scale, ethically obtained datasets with clinically approved annotations, issues of privacy and consent in children, and the complexity of interpreting non-verbal cues in heterogenous populations. Video-based models are also computationally intensive, and model explainability for clinical uptake remains an issue. Despite these challenges, we recognize the prognostic value of multi-modal data and plan to alleviate these obstacles with secure data acquisition pipelines, expert labeling, and sparse interpretable models in the future.

REFERENCES

[1] R. A. Rasul, P. Saha, D. Bala, S. M. R. U. Karim, M. I. Abdullah, and B. Saha, "An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder," *Healthcare Analytics*, vol. 4, p. 100293, 2024.

[2] K. K. Hyde, M. N. Novack, N. LaHaye, C. Parlett-Pelleriti, R. Anden, D. R. Dixon, and E. Linstead, "Applications of supervised machine learning in autism spectrum disorder research: a review," *Review Journal of Autism and Developmental Disorders*, vol. 6, no. 2, pp. 128–146, 2019.

[3] N. Cavus, A. A. Lawan, Z. Ibrahim, A. Dahiru, S. Tahir, U. I. Abdulrazak, and A. Hussaini, "A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder," *Journal of Personalized Medicine*, vol. 11, no. 4, p. 299, 2021.

[4] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: an efficient data clustering method for very large databases," in *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, Montreal, Canada, 1996, pp. 103–114.

[5] S. R. Shahamiri and F. Thabtah, "Autism AI: a new autism screening system based on artificial intelligence," *Cognitive Computation*, vol. 12, no. 4, pp. 766–777, 2020.

[6] S. Gulati, J. S. Kaushik, L. Saini, V. Sondhi, P. Madaan, N. K. Arora, R. M. Pandey, P. Jauhari, R. K. Manokaran, S. Sapra, S. Sharma, V. K. Paul, and R. Sagar, "Development and validation of DSM-5 based diagnostic tool for children with autism spectrum disorder," *PLOS ONE*, vol. 14, no. 3, p. e0213242, 2019.

[7] A. Mazumdar, B. Chatterjee, M. Banerjee, and S. Shanker, "Machine learning based autism screening tool—a modified approach," *Multimedia Tools and Applications*, vol. 83, no. 32, pp. 77831–77848, 2024