Using federated learning for detecting autism in children

Maddala Kranthi[1], Saswati Debnath[2], Priyadharsini[1], R. Venkatesan[1,*]

[1]Division of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India

[2]Department of computer science and engineering, Alliance University, Bangalore, India

Emails: maddalakranthi22@karunya.edu.in; debnath.saswati123@gmail.com; priyadharsini@karunya.edu; rlvenkei2000@gmail.com

Abstract

Identifying Autism early in children is vital for ensuring more precise developmental support and effective therapeutic interventions. Traditional diagnostic approaches are frequently delayed, and data privacy concerns limit the availability of broad, multi-institutional datasets required for effective machine learning models. To address these limitations, this study proposes a CNN-LSTM-based autism detection model for children using Federated Learning (FL). In the model, temporal and spatial information is extracted from the facial CNNs are highly adept at using convolutional filters to extract spatial features from images. LSTM networks are a specific type of Recurrent Neural Network (RNN) that is ideal for processing time-series or sequences because it can identify long-term relationships in sequential data. This architecture uses CNN layers to extract spatial information from important indications that are important for detecting ASD, like eye patterns, gestures, and facial expressions. After that, these features are sent to LSTM layers, which examine the time-dependent and sequential behavioral patterns associated with autism. Federated Learning allows the locally to train the model on its own dataset locally, sharing only model updates with a central server, thereby preserving data privacy while promoting diverse data contributions. According to experimental results using the proposed techniques, the federated CNN-LSTM model performs 4.3% better than the conventional centralized models because it has less overfitting and is more resilient to a range of data distributions. The model's performance metrics further highlight its reliability, accuracy, precision, recall, and F1-Score values reaching 98.90%, 97.80%, 98.05%, and 98%, respectively, showing its potential for reliable ASD detection in children across diverse populations.

Keywords: Autism; Convolutional Neural Networks; Long Short-Term Memory; Children; Federated Learning

1. Introduction

Identifying Autism early in children is vital for ensuring more precise developmental support and effective therapeutic interventions. Traditional diagnostic approaches are frequently delayed, and data privacy concerns limit the availability of broad, multi-institutional datasets required for effective machine learning models. To address these limitations, this study proposes a CNN-LSTM-based autism detection model for children using Federated Learning (FL). In the model, temporal and spatial information is extracted from the facial CNNs are highly adept at using convolutional filters to extract spatial features from images. LSTM networks are a specific type of Recurrent Neural Network (RNN) that is ideal for processing time-series or sequences because it can identify long-term relationships in sequential data. This architecture uses CNN layers to extract spatial information from important indications that are important for detecting ASD, like eye patterns, gestures, and facial expressions. After that, these features are sent to LSTM layers, which examine the time-dependent and sequential behavioral patterns associated with autism. Federated Learning allows the locally to train the model on its own dataset locally, sharing only model updates with a central server, thereby preserving data privacy while promoting diverse data contributions. According to experimental results using the proposed techniques, the federated CNN-LSTM model performs 4.3% better than the conventional centralized models because it has less overfitting and is more resilient to a range of data distributions. The model's performance metrics further highlight its reliability, accuracy, precision, recall, and F1-Score values reaching 98.90%, 97.80%, 98.05%, and 98%, respectively, showing its potential for reliable ASD detection in children across diverse populations.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

Autism Spectrum Disorder (ASD) manifests as a complicated neurodevelopmental disease with a wide range of symptoms. ASD is marked by difficulty in social relationships, communication obstacles, and repetitive behaviors, the severity of which varies greatly across people. Early and precise diagnosis is required for successful therapies that can assist children in developing crucial social, cognitive, and emotional abilities. According to the World Health Organization, 1 in 160 children globally have been diagnosed with ASD, and this number has been growing, perhaps due to improved diagnostic methods and greater awareness (WHO, 2022). However, traditional diagnostic approaches are primarily clinical, relying on observable behaviors and caregiver-reported information. This often leads to a delayed diagnosis, typically not occurring until the child is around four years old or older (Baio et al., 2018). Recent advanced learning of Machine Learning (ML) offers promising methods for detecting autism at earlier stages by identifying patterns in behavioral, physiological, and environmental data collected through IoT devices and sensors. However, the CNN-LSTM architectures for ASD detection highlight several gaps. One key limitation is the lack of multimodal data integration. Studies typically rely on a single modality, such as video or sensor data, rather than combining visual, auditory, and physiological signals, which could enhance accuracy (Gupta et al., 2024). Additionally, CNN-LSTM models often require large labeled datasets, which are scarce in the ASD domain, making it challenging to train robust models. Another gap is the limited focus on model interpretability with spatial-temporal patterns that are most indicative of ASD, especially for clinical applications (Li et al., 2024). Addressing these gaps could significantly improve the effectiveness and usability of CNN-LSTM models in early ASD detection.

Despite the potential of ML, a significant barrier is the availability of large, diverse datasets necessary for robust model training. Privacy concerns and data-sharing limitations across institutions restrict data access, especially in sensitive domains such as healthcare. Federated Learning has emerged as a viable solution, enabling to collaboratively train models on decentralized data without transferring sensitive information to a central server (Rasul et al., 2024). Utilizing FL, this study proposes a CNed ASD detection framework that preserves privacy by keeping data local while utilizing aggregated model updates to enhance performance across institutions. This approach ensures model robustness across diverse data sources, potentially reducing biases and improving generalizability in ASD diagnosis.

The proposed research aims to address the limitations of centralized ML models by introducing a Federated CNN-LSTM framework for early ASD detection in children. This approach will improve accuracy and privacy, thus facilitating broader adoption in clinical and educational settings. The novel contribution of the research is as follows.

i. Combining spatial and temporal features
- Spatial characteristics, such as motions and facial expressions, are extracted using CNNs.
- Temporal aspects, such as behavioural trends across time, are captured by LSTMs.

ii. Sequential modelling of behaviour
- CNNs capture individual behaviour patterns (spatial data).
- LSTMs model how these behaviours change over time (temporal data).

iii. Improved generalization and reduced overfitting

iv. Enhanced feature representation
- CNNs extract abstract features that are processed by LSTMs to understand behaviour sequences more effectively.

v. Scalability and Flexibility
- The architecture can easily adapt to new types of multimodal data or sensors.

The paper is organized as follows: Section 2 contains a review of relevant literature; Section 3 outlines the methodology for the research; Section 4 shows the findings; and Section 5 summarizes the research.

This article gives linear model, which is the direct simplex method using neutrosophic logic, the logic that is the new vision of modelling and is designed to effectively address the uncertainties inherent in the real world founded by the Romanian mathematician Florentine Smarandache [1, 2]. In addition to that, Ahmed A. Salama presented the theory of neutrosophic classical categories as a generalization of the theory of classical categories [12,20], also, he developed, introduced, and formulated new concepts in the various disciplinary of mathematics, statistics, computer science by neutrosophic theory [17,18,19,22,28].

2. Related Work

It is well known that to get an optimal solution for any linear programming problem using the direct simplex algorithm should be processed to be in standard form, the simplex method for solving an LP problem requires the problem to be expressed in the standard form. Nevertheless, not all LP problems appear in the standard form. In many cases, some of the constraints are expressed as inequalities rather than equations.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

There are several difficulties in detecting and diagnosing ASD, especially when employing cutting-edge technology like ML and FL. Tackling these challenges involves both technical and practical aspects, and is vital for advancing the effectiveness and accessibility of healthcare solutions.

ML models, especially Deep Learning models (Chen et al., 2024) such as CNN and LSTM require diverse datasets to train effectively. However, in the case of ASD detection, obtaining such datasets is difficult due to privacy concerns, data fragmentation across institutions, and the sensitivity of health-related data. The lack of comprehensive, labeled datasets that include a wide range of ASD symptoms and behavioral data across different age groups, genders, and socio-economic backgrounds limits the generalizability and robustness of ML models (Zhu et al., 2023). Health data particularly that related to children is highly sensitive. This creates a significant barrier to data sharing, as traditional data centralization methods can expose personal information, raising privacy risks. Even though the adoption of Federated Learning has alleviated some concerns by enabling decentralized model training, data privacy remains a challenge. Ensuring that the models are adequately protected from potential breaches while complying with privacy laws such as HIPAA in the U.S. and GDPR in Europe requires continuous innovation in secure computation methods (Caruana et al., 2015).

The quality of data used for training ML models is often inconsistent. For ASD, data can come from multiple sources, including behavioral videos, speech recordings, and clinical assessments, each differing in format, resolution, and accuracy. This inconsistency can introduce noise and bias, hindering the ability of the model to learn accurate and reliable features for detecting ASD (Zhu et al., 2023). Preprocessing techniques to standardize the data and remove noise are often necessary but time-consuming and computationally expensive. Improving the explainability of ML models is a key challenge, and there is ongoing research into developing more interpretable AI techniques for healthcare applications (Caruana et al., 2015). Integrating AI-based tools into clinical workflows remain a challenge. Healthcare providers are often hesitant to adopt new technologies due to concerns about the reliability of the models, as well as resistance to changing established diagnostic practices. Additionally, training clinicians to use ML models effectively requires time, resources, and ongoing support. Even when machine-learning models show promise, they must be adapted to fit into the existing healthcare infrastructure, which is often outdated and not designed for integrating advanced technologies (Valizadeh et al., 2024).

In many cases, the data available for training machine-learning models is imbalanced, with a significantly lower number of ASD cases compared to neurotypical cases. This class imbalance can cause ML models to favor the majority class, leading to poor sensitivity for detecting ASD, especially in early-stage cases. While methods like sophisticated loss functions, undersampling, and oversampling can aid in resolving this problem, they might not completely remove the effect of data imbalance on model performance (Chawla et al., 2002). When a machine-learning model becomes overfit to the training set, it captures noise instead of the underlying patterns that apply to new data. Given the variability of ASD symptoms and the often-small dataset sizes available, ML models are at risk of overfitting, leading to poor performance on unseen data. Ensuring that the model generalizes well across different demographics and clinical settings requires careful regularization techniques and validation on diverse datasets (Amirbay et al., 2024). ASD often presents with developmental changes over time, making it essential to analyze the temporal and sequential aspects of the data. Longitudinal data, such as video recordings or speech patterns over a period, is required to track developmental progress or regression. However, processing such temporal data introduces additional complexity, as Bi-LSTM models, while suitable for sequence prediction, still require extensive data preprocessing and hyperparameter tuning to handle long sequences effectively (Khan et al., 2024).

Future research can significantly advance the early diagnosis and detection of ASD using ML and FL techniques by tackling these issues. Progress in these areas could lead to more accessible, accurate, and privacy-preserving diagnostic tools for healthcare providers, ultimately improving outcomes for children with ASD

3. Methodology

A neurodevelopmental disorder, autism spectrum disorder (ASD) causes a variety of behavioral, communicative, and social interaction issues. Effective intervention depends on early and precise detection, but conventional methods of diagnosing autism mostly rely on direct clinical observation, which can be subjective and resource-intensive. The proposed model, which is built on data acquired from many datasets, uses Federated Learning to provide a scalable, privacy-preserving method for detecting autism in children. As shown in Figure 1, facial behavioral data is collected on local devices from sources like hospitals, family smartphones, and schools, ensuring privacy by keeping data stored locally. Each device pre-processes this data with cleaning and feature extraction, after which a local model is trained to recognize autism-related patterns.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

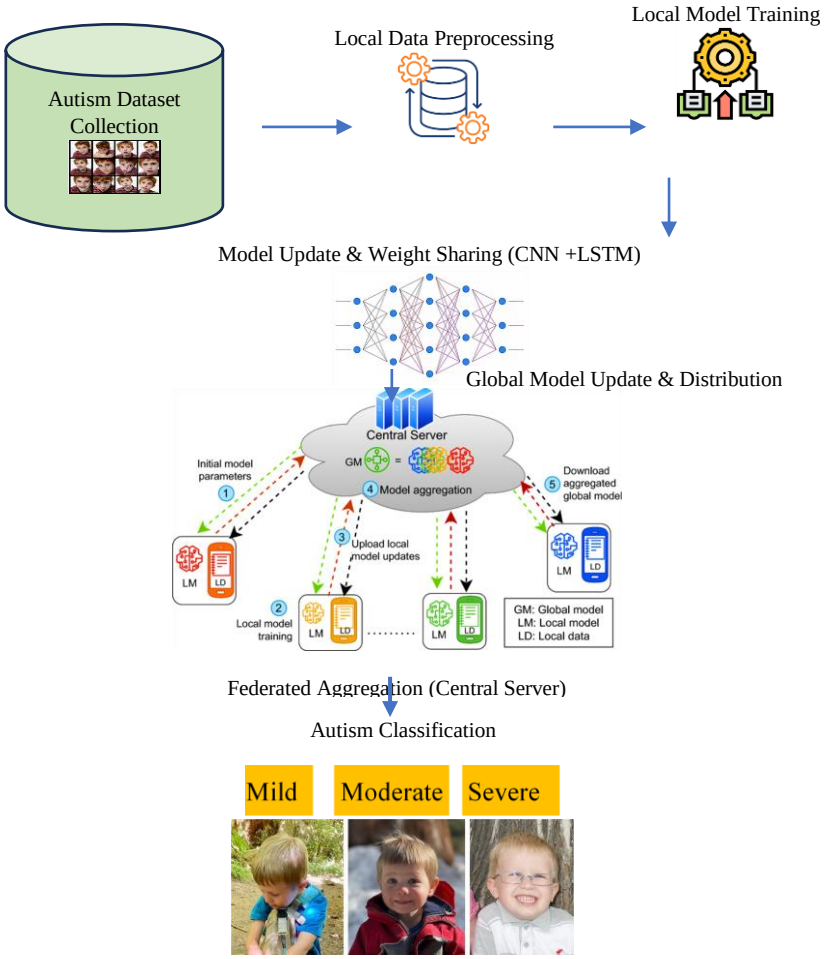


Figure 1. Architecture of the proposed ASD model using Federated Learning

A global model that reflects collective knowledge across institutions is created by securely transmitting model updates and weights to a central server instead of sharing raw data. After that, the global model is redistributed to every local device, allowing for iterative refinement as more local training, aggregation, and redistribution cycles take place until the model converges. Finally, the trained model is deployed for inference, detecting autism indicators in new data inputs while a feedback loop continually refines its performance, making it a dynamic and adaptable tool for early autism detection across diverse environments. A number of interrelated procedures make up the proposed study, which protects data privacy and makes it possible to develop a reliable, broadly applicable autism detection model. The available datasets provide diverse facial imaging data that facilitate the study of autism-related features in children.

Table 1: Patient distribution across various ASD facial imaging datasets

Dataset	**ASD Patients**	**Non-ASD Patients**
CHOP Facial Features Dataset	200	180
ASD-FID Dataset	150	150
Stanford Children's Health Facial Imaging Dataset	300	320
DDG2P Dataset	100	100
Autistic Children Facial Image Data Set	1327	1327

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

Table 1 summarizes the distribution of patients in various facial imaging datasets related to ASD. It compares the number of ASD and non-ASD patients across five different datasets: CHOP Facial Features, ASD-FID, Stanford Children's Health Facial Imaging, DDG2P, and Autistic Children Facial Image Data. Each dataset includes a specific number of ASD and non-ASD patients, reflecting the diversity of data used in ASD-related research. As shown in Figure 2, the CHOP facial features dataset focuses on 3D facial structures and morphological differences between ASD and non-ASD groups, highlighting facial landmarks indicative of ASD traits.

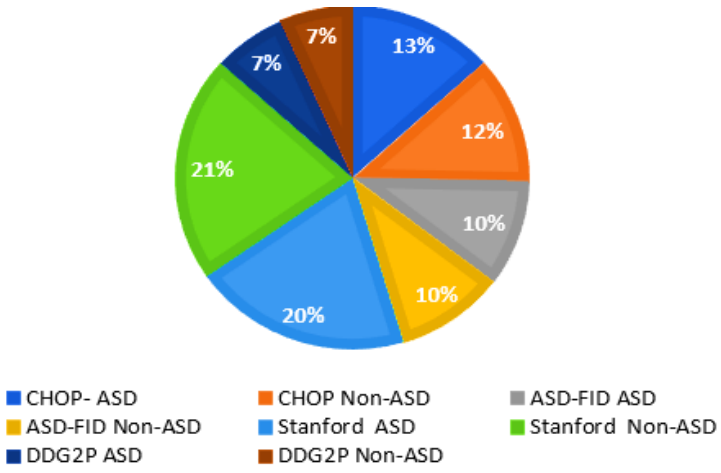


Figure 2. Distribution of datasets within classes.

The ASD-FID Dataset offers 2D images of ASD and neurotypical children, allowing for comparative analysis of facial expressions and features relevant to ASD markers. The Stanford Children's Health facial imaging dataset supports emotion recognition and facial expression analysis in children with autism, which can aid in understanding autism-specific emotional cues. Lastly, the DDG2P Dataset, with a focus on rare developmental disorders, provides facial and genetic data to study dysmorphology patterns in ASD. Together, these datasets contribute to a comprehensive approach to examining facial morphology and expression patterns associated with ASD in children.

3.2 Data Pre-processing

Pre-processing steps include data cleaning, where noise and irrelevant information are filtered out. Data entries with incomplete or inconsistent records are removed and imputed based on local heuristics. If a feature has a large amount of missing data, then it is removed. For features with minor missing values, impute the missing values using a method of mean imputation as shown in equation (1).

$$x_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j} \quad (1)$$

Where x_j is the imputed value for feature j when it has missing data points; data points are denoted for feature j that have been observed as non-missing values.

Additionally, normalization techniques are applied to standardize the scale of data features to ensure comparability across different devices and institutions. It is given in equation (2).

$$f_j' = \frac{f_j - \min(f_j)}{\max(f_j) - \min(f_j)} \quad (2)$$

where $f_j$is normalized to a range of [0, 1] with f_j' representing the normalized feature. The minimum value of features $f_j$across all samples is $\min(f_j)$and $\max(f_j)$.

This scaling enhances the performance of ML algorithms, especially those sensitive to feature scales by reducing the potential impact of large disparities across different features.

3.3 Local Model Training for Feature Extraction

The proposed system employs CNN with LSTM networks for local training to find the autism-related patterns within its data.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

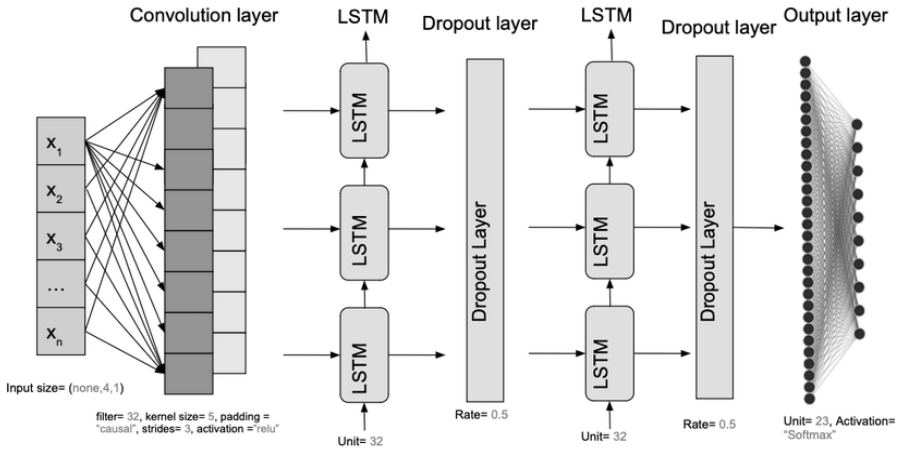


Figure 3. The proposed 1D CNN-LSTM architecture

As shown in Figure 3, this combined CNN-LSTM architecture is beneficial for learning spatial patterns from biometric markers such as eye-tracking data and sequential dependencies, which is vital for autism detection. The spatial characteristics are extracted from input data via the CNN component x_d on device *d* where each data point could be an image or frame representing eye gaze position. Each convolutional layer in the CNN performs two primary operations: convolution and pooling.

3.3.1 Convolution Operation

For an input image $x \in x_d$ denote the output from a convolutional layer as per equation (3).

$$f_{i,j} = \sigma\left(\sum_{m=1}^{M}\sum_{n=1}^{N} x_{i+m,j+n} \cdot w_{m,n} + b\right) \quad (3)$$

For a given output feature map $f_{i,j}$, the value at the position *(i,j)* is determined by the weighted sum of input values in the receptive field, where each value in the receptive field $x_{i+m,j+n}$corresponds to an input pixel at position *(i+m, j+n)* in the input image. $w_{m,n}$is the convolutional kernel's weight at location *(m,n)*.B is the bias. $\sigma(\cdot)$ is the activation function, often $ReLU(\sigma(z) = \max(0, z))$.

3.3.2 Pooling Operation

After convolution, pooling reduces the feature map's spatial dimensions to retain essential features while reducing complexity as given in equation (4).

$$f'_{i,j} = \max_{(p,q)\in R} f_{i+p,j+q} \quad (4)$$

where:

- $f'_{i,j}$ is the feature map that is produced following pooling.
- *R* is the pooling region.
- Max-pooling operation takes the maximum value in *R*.

The CNN's final layer outputs a feature representation F_d that captures relevant spatial patterns in the data.

3.3.3 LSTM for Temporal Dependency

The feature vector sequence is processed by the LSTM component as $F_d = [f_{d,1}, f_{d,2}, \ldots, f_{d,r}]$, where each $f_{d,t}$ is the CNN-extracted feature at time *t*. This allows the model to capture patterns over time. As shown in Figure 4, Long-term memory is transported throughout the network by the cell state. It is essential to the LSTM's ability to maintain long-term dependencies and is updated at every time step.The input gate and forget gate work together to update the cell state.

Input Gate:

$$i_t = \sigma\left(W_i \cdot f_{d,t} + U_i \cdot h_{t-1} + b_i\right) \quad (5)$$

Forget Gate:

$$f_t = \sigma\left(W_f \cdot f_{d,t} + U_f \cdot h_{t-1} + b_f\right) \quad (6)$$

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

Cell State Update:

$$\tilde{c_t} = \tanh(W_c \cdot f_{d,t} + U_c \cdot h_{t-1} + b_c) \quad (7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c_t} \quad (8)$$

Output Gate:

$$O_t = \sigma(W_0 \cdot f_{d,t} + U_0 \cdot h_{t-1} + b_o) \quad (9)$$

Hidden State:

$$h_t = O_t \odot \tanh(c_t) \quad (10)$$

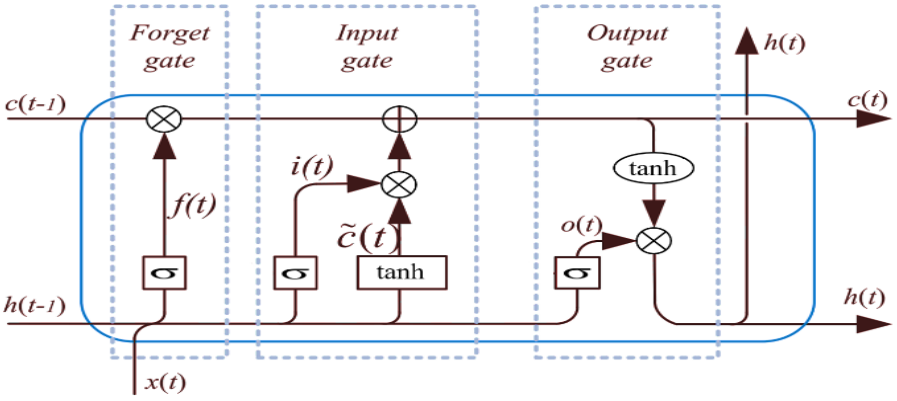


Figure 4. LSTM model.

where:

- $f_{d,t}$: Feature vector from CNN for time step *t*.
- At time step t, the learnt temporal pattern up to *t* is represented by the hidden state, h_t .
- Long-term data is stored in the cell state at time step *t*, or c_t.
- i_1, f_t, O_t: Input, forget, and output gates, respectively, each controlling different aspects of the LSTM's memory.
- W_i, W_f, W_c, W_o: Weight matrices for the input $f_{d,t}$.
- U_i, U_f, U_c, U_o: Weight matrices for the hidden state h_{t-1}.
- b_i, b_f, b_c, b_o: Bias terms for each gate.
- σ: Sigmoid activation function used in the gates to keep values between 0 and 1.
- $tanh$: Hyperbolic tangent function, used to create values in [−1,1].

The LSTM produces a final hidden state h_T, which is a summary of the temporal information across the sequence.

3.4 Model Update & Weight Sharing

Following training, each local model produces updates that represent the patterns discovered in the local dataset as model weights (θ_d). Equation (11) illustrates how these model updates are sent to a central server rather than exchanging raw data.

$$\Delta\theta_d = \theta_d^{(t)} - \theta_d^{(t-1)} \quad (11)$$

Each update is securely encrypted and transmitted to prevent unauthorized access during transfer. By sharing only the model weights, Federated Learning adheres to strict privacy standards. Additionally, this step reduces the risk of data leakage, as even if intercepted, the model weights would be less informative than raw data.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

3.5 Federated Aggregation

To generate a global model, the central server aggregates the model updates it gets from every participating device. Methods such as Federated Averaging, which averages the weights from each local model and weights them according to the size of the dataset or significance of each device, are commonly used to accomplish aggregation. This process creates a unified model that captures patterns from diverse sources, enhancing its generalization across various demographics and data sources.

$$\theta_{global} = \frac{1}{D}\sum_{d=1}^{D}\theta_d \quad (12)$$

where *D* represents all of the devices taking part in the round and aggregation using a weighted average is dθ_d. This aggregation step allows the global model to integrate without directly accessing any individual's data. Federated Aggregation builds a comprehensive model to recognize autism-related patterns across different settings, from clinical environments to home-based data sources.

3.6 Global Model Update & Distribution

After aggregation, the central server distributes the updated global model θ_{global} back to each local device. The updated model incorporates knowledge from a broader dataset, representing patterns learned from all datasets. Each local device d receives θ_{global} and integrates it with local training to improve its detection capabilities. Distributing the updated model to local devices, Federated Learning creates a cycle of iterative model enhancement, where each round of training and aggregation contributes to a progressively more accurate and robust model. Until the global model reaches a predetermined degree of accuracy or convergence, this cycle keeps going.

3.7 Iterative Training

The training and aggregation cycle is repeated iteratively to refine the model continuously. Each iteration incorporates new data and updates from local devices, which allows the global model to adapt to evolving patterns and maintain high performance. This iterative process continues until model convergence, where additional iterations yield minimal improvements. Stop iterations when the change in global model weights $\left\|\theta_{global}^{(t)} = \theta_{global}^{(t-1)}\right\|$is below a small threshold ∈. Iterative training is essential for achieving a high level of accuracy and ensuring that the model remains relevant over time. This approach enables the autism detection model to utilize the ongoing data collection, adapt to new behaviours, and improve detection rates.

3.8 Inference & Detection Phase

Once the global model θ_{global} reaches convergence, it is deployed across local devices for inference; enabling real-time autism detection based on new data inputs as follows,

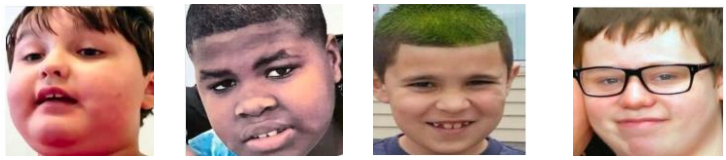
$$\hat{y} = M(x; \theta_{global}) \quad (13)$$

By utilizing the aggregate knowledge of all participants, the inference step enables each local device to apply the trained model independently. Determine possible autism signs if $\hat{y} > T$, where *T* is a decision threshold, based on the model output yˆ. This decentralized deployment ensures that the model performs consistently across various settings while continuing to protect individual data privacy.

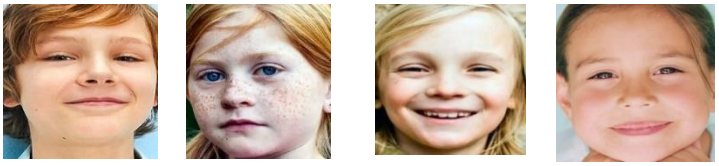
4. Results

Using a CNN-LSTM model for facial feature extraction from images in an ASD dataset strengthens CNNs and LSTMs to classify individuals with ASD from those without. Important facial landmarks, emotions, and other patterns that can help distinguish people with ASD are first identified by applying the CNN to extract essential spatial characteristics from face images. Several convolutional layers are used to train these features, followed by pooling layers to reduce spatial dimensions and retain important data. Following CNN processing, the retrieved features are transformed into sequences and fed via LSTM layers, which, in the case of sequences of images, capture temporal associations. The LSTM layers enable the model to understand dynamic patterns across time, which is necessary for identifying behavioral signals associated with ASD. This model allows for the automated classification of ASD based on facial features, combining spatial and temporal information to enhance accuracy in detecting the differences between the two groups.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025



(a)



(b)

Figure 5. Facial Feature analysis results of the proposed model. (a) ASD (b) non-ASD

The findings of the proposed model's face feature analysis are shown in Figure 5 under two different categories: (a) ASD and (b) non-ASD. In part (a), the facial features of individuals with ASD are analyzed, highlighting unique patterns or abnormalities in facial expressions, movements, or landmark positioning that the model has identified. These features reflect the social and behavioral characteristics associated with ASD, such as differences in facial expressions, gaze patterns, or facial symmetry. In part (b), the facial features of non-ASD individuals are shown, where the model captures more typical facial characteristics without the distinctive patterns observed in ASD.

The comparison between these two categories allows for a visual understanding of the model to differentiate between the facial traits of individuals with ASD and those without, based on the learned spatial and temporal features. This analysis provides insights into the potential biomarkers and visual cues to distinguish ASD from non-ASD conditions. To evaluate the performance of an autism detection model using Federated Learning, the following metrics are used.

4.1 Accuracy

The percentage of correctly categorized samples is known as accuracy.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

4.2 Precision

The capacity of the model to accurately identify every real positive case is known as recall.

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

4.3 Recall

The capacity of the model to accurately identify every real positive case is known as recall.

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

4.4 F1 Score

By calculating the harmonic mean of the two criteria, the F1 score strikes a compromise between recall and accuracy.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (17)$$

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

Table 2: Performance metrics of the proposed model.

Performance Metrics	Values (%)
Accuracy	98.90
Precision	97.80
Recall	98.05
F1-Score	98

As shown in Table 2, the proposed ASD model demonstrates strong classification performance across key metrics. High accuracy indicates that the proposed method correctly classifies nearly all cases. The performance of several ASD models utilizing various machine-learning techniques is compared in Table 3. With the accuracy (98.9%) and F1score (0.98), the proposed Federated Learning Model stands out as providing strong prediction capabilities while protecting data privacy across decentralized sources.

While competing models such as the Random Forest Classifier, CNN, and SVM demonstrate strong performance, their accuracy and F1 scores are inferior to those of the proposed model. Specifically, CNN excels in image feature extraction (96.2%), SVM is effective with smaller datasets (95.8%), and Random Forest delivers good interpretability (97.3%). The least accurate (94.7%) method for capturing temporal patterns in sequential data is RNN. Lastly, the Hybrid CNN-RNN Model combines the strengths of both architectures, achieving a balanced performance (96.8%). Overall, while the competitive models each have strengths in specific areas, the proposed Federated Learning approach provides superior performance, particularly in decentralized and privacy-sensitive applications.

As shown in Table 2, the proposed ASD model demonstrates strong classification performance across key metrics. High accuracy indicates that the proposed method correctly classifies nearly all cases. The performance of several ASD models utilizing various machine-learning techniques is compared in Table 3. With the accuracy (98.9%) and F1score (0.98), the proposed Federated Learning Model stands out as providing strong prediction capabilities while protecting data privacy across decentralized sources. While competing models such as the Random Forest Classifier, CNN, and SVM demonstrate strong performance, their accuracy and F1 scores are inferior to those of the proposed model. Specifically, CNN excels in image feature extraction (96.2%), SVM is effective with smaller datasets (95.8%), and Random Forest delivers good interpretability (97.3%). The least accurate (94.7%) method for capturing temporal patterns in sequential data is RNN. Lastly, the Hybrid CNN-RNN Model combines the strengths of both architectures, achieving a balanced performance (96.8%). Overall, while the competitive models each have strengths in specific areas, the proposed Federated Learning approach provides superior performance, particularly in decentralized and privacy-sensitive applications.

Table 3: The performance of several ASD models utilizing various machine-learning techniques

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN (Alsaidi et al., 2024)	96.20	95.50	94.70	95.00
SVM (Qiu et al., 2024)	95.80	94.20	95.10	95.00
Random Forest Classifier (Gelmez et al., 2024)	97.30	96.10	96.80	96.00
RNN (Jugunta et al., 2023)	94.70	93.40	94.00	94.00
Hybrid CNN-RNN (Qiu et al., 2024)	96.80	95.90	96.20	96.00
Gradient Boosting Machine (GBM) (Mittal et al., 2024)	96.50	96.00	95.70	96.00
Proposed Model (Federated Learning)	98.90	97.80	98.05	98.00

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

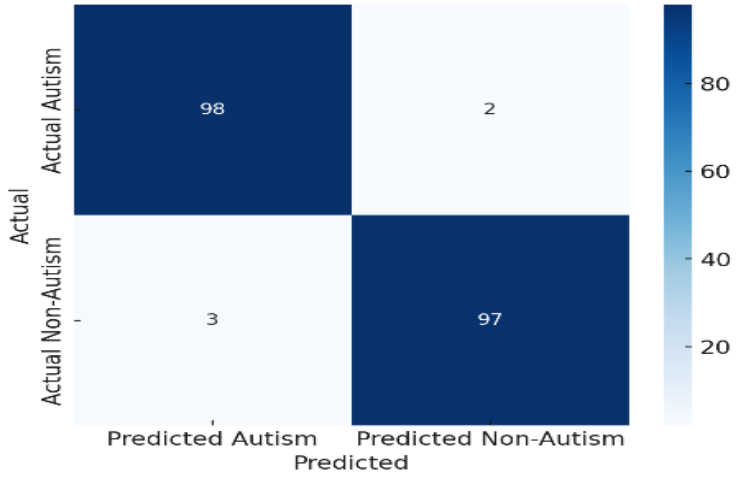


Figure 6. Confusion matrix of the proposed ASD model.

Figure 6 provides the information on the model's categorization performance. The matrix shows that the model correctly predicted 98 cases of autism and 97 cases of non-autism with just two false negatives and three false positives.

While recent methodologies of ASD detection have utilized various ML approaches, several research gaps and limitations persist. Convolutional Neural Networks (CNNs), as employed by Alsaidi et al., (2024), show strong feature extraction capabilities but struggle with interpretability, making clinical application challenging. Similarly, Support Vector Machines (SVMs) in Qiu et al., (2024) lack scalability with large, complex datasets. Random Forest models, used by Gelmez et al., (2024), demonstrate reliability with behavioural data but may be less effective for real-time analysis due to computational constraints. Jugunta et al., (2024) apply Recurrent Neural Networks (RNNs) for sequential behavioural data, yet RNNs often face issues with long-term dependency handling, which can affect the accuracy of extended behavioural sequences. Qiu et al., (2024) combine CNNs and RNNs, yet this hybrid model complicates real-world deployment. Finally, Mittal et al., (2024) utilize Gradient Boosting Machines that are prone to overfitting, particularly with limited training data. These challenges highlight the need for models that balance interpretability, scalability, efficiency, and generalizability for effective ASD detection in diverse clinical settings.

The proposed CNN-LSTM-based ASD model for children using Federated Learning marks a significant advancement in early ASD detection by addressing challenges in data privacy, model generalization, and comprehensive feature extraction. Traditional ASD diagnostic processes often suffer delays and rely on limited datasets, but the proposed model overcomes these barriers by leveraging federated learning, allowing multi-institutional collaboration without compromising data privacy. This federated approach enables each dataset to train the model locally, only sharing model updates with a central server for aggregation into a global model, thus ensuring data security while benefiting from the diversity of data sources. This model structure reduces biases commonly seen in centralized models trained on homogenous datasets, enhancing its applicability across varied populations. The combined CNN-LSTM architecture of the model proves instrumental in capturing both spatial and temporal features from the image dataset. CNN layers effectively extract spatial features including eye movements, gestures, and facial expressions, all of which are acknowledged markers of ASD. While the LSTM layers effectively, model the sequential patterns of behaviour over time, offering a detailed temporal perspective essential for ASD detection. This fusion of CNN and LSTM networks enables a comprehensive analysis that is capable of capturing the nuanced visual-spatial and time-dependent behavioural indicators associated with autism.

Moreover, the federated learning framework supports high model performance through the aggregation of diverse, institution-specific datasets, resulting in a model with improved generalization and robustness. Experimental results indicate that this federated CNN-LSTM model outperforms traditional centralized models. The improvement is attributed to federated learning's ability to minimize overfitting and reduce susceptibility to model biases, which are common when using datasets that lack variability. The Federated Learning-based ASD detection model could be integrated into real-world applications to facilitate early screening processes in clinical settings. The model's ability to operate on multimodal data makes it suitable for deployment in settings where behavioural data from children can be collected through sensors or devices, allowing for a non-invasive, continuous screening process that does not interfere with children's day-to-day activities. This flexibility also opens avenues for global deployment, where each region can contribute data without risking patient privacy, enabling a more equitable and accessible approach to ASD detection.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

5. Conclusion

The proposed CNN-LSTM Federated Learning model for autism detection demonstrates strong performance in identifying autism-related patterns across diverse biometric datasets, with each device independently training a model that respects data privacy. The model exhibited robust predictive performance across key metrics, achieving a high accuracy of 98.9%. Its precision (97.8%) and recall (98.05%) underscore its effectiveness in accurately identifying autism-related behaviours while minimizing false negatives, which is critical for enabling early intervention. A balanced trade-off between recall and precision is shown in an F1 score of 98%. These metrics collectively confirm that the CNN-LSTM architecture, supported by Federated Learning, is an effective and privacy-preserving approach for autism detection, adaptable to different data sources without compromising patient confidentiality.

References

[1] M. Alsaidi, N. Obeid, N. Al-Madi, H. Hiary, and I. Aljarah, "A Convolutional Deep Neural Network Approach to Predict Autism Spectrum Disorder Based on Eye-Tracking Scan Paths," *Information*, vol. 15, no. 3, p. 133, 2024, doi: 10.3390/info15030133.

[2] A. Amirbay et al., "Development of an algorithm for identifying the autism spectrum based on features using deep learning methods," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 5, pp. 2088–8708, 2024, doi: 10.11591/ijece.v14i5.pp5513-5523.

[3] Autism Spectrum Disorder Facial Image Dataset (ASD-FID), [Online]. Available: https://drive.google.com/drive/folders/1XQU0pluL0m3TIlXqntano12d68peMb8A?usp=sharing. [Accessed: 14-Nov-2024].

[4] Autistic Children Facial Image Data Set, [Online]. Available: https://www.kaggle.com/datasets/imrankhan77/autistic-children-facial-data-set. [Accessed: 14-Nov-2024].

[5] J. Baio, L. Wiggins, and D. L. Christensen, "Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014," *MMWR Surveill. Summ.*, vol. 67, no. SS-6, pp. 1–23, 2018, doi: 10.15585/mmwr.ss6706a1.

[6] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2015, pp. 1721–1730, doi: 10.1145/2783258.278861.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[8] S. Chen, M. Jiang, and Q. Zhao, "Deep Learning to Interpret Autism Spectrum Disorder Behind the Camera," *IEEE Trans. Cogn. Dev. Syst.*, 2024, doi: 10.1109/tcds.2024.3386656.

[9] P. Gelmez, T. E. Karakoc, and O. Ulucan, "Autism Spectrum Disorder and Atypical Brain Connectivity: Novel Insights from Brain Connectivity-Associated Genes by Combining Random Forest and Support Vector Machine Algorithm," *OMICS: J. Integr. Biol.*, 2024, doi: 10.1089/omi.2024.016.

[10] S. Gupta et al., "Enhancing Autism Spectrum Disorder Classification with Lightweight Quantized CNNs and Federated Learning on ABIDE-1 Dataset," *Mathematics*, vol. 12, no. 18, p. 2886, 2024, doi: 10.3390/math12182886.

[11] S. B. Jugunta et al., "Unleashing the Potential of Artificial Bee Colony Optimized RNN-Bi-LSTM for Autism Spectrum Disorder Diagnosis," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 11, 2023, doi: 10.14569/IJACSA.2023.0141173.

[12] K. Khan and R. Katarya, "WS-BiTM: Integrating White Shark Optimization with Bi-LSTM for Enhanced Autism Spectrum Disorder Diagnosis," *J. Neurosci. Methods*, vol. 110319, 2024, doi: 10.1016/j.jneumeth.2024.110319.

[13] C. Li et al., "Multimodal foundation models: From specialists to general-purpose assistants," *Found. Trends® Comput. Graph. Vis.*, vol. 16, no. 1-2, pp. 1–214, 2024, doi: 10.1561/0600000110.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025

[14] C. Lord et al., "The Autism Diagnostic Observation Schedule (ADOS): A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism," *J. Autism Dev. Disord.*, vol. 30, no. 3, pp. 205–223, 2020, doi: 10.1023/A: 1005379700087.

[15] K. Mittal, K. S. Gill, D. Upadhyay, S. Dangi, and G. R. Kumar, "Applying Machine Learning and the Gradient Boosting Classification Method for Evaluating the Probability of Autism," in *Proc. IEEE 9th Int. Conf. Converg. Technol. (I2CT)*, 2024, pp. 1–5, doi: 10.1109/I2CT61223.2024.10544362.

[16] L. Qiu and J. Zhai, "A hybrid CNN-SVM model for enhanced autism diagnosis," *PLoS One*, vol. 19, no. 5, p. e0302236, 2024, doi: 10.1371/journal.pone.0302236.

[17] R. A. Rasul et al., "An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder," *Healthc. Anal.*, vol. 5, p. 100293, 2024, doi: 10.1016/j.health.2023.100293.

[18] Stanford Children's Health Facial Imaging Dataset, [Online]. Available: https://aimidatasetindex.stanford.edu/datasets-home. [Accessed: 14-Nov-2024].

[19] A. Valizadeh et al., "Automated diagnosis of autism with artificial intelligence: State of the art," *Rev. Neurosci.*, vol. 35, no. 2, pp. 141–163, 2024, doi: 10.1515/revneuro-2023-0050.

[20] World Health Organization (WHO), "Autism Spectrum Disorders," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders. [Accessed: 2022].

[21] F. L. Zhu et al., "A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name," *Front. Psychiatry*, vol. 14, p. 1039293, 2023, doi: 10.3389/fpsyt.2023.1039293.

[22] N. K. Kumar et al., "Enhanced Alzheimer's disease Prediction through Advanced Imaging: A Study of Machine Learning and Deep Learning Approaches," in *Proc. 5th Int. Conf. Smart Electron. Commun. (ICOSEC)*, Trichy, India, 2024, pp. 1177–1182, doi: 10.1109/ICOSEC61587.2024.10722116.

DOI: https://doi.org/10.54216/FPA.190216
Received: December 09, 2024 Revised: February 11, 2025 Accepted: March 04, 2025