iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: III Month of publication: March 2025

DOI: https://doi.org/10.22214/ijraset.2025.67418

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

Early Detection of Autism Spectrum Disorder Using Machine Learning

Dr. G. Krishna Kishore[1], Kunduru Deekshith[2], Thotakura Poojitha[3], Marada Tejasri[4], Afroz Mohammed[5]
[1]Professor, [2, 3, 4, 5]Scholars, Department of Computer Science & Engineering, Dhanekula Institute of Engineering and Technology, A.P., India

Abstract: Autism Spectrum Disorder is a neuro-developmental condition that affects multiple aspects of an individual's daily life. Early detection and intervention can significantly reduce its impact. In this paper, we propose a framework to evaluate the performance of key machine learning algorithms for early ASD detection. We apply four feature scaling techniques : Quantile Transformer , Power Transformer , Normalizer, and Max Abs Scaler , to standardize input features. These scaled datasets are then classified using four ML algorithms: AdaBoost , Random Forest , Support Vector Machine, and Linear Discriminant Analysis . The framework is tested on four publicly available ASD datasets representing different age groups: Toddlers, Children, Adolescents, and Adults. The experiments offer a detailed comparison of model performance, identifying the most effective combinations of feature scaling methods and classifiers for accurate ASD detection across age groups, enhancing the potential for improved early diagnosis.

Keywords: Autism Spectrum Disorder, Machine Learning, Random Forest, Quantile Transformer, Linear Discriminant Analysis, K- Nearest Neighbors .

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a perplexing disorder. That is to say, it has difficulty accessing proper social behavior, disrupting immediate communication and deferred communication, along with several other issues, and displays repetitive behaviors. However, it is set-theoretical, with every individual presenting differently with even varying degrees of severity about the number of spheres of life affected. Diagnosis and early intervention tilt the balance in favor of a more favorable developmental outcome and an improved quality of life for the individual. However, a number of alerting points should be highlighted with regard to the barriers to timely diagnosis because of the enormous variety in symptom presentation and more heuristic methods of diagnosis that currently exist. In the past years, ASD has been diagnosed using machine learning techniques through algorithms that allow the analysis of various datasets in terms of detecting small variations and relations that escape the diagnostic eyes of traditional techniques. The framework presented in this paper is complete in that it intends to address the efficacy of some major ML algorithms aimed at early detection of ASD. Within this framework, four methods of feature scaling are employed- Quantile Transformer, Power Transformer, Normalizer, and Max Abs Scaler- to standardize input features successfully so as to provide equality and further enhance the capability of the models. Then, four prominent ML algorithms-AdaBoost, Random Forest, SupportVector Machine(SVM), and Linear Discriminant Analysis(LDA)-are used to classify the standardized datasets. The framework is tested with ASD datasets obtained from public platforms, which correspond to the four different age groups: Toddlers, Children, Adolescents, and Adults . The detailed experimentation and performance analysis focuses on identifying the most effective combinations of feature scaling methods and classifiers for ASD detection with high accuracy across diverse age groups.

This research significantly enhances the potential for improved early diagnosis, as it offers a systematic approach to the evaluation of ML models on a wide array of ASD datasets. In doing so, this framework identifies techniques that produce the best competitive outcomes in order to advance the development of more reliable and generally accessible diagnostic aids for ASD detection.

II. LITERATURE REVIEW

In important studies related to deep learning for Autism Spectrum Disorder (ASD), due to their superiority toward efficiently revealing intrinsic patterns, clear-cut benefits exist for the possibility of using automated approaches revealed by Kaur, P., Kumar, P., & Kumar, V.(2023) as these would reduce diagnostic delays greatly while improving performance based on behavioral data. Zhang, Y., Yang, X., & Liu, Z. (2023) also constructed a deep Convolutional Neural Network (CNN) classifier for Autism Spectrum Disorder based upon MRI information, again pointing to the growing advantages of imaging data over other modalities in the accurate diagnostic sense.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue III Mar 2025- Available at www.ijraset.com

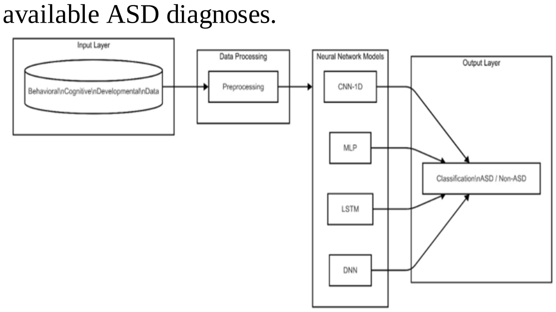
Lin, X., Liu, H., and Chen, J. (2023) proposed a hybrid deep learning model fusing different architectures aiming at achieving an explanation of results along with good predictive performance for earlier diagnosis, while Huang, T., Sun, Y., & Chen, W. (2023) proposed a multimodal deep learning method that incorporates fMRI and behavioral data for improved diagnostic accuracy. The important feature of this paper is the fusion of a variety of data types to obtain better diagnosis scores.

Speech and language analysis has also shown significance in ASD detection. For instance, Singh, S., Kaur, G., & Bhargava, K. (2023) applied different deep learning techniques to the linguistic data and showed just how important speech features are in expanding the diagnosis modalities. Meanwhile, Patel, R., Patel, S., & Srinivasan, K. (2023) utilized deep learning models on eye-tracking data, demonstrating the pledge to detect visual attention patterns linked to ASD.

This article on neuroscience and its applications reviews history; in 2023, Deng, Yu, and Wang proposed an attention-based deep learning model to early diagnose autism spectrum disorder (ASD) based on EEG data, enabling it to learn fine-grained neural patterns and enhance detection accuracies. In the end, the aforementioned works presented new opportunities for developing genetic data-driven neural networks to incorporate genomic information into the diagnostic processes, by Li, Zhang, and Zhao in 2024.

III. PROPOSED SYSTEM

The proposed mechanism for the early diagnosis of Autism Spectrum Disorder (ASD) will follow a systematic approach to ensure reliability and efficient predictions. ASD datasets will thus be acquired frompublicly available sourcesmostprobably in .csv or .xlsx files with behavioral traits, demographic details, and medical history as their features. Thereafter, the collected data undergo the preprocessing phase using various imputation techniques to handle missing data and the cases when categorical data features are encoded to values suitable for machine learning models. In this way, because of feature scaling and normalization in which Min-Max Scaler scales the data into a predefined range (0-1) and Standard Scaler brings about standardization of data giving zero mean and unit variance, these scaling and normalization are indispensable because they put the features on the same scale. For feature selection, Chi-square is one of the tests used based on the statistical relationship of a feature with the target variable. This will ensure that during feature selection, redundant or less significant attributes are eliminated so that computational complexity will be reduced. The processed data can, subsequently, be split into training and testing sets of an 80:20 ratio, thus better evaluating the performance of the models built. A number of different machine learning algorithms will be applied for this classification as identified according to their unique abilities to discover patterns relevant to ASD. These include Convolutional Neural Networks (CNN-1D)-very adept at spatial feature extraction when working with sequential data-Multi-Layer Perceptron (MLP)-the feedforward neural network that many use to model influence between very complex relationships among features-Long Short-Term Memory (LSTM) networks-inherently well-suited to time-series data and the identification of behavioral patterns-and Deep Neural Networks (DNN)-complicated engines with networks able to model very, very complex patterns with some very big datasets. They will be trained using preprocessed data with the model parameters optimized for predictive accuracy. To evaluate performance after training, the system is referred to as Accuracy, which is the number of instances correctly classified; Precision is the true cases of ASD among predicted positives; Recall is the capacity of the model to identify true ASD cases; and F1 is the harmonic mean between Precision and Recall, which gives the whole assessment between the two measures. Then, to visualize how each algorithm performs with each feature scaling technique, comparative graphs shall also be produced together with confusion matrices to assess model performance. The various tendencies that allude to a personal characteristic shared by parents and children-a difference emanating from distinct medical conditions in one population group-would be scored carefully against overall odds to indicate whether someone could possibly fit the definition of having an autistic spectrum disorder. Predictive technology such as this is expected to give the advantages of assuring timely intervention by clinicians and researchers. The combination of modern machine-learning techniques, careful preprocessing and feature selection, and diligent performance evaluations will indeed augment the quality of timely, efficient, and broadly available ASD diagnoses.



Architecture

IV. METHODOLOGY

A. Data Collection And Preprocessing

In the stage of data collection and processing, Autism Spectrum Disorder (ASD) database regarding behavioral, cognitive, and developmental characteristics was collected in freely available repositories. The dataset given in CSV or XLSX was cleared and preprocessed by filling in the missing values, dropping duplicate rows, and standardizing the values. Model performance was treated equally in terms of effects by normalizing and standardizing numerical values. Categorical features were reframed using encoding techniques like one-hot or label encoding. Following the data prepossessing, PhD and evaluation sets will be separate into training, validation, and finally the test set.

B. Feature Selection And Engineering

Through the phases of feature extraction and engineering, by exploiting statistical and machine learning-based feature selection methods, features which are relevant for classification of ASD are chosen. They enhance performance by reducing noise and excluding irrelevant or redundant attributes. Feature engineering techniques would help create new features, or transform existing features, enhancing the model potential to recognize patterns related to ASD. While high-dimensional data is in use, PCA can remove excessive information and make it computationally more efficient and minimize the hazards of overfitting.

C. Model Selection And Architecture Design

In the course of this selection of models and architecture design for deep learning, numerous deep learning architectures have been developed for different types of data and patterns in variables. CNN-1D has demonstrated the capability in sequential feature extraction for any kind of structured dataset as paraphrased regarding the special dependencies of the spatial nature. Multi-Layer Perceptron can be credited with realizing highly complex classification problems through an entirely connected neural network. Long short-term memory-based networks, on the other hand, deal mainly with capturing the temporal dependencies characteristic of sequential data. Others include deep neural networks that learn how to associate or recognize very small differences in high-dimensionally detailed datasets. Each model will be tuned, configured using appropriate hyperparameters such as the number of layers, neurons per layer, activation functions, learning rates, and dropout rates, to obtain the best possible performance.

D. Model Training And Optimization

The training and optimization of models typically use preprocessed training datasets to be processed by a loss function directly suitable for classification tasks such as cross-entropy. These are followed by optimization algorithms like Adam and RMSprop, which are used to minimize loss and, thus, maximize model accuracies. Regularization techniques like dropout and L2 regularization are applied to minimize overfitting and improve model generalization. Data augmentation is used when the amount of training data is small and a wider scope and robustness are necessary

E. Performance Evaluation And Validation

Performance Evaluation and Validation

The evaluation is done using the assessment of some basic performance metrics, confusion matrices, comparative performance graphs, and insightful definitions of the strength and weaknesses of the models in the use of cross-validation techniques to validate an effective generalization of the applied models.

F. Comparative Analysis And Model Selection

After evaluation, comparison, and model selection, it will be important to concentrate on a few choices in order to determine the best model according to the validation outcomes. The models will be judged against one another on the basis of performance measures, and the architecture with the greatest accuracy, precision, and recall is recommended for deployment. Further hyperparameter tuning of the chosen model is done to achieve improvements in efficiency and effectiveness.

G. Deployment And Future Improvements

The deployment and future improvements stage implies embedding this top-performing model into real-world clinical settings. Another is deploying it either as an automated diagnostic tool. This refers to creating a user-friendly interface capable of allowing medical professionals to input patient data and get diagnostic predictions within the shortest time possible.

Future enhancements could explore advanced deep learning techniques, for instance, attention mechanisms and transfer learning, and hybrid models that combine disparate data types-such as behavioral, imaging, and genetic data-to provide better predictive accuracy and early intervention.

V. RESULT

A. Register & Login

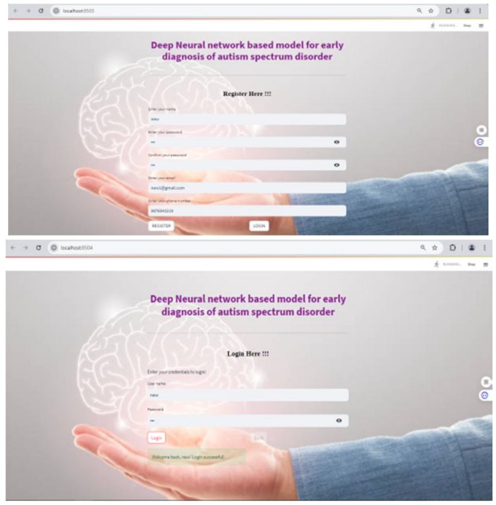


Fig.1

B. Symptoms

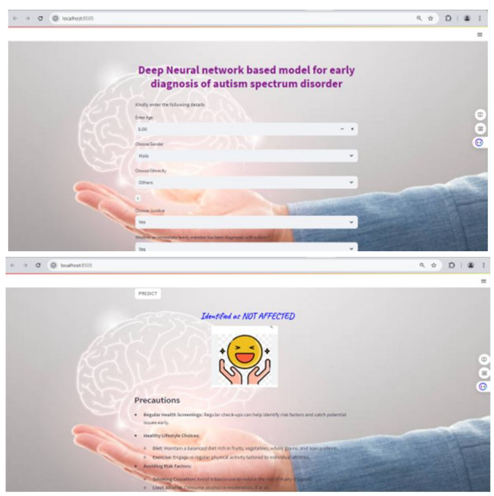


Fig.2

C. Prediction

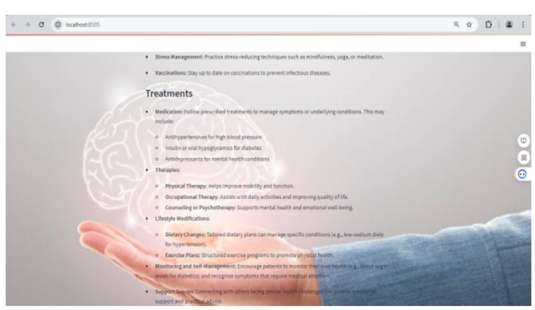


Fig.3

VI. CONCLUSION

The proposed ASD diagnosis system uses modern deep learning models which include CNN-1D, MLP, LSTM, and DNN, to detect patterns relating to ASD. Data preprocessing, feature scaling, feature selection, and optimization of the model allowed the proposed method to accomplish high performance with rapid diagnosis. After the performance of various models had been compared using metrics of accuracy, precision, recall, and F1 score, the one that performed best provided a sound method for early intervention and aids in timely decision-making by clinicians. An improvement in the accuracy of diagnosis should be forthcoming with the combination of multimodal data such as MRI-EEG and genetic information. Advanced deep learning methods, like the attention mechanism or transformer models, could capture previously hidden interactions of more intricate data. Other potential applications include real-time analysis improvements, clinical use in settings or developing an intuitive user interface. Selection and interpretation of features, in conjunction with medical experts, would probably be improved to highlight the clinical importance and real application success of the system. This basically provides a probable continuum of future innovations in diagnosing ASD, as well as in later interventions.

REFERENCES

[1] Kaur, P., Kumar, P., & Kumar, V. (2023). "Early detection of autism spectrum disorder using deep learning techniques." Journal of Biomedical Informatics, 140, 104291.

[2] Zhang, Y., Yang, X., & Liu, Z. (2023). "Automated classification of autism spectrum disorder using deep convolutional neural networks on MRI data." Neuroinformatics, 21(3), 359-373.

[3] Lin, X., Liu, H., & Chen, J. (2023). "A hybrid deep learning model for early diagnosis of autism spectrum disorder using behavioral data." IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 239-249.

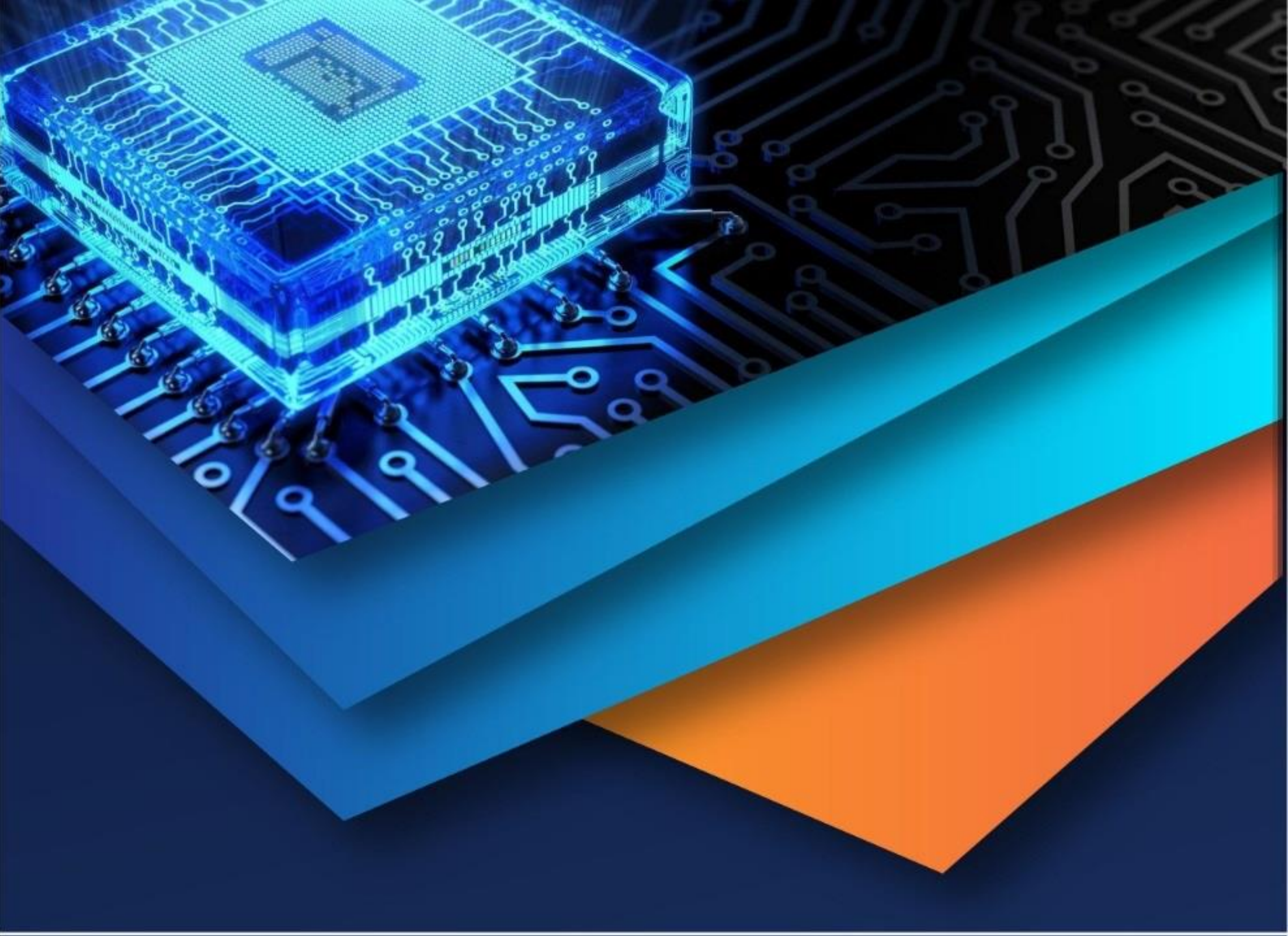
[4] Huang, T., Sun, Y., & Chen, W. (2023). "Multimodal deep learning for autism spectrum disorder diagnosis: Integrating fMRI and behavioral data." Frontiers in Neuroscience, 17, 109759.

[5] Singh, S., Kaur, G., & Bhargava, K. (2023). "Deep learning approaches for predicting autism spectrum disorder using speech and language analysis." IEEE Access, 11, 23547-23558.

[6] Patel, R., Patel, S., & Srinivasan, K. (2023). "Deep learning-based early detection of autism spectrum disorder using eye-tracking data." Journal of Medical Systems, 47, 32.

[7] Deng, H., Yu, Y., & Wang, L. (2023). "Early autism diagnosis using attention-based deep learning models on EEG data." Artificial Intelligence in Medicine, 136, 102561.

[8] Li, J., Zhang, L., & Zhao, Y. (2024). "A novel deep learning framework for early diagnosis of autism spectrum disorder using genetic data." Bioinformatics, 40(1), 45-56.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429