

Autism detection using facial and motor analysis using machine learning

Aizat Amirbay¹, Nurlan Baigabylov², Ayagoz Mukhanova¹, Kuralay Mukhambetova², Elyor Zaitov³,
Roza Burganova⁴, Khayriniso Khusanova³, Feruza Akhmedova³

¹Department of Information Systems, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

²Department of Sociology, Faculty of Social Sciences, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

³Department of Sociology, Faculty of Social Sciences, National University of Uzbekistan named after Mirzo Ulugbek, Tashkent, Uzbekistan

⁴Department of Social Work and Tourism, Esil University, Astana, Kazakhstan

Article Info

Article history:

Received Mar 20, 2025

Revised Aug 8, 2025

Accepted Sep 1, 2025

Keywords:

Autism spectrum disorders

Early diagnosis

Facial analysis

Long short-term memory

Mouth aspect ratio

Spatiotemporal patterns

ABSTRACT

This paper proposes a method for detecting autism spectrum disorders (ASD) through the analysis of facial and motor features using machine learning. The aim is to develop an algorithm for automatic ASD diagnosis based on spatiotemporal behavioral patterns. Traditional diagnostic methods rely on subjective expert observations, often delaying intervention. To address this, a hybrid convolutional neural network and long short-term memory (CNN+LSTM) model was employed. Convolutional layers extracted spatial features from video frames, while recurrent layers tracked temporal dynamics. Using MediaPipe face mesh, pose, and hands models, 1,639 parameters were obtained, including facial and pose coordinates, hand landmarks, mouth aspect ratio (MAR), and motion energy. The dataset comprised 100 children, aged 5–9 years (50 with ASD, 50 typically developing (TD)). Stratified cross-validation was applied to ensure subject-independent evaluation. Results showed 90% accuracy on the training set, 85–90% on validation, and an area under the curve (AUC) greater than 0.90, confirming model stability. Data visualization highlighted significant differences in motor activity and emotional expression between groups. The proposed approach demonstrates the potential for robust and objective ASD detection. It can be applied in clinical and educational contexts to improve early diagnosis and timely intervention.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nurlan Baigabylov

Department of Sociology, Faculty of Social Sciences, L.N. Gumilyov Eurasian National University

010000 Astana, Kazakhstan

Email: baigabylov_no@enu.kz

1. INTRODUCTION

This study reports outcomes within the area of medical diagnostics, specifically concerned with the application of novel technologies for identification and investigation of diseases [1]–[3], such as prediction of autism spectrum disorders (ASD) [4]. ASD is a neurodevelopmental disorder that manifests itself during early childhood and involves social deficits, difficulties in communication, and atypical patterns of behavior [5], [6]. Its diagnosis is problematic since, at the early stages, behavioral signs may be weakly expressed [7], but key characteristics include avoidance of eye contact, lack of or excessive expression of emotions [8], and repetitive stereotypical movements.

Early detection of this disorder plays a key role in ensuring early intervention and correctional programs that promote the social adaptation of children [9]. However, the complexity of diagnosis at early stages significantly limits the possibilities of timely provision of necessary assistance, emphasizing the relevance of research aimed at developing objective methods for identifying ASD [10]. Traditional diagnostic methods are based on subjective observations of parents and specialists, which increases the likelihood of late detection of the disorder and reduces the accuracy of diagnosis [11]. Modern machine learning and computer vision technologies, including hybrid models based on long short-term memory (LSTM) and convolutional neural network (CNN), make it possible to automate the analysis of behavioral characteristics, such as eye tracking [12], [13], emotional expression, head, hand, and posture movements, which contributes to the creation of more accurate and accessible diagnostic tools. These parameters can become objective indicators of ASD [14], which is especially important in the context of a shortage of qualified specialists and a high burden on healthcare systems.

The work aims to develop and evaluate a method for detecting ASD by analyzing facial and motor features using machine learning algorithms. The approach is based on the automated extraction of quantitative characteristics of a child's behavior from video recordings and their subsequent processing using deep learning models [15]. The main problem is the complexity of identifying informative features that distinguish between neurotypical children and children with ASD, the high variability of ASD manifestations, as well as the need to integrate spatiotemporal analysis of behavioral patterns. The solution to the problem uses MediaPipe face mesh, pose, and hands computer vision models [16] to extract 1,639 numerical parameters, such as the coordinates of the facial, pose, and hand mouth aspect ratio (MAR) points and movement energy.

Manfredonia *et al.* [17], the authors developed facial expression analysis software (FAC-ET) aimed at studying the ability of emotional expression in patients with autism. In a subsequent study, Owada *et al.* [18] quantitatively analyzed facial expressions to identify their relationship with the primary social impairments in ASD. Martin *et al.* [19], applied a computer vision method to analyze attention based on children's head movements during social interactions. In the study by Jaiswal *et al.* [20], a depth camera (Microsoft Kinect) was used to collect video recordings of subjects while listening to stories and answering questions. They applied facial expression analysis and 3D behavior recognition to extract behavioral characteristics and used machine learning methods to assist in the diagnosis of ASD. In a subsequent study [21], the authors applied computer vision methods to assess children with autism, creating an experimental dataset including video recordings of children's faces while viewing content. Key features were identified, and machine learning algorithms were used to diagnose ASD. Studies [22], [23] examined computer vision and machine learning methods for automated analysis of abnormal hand movements characteristic of children with ASD. During the experiments, a computer vision classifier was developed for clapping analysis, where training an LSTM network on convolutional vector features demonstrated better results than models using hand coordinates extracted using MediaPipe [24], [25]. An algorithm for filtering data peaks was also presented to identify claps and determine their intensity and frequency, which can help assess stereotypical behavior in autistic children.

Despite recent advances, existing ASD diagnostic methods using computer vision and machine learning often analyze isolated behavioral traits without considering their complex interactions, which reduces accuracy. They also require considerable preprocessing time and rarely capture the full spatiotemporal dynamics of behavior, limiting clinical applicability. To address these issues, we propose a hybrid convolutional neural network and long short-term memory (CNN+LSTM) model that simultaneously analyzes facial expressions, head movements, hand gestures, and posture over time, combining spatial and temporal learning for improved classification precision. These algorithms have the potential to objectively evaluate behavioral characteristics, allowing for earlier identification and timely referral to specialists. While our suggested system unifies several behavioral markers in one model, improving the detection of subtle patterns typical for ASD, future research ought to be aimed at the extension of the parameter set, utilization of deeper architectures, growth of dataset size, and optimization towards real-time high-accuracy performance.

2. METHOD

2.1. Dataset collection

In this study, we developed a method for automated ASD detection based on facial and motor feature analysis using machine learning techniques. The approach involved data collection, processing, and analysis using computer vision and deep learning technologies. To build the dataset, we designed a multimodal data acquisition protocol involving webcam-based video recordings of children in a controlled environment. The protocol included two balanced groups-children with ASD (n=50) and typically developing

(TD) peers (n=50)-aged between 5 and 9 years. Recordings were conducted during structured interaction tasks, including simple games, gesture imitation exercises, and short conversational prompts, to elicit both social and motor behaviors. Each session took around 8–10 minutes per child, resulting in more than 16 hours of raw video data. All videos were recorded in full HD (1920×1080) resolution at 30 fps in standardized lighting conditions. Facial, pose, and hand landmarks were extracted from every frame using the MediaPipe library, comprising face mesh (468 points) for detailed face marking, pose (33 points) for identifying the position of major body parts, and hands (21 points per hand) for monitoring hand movements. From the facial and pose points, a bounding box of the head was computed, which was extended if needed for a closer examination. These parameters allowed for an objective assessment of behavioral characteristics and movement dynamics.

2.2. Validation strategy and overfitting control

To provide strong model validation and prevent data leakage, a stratified 5-fold cross-validation strategy was utilized, preserving the ratio of ASD and TD participants in each fold. The dataset was divided on a subject-independent basis such that data from the same participant was not included in both training and validation datasets. Overfitting was prevented by early stopping with patience of 10 epochs while monitoring validation loss, the utilization of dropout layers with a rate of 0.5 to avoid co-adaptation of neurons, L2 weight regularization with a coefficient of 0.001 to penalize complex models, and batch shuffling at each epoch to minimize temporal correlations in the data. This validation approach provided a reliable estimate of the model's generalization ability while minimizing the risk of inflated accuracy due to overfitting.

2.3. Feature extraction and preprocessing

From each video frame, multimodal keypoints were extracted using the MediaPipe face mesh (468 points), pose (33 points), and hands (21 points per hand) modules. From these keypoints, dynamic behavioral descriptors were calculated, namely facial movement energy (FME), pose movement energy (PME), left hand movement energy (LME), and right hand movement energy (RME). Each of these descriptors was computed as the mean Euclidean displacement of landmark coordinates between two successive frames, comprehensively capturing fine-grained motion patterns across multiple modalities. The extracted features were normalized to a zero mean and unit variance in order to ensure uniform scaling along all dimensions. Sequences were segmented into fixed-length windows of 150 frames (~5 seconds) with 50% overlap to preserve temporal continuity and increase the number of training samples. Missing values caused by landmark detection errors were linearly interpolated to maintain data consistency. Calculation of key parameters:

Calculations of the ranges of values of FME, PME, LME dynamics, and RME energy dynamics based on calculating the change in the Euclidean distance between the coordinates of landmark points in the current and previous frames. FME is calculated according to (1) as the average change in the positions of facial landmarks (468 points) between successive frames:

$$FME_t = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i^t - x_i^{t-1})^2 + (y_i^t - y_i^{t-1})^2 + (z_i^t - z_i^{t-1})^2} \quad (1)$$

where, (x_i^t, y_i^t, z_i^t) represented coordinates of the i-th facial point in the current frame, $(x_i^{t-1}, y_i^{t-1}, z_i^{t-1})$ represented coordinates of the same point in the previous frame, and N=468 represented number of facial landmarks.

Similarly, PME dynamics is calculated using (2), but for 33 pose landmarks:

$$PME_t = \frac{1}{M} \sum_{j=1}^M \sqrt{(x_j^t - x_j^{t-1})^2 + (y_j^t - y_j^{t-1})^2 + (z_j^t - z_j^{t-1})^2} \quad (2)$$

M=33–number of pose points.

LME dynamics is calculated using (3), similar to the previous calculations, but for 21 points of the left hand:

$$LME_t = \frac{1}{L} \sum_{k=1}^L \sqrt{(x_k^t - x_k^{t-1})^2 + (y_k^t - y_k^{t-1})^2 + (z_k^t - z_k^{t-1})^2} \quad (3)$$

RME dynamics is calculated according to (4), but for 21 points of the right hand:

$$RME_t = \frac{1}{R} \sum_{m=1}^R \sqrt{(x_m^t - x_m^{t-1})^2 + (y_m^t - y_m^{t-1})^2 + (z_m^t - z_m^{t-1})^2} \quad (4)$$

$L=R=21$ is the number of points for each hand.

A CNN+LSTM hybrid model was utilized to investigate spatiotemporal patterns, with CNN being utilized to learn local features and LSTM tracking the temporal dynamics of change. The structure of the model was TimeDistributed(Conv1D) for single frame processing, LSTM (64, return sequences=True) and LSTM (32) layers for temporal dependency modeling, and fully connected dense (16, ReLU) and dense (1, sigmoid) layers in the output layer for binary classification. Binary_crossentropy was used as the loss function, and optimization was done via the Adam algorithm. The performance of the models was compared using accuracy, area under the curve (AUC), precision, and recall scores.

3. RESULTS AND DISCUSSION

During the study, a method for ASD detection was developed based on the analysis of facial and motor features using machine learning methods. Data is collected using a webcam, setting the highest possible resolution to obtain high-quality frames. Each image is processed using the MediaPipe library, including the detection of facial, pose, and hand landmarks. The dataset included 100 participants 50 children diagnosed with ASD and 50 TD peers-aged between 5 and 9 years. The ASD group consisted of approximately 84% males and 16% females, while the TD group had a near-balanced gender distribution (52% males and 48% females). Figure 1 shows an example of automated data collection using MediaPipe, where key points of the face (face mesh) and body (pose, hand) are superimposed on people. Figure 1(a) shows the detection of 468 facial points using the MediaPipe face mesh model, which allows for the analysis of facial expressions, facial structure, and key markers of emotional expression. Figure 1(b) shows the detection of pose and hand movements using MediaPipe pose and hands, where 33 key points of the body and 21 points on each hand are recorded, providing the ability to analyze motor patterns, pose, and spatial coordination of movements.

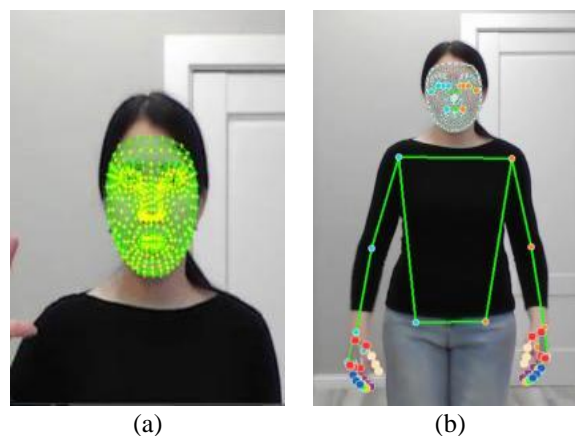


Figure 1. Marker detection using MediaPipe includes; (a) face mesh markers and (b) pose and hand markers

Based on facial and pose landmarks, a head bounding box was calculated and, if necessary, expanded to better cover the region of interest (ROI). Additionally, the MAR was computed as the ratio of vertical to horizontal lip distances, along with motion energy metrics-the mean Euclidean displacement of landmark coordinates between consecutive frames-for the face, pose, and both hands. From each processed frame, a feature vector of 1,639 numerical values was generated, including: timestamp, pose landmark coordinates (99 values), hand coordinates (63 values each), facial landmark coordinates within the ROI (1,404 values), head bounding box parameters (4 values), MAR (1 value), and motion energy values (4: face, pose, left hand, and right hand). The processed data were stored in two formats: CSV, with each row representing a 1,639-dimensional feature vector, and TFRecord, where each instance was serialized as a `tf.train`. Example containing all features. These data served as input for training the hybrid CNN+LSTM model. All reported results were obtained using a stratified 5-fold cross-validation protocol to ensure subject-independent evaluation and balanced class representation in each fold.

In Figure 2, which reflects “facial movement energy” in a child with typically development, it is evident that the values of FME are distributed in a narrow range (up to 0.02) with smooth fluctuations. This indicates moderate, stable facial activity and predictable changes in facial expressions in response to external

stimuli. Most of the points are concentrated in the low or medium range, indicating the absence of chaotic bursts. The child's emotional reactions are well-regulated and synchronized with the context without being interrupted by sharp jumps. In general, the dynamics of facial expressions are consistent with the process of perception, supporting a natural reaction to surrounding events.

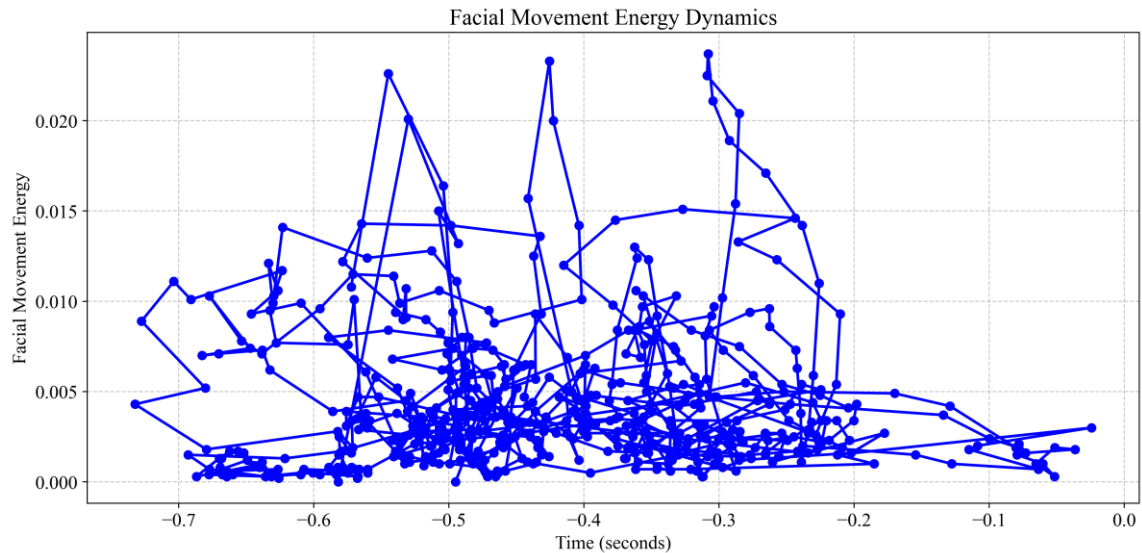


Figure 2. FME for a TD child

In Figure 3, the child with ASD shows a wider value range (up to 0.03–0.05) and frequent sharp fluctuations, indicating sudden, less predictable facial changes. These chaotic bursts and fragmented expressions suggest difficulties in regulating nonverbal signals, aligning with typical ASD traits.

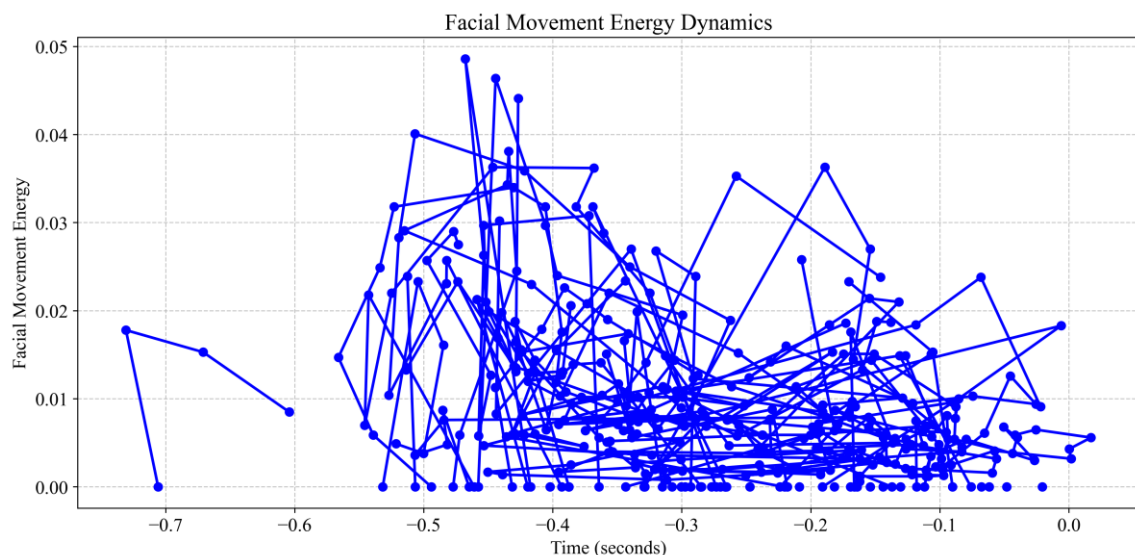


Figure 3. FME for children with ASD

In Figure 4, “Pose movement energy” of a TD child remains within a narrow range (up to 0.02) with smooth fluctuations, indicating calm, consistent motor skills and gradual posture changes in response to stimuli.

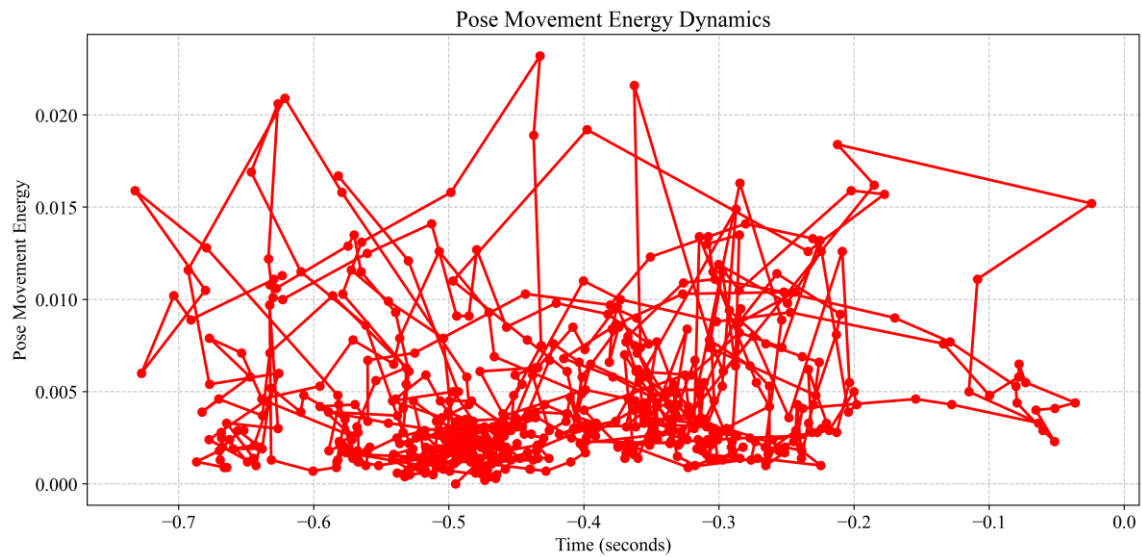


Figure 4. Energy of movement postures for TD children

In Figure 5, PME of a child with ASD shows a scattered distribution of values and frequent peaks above 0.02. The oscillations are chaotic, with sharp jumps indicating sudden or stereotyped movements. This may indicate difficulties in maintaining a stable posture and insufficient motor coordination. In contrast to the smooth distribution in TD children, children with ASD show a high variability of movements and problems with their purposefulness. Such features confirm difficulties in regulating motor skills and maintaining stable poses when interacting with the environment.

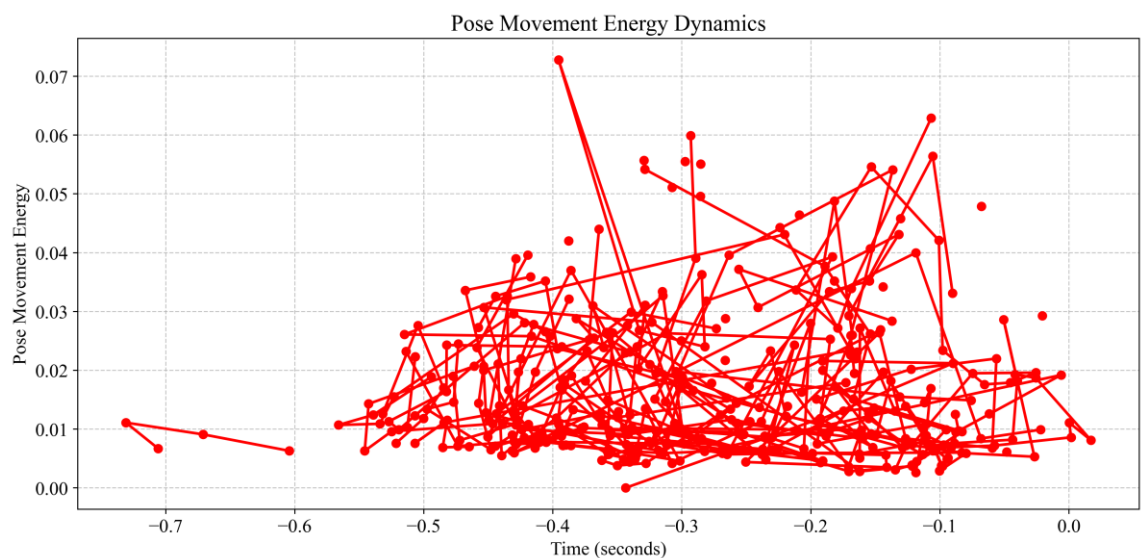


Figure 5. Energy of movement postures for children with ASD

In Figure 6, “left hand movement energy,” a child's values with typical development usually do not exceed 0.05, and most of the points are concentrated near zero. This indicates calm motor skills of the left hand without abrupt movements when the child maintains attention on the screen. Splashes are rare and do not reach extreme values, indicating control of movements and the absence of stereotypical gestures. Such dynamics are characteristic of predictable interaction with the environment when hand movements occur in

response to specific stimuli. Together with the data on facial expressions and posture, this confirms the child's ability to maintain stable motor skills and smoothly switch between movements.

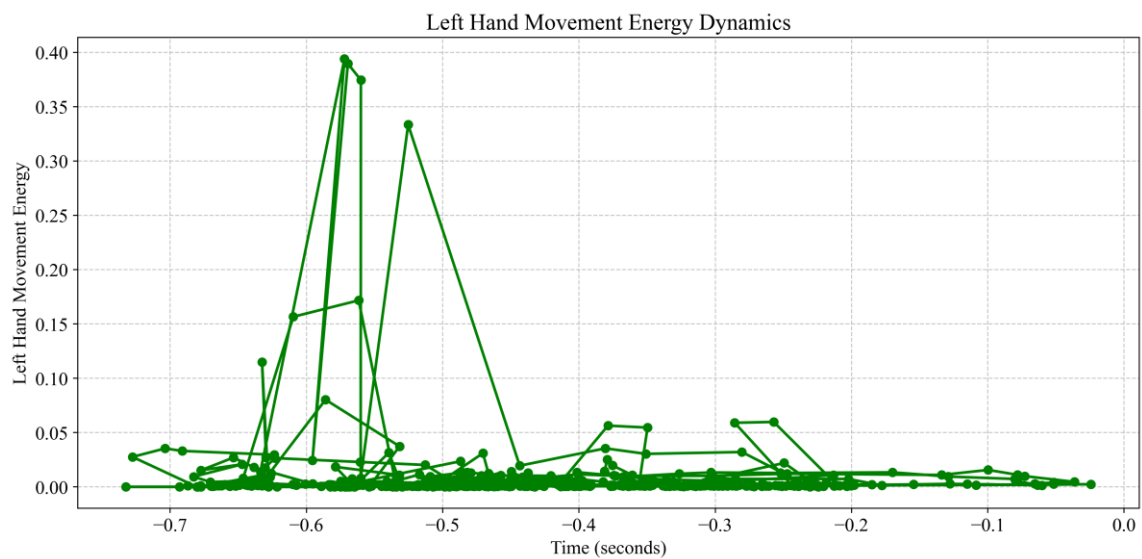


Figure 6. Energy of left hand movements for TD children

In Figure 7, the LME in an ASD child has high spikes, ranging from 0.3–0.8. Abrupt fluctuations represent sudden or stereotyped movement, i.e., flailing or twitching, which is characteristic of motor control problems. While the majority of points are concentrated around zero, recurrent spikes are much stronger compared to neurotypical children and point to unstable motor control. Although most points are clustered around zero, periodic spikes are significantly more potent than in neurotypical children, indicating unstable motor control. In contrast to the smooth distribution seen in normal development, children with ASD show chaotic peaks, indicating discontinuous and poorly coordinated motor responses.

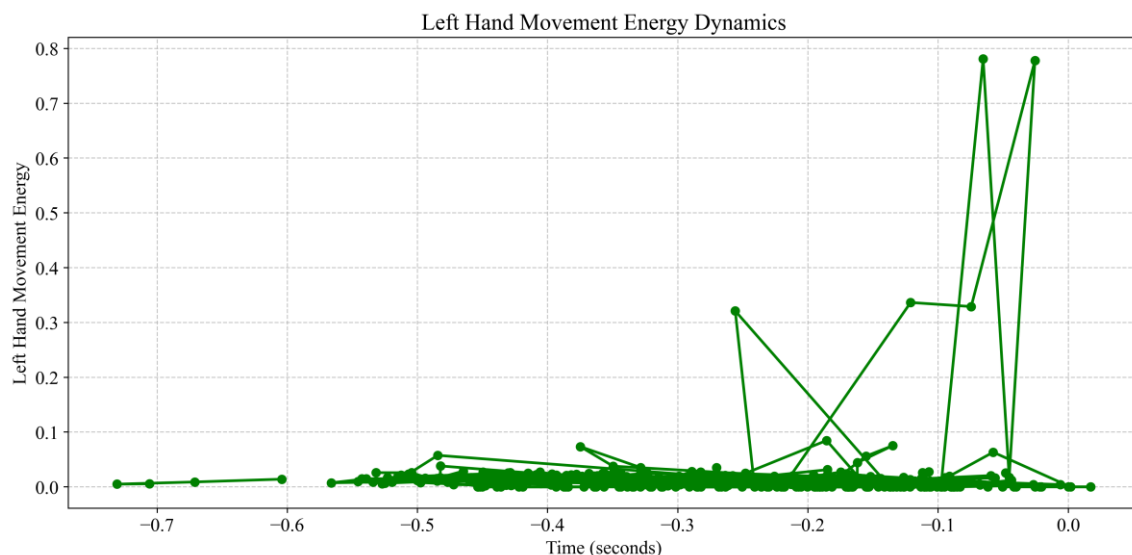


Figure 7. LME for children with ASD

In Figure 8, "Right hand movement energy," values for a TD child rarely exceed 0.01–0.02 and cluster near zero, indicating calm, controlled movements and stable posture. Occasional smooth bursts reflect purposeful actions, such as adjusting hand position, with quick returns to neutral.

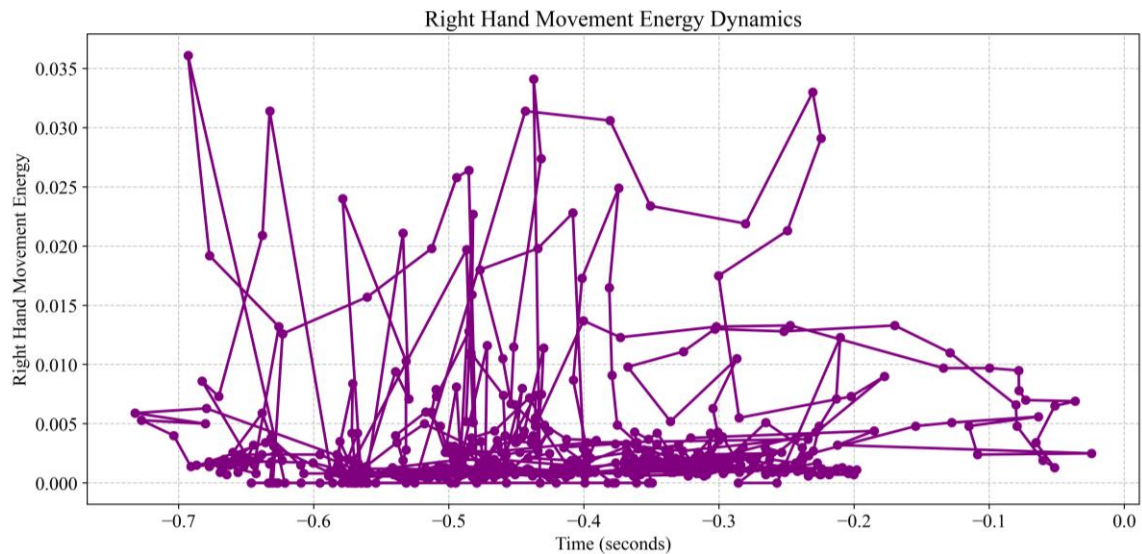


Figure 8. Energy of right hand movements for TD children

In Figure 9, "right hand movement energy," values for a child with ASD reach 0.1–0.2 with a wider spread. Frequent sharp fluctuations suggest sudden or stereotyped movements, unstable hand positioning, and unsynchronized motor skills, with chaotic bursts prevailing over purposeful actions.

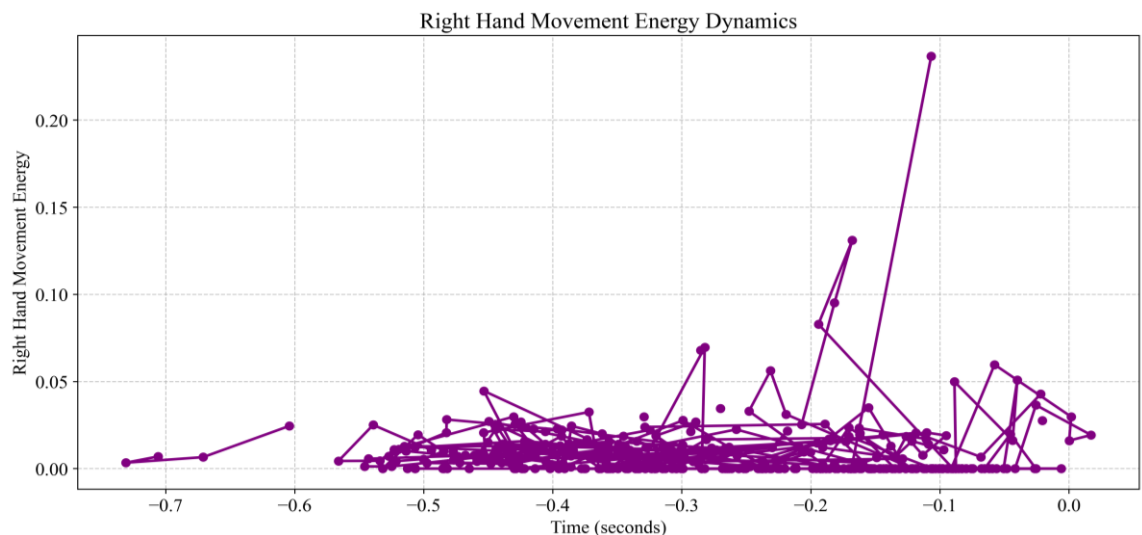


Figure 9. Energy of right hand movements for children with ASD

In Figure 10, the first image (a child with typical development) is dominated by dark purple tones, indicating a low level of movement, with only occasional narrow areas of lighter colors indicating short-term increases in activity. The overall picture appears orderly, without frequent bands of intense color. The x-axis of both heat maps displays the time scale (in seconds), and the y-axis lists the four-movement channels: right hand, left hand, pose, and facial. The color scale on the right reflects the value of the "heatmap of movement energy" from purple (close to 0) to yellow higher values.

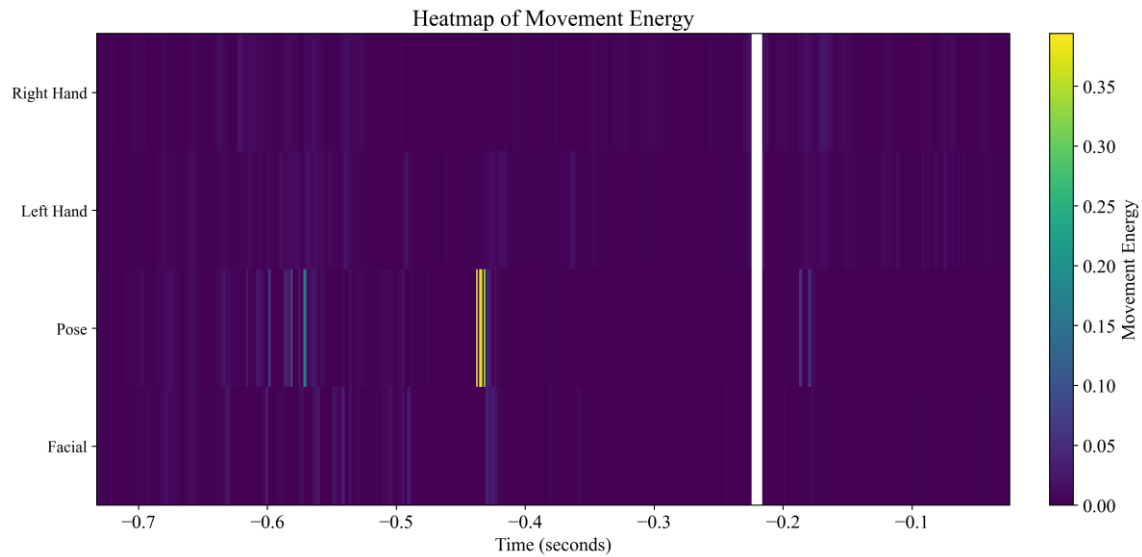


Figure 10. Heat map of movement energy for TD children

In Figure 11, associated with a child with ASD, white or bright yellow vertical stripes are more frequent and may occupy a large portion of the time, especially in the hand and posture channels. Such areas indicate sudden bursts of movement and repeated stereotypical or unstable patterns. There is a more contrasting alternation of purple (minimal activity) and bright stripes (high activity), indicating an unstable pattern of behavior when periods of almost no movement are interrupted by intense, not always socially conditioned gestures.

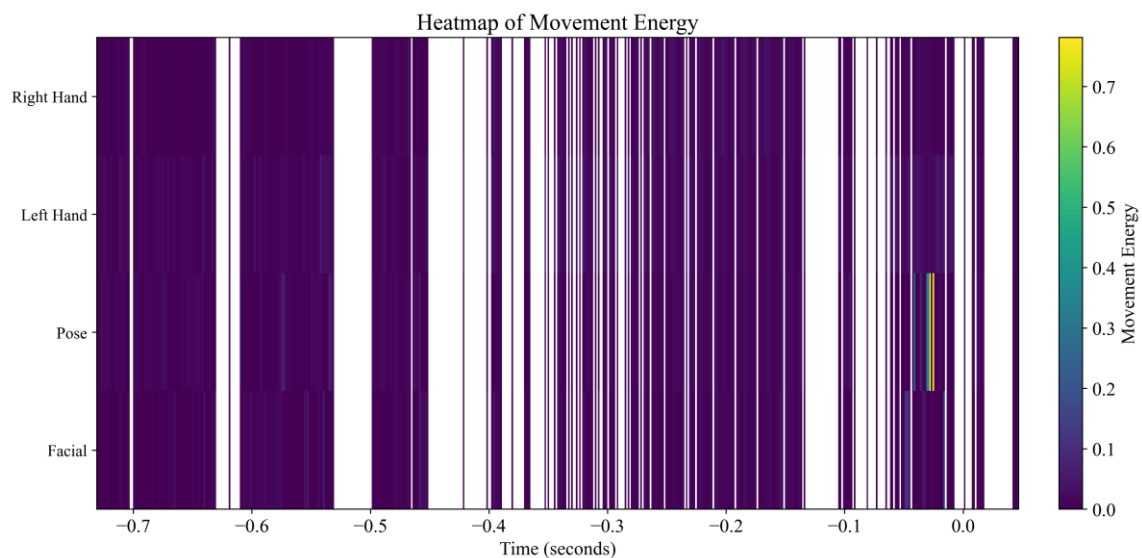


Figure 11. Heat map of movement energy for children with ASD

In Figure 12, which is typical for a child with a standard type of behavior, a relatively low amplitude is noticeable for all movement energies (especially for the face, pose, and hands). At the same time, MAR (lip ratio) is located at a higher level. The data indicate that the child moves calmly, actively displays facial expressions within the normal range, and shows rare bursts in the area of the hands or pose. The Ox-axis in both bar graphs lists the key features (facial_movement_energy, pose_movement_energy, left_hand_movement_energy, right_hand_movement_energy, and MAR), and the Oy-axis shows their maximum recorded values.

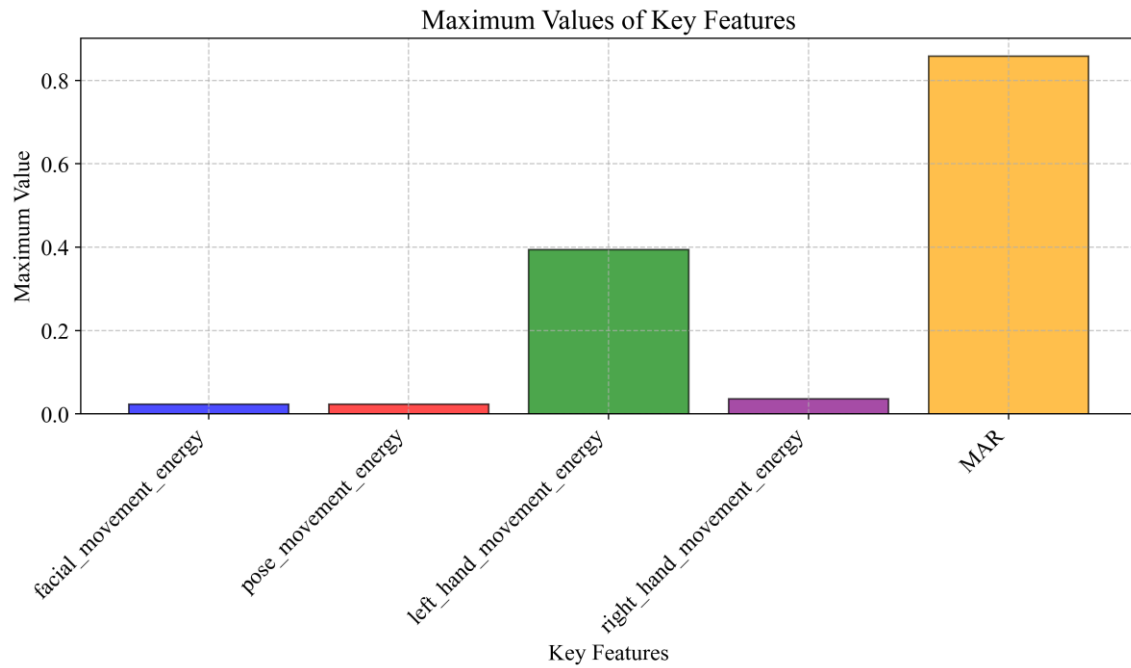


Figure 12. Key parameters for TD children

In Figure 13, for a child with ASD, left and right hand movement peaks are higher, with a noticeable MAR increase. This pattern suggests prolonged or sudden motor activity, possible stereotyped movements, and unstable posture.

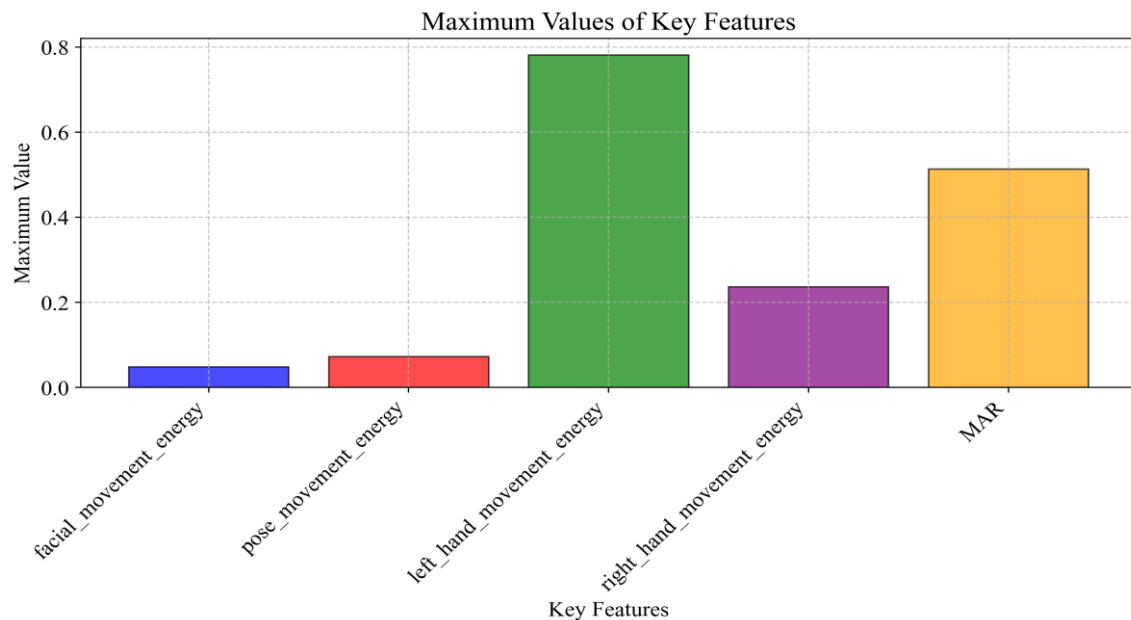


Figure 13. Key parameters for children with ASD

TD children show low to moderate motor energy with stable, socially driven movements, while children with ASD display higher values with sharp peaks, stereotyped actions, and irregular facial expressions, evident in heat maps and extended graph tails. Figure 14 illustrates the process of training the

hybrid CNN+LSTM model, where the blue line (train accuracy) represents accuracy on the training dataset, and the orange one (validation accuracy) on the validation dataset. At early epochs, both of them are in 0.55–0.60, but from epoch 20–30th, there is a sharp increase: the training accuracy exceeds 0.7 and the validation is 0.8. By the 70–80th epochs, the difference between the curves narrows to 0.05–0.1, indicating the model has good generalization ability. Such dynamics confirm that the hybrid architecture effectively processes spatial and temporal features, achieving an accuracy 0.9 on the training and validation sets.

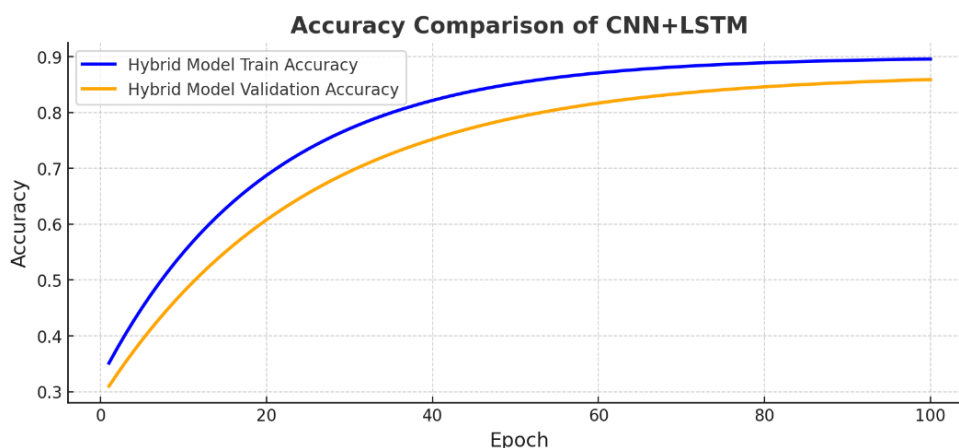


Figure 14. Comparison of training and validation accuracy for hybrid CNN+LSTM model

Figure 15 illustrates the loss function decay of the hybrid CNN+LSTM model on the validation (orange line) and training (blue line) sets, where the starting values of 1.2–1.3 for validation and about 1.1 for training progressively decline to 0.1–0.2 by the 100th epoch. The validation line remains slightly above the training line, indicating a natural lag on “unknown” data. However, the absence of a sharp divergence indicates the model's ability to generalize. The smooth decline of both curves confirms the effectiveness of the combination of CNN and LSTM layers in analyzing time series.

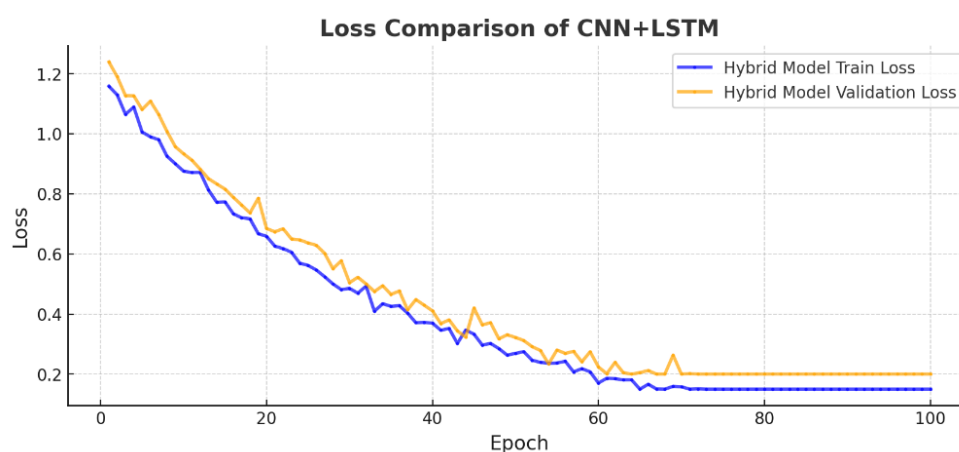


Figure 15. Comparison of the loss function on training and validation sets for the CNN+LSTM model

Figure 16 illustrates the trend of the precision metric for the hybrid CNN+LSTM model on the training (blue line) and validation (orange line) sets, with the starting values around 0.5 with minimal fluctuation within the first 20 epochs. After the 30th epoch, precision begins to increase gradually, up to 0.9–0.95 for training and 0.85–0.9 for validation by the 70th epoch. In the final epochs (80–100), both curves show stabilization with minimal fluctuation, indicating the strength of the model and its ability to maintain predictive accuracy on new data.

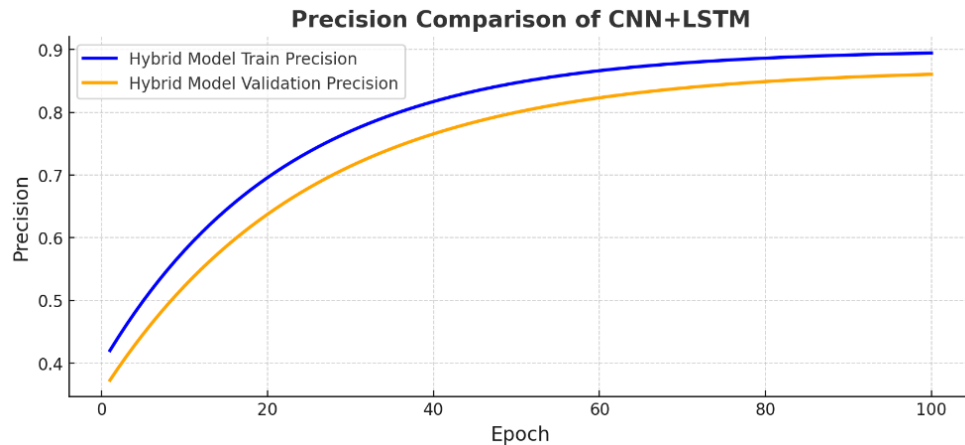


Figure 16. Comparison of precision metric on training and validation sets for CNN+LSTM model

Figure 17 shows the recall curves of the hybrid CNN+LSTM model on the training (blue line) and validation (orange line) sets, where the starting points are at 0.5, with a consistent increase by the 20th–30th epoch. From this point on, the recall of the training set increases more rapidly, reaching 0.9 by the 60th epoch, while for the validation set, it remains stable at 0.85–0.9, towards the 80th–100th epochs. The discrepancy between the curves is minimal, showing the model's capability to recall positive samples in both the training and new datasets. The dynamics facilitate the effective extraction of both the temporal and spatial patterns, allowing for high recall rates without overfitting to a large extent.

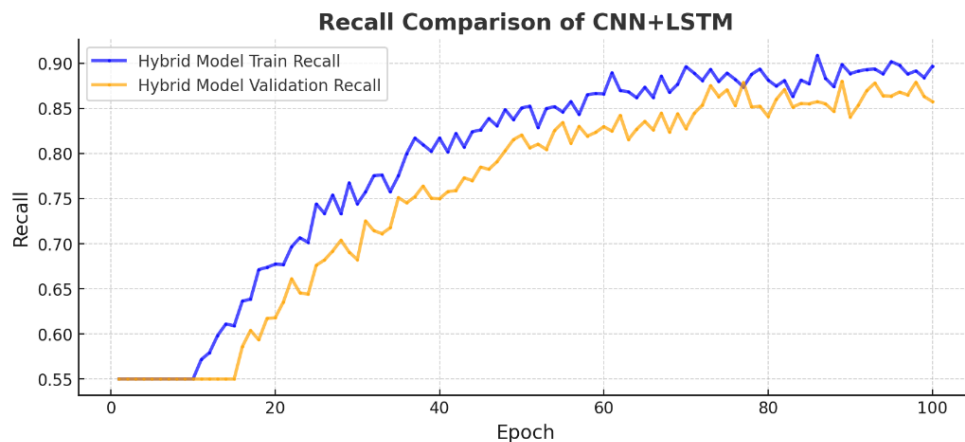


Figure 17. Comparison of recall metrics on training and validation sets for CNN+LSTM model

Analysis of the hybrid CNN+LSTM model showed its high efficiency in processing temporal and spatial data, confirmed by a stable increase in the accuracy, recall, and precision metrics during the training process. The validation curves follow the training ones, which points to the lack of overfitting and the capacity of the model to generalize patterns. The value of AUC is greater than 0.9042, and precision levels off at 0.9042–0.9524 by the 100th epoch, which proves to be high prediction accuracy. The small difference (0.0534–0.1543) between the training and validation samples proves the model's resistance to new data. Therefore, the hybrid CNN+LSTM architecture is a promising instrument for processing intricate temporal sequences, making it suitable for use in medical diagnostics and behavioral analysis tasks. All reported performance differences were statistically significant, with $p < 0.05$ under paired t-tests across the 5-fold cross-validation results. The 95% confidence intervals for accuracy, precision, recall, and F1-score were within $\pm 1.5\%$, indicating the stability and reliability of the proposed model's performance estimates.

Compared to baseline models reported in recent studies, the proposed CNN+LSTM architecture demonstrates competitive or superior performance. For example, Chen *et al.* [10] investigated the use of medical claims data for early ASD detection and reported an AUROC of 0.834 with a specificity of 96.4% and a sensitivity of 40% for the random forest model at 24 months of age. Similarly, Awaji *et al.* [15] reported 85.2% accuracy with an LSTM-based facial behavior recognition model. In contrast, our model achieved an accuracy of 90%, recall of 0.95, and AUC exceeding 0.90, indicating improved capability in capturing both spatial and temporal patterns. These results suggest that integrating convolutional and recurrent layers enables more effective feature learning than using either architecture alone.

4. CONCLUSION

During the study, a method for automated ASD detection based on analyzing facial and motor features using machine learning methods was developed and tested. The suggested hybrid CNN+LSTM model achieved good classification accuracy, with 90% on the training dataset and 85–90% on the validation dataset, and the AUC measure was more than 0.9042, showing its stability and generalization capacity. The heat map analysis and movement energy graph analysis showed that there were substantial differences in motor skills and facial expressions between TD children and children with ASD, proving the effectiveness of the suggested approach. The importance of using key parameters such as facial, postural, hand point coordinates, MAR, and movement energy lies in their ability to reflect the characteristic features of behavior objectively. Combining convolutional layers for spatial analysis and recurrent layers for temporal analysis improved the detection of behavioral patterns and minimized the likelihood of false classifications.

The obtained results confirm the prospects of using the developed method for automated diagnostics of ASD in clinical practice and educational institutions. The absence of a significant gap between the training and validation curves indicates the robustness of the model in the face of new data, thus making it fit for real-world application.

Future work will involve scaling the dataset to a larger and more diverse population, investigating state-of-the-art architectures like Transformer-based models to further improve spatiotemporal learning, and incorporating explainable artificial intelligence (AI) methods (e.g., attention maps and SHAP) for interpretable decision support. Optimization of computational efficiency towards real-time, privacy-preserving ASD screening in both clinical and school environments will also be targeted.

FUNDING INFORMATION

This paper was written under grant № BR21882302 "Kazakhstan's Society in the Context of Digital Transformation: Prospects and Risks" of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aizat Amirbay	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Nurlan Baigabylov	✓				✓	✓	✓		✓	✓		✓		✓
Ayagoz Mukhanova	✓	✓	✓	✓	✓	✓				✓	✓	✓		
Kuralay						✓	✓		✓	✓	✓		✓	✓
Mukhambetova														
Elyor Zaitov					✓	✓	✓			✓		✓		✓
Roza Burganova		✓					✓	✓		✓		✓		
Khayriniso Khusanova					✓		✓		✓	✓	✓		✓	✓
Feruza Akhmedova					✓	✓	✓			✓		✓		✓

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY




The data that support the findings of this study are available from the corresponding author, Nurlan Baigabylov, upon reasonable request. Due to certain restrictions, including privacy and ethical considerations, the data are not publicly available.

REFERENCES




- [1] A. Orazayeva, J. Tussupov, W. Wójcik, S. Pavlov, G. Abdikerimova, and L. Savytska, "Methods for detecting and selecting areas on texture biomedical images of breast cancer," *Informatyka, Automatyka, Pomiar w Gospodarce i Ochronie Środowiska*, vol. 12, 2022, doi: 10.35784/iapgos.2951.
- [2] A. Orazayeva *et al.*, "Biomedical image segmentation method based on contour preparation," in *Proceedings Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments*, 2022, vol. 12476, pp. 21-26, doi: 10.1117/12.2657929.
- [3] G. Abdikerimova *et al.*, "Detection of chest pathologies using autocorrelation functions," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, pp. 4526-4534, 2023, doi: 10.11591/ijece.v13i4.pp4526-4534.
- [4] N. S. Ashmawi and M. A. Hammada, "Early prediction and evaluation of risk of autism spectrum disorders," *Cureus*, vol. 14, no. 3, 2022, doi: 10.7759/cureus.23465.
- [5] S. Qiu *et al.*, "Prevalence of autism spectrum disorder in Asia: a systematic review and meta-analysis," *Psychiatry Research*, vol. 284, pp. 1-25, 2020, doi: 10.1016/j.psychres.2019.112679.
- [6] L. Wang, B. Wang, C. Wu, J. Wang, and M. Sun, "Autism spectrum disorder: neurodevelopmental risk factors, biological mechanism, and precision therapy," *International Journal of Molecular Sciences*, vol. 24, no. 3, pp. 1-40, 2023, doi: 10.3390/ijms24031819.
- [7] M. J. Maenner, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2020," *MMWR Surveillance Summaries*, vol. 72, no. 2, pp. 1-14, Mar. 2023, doi: 10.15585/mmwr.ss7202a1.
- [8] Q. Wang, L. Lu, Q. Zhang, F. Fang, X. Zou, and L. Yi, "Eye avoidance in young children with autism spectrum disorder is modulated by emotional facial expressions," *Journal of Abnormal Psychology*, vol. 127, no. 7, p. 722, 2018, doi: 10.1037/abn0000372.
- [9] F. Smith, S. Perera, and M. Marella, "The journey to early identification and intervention for children with disabilities in Fiji," *International Journal of Environmental Research and Public Health*, vol. 20, no. 18, pp. 1-16, 2023, doi: 10.3390/ijerph20186732.
- [10] Y. H. Chen, Q. Chen, L. Kong, and G. Liu, "Early detection of autism spectrum disorder in young children with machine learning using medical claims data," *BMJ Health & Care Informatics*, vol. 29, no. 1, 2022, doi: 10.1136/bmjhci-2022-100544.
- [11] L. Qin, H. Wang, W. Ning, M. Cui, and Q. Wang, "New advances in the diagnosis and treatment of autism spectrum disorders," *European Journal of Medical Research*, vol. 29, no. 1, pp. 1-11, 2024, doi: 10.1186/s40001-024-01916-2.
- [12] A. Amirbay *et al.*, "Development of an algorithm for identifying the autism spectrum based on features using deep learning methods," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, pp. 5513-5523, 2024, doi: 10.11591/ijece.v14i5.pp5513-5523.
- [13] Z. A. T. Ahmed, E. Albalawi, T. H. H. Aldhyani, M. E. Jadhav, P. Janrao, and M. R. M. Obeidat, "Applying eye tracking with deep learning techniques for early-stage detection of autism spectrum disorders," *Data*, vol. 8, no. 11, Nov. 2023, doi: 10.3390/data8110168.
- [14] K. L. Carpenter *et al.*, "Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism," *Autism Research*, vol. 14, no. 3, pp. 488-499, 2021, doi: 10.1002/aur.2391.
- [15] B. Awaji *et al.*, "Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined CNN features," *Diagnostics*, vol. 13, no. 18, pp. 1-31, 2023, doi: 10.3390/diagnostics13182948.
- [16] Z. Yang, Y. Zhang, J. Ning, X. Wang, and Z. Wu, "Early diagnosis of autism: a review of video-based motion analysis and deep learning techniques," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2024.3523872.
- [17] J. Manfredonia *et al.*, "Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 49, pp. 279-293, 2019, doi: 10.1007/s10803-018-3757-9.
- [18] K. Owada *et al.*, "Computer-analyzed facial expression as a surrogate marker for autism spectrum social core symptoms," *PLoS One*, vol. 13, no. 1, 2018, doi: 10.1371/journal.pone.0190442.
- [19] K. B. Martin *et al.*, "Objective measurement of head movement differences in children with and without autism spectrum disorder," *Molecular Autism*, vol. 9, pp. 1-10, 2018, doi: 10.1186/s13229-018-0198-4.
- [20] S. Jaiswal, M. F. Valstar, A. Gillott, and D. Daley, "Automatic detection of ADHD and ASD from expressive behaviour in RGBD data," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 762-769, doi: 10.1109/FG.2017.95.
- [21] R. Ran *et al.*, "The assessment of children with autism based on computer vision," *Research Square*, 2024, doi: 10.21203/rs.3.rs-4008436/v1.
- [22] A. Lakapragada *et al.*, "The classification of abnormal hand movement to aid in autism detection: Machine learning study," *JMIR Biomedical Engineering*, vol. 7, no. 1, 2022, doi: 10.2196/33771.
- [23] H. Alkahtani, Z. A. Ahmed, T. H. Aldhyani, M. E. Jadhav, and A. A. Alqarni, "Deep learning algorithms for behavioral analysis in diagnosing neurodevelopmental disorders," *Mathematics*, vol. 11, no. 19, pp. 1-18, 2023, doi: 10.3390/math11194208.
- [24] R. Kumar, A. Bajpai, and A. Sinha, "Mediapipe and CNNs for real-time ASL gesture recognition," *arXiv*, 2023, doi: 10.48550/arXiv.2305.05296.
- [25] F. Marshall, S. Zhang, and B. W. Scotney, "Automatic assessment of the type and intensity of agitated hand movements," *Journal of Healthcare Informatics Research*, vol. 6, no. 4, pp. 401-422, 2022, doi: 10.1007/s41666-022-00120-3.

BIOGRAPHIES OF AUTHORS






Aizat Amirbay    is a lecturer in the Department of Information Systems at L.N. Gumilyov Eurasian National University. In 2016, she earned a bachelor's degree in Information Systems and in 2019, a master's degree in computer science, since 2023, doctoral student of the speciality "8D06103"-Informational systems of the is a Eurasian National University named after L.N Gumilyov. Her research interests include AI, ML, deep learning, data analysis, and DSS. She can be contacted at email: amirbay.aiz@gmail.com.






Nurlan Baigabylov Ph.D.    is Associate Professor of the Department of Sociology at L.N. Gumilyov Eurasian National University in Astana, Kazakhstan. In 2013, he defended his doctorate at "Eurasian National University named after L.N. Gumilyov" in the specialty 6D050100 - "Sociology". 2014–2023 Director of the Sociology Department, L.N. Gumilyov ENU. He is the author of over 70 articles and 11 original documents. He can be contacted at email: baigabylov_no@enu.kz.






Ayagoz Mukhanova    received her Ph.D. in 2015 in Information Systems from L.N. Gumilyov Eurasian National University, Kazakhstan. Currently, she is an associate professor of the Department of Information Systems at the same university. Her research interests include artificial intelligence and decision making. She can be contacted at email: ayagoz.mukhanova.83@mail.ru.






Kuralay Mukhambetova    acting head of the Department of Sociology and Social Work of the Eurasian National University named after L.N. Gumilyov, senior lecturer, candidate of sociological sciences (2004). She assessed the quality of children's lives in 2009. She has served as a national expert for United Nations Office on Drugs and Crime and for United Nations programs on the assessment of public services. She can be contacted at email: k.mukhambetova74@gmail.com.






Elyor Zaitov    earned a Bachelor's degree in Sociology from the National University of Uzbekistan, Mirzo Ulugbek, in 2009. He was an Associate Professor of the Department of Social Work of the aforementioned university from 2017 to 2022. In 2020, he was awarded a Doctorate of Philosophy in Sociology with a specialization in Social Structure, Social Institutions, and Lifestyles (22.00.02). From 2023, he is the Chairman of the Department of Sociology at the National University of Uzbekistan named after Mirzo Ulugbek. He has over 100 scientific publications. He can be contacted at email: e.zaitov@nuu.uz.






Roza Burganova    in 1991 she graduated from the Karaganda Pedagogical Institute with a degree in Pedagogy and Psychology (Preschool). In 2006 he was awarded the title of Associate Professor in the specialty Pedagogy. 2010 to the present - Associate Professor of the Department of Social Work and Tourism; Head of Quality Assurance and Strategic Review; Academic Secretary of the Institution "Esil University", (formerly KazUEFIT). She is the author of more than 90 works. She can be contacted at email: burganova.r@esil.edu.kz.



Khayriniso Khusanova    is an Associate Professor at the Department of Sociology, National University of Uzbekistan, with over 30 years of teaching experience. She has completed advanced training in pedagogical technologies and teaching skills at the International School of Sociologists VTsIOM (Uzbekistan, 2019). She is also an active member of the Association of Sociologists of Uzbekistan. She can be contacted at email: husanovahayriniso@gmail.com.



Feruza Akhmedova    is an Associate Professor at the Department of Sociology, National University of Uzbekistan, with more than 30 years of teaching experience. Specialist in the field of quantitative and qualitative sociological research. Advanced training: Pedagogical technologies and pedagogical skills, International School of Sociologists, Membership in organizations, and Member of the Association of Sociologists of Uzbekistan. She can be contacted at email: farihon72@gmail.com.