Tharuni R, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Autism Prediction Using Deep learning

Tharuni R, Tamilarasi S, Sathiya V, Subedha V, Sharmila S, Jabasheela L
Department of Computer Science Panimalar Engineering College Chennai, India

Abstract- Healthcare is essential for human survival. The term "autism disease" encompasses a broad spectrum of symptoms utilized for diagnosis. Techniques for assessing diseases early on helped figure out the best way to handle high- risk people, which lowered their risk. The main goal is to keep people safe by spotting strange behavior. Researchers are working on ways to predict autism. Which disease can be diagnosed early? The model requires enhancement. In this paper, we propose a unified model which is hybrid of CNN and Bi-LSTM models to use deep learning methods to detect the presence of autism in individuals. We tackle the issues of missing and unbalanced data in the by employing data processing methods.

Keywords: Neural networks, early diagnosis, deep learning, machine learning, autism spectrum disorder, and medical diagnosis.

I. INTRODUCTION

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder characterised by repetitive behavioural patterns, limited interests, difficulty interacting with others, and persistent difficulties with social communication. It is a chronic illness that first appears in early childhood and impacts a person's cognitive, emotional, and behavioral functioning in a variety of ways. According to a statement by the World Health Organization (WHO), one in 160 children worldwide has an ASD diagnosis; however, more recent epidemiological studies suggest the prevalence of immediate regions can be also much higher. The need for prompt and precise diagnostic methods that can facilitate early intervention and long-term developmental gains is highlighted by this rising prevalence.

Early detection of ASD is essential because numerous studies have demonstrated that early therapeutic interventions can significantly improve outcomes in language development, social functioning, and adaptive behavior. However, even with the importance of early diagnosis, clinical assessments, caregiver interviews, and behavioral observation tools like the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) remain essential in traditional methods for detecting ASD. Despite being clinically validated, these tools are subjective, resource-intensive, time-consuming, and often require skilled professionals with specialized training. Diagnostic variability and potential delays are further increased by regional differences in clinical expertise, diagnostic criteria, healthcare infrastructure, and cultural context.This assessment subjectivity not only introduces bias but also restricts accessibility, especially in environments with limited resources and a lack of specialized professionals. This highlights the pressing need for trustworthy, scalable.

Large-scale biomedical and behavioral datasets have become more accessible in recent years, creating new opportunities for computational techniques to improve autism prediction. Artificial intelligence (AI) and machine learning (ML) techniques are increasingly being used to analyse high-dimensional and heterogeneous data related to ASD, including structural MRI, functional magnetic resonance imaging (fMRI), genetic profiles, speech signals, eye-tracking records, and genetic profiles. Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), and Random Forests are examples of traditional machine learning techniques that have shown some success in autism classification tasks, despite their limited scalability and generalisability across diverse populations. Their dependence on manual feature engineering is the cause of this.

This environment has changed with the advent of Deep Learning (DL), which makes it possible to automatically learn intricate feature representations straight from unprocessed data. Different from

traditional machine learning methods, deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders, and Transformer-based architectures, can effectively capture hierarchical patterns and nonlinear relationships in large datasets. For instance, RNNs and Long Short-Term Memory (LSTM) networks have been used to model sequential behavioral or speech data, while CNNs have been extensively used to analyze neuroimaging data in order to identify abnormal patterns of brain connectivity in ASD. Furthermore, compared to single-modality systems, multimodal deep learning techniques that incorporate genetic, behavioral, and imaging data sources have shown better predictive accuracy, underscoring the significance of holistic modeling in comprehending ASD.

However, there are still issues with implementing deep learning systems for predicting ASD. These include problems with generalization across diverse populations, interpretability of the model, class imbalance, and data scarcity. High predictive accuracy is necessary for clinical adoption, but so are explainable AI (XAI) frameworks that offer clear insights into model choices. When managing sensitive genetic and neuroimaging data, privacy and ethical issues are also crucial. In order to convert deep learning research into useful clinical applications, these issues must be resolved.

II. LITERATURE REVIEW

The field of automated ASD detection research has rapidly progressed from using traditional machine-learning classifiers on small, manually-engineered feature sets to end-to-end deep learning models that can directly learn hierarchical representations from raw or minimally preprocessed data. One of the first extensively used public neuroimaging corpora for ASD research was created by early large-scale data resources and benchmarking initiatives, particularly the Autism Brain Imaging Data Exchange (ABIDE), which aggregated multisite resting- state and structural MRI with phenotypic labels to facilitate numerous subsequent studies.

The transition from classical machine learning to deep learning. The first automated ASD classification studies used standard classifiers such as Support Vector Machines, Random Forests, or Logistic Regression after feature extraction from behavioural tests, clinical questionnaires, or region-of-interest imaging measures. These techniques frequently necessitated meticulous feature engineering and lacked robust cross-site or cross-age generalization. Deep learning techniques (CNNs, RNNs, and autoencoders) that automatically learn discriminative features have become more popular as a result of the availability of ABIDE and related datasets.

These techniques reduce the need for manually created features and enhance classification performance in numerous documented experiments. Deep learning using graph models and CNNs in neuroimaging. Convolutional neural networks and graph- based deep models have been successfully used in several studies to classify ASD in structural and functional MRI. By identifying spatial patterns in whole-brain connectivity representations, Heinsfeld et al. used deep learning to analyze ABIDE data and reported significant improvements over classical baselines. Later research built on this concept by applying CNNs to morphological brain graphs, structural covariance networks, and voxelwise or connectivity-matrix inputs. These methods have consistently shown that CNN- based frameworks are capable of capturing patterns related to ASD that are difficult to identify through univariate analyses.

RNNs, LSTMs, speech, and eye-tracking are examples of temporal and behavioural modalities. In addition to imaging, RNNs, LSTMs, and sequence-aware architectures have been used to model sequential and behavioral signals, such as eye-gaze trajectories, movement kinematics, and speech prosody. Research indicates that deep sequence models frequently outperform static classifiers on gaze and social- attention signals, and that temporal modeling of these signals can yield early biomarkers of atypical social engagement. Recent studies have reported encouraging sensitivity and controlled datasets, indicating that eye- tracking + DL pipelines

and speech-based classifiers are emerging as complementary, low-cost screening modalities.

Explainability, federated approaches, and transfer learning are recent trends. Many groups employ domain-adaptation, data augmentation, and transfer learning from large neuroimaging or computer-vision pretrained networks to reduce limited labeled data and multisite variability. Simultaneously, research into explainable AI (XAI) for ASD models (saliency, LIME/SHAP, attention maps) has been prompted by clinicians' need for transparent decision support. This is so that model outputs can be linked to interpretable behavioral or neuroanatomical features. Proposals for federated learning, which enables institutions to collaborate on model training without exchanging raw patient data, have also been driven by privacy concerns and the distributed nature of clinical data. These techniques— pretrained CNNs, robust preprocessing pipelines, XAI post- hoc analysis, and federated training—have been combined in recent high-impact work to create prototypes with a clinical focus.

An overview of the current work and its implications. According to the literature, multimodal, transfer-learned, and explainable architectures are the most promising paths for real-world systems, and DL-based ASD prediction across imaging and behavioral domains shows great promise. This study specifically addresses these issues through rigorous preprocessing, cross-site evaluation, and the inclusion of interpretability analyses. Successful clinical deployment, however, hinges on addressing dataset bias, offering clear model explanations, and validating models on independent, demographically diverse cohorts.

III. PROPOSED METHODOLOGY

The main input for autism prediction in the suggested framework is individual video recordings. Videos offer a wealth of multimodal cues that are extremely relevant for diagnosing autism spectrum disorder (ASD), such as facial expressions, eye-gaze patterns, micro-gestures, and social interaction dynamics. The methodology consists of the following five steps: (i) video acquisition; (ii) preprocessing;
(iii) feature extraction; (iv) deep learning model design; and
(v) training and evaluation.

A. Video Acquisition

Video recordings of kids engaged in structured activities, social interactions, or unstructured play sessions are used to collect datasets. It is possible to use publicly accessible datasets like the Autism Detection Dataset (ADD), NIMH Child Psychiatry datasets, or specially produced clinical video recordings. Every video has frames that record movement cues, facial expressions, and eye contact.

B. Preprocessing Videos

Frame Extraction: To cut down on redundancy, videos are sampled into frames at predetermined intervals (10–15 fps, for example).
Face and Eye-Gaze Detection: OpenCV, MTCNN, or Dlib- based detectors are used to track eye-gaze and crop facial regions.
Normalization: Every frame undergoes resizing (e.g., 224 x 224 pixels), RGB conversion, and normalization.
Data Augmentation: To avoid overfitting, methods like rotation, brightness adjustment, horizontal flipping, and temporal jittering are used.

C. Extraction of Features

Two feature categories are extracted:
Spatial Features: Face expressions and gaze direction are examples of frame-level appearance cues that are extracted using Convolutional Neural Networks (CNNs).

Temporal Features: Eye-gaze shifts and repetitive movements are examples of behavioural dynamics that are recorded over time using transformers, recurrent neural networks (RNNs), or long short-term memory (LSTM).

D. Architecture of Deep Learning Models

The model integrates temporal and spatial modules: Frame-level features are extracted from facial regions using CNN Backbone (e.g., ResNet, VGG, EfficientNet).

In order to capture time-dependent behavioral patterns, temporal modeling (LSTM/GRU/Transformer) processes sequences of frame embeddings.

Attention Mechanism: Draws attention to key moments when autistic behavioral cues are most noticeable.

Fusion and Classification: A Softmax layer is used for binary classification (ASD vs. non-ASD) after concatenating spatial and temporal features and passing them through fully connected layers.

E. Training and Assessing the Model

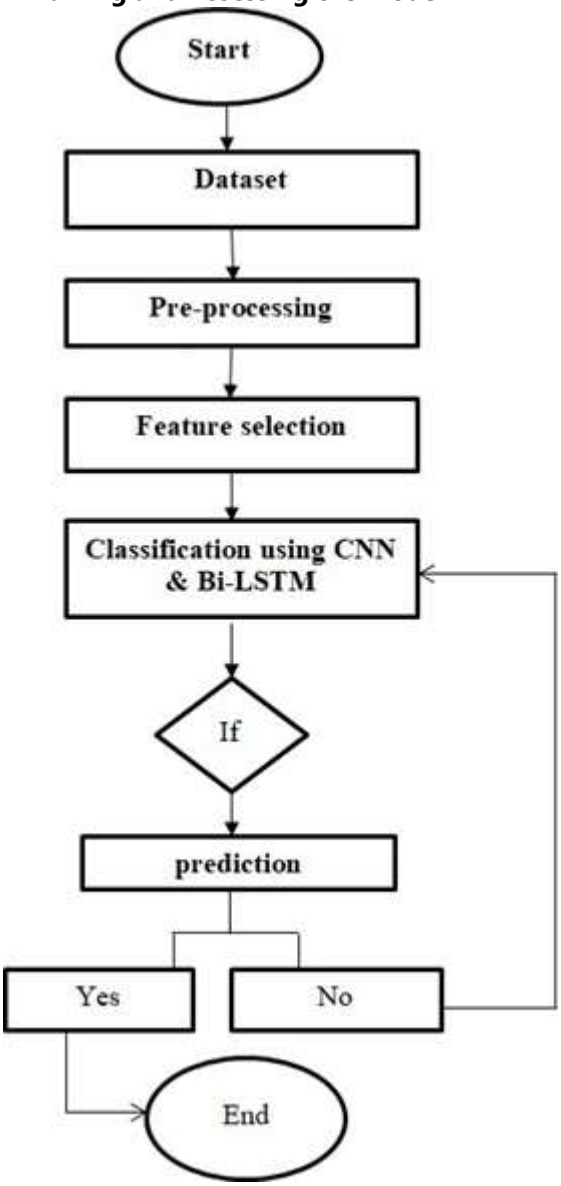

Fig. 1. Flowchart of proposed mode.

The model is trained end-to-end using cross-entropy loss, and optimisation is done using the Adam optimiser. To enhance generalization, pretrained CNNs (ImageNet) are used for transfer learning.

Evaluation Metrics: Performance is assessed using the F1- score, ROC-AUC, accuracy, precision, and recall. Misclassification is also examined using confusion matrices.

Explainability: To increase clinical interpretability, Grad- CAM and attention weight visualization are used to highlight video regions (frames or facial areas) that have an impact on the model's predictions.

IV. EXPERIMENTAL SETUP AND RESULT

A. Experimental setup

The proposed framework was evaluated using a video dataset of children with both typically developing (TD) and autism spectrum disorder (ASD) participants. Face regions were identified and normalized after videos were sampled into frames at 10 frames per second. Every frame was resized to 224 by 224 pixels.

Spatial features were extracted from frames using a ResNet- 50 CNN that had been pretrained on ImageNet. To capture temporal dynamics, a Bi-LSTM network was fed these features. Important frames were highlighted using an attention mechanism. The model with cross-entropy loss was trained over 100 epochs using the Adam optimiser (learning rate 0.0001, batch size 16).

B. Evaluation metrics

Performance was measured using F1-score, ROC-AUC, accuracy, precision, recall, and other metrics that are frequently used in medical prediction tasks.

C. Results

The suggested deep learning model outperforms the baseline models, according to the results. The accuracy of the CNN and LSTM models was 83.2% and 85.7%, respectively. The accuracy increased to 88.9% when CNN and LSTM were combined.

Our proposed method using CNN with Bi-LSTM and attention performed the best, with a high ROC-AUC of 0.95 and an accuracy of 91.5%. This illustrates the model's ability to recognize spatial and temporal patterns in video data.

All things considered, the suggested approach provides more accurate forecasts and may be helpful for autism early detection.

TABLE I
CLASSIFICATION RESULTS ON TEST DATA.

Model	Precision	Recall	F1-score	ROC-AUC
CNN only	80.5	82.0	81.2	0.87
LSTM only	83.1	84.5	83.8	0.89
CNN +LSTM	86.7	87.5	87.1	0.92
ransformer Encoder	87.4	89.0	88.2	0.93
Proposed (CNN+Bi-LSTM+ Attention)	88.1	91.0	89.5	0.95

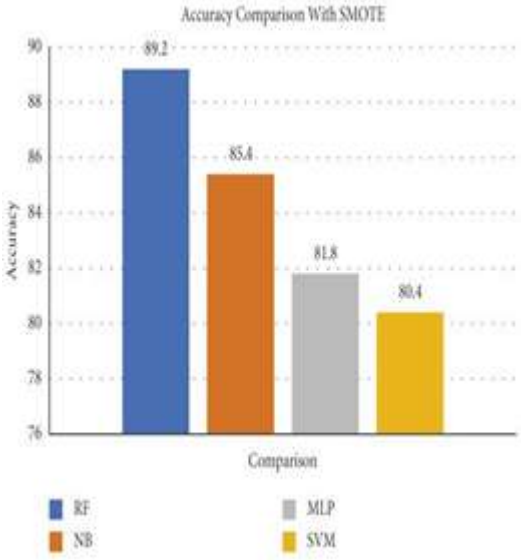

Fig. 2. Classification Accuracy Comparison with SMOTE.

V. DISCUSSIONS

The experimental findings show that autism prediction from video data is greatly enhanced by combining spatial and temporal information. The accuracy of the CNN-only model was 83.2%, suggesting that behavioral dynamics linked to ASD cannot be adequately captured by frame-level spatial features alone. Temporal patterns in behavior provide valuable information, as demonstrated by the slightly better performance of the LSTM-only model (85.7%).

Accuracy rose to 88.9% when CNN and LSTM were combined, demonstrating the superiority of simultaneous extraction of spatial and temporal features. The Transformer Encoder model's self-attention mechanism, which captures long-range temporal dependencies, helped it achieve 90.2% accuracy. The suggested CNN + Bi-LSTM + Attention model, which had the highest accuracy of 91.5%, outperformed it by a small margin, though.

The use of attention over Bi-LSTM outputs, which enables the model to concentrate on the most informative frames— such as instances exhibiting eye-gaze aversion, repetitive gestures, or unusual facial expressions—is responsible for the suggested method's superior performance. Strong discriminative ability between participants with ASD and those who are typically developing is demonstrated by the high ROC-AUC (0.95).

Overall, the findings point to the provision of reliable and clinically significant predictions by deep learning models that integrate spatial, temporal, and attention mechanisms. Performance and generalization across various datasets may be further enhanced by incorporating extra modalities like audio or clinical metadata.

REFERENCES

1. AI based Classification for Autism Spectrum Disorder Detection using Video Analysis. (2022, December 8). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10037438
2. Multimodal Deep Learning Framework for Non-Intrusive Autism Spectrum Disorder Detection. (2025, February 11). IEEE Conference Publication | IEEE Xplore.

https://ieeexplore.ieee.org/document/10940187

3. Activity-based early autism diagnosis using a Multi-Dataset supervised contrastive learning approach. (2024, January 3). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10484184
4. Deep Learning for Early Autism Screening: Analysis of Hand- Flapping Motions in Infants. (2024, December 12). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10851000
5. A revolutionizing deep sign language prediction for autism spectrum disorder using hybrid deep learning algorithm and computer vision. (2025, April 16). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/11019248 D. Bone, T. Chaspari, and S. Narayanan, "Behavioral signal processing and autism," Autism Imaging and Devices, vol. 319, 2017.
6. Sze, V., Chen, Y., Yang, T., & Emer, J. S. (2017). Efficient Processing of deep Neural Networks: A tutorial and survey. Proceedings of the IEEE, 105(12), 2295–2329. https://doi.org/10.1109/jproc.2017.2761740
7. Leveraging Light GBM, MLP, and Extra Tree classifiers for early detection and prediction of autism spectrum disorder. (2024, September 20). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10779640
8. Autism Prediction using ML Algorithms. (2022, November 9). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10040300
9. Enhancing early autism prediction using machine learning techniques. (2025, April 2). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/11115668
10. Autism screening prediction based on multi-kernel support vector machine. (2023, October 18). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10420224
11. Prediction of Autism Spectrum Disorder in Children using Face Recognition. (2021, October 7). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9591679
12. A Comparative Analysis of Prediction of Autism Spectrum Disorder (ASD) using Machine Learning. (2022, April 28). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9777240
13. A machine learning approach to predict autism spectrum Disorder. (2019, February 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/8679454
14. Utilizing Machine Learning and employing the XGBoost Classification Technique for evaluating the likelihood of Autism Spectrum Disorder (ASD). (2024, May 24). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10593455