# Hybrid deep learning architecture for autism spectrum disorder detection from gait kinematic data

**Gulzat Ziyatbekova[1,2,3], Svetlana Beglerova[4], Dinara Zhukenova[5], Alokhon Alikarieva[6], Madi Akhmetzhanov[7], Nuriddin Alikariev[6], Quvvatali Rakhimov[8], Yersultan Tulebayev[9]**
[1]Department of Computer Engineering, Faculty of Information Technology, Almaty Technological University, Almaty, Kazakhstan
[2]Faculty of Information Technology and Artificial Intelligence, Al-Farabi Kazakh National University, Almaty, Kazakhstan
[3]Institute of Information and Computational Technologies SC MSHE RK, Almaty, Kazakhstan
[4]Department of Information Systems, Faculty of Technology, Taraz University named after M.Kh.Dulaty, Taraz, Kazakhstan
[5]Department of Power Engineering, Automation and Computer Engineering, Institute of Industrial Technologies, PH PEI "West Kazakhstan Innovation and Technological University", Uralsk, Kazakhstan
[6]Department of Sociology, Faculty of Social Sciences, National University of Uzbekistan named Mirzo Ulugbeka, Toshkent, Uzbekistan
[7]Department of Applied Informatics and Programming, Taraz University named after M.Kh.Dulaty, Taraz, Kazakhstan
[8]Department of Applied Mathematics and Informatics, Fergana State University, Fergana, Uzbekistan
[9]Department of Computer Engineering, Astana IT University, Astana, Kazakhstan

**Article Info**

***Article history:***

Received Oct 5, 2025
Revised Mar 12, 2026
Accepted Apr 19, 2026

***Keywords:***

Autism spectrum disorder
Bidirectional long short-term memory
Deep learning
Gait analysis
Hybrid architecture
Kinematic data

**ABSTRACT**

Early detection of autism spectrum disorder (ASD) is essential for enabling timely interventions that significantly impact developmental outcomes. Traditional diagnostic procedures often rely on clinical observations and subjective assessments, which can reduce objectivity and delay diagnosis. This study presents an automated classification approach for detecting ASD in children using gait kinematic data obtained through video-based skeletal tracking. We propose a hybrid deep learning architecture that combines one-dimensional convolutional layers (Conv1D), bidirectional long short-term memory (BiLSTM) networks, and a multi-head attention mechanism to capture complex spatiotemporal patterns in motion data. The dataset includes 100 participants—50 diagnosed with ASD and 50 typically developing (TD) peers. Preprocessing steps included Euclidean norm transformation, logarithmic scaling, Z-score normalization, and sliding window segmentation. The proposed model achieved 94.9% accuracy, 0.91 recall, 0.86 precision, and an area under the curve (AUC) of 0.97, outperforming a baseline long short-term memory (LSTM) architecture. These findings demonstrate the potential of gait-based kinematic features and hybrid neural networks for objective and reproducible ASD screening. The developed system can contribute to AI-assisted decision support tools in clinical and educational environments, enhancing diagnostic accuracy, and supporting inclusive developmental care.



***Corresponding Author:***

Svetlana Beglerova
Department of Information Systems, Faculty of Technology, Taraz University named after M.Kh.Dulaty
080000, Taraz, Kazakhstan
Email: nbaygabylov@gmail.com

## 1. INTRODUCTION

In recent years, autism spectrum disorder (ASD) is a complex neurodevelopmental condition characterized by persistent challenges in social communication, restricted interests, and repetitive patterns of behavior [1], [2]. According to the World Health Organization, approximately one in every hundred children worldwide is affected by ASD [3]. Early and accurate identification of ASD plays a crucial role in initiating timely behavioral interventions, improving social inclusion, and enhancing developmental outcomes [4].

However, conventional diagnostic methods such as the autism diagnostic observation schedule (ADOS) and the childhood autism rating scale (CARS) rely heavily on behavioral observation and expert evaluation, which are time-consuming, subjective, and prone to interrater variability [5], [6]. This has created a strong demand for objective, automated, and scalable diagnostic tools that can assist clinicians in early detection.

One promising direction for objective ASD assessment lies in the analysis of motor behavior, particularly gait kinematics, which reflects underlying neuromotor development. Prior studies have reported that children with ASD often exhibit atypical gait patterns, including asymmetrical movement, instability, and reduced coordination [7]–[9]. With the advent of modern motion-capture technologies such as the Microsoft Kinect sensor, it has become possible to record three-dimensional body-joint coordinates in a non-invasive and cost-effective way [10], [11]. These data provide opportunities for applying artificial intelligence (AI) and deep learning techniques to extract meaningful biomechanical features and perform automated classification of movement patterns.

Recent studies have increasingly focused on applying AI techniques to ASD detection using various behavioral and physiological signals. From the perspective of data modality, prior works can be broadly categorized into video-based, sensor-based, and multimodal approaches. For example, Li *et al.* [12] proposed a convolutional neural networks-bidirectional long short-term memory (CNN-BiLSTM) framework for multimodal gait analysis, demonstrating the effectiveness of combining spatial and temporal feature extraction. Similarly, Cicirelli *et al.* [13] emphasized the importance of gait analysis as a biomarker in neurological disorders, while Gao *et al.* [14] introduced an long short-term memory (LSTM)-CNN fusion model for abnormal gait recognition, achieving improved performance in sequential motion analysis. In addition, Yang *et al.* [15] provided a comprehensive review of deep learning approaches for ASD detection based on video analysis, highlighting the growing role of computer vision techniques in behavioral assessment. From the perspective of model type, traditional machine learning approaches have also been widely applied in movement-based ASD detection. Lee *et al.* [16] utilized support vector machines (SVM) for postural and motion analysis, while previous studies [17] explored machine learning techniques for detecting ASD based on facial and motor features. Although these methods offer interpretability, they are limited in their ability to capture complex nonlinear relationships and long-term temporal dependencies in high-dimensional kinematic data.

More recent studies have increasingly focused on deep learning architectures capable of modeling sequential data. Zhang *et al.* [18] highlighted the effectiveness of LSTM-based models for temporal sequence learning, whereas Al-Selwi *et al.* [19] proposed attention-based neural networks to enhance feature representation in time-series data. Zakaria *et al.* [20] reported that many gait-based ASD detection approaches do not jointly model spatial and temporal dependencies within a unified analytical framework, which limits their effectiveness in capturing complex motion dynamics. Peya *et al.* [21] investigated electroencephalography (EEG-based) methods for autism detection; however, these approaches require intrusive data acquisition procedures that restrict their applicability in large-scale and naturalistic screening scenarios. Trăscău *et al.* [22] emphasized the importance of effective spatiotemporal feature representation for accurate skeletal data analysis, noting that this aspect is often insufficiently addressed in existing models. Zhu *et al.* [23] introduced a multimodal machine learning framework for early autism screening, demonstrating improved diagnostic performance but also increased system complexity and computational overhead. Kolaghassi *et al.* [24] further demonstrated the potential of deep learning models for forecasting gait trajectories, highlighting the capability of neural architectures to learn complex motion dynamics. Similarly, Rad *et al.* [25] applied deep learning techniques to detect stereotypical motor movements associated with ASD, confirming the advantages of deep neural networks for modeling nonlinear behavioral patterns.

Therefore, there remains a need for a unified and efficient model capable of simultaneously capturing spatial structures, temporal dynamics, and salient motion patterns in a non-invasive and computationally efficient manner. To address these challenges, this study proposes a hybrid deep learning architecture that integrates one-dimensional convolutional layers (Conv1D) for spatial feature extraction, BiLSTM networks for bidirectional temporal modeling, and a multi-head attention mechanism to focus on the most informative segments of gait sequences. By combining these complementary components within a single framework, the proposed model effectively captures both local spatial features and long-range temporal dependencies in gait kinematic data. Experimental results demonstrate the effectiveness of the proposed approach, achieving an accuracy of 94.9% and outperforming a baseline LSTM model. Overall, the proposed framework provides a robust, non-invasive, and scalable solution for AI-assisted early ASD screening.

## 2. METHOD

### 2.1. Dataset collection

The present study utilized an open-access three-dimensional gait dataset obtained from the University of Babylon, Iraq, recorded using the Microsoft Kinect v2 depth sensor. This optical motion-capture device enables real-time tracking of three-dimensional skeletal coordinates of 25 body joints without the need for physical markers. The non-contact nature of this technology makes it highly suitable for behavioral research involving children, as it minimizes discomfort and preserves ecological validity.

The dataset comprises gait recordings of 100 children aged between 4 and 14 years, including 50 children clinically diagnosed with ASD and 50 typically developing (TD) peers. All recordings were fully anonymized - participants' faces were blurred, and no personal identifiers were retained. Informed consent had been obtained from parents or guardians prior to data collection. Each subject walked along a straight path in a controlled laboratory setting that simulated natural walking conditions. From each video recording, 25 key skeletal joints were extracted, representing major anatomical landmarks: head, neck, shoulders, elbows, wrists, hands, pelvis, knees, ankles, and feet. The sampling rate of the Kinect v2 camera was approximately 30 Hz, with a time resolution of 33 milliseconds per frame. For each participant, one complete gait cycle was identified, defined as the interval between two successive contacts of the same foot with the ground. After data augmentation, the total number of gait cycles amounted to 2,983, including 1,565 cycles from ASD children and 1,418 cycles from TD participants. The dataset was balanced, preprocessed, and normalized before being used for model training. The dataset is publicly available via the Dryad repository: https://datadryad.org/stash/dataset/, doi: 10.5061/dryad.s7h44j150 [26].

## 2.2. Rationale for a method selection

The architecture was specifically designed to capture both the temporal and spatial dynamics of human motion. To achieve this, a combination of convolutional, recurrent, and attention-based mechanisms was implemented. The preprocessing pipeline included normalization, logarithmic transformation, and the sliding window technique to segment raw kinematic data into fixed-length sequences suitable for input to deep learning models. The model architecture, activation functions, regularization strategies, and training parameters were selected to balance model expressiveness and generalization performance. For the classification task, two primary deep learning models were employed. One of the most prominent examples of this is the LSTM network, a variant of the recurrent neural network that is particularly successful in learning sequential knowledge from temporal relations.

The suggested LSTM-based model consists of two consecutive LSTM layers with 64 and 32 units, respectively, and a dropout layer with a rate of 0.3 to avoid overfitting. A dense layer containing 16 rectified linear unit (ReLU)-activated neurons is applied prior to the final out-put layer consisting of a single sigmoid-activated neuron for binary classification. The model is optimized using the Adam algorithm and trained with a binary cross-entropy loss function. Performance is evaluated on metrics such as accuracy, area under the curve (AUC), precision, and recall. As indicated in Figure 1, the input layer accepts tensors of dimension (time_steps=19, features=25) for a time sequence of 19 steps and 25 features for the normalized 3D coordinates of body joints. This design allows the model to capture hidden patterns in movement dynamics that may be indicative of ASD.

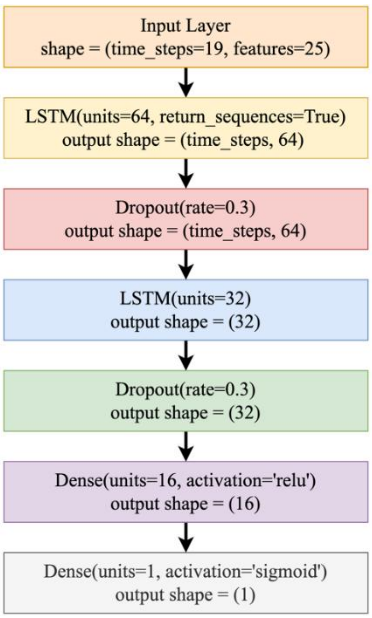



Figure 1. LSTM-based recurrent neural network architecture for time series classification

The architecture of the baseline model begins with an LSTM layer consisting of 64 units, configured to return the full sequence of outputs (return_sequences=True), thereby preserving temporal dependencies for subsequent layers. A dropout layer with a probability of 0.3 follows, mitigating overfitting by randomly deactivating a portion of connections during training. The second LSTM layer, with 32 units, extracts higher-level temporal abstractions without returning sequences, effectively reducing feature dimensionality. Another dropout layer is applied for enhanced regularization. A full connection layer of 16 neurons, with ReLU activation, then non-linearly maps the learned features, thereby enabling generalization. The last output layer is a single neuron with sigmoid activation, giving a probability that signifies the absence or presence of ASD. This LSTM-based architecture effectively captures temporal dynamics with the guarantee of resistance against overfitting, which is important in the case of high-dimensional data using few samples.

In comparison, Figure 2 illustrates the architecture of suggested hybrid deep neural network that can extract and recognize complex spatiotemporal features from sequences of skeletal joint coordinates. This multi-layered model includes Conv1D, BiLSTM, and a multi-head attention mechanism. The convolutional layer extracts local spatial patterns within each time slice, the BiLSTM units extract bidirectional temporal relations, and the attention mechanism selectively high-lights informative regions of the input. Together, this enables the model to take ad-vantage of spatial structure, temporal context, and semantic salience simultaneously and enhance its discrimination between normal and abnormal gait patterns in ASD.

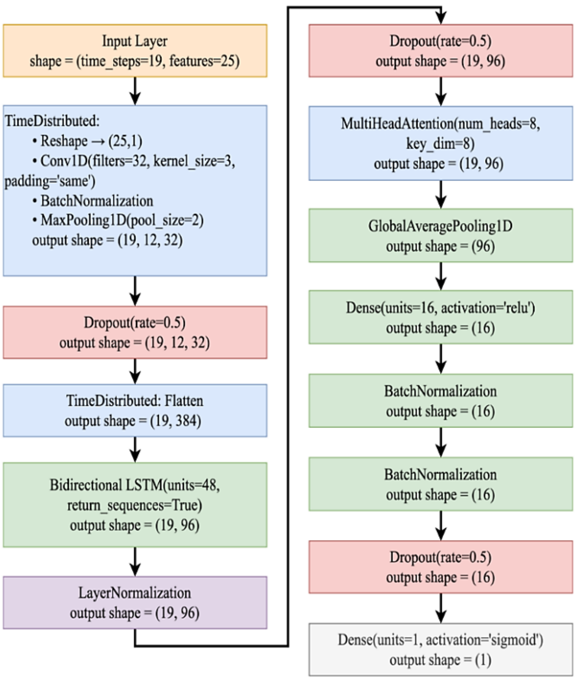



Figure 2. Hybrid neural network architecture based on Conv1D, BiLSTM, and multi-head attention for temporal sequence analysis

The proposed model architecture begins with an input layer that accepts data in the form of normalized joint coordinates. This is followed by a reshaping operation to add an additional dimension required for convolutional processing. A Conv1D with 32 filters and a kernel size of 3 is applied to extract local pat-terns within each temporal window. Batch normalization and a MaxPooling1D operation are used to improve training stability and reduce feature dimensionality. To mitigate overfitting, a dropout layer with a rate of 0.5 is applied at this stage. Subsequently, the features are flattened and passed to a BiLSTM layer with 48 units, which models temporal dependencies in both forward and backward directions, offering a comprehensive view of movement dynamics. The output is then normalized using a layer normalization layer and subjected to another dropout operation. A key component of the architecture is the multi-head attention mechanism with eight heads, which identifies salient temporal and spatial dependencies in the data - an essential aspect for neurodevelopmental classification tasks. The resulting attention-weighted features are aggregated using a global average pooling layer and passed through a fully connected (Dense) layer with 16 ReLU-activated units, followed by batch normalization and dropout regularization. The final output layer consists of a single neuron with sigmoid activation, producing a probability score indicating class membership (ASD or TD).

The key hyperparameters of the proposed hybrid architecture were selected through preliminary empirical experiments and validation-based tuning. Several configurations were evaluated to balance model complexity, training stability, and generalization capability. The Conv1D layer employs 32 filters with a kernel size of three, which was found sufficient for capturing local spatial dependencies between skeletal joints without introducing excessive model complexity. The BiLSTM layer contains 48 hidden units, providing adequate capacity for modeling bidirectional temporal relationships within gait sequences while avoiding overfitting on the relatively limited dataset. The multi-head attention mechanism utilizes eight attention heads, enabling the model to learn multiple complementary temporal representations and capture complex dependencies across different segments of motion data. A dropout rate of 0.5 was applied as a regularization strategy to mitigate overfitting and improve generalization performance during training. The final hyperparameter configuration was selected based on validation performance and training stability across multiple experimental runs.

This hybrid Conv1D–BiLSTM–multi-head attention model processes sequential joint coordinate data and combines local spatial feature extraction, long-range temporal modeling, and attention-based focus on informative segments. The training parameters included the Adam optimizer (learning rate=0.0005), binary cross-entropy loss, and evaluation metrics such as AUC, accuracy, precision, and recall. Training was performed for up to 50 epochs with a batch size of 64, using early stopping based on validation AUC and class weighting to address data imbalance.

## 3. RESULTS AND DISCUSSION

The experimental results demonstrate the effectiveness of the proposed hybrid deep learning model in classifying ASD based on gait kinematic data. Performance metrics such as accuracy, precision, recall, and AUC indicate significant improvement over the baseline LSTM model, confirming the superior capability of the proposed architecture to capture complex spatiotemporal dependencies in human motion. The three-dimensional joint coordinates extracted from the Kinect v2 sensor served as key input features, representing the dynamic patterns of movement across the human body. These kinematic features provided a comprehensive foundation for analyzing gait behavior and developing reliable classification models for ASD detection.

### 3.1. Model performance evaluation

The use of the full set of skeletal joints ensures high spatial resolution for gait analysis and enables deep learning models to capture subtle motor patterns characteristic of ASD. Moreover, the inclusion of joints such as wrists, fingertip points, and shoulder joints allows the detection of micro-movements that may play a crucial role in differentiating behavioral trajectories between children with ASD and their TD peers. Figures 3 and 4 illustrate the training and validation processes of two deep learning models: the baseline LSTM and the proposed hybrid architecture (Conv1D-BiLSTM-multi-head attention).

An examination of the learning curves of the baseline LSTM model shows successful learning of patterns across the training set, as demonstrated by a steady decrease in loss and improvement in major performance metrics such as accuracy, recall, precision, and AUC. However, when evaluated on the validation set, signs of overfitting are observed, including fluctuations in the loss function, inconsistencies in recall and precision values, and a decline in generalization capability starting from approximately the 10th epoch. Although initial performance is marked by high AUC values (reaching 0.95) and accuracy percentages (up to 83%) in the early stages of training, final values after 50 epochs demonstrate lower stability. Specifically, accuracy dropped to approximately 79%, AUC decreased to 0.88, and recall fluctuated between 0.65 and 0.75, indicating too low a sensitivity to the positive class. These results point out the limitations of the baseline LSTM model in the context of a small dataset along with high variability in motor patterns.

This substantiates the need for more advanced architectures and regularization techniques to improve classification robustness and enhance ASD diagnosis. Figure 4 displays a series of diagrams that show the variation in core training and validation performance measures for the proposed hybrid neural network, which uses a Conv1D, a BiLSTM, and a multi-head attention mechanism. These diagrams provide an in-depth understanding of the model's performance on training and testing datasets, hence facilitating the identification of potential generalization issues.

An analysis of the training results for the hybrid model demonstrates its high effectiveness in the automated classification of ASD based on gait kinematic data. The loss curves show a consistent and stable reduction in error on the training set, accompanied by a similar trend on the validation set, indicating effective optimization and a low risk of overfitting. Validation accuracy ranges from 85% to 95%, remaining stable throughout training, which reflects strong generalization capability. The AUC values range from 0.85 to 0.96, demonstrating high discriminative power across all epochs. Precision and recall on the training set consistently exceed 0.82 and 0.90, respectively, while on the validation set they reach 0.75–0.85 and 0.65–0.75, confirming reliable classification performance. During final evaluation, the model achieved an accuracy of 94.9%, with validation loss decreasing to 0.32-0.36. A slight drop in AUC was observed, which may be attributed to re-execution variability. The use of class weights improved sensitivity without compromising overall accuracy, further confirming the robustness of the proposed architecture in handling imbalanced and variable motor patterns. Table 1 presents the quantitative comparison of key performance metrics between the baseline and proposed models.

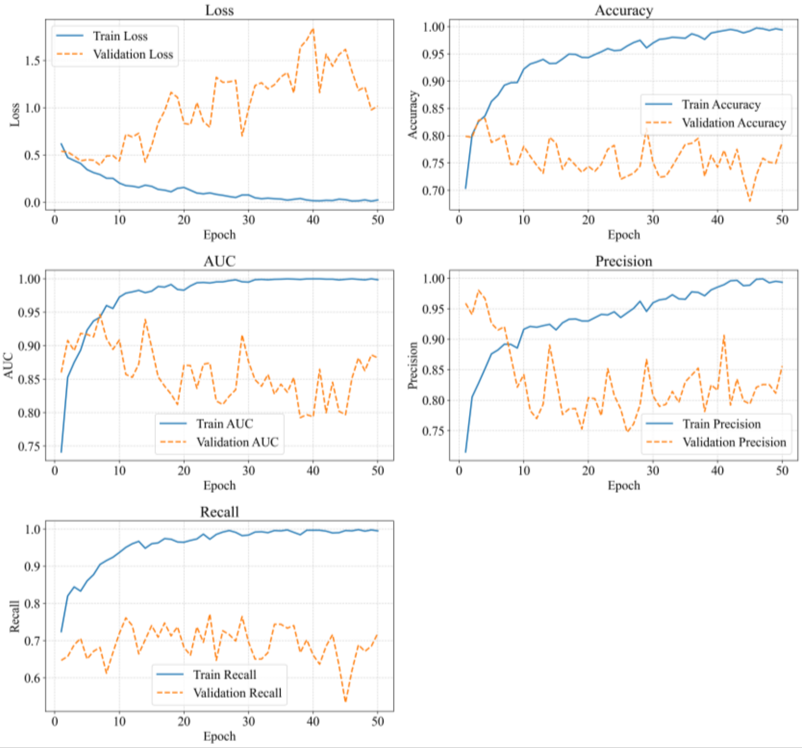



Figure 3. Dynamics of key training and validation metrics of the LSTM model across epochs

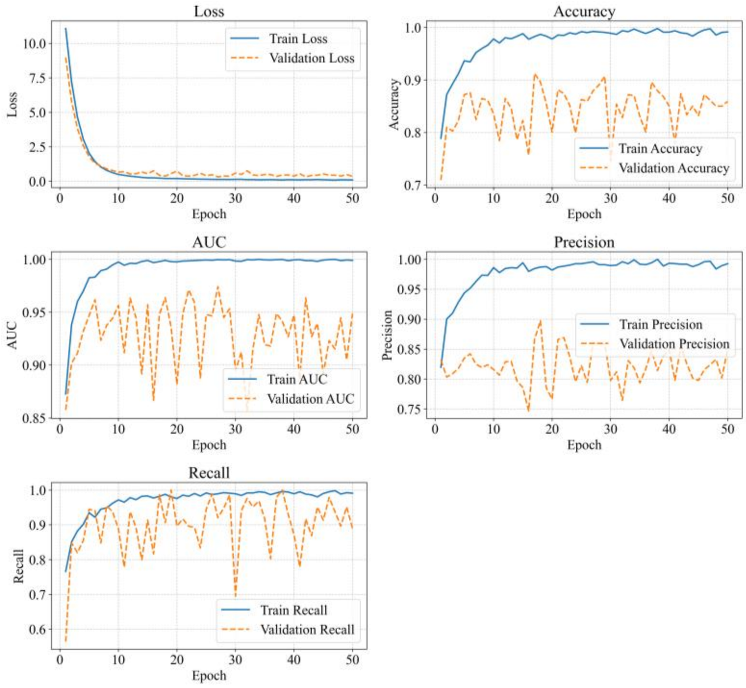



Figure 4. Training and validation metrics of the hybrid model over 50 epochs

Table 1. Performance comparison metrics of LSTM model and hybrid architecture (Conv1D–BiLSTM–multi-head attention)

| Metric | LSTM | Hybrid model |
|---|---|---|
| Accuracy | 0.91 | 0.95 |
| AUC | 0.93 | 0.97 |
| Precision | 0.82 | 0.87 |
| Recall | 0.88 | 0.91 |
| Validation loss | 0.41 | 0.33 |

The results in Table 1 indicate that the proposed hybrid model outperforms the baseline LSTM across all evaluation metrics, including accuracy, recall, precision, and AUC. These improvements suggest enhanced generalization capability and more robust classification performance. To further evaluate the discriminative capability of the models, receiver operating characteristic rank one computing (ROC) analysis was conducted. Figure 5 presents a visual comparison of the ROC curves for the baseline LSTM model and the proposed hybrid architecture. The results show that the hybrid model consistently achieves higher true positive rates across most classification thresholds while maintaining lower false positive rates. The area under the ROC curve AUC increases from 0.94 for the baseline LSTM to 0.96 for the hybrid model, further confirming the improved discriminative performance of the proposed architecture.

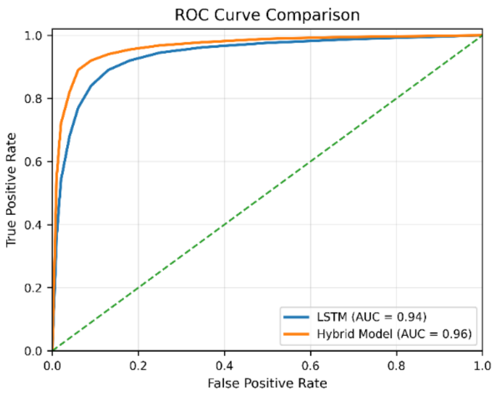


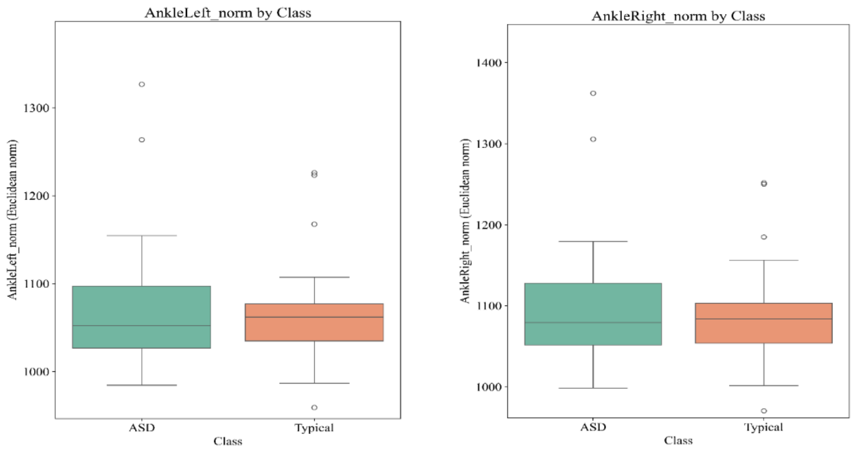

Figure 5. ROC curve comparison between the baseline LSTM model and the proposed hybrid architecture

To further interpret the observed performance improvements, a detailed analysis of the proposed hybrid architecture is provided below. From a feature learning perspective, the Conv1D layers extract local spatial patterns from joint coordinate sequences, enabling the identification of relationships between neighboring joints and short-term motion variations. This aligns with previous studies demonstrating the effectiveness of convolutional architectures in motion analysis [12], [24]. The BiLSTM component further enhances temporal modeling by capturing long-range dependencies in both forward and backward directions, providing a more comprehensive representation of sequential dynamics compared to standard LSTM models [19]. Together, these components enable the model to learn rich spatiotemporal representations of gait patterns.

Furthermore, the multi-head attention mechanism allows the model to focus on the most informative segments of gait sequences, improving feature discrimination, and reducing the influence of irrelevant variations. Attention-based approaches have been shown to enhance both interpretability and classification performance in time-series analysis [18], [22]. By integrating Conv1D, BiLSTM, and attention within a unified framework, the proposed model jointly learns spatial structures, temporal dynamics, and salient motion patterns. This directly addresses a key limitation of previous methods, where these aspects were modeled separately or insufficiently integrated [20]. Moreover, the model demonstrates robustness to variability in motor behavior, which is characteristic of ASD-related movement patterns [25]. This contributes to more stable and reliable classification outcomes. Overall, the findings indicate that the proposed hybrid architecture provides superior spatiotemporal representation capability, leading to improved classification accuracy and stronger generalization performance compared with single-branch recurrent models.

### 3.2. Kinematic feature analysis and interpretation

To gain deeper insights into the spatial characteristics of motor behavior in children with ASD versus TD peers, we conducted a comparative visualization of normalized joint coordinate distributions across key anatomical landmarks. Figures 6-15 present boxplots for each selected joint, including points on the lower and upper extremities, trunk, and head. These visualizations reflect the range, median, interquartile spread, and presence of outliers in the Euclidean-normalized positions, computed from 2D pose data over temporal sequences. Great significance is placed on the occurrence of extreme values and a large range of distribution, which can reflect atypical or less stable motor control, a feature frequently linked to motor phenotypes in ASD. This analysis enables a qualitative comparison of motion variability and highlights joints or regions that may serve as potential biomarkers in the classification process. The patterns observed in these figures form the basis for subsequent interpretation and justification of feature selection for deep learning classification models. Figures 6(a) and (b) show boxplots of normalized Euclidean coordinates for the left and right ankles in children with ASD and TD peers, respectively.

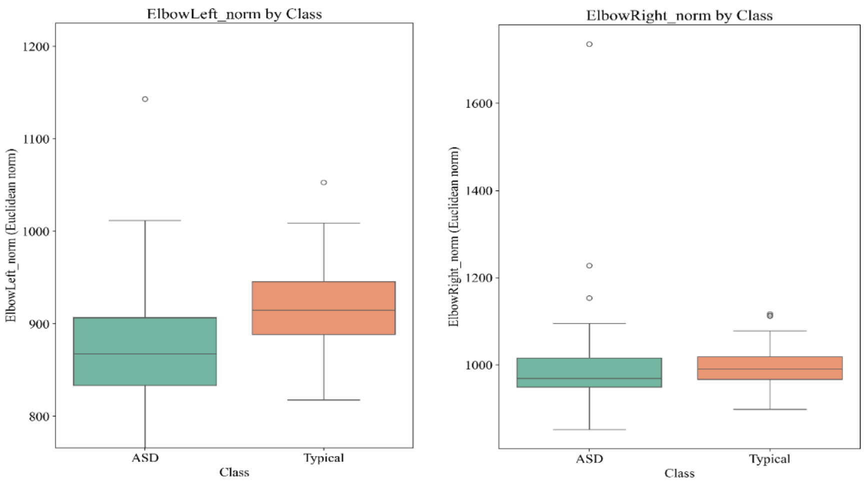



(a) (b)

Figure 6. Distribution of the Euclidean norm of ankle coordinates by class; (a) left ankle and (b) right ankle

As illustrated in Figure 6(a), the left ankle in the ASD group exhibits a wider distribution and multiple outliers, indicating increased variability and instability in gait patterns. Similarly, Figure 6(b) shows that the right ankle also demonstrates a broader spread in the ASD group compared to the more compact and consistent distributions observed in TD children. These findings suggest reduced motor stability and coordination in lower-limb movements among children with ASD.

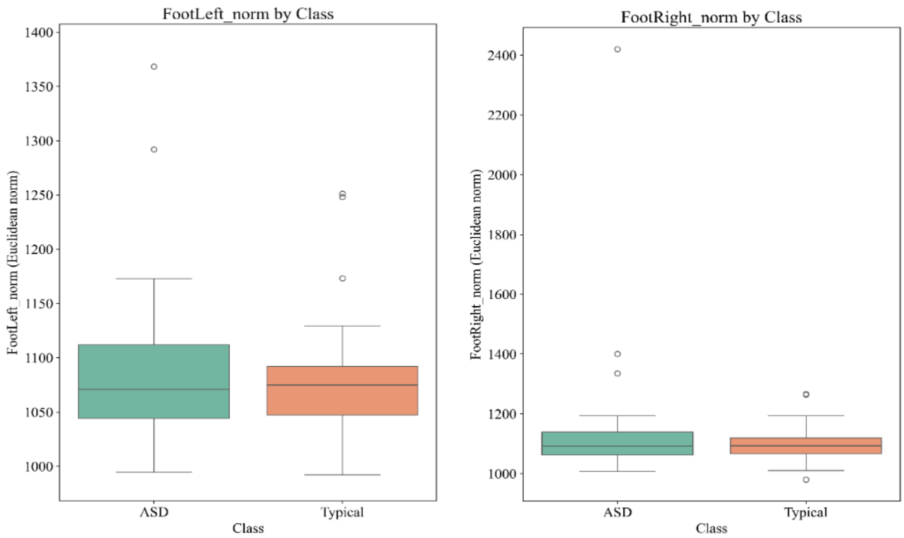
Figures 7(a) and (b) present boxplots of normalized Euclidean values for the left and right elbows in children with ASD and TD peers. As shown in Figure 7(a), both groups exhibit comparable ranges for the left elbow; however, the ASD group displays a slightly higher number of outliers. Figure 7(b) further demonstrates that the right elbow in the ASD group shows increased variability, indicating individual cases of pronounced movement deviations. These observations suggest mild asymmetry and variability in upper-limb motion among children with ASD.




(a) (b)

Figure 7. Distribution of the Euclidean norm of elbow coordinates by class; (a) left elbow and (b) right elbow

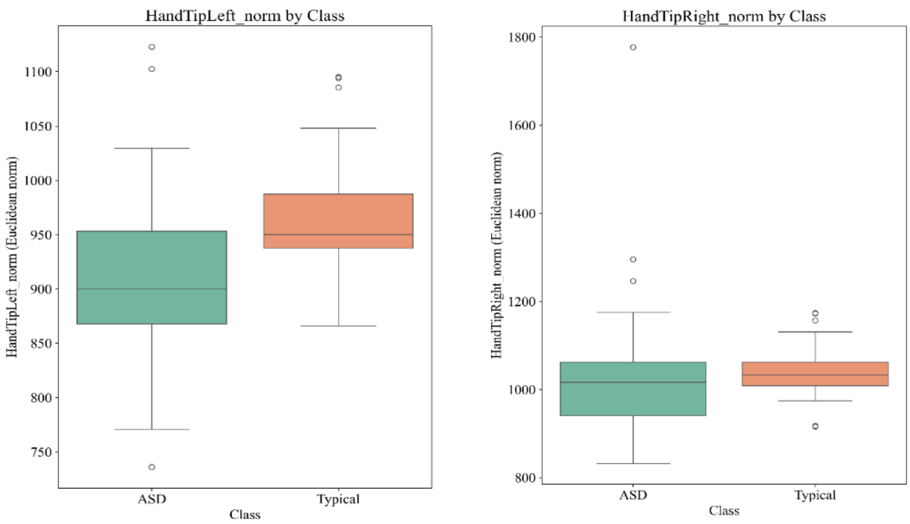
Figures 8(a) and (b) illustrate boxplots of normalized Euclidean coordinates for the left and right feet in children with ASD and TD peers. In Figure 8(a), the left foot in the ASD group exhibits a wider range of values and numerous outliers, indicating coordination deficits and gait instability. Figure 8(b) shows asymmetrical patterns in the right foot, further emphasizing distinct motor activity characteristics in children with ASD. These findings highlight the diagnostic relevance of kinematic parameters such as movement variability and amplitude in identifying autism-related motor anomalies.



(a) (b)

Figure 8. Distribution of the normalized Euclidean values of ankle coordinates by class; (a) left foot and (b) right foot

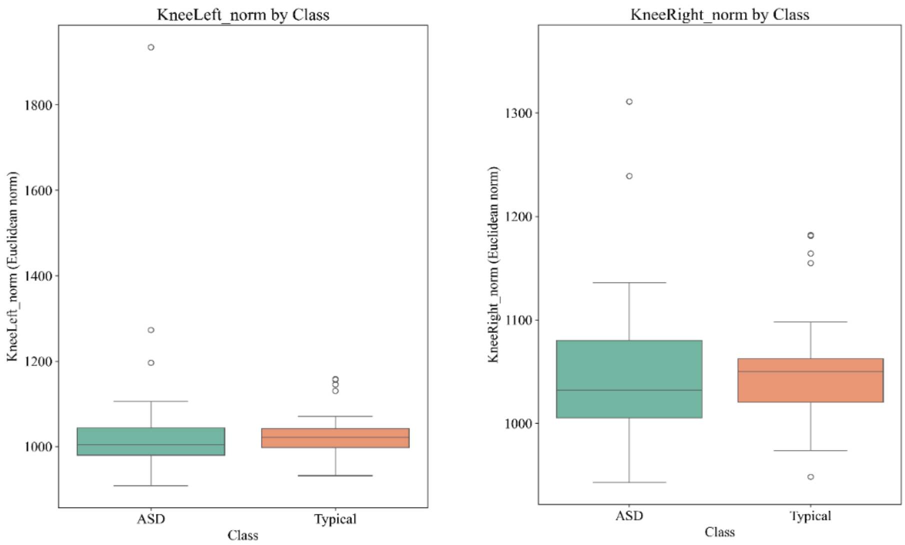
Figures 9(a) and (b) present boxplots of normalized Euclidean coordinate values for the left- and right-hand tips in children with ASD and TD peers, respectively. As shown in Figure 9(a), both groups exhibit relatively narrow value ranges, with slightly higher median values observed in the TD group. The left hand demonstrates a clearer separation between groups, where TD children exhibit more clustered and stable movements, while the ASD group shows multiple outliers indicating increased motor variability and reduced control. Figure 9(b) shows a similar pattern for the right hand, although the differences between the groups are less pronounced.



(a) (b)

Figure 9. Distribution of the normalized Euclidean values of hand tip coordinates by class; (a) left hand tip and (b) right hand tip

Figures 10(a) and (b) show boxplots of normalized Euclidean coordinate values for the left and right knee joints in children with ASD and TD peers, respectively. As illustrated in Figure 10(a), the median values for both groups are similar, indicating comparable average knee positions. However, the ASD group exhibits a wider range of values and a higher number of outliers, reflecting increased variability in movement patterns. Figure 10(b) further demonstrates that the variability is more pronounced for the right knee, where the ASD group shows a greater dispersion of values compared to the TD group. This increased variability suggests irregularities in lower-limb coordination and supports evidence of atypical motor control among children with ASD.




(a) (b)

Figure 10. Distribution of the normalized Euclidean values of knee joint coordinates by class; (a) left knee and (b) right knee

Figure 11 presents boxplots of normalized mid-spine coordinates (MidSpine_norm) for children with ASD and TD peers. Both groups exhibit similar median values, indicating comparable average mid-spine positions. However, the ASD group displays several outliers above the upper quartile, suggesting increased postural variability and irregular movement control. These deviations point to unique motor characteristics and possible instability in trunk coordination among children with ASD.

Figures 12(a) and 12(b) show boxplots of normalized Euclidean coordinate values for the left and right shoulder joints in children with ASD and TD peers, respectively. As shown in Figure 12(a), both groups exhibit similar median values, indicating comparable average shoulder positioning; however, the ASD group presents several outliers, suggesting possible motor control irregularities. Figure 12(b) shows a more compact distribution, reflecting greater stability and symmetry, particularly in the TD group. Although group-level differences are limited, the observed individual variability highlights the relevance of shoulder kinematic features for ASD-related motor analysis.

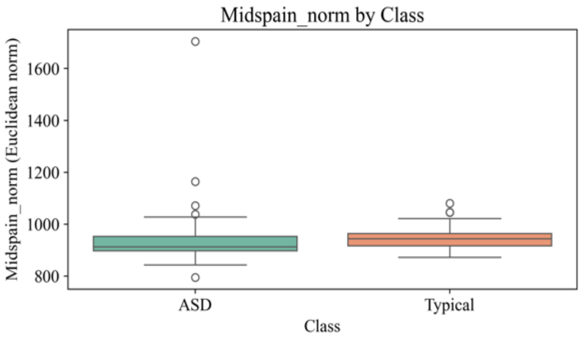
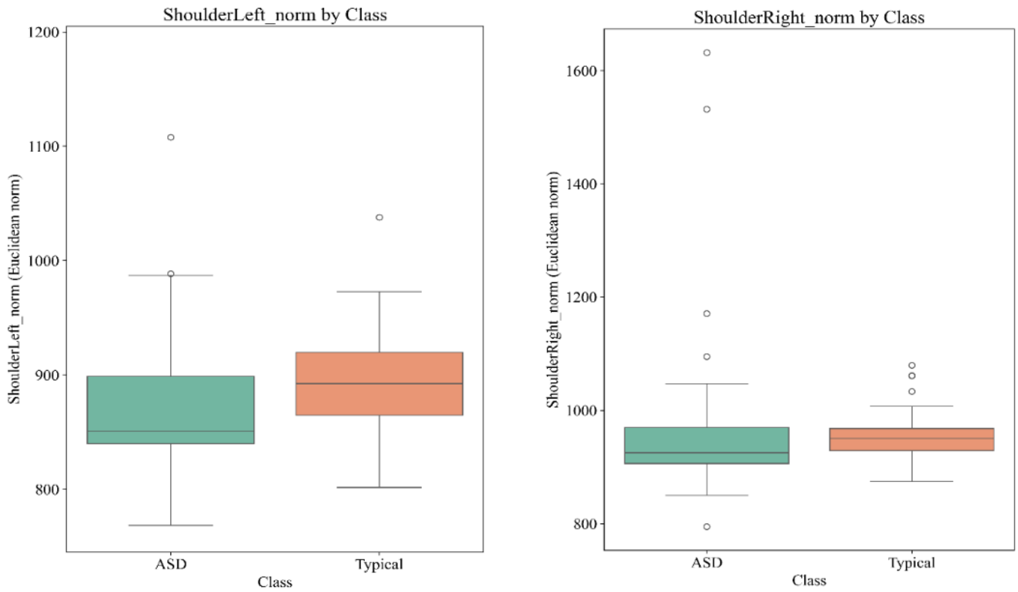



Figure 11. Normalized values of the mid-spine by class




(a) (b)

Figure 12. Distribution of the normalized Euclidean values of shoulder joint coordinates by class; (a) left and (b) right shoulder

Figure 13 shows boxplots of normalized shoulder joint positions for children with ASD and TD peers. Both groups have similar mean values, but the ASD group exhibits greater dispersion and several extreme outliers, indicating reduced postural stability and atypical motor control. These deviations highlight the diagnostic relevance of abnormal shoulder movement patterns in identifying motor irregularities associated with ASD.

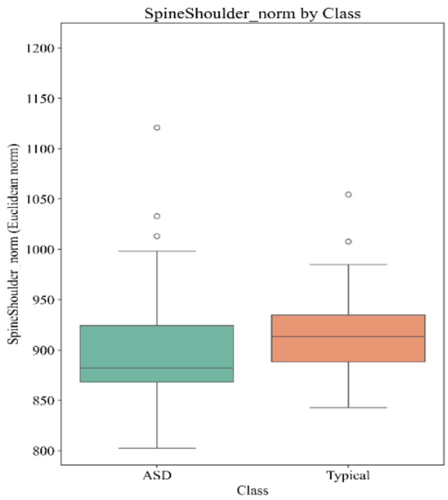
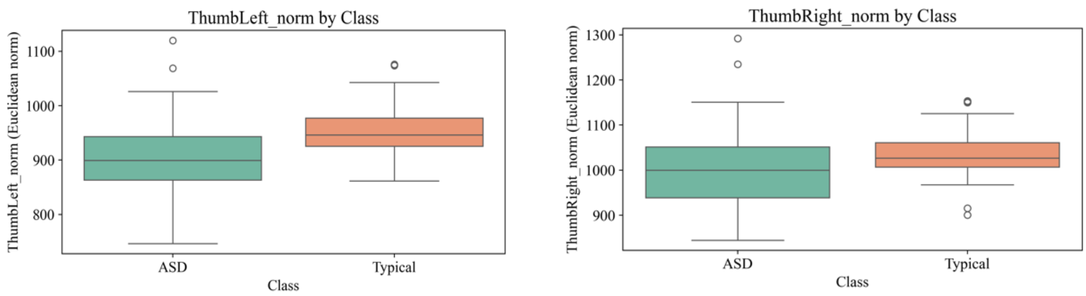



Figure 13. Normalized values of the spine shoulder joint

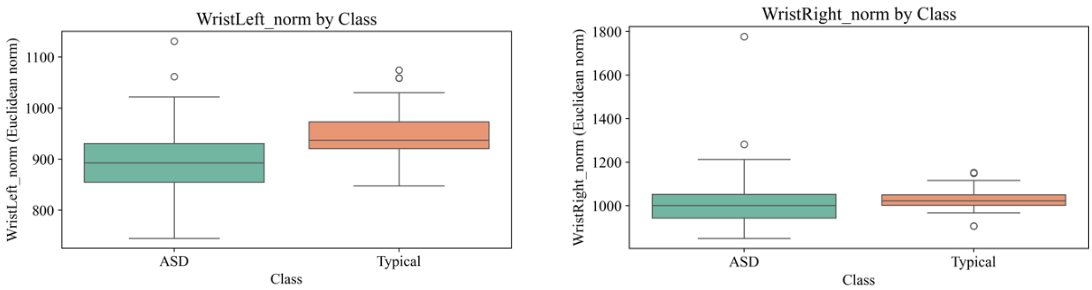
Figures 14(a) and (b) present boxplots of normalized Euclidean values for the left and right thumbs in children with ASD and TD peers, respectively. As shown in Figure 14(a), the TD group exhibits higher median values, indicating greater stability and coordination in fine motor movements, whereas the ASD group shows increased variability and multiple outliers. Figure 14(b) demonstrates a similar pattern, with the ASD group displaying wider dispersion, suggesting reduced precision in thumb movements. These findings highlight the relevance of fine motor features for ASD-related motor assessment.




(a) (b)

Figure 14. Distribution of the normalized position indicators of the thumb by class; (a) left thumb and (b) right thumb

Figures 15(a) and (b) show boxplots of normalized Euclidean values for the left and right wrists in children with ASD and TD peers, respectively. As illustrated in Figure 15(a), the TD group demonstrates higher median values and lower variability, indicating more stable motor control, whereas the ASD group exhibits increased dispersion. Figure 15(b) shows that although both groups have similar median values, the ASD group presents a wider range and more outliers, suggesting greater instability and variability in wrist movement patterns.




(a) (b)

Figure 15. Distribution of the normalized wrist position values by class; (a) left wrist and (b) right wrist

The obtained results demonstrate the high effectiveness of the proposed hybrid architecture, Conv1D–BiLSTM–multi-head attention for the task of automatic classification of children with ASD based on gait kinematic data. Compared to the baseline LSTM model, the proposed approach provides substantial improvements in accuracy, robustness, and generalization capability. The boxplot analysis of key joints revealed distinctive movement patterns in the ASD group, particularly in the lower limbs and wrists. The comprehensive investigation of coordinate distributions confirmed the hypothesis regarding the presence of specific motor features that may serve as potential biomarkers for early autism diagnosis. These findings emphasize the importance of a thorough analysis of spatiotemporal motion characteristics and highlight the promise of deep learning methods in clinical applications for autism detection and monitoring.

### 3.3. Limitations and future work

This study has several limitations that should be acknowledged. Although the dataset contained balanced groups of children with ASD and TD peers, the overall sample size remained relatively limited, which may affect the generalizability of the obtained results. In addition, the extraction of gait cycles was performed manually or semi-automatically, which restricts the applicability of the proposed approach for real-time clinical screening scenarios. Furthermore, the current framework relies exclusively on gait kinematic data. While these features provide valuable insights into motor behavior, they may not fully capture the broader spectrum of behavioral indicators associated with ASD.

Future research should focus on expanding the dataset to include a larger and more diverse population in order to improve model robustness and generalization capability. Another important direction involves the development of fully automated gait segmentation pipelines that would enable real-time analysis in clinical or educational environments. Additionally, integrating multimodal behavioral data-such as facial expressions, speech characteristics, and physiological signals-could further enhance diagnostic accuracy and provide a more comprehensive representation of neurodevelopmental patterns. Future studies may also explore lightweight deep learning architectures and edge-computing solutions to facilitate practical deployment, while improving model interpretability to support clinical decision-making.

## 4. CONCLUSION

This study presented a hybrid deep learning framework for automated detection of ASD using gait kinematic data obtained from a non-invasive motion capture system. The proposed architecture integrates Conv1D, BiLSTM, and multi-head attention mechanisms, enabling effective extraction of spatial and temporal features associated with ASD-related motor patterns. Experimental results demonstrate that the proposed model achieves high classification performance, reaching an accuracy of up to 95% and an AUC exceeding 0.97, while consistently outperforming a baseline LSTM architecture. The results highlight the potential of gait-based behavioral analysis as an objective and non-invasive approach for early ASD screening and demonstrate the value of hybrid deep learning models for spatiotemporal motion analysis. Overall, the proposed framework provides a promising foundation for the development of intelligent decision-support tools in neurodevelopmental diagnostics.


## FUNDING INFORMATION

This research received no external funding.


## 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 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gulzat Ziyatbekova | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | | ✓ | |
| Svetlana Beglerova | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | |
| Dinara Zhukenova | | | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ | ✓ | | ✓ |
| Alokhon Alikarieva | ✓ | | | | | ✓ | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| Madi Akhmetzhanov | | | | | ✓ | ✓ | | | | ✓ | | ✓ | | ✓ |
| Nuriddin Alikariev | | ✓ | | | | | ✓ | ✓ | | ✓ | | ✓ | | |
| Quvvatali Rakhimov | ✓ | | ✓ | | ✓ | | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| Yersultan Tulebayev | | ✓ | ✓ | | ✓ | ✓ | | | | ✓ | | ✓ | | ✓ |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The dataset is publicly available at the Dryad digital repository (https://doi.org/10.5061/dryad.s7h44j150) [26]. Processed data and code used in this study are available from the authors upon reasonable request.

## REFERENCES


[1] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 5th ed. Washington, DC, USA: American Psychiatric Association Publishing, 2013.

[2] M. J. Maenner *et al.*, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network," *Surveillance Summaries*, vol. 70, no. 11, pp. 1–16, 2021.

[3] U. J. Ganai, A. Ratne, B. Bhushan, and K. S. Venkatesh, “Early detection of autism spectrum disorder: gait deviations and machine learning,” *Scientific Reports*, vol. 15, no. 1, Jan. 2025, doi: 10.1038/s41598-025-85348-w.

[4] 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, Jan. 2023, doi: 10.3390/ijms24031819.

[5] Y. Ohmoto, K. Terada, H. Shimizu, H. Kawahara, R. Iwanaga, and H. Kumazaki, “Machine learning’s effectiveness in evaluating movement in one-legged standing test for predicting high autistic trait,” *Frontiers in Psychiatry*, vol. 15, Oct. 2024, doi: 10.3389/fpsyt.2024.1464285.

[6] S. A. Yazdi, A. Janghorbani, and A. Maleki, “Diagnosis of autism in children based on their gait pattern and movement signs using the Kinect sensor,” *Journal of Medical Signals & Sensors*, vol. 14, no. 10, Oct. 2024, doi: 10.4103/jmss.jmss_19_24.

[7] S. M. S. Aljabiri and M. M. Hamdan, “Analyzing lower body movements using machine learning to classify autistic children,” *Biomedical Signal Processing and Control*, vol. 94, Aug. 2024, doi: 10.1016/j.bspc.2024.106288.

[8] S. Raj and S. Masood, “Analysis and detection of autism spectrum disorder using machine learning techniques,” *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020, doi: 10.1016/j.procs.2020.03.399.

[9] R. Ranjan, D. Ahmedt-Aristizabal, M. A. Armin, and J. Kim, “Computer vision for clinical gait analysis: A gait abnormality video dataset,” *IEEE Access*, vol. 13, pp. 45321–45339, Jul. 2025, doi: 10.1109/ACCESS.2025.3545787.

[10] T. Barami *et al.*, “Automated identification and quantification of stereotypical movements from video recordings of children with ASD,” *bioRxiv*, Mar. 2024, doi: 10.1101/2024.03.02.582828.

[11] A. Bawa, K. Banitsas, and M. Abbod, “A review on the use of Microsoft Kinect for gait abnormality and postural disorder assessment,” *Journal of Healthcare Engineering*, vol. 2021, pp. 1–19, 2021, doi: 10.1155/2021/4360122.

[12] J. Li, W. Liang, X. Yin, J. Li, and W. Guan, “Multimodal gait abnormality recognition using a convolutional neural network–bidirectional long short-term memory (CNN-BiLSTM) network based on multi-sensor data fusion,” *Sensors*, vol. 23, no. 22, Nov. 2023, doi: 10.3390/s23229101.

[13] G. Cicirelli, D. Impedovo, V. Dentamaro, R. Marani, G. Pirlo, and T. R. D’Orazio, “Human gait analysis in neurodegenerative diseases: A review,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 229–242, Jan. 2022, doi: 10.1109/JBHI.2021.3092875.

[14] J. Gao, P. Gu, Q. Ren, J. Zhang, and X. Song, “Abnormal gait recognition algorithm based on LSTM–CNN fusion network,” *IEEE Access*, vol. 7, pp. 163180–163190, 2019, doi: 10.1109/ACCESS.2019.2950254.

[15] 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*, vol. 13, pp. 2903–2928, 2025, doi: 10.1109/ACCESS.2024.3523872.

[16] P. Lee, T.-B. Chen, H.-Y. Lin, L.-R. Yeh, C.-H. Liu, and Y.-L. Chen, “Integrating OpenPose and SVM for quantitative postural analysis in young adults: A temporal–spatial approach,” *Bioengineering*, vol. 11, no. 6, May 2024, doi: 10.3390/bioengineering11060548.

[17] A. Amirbay *et al.*, “Autism detection using facial and motor analysis using machine learning,” *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 5, pp. 3985–4000, Oct. 2025, doi: 10.11591/eei.v14i5.10319.

[18] S. Zhang, J. Zhang, W. Song, L. Yang, and X. Zhao, “Hierarchical-attention-based neural network for gait emotion recognition,” *Physica A: Statistical Mechanics and its Applications*, vol. 637, Mar. 2024, doi: 10.1016/j.physa.2024.129600.

[19] S. M. Al-Selwi *et al.*, “RNN-LSTM: From applications to modeling techniques and beyond—Systematic review,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 5, Jun. 2024, doi: 10.1016/j.jksuci.2024.102068.

[20] N. K. Zakaria, N. M. Tahir, and R. Jailani, “Experimental approach in gait analysis and classification methods for autism spectrum disorder: A review,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 3995–4005, Jun. 2020, doi: 10.30534/ijatcse/2020/224932020.

[21] Z. J. Peya, M. A. H. Akhand, J. F. Srabonee, and N. Siddique, “Autism Detection from 2D Transformed EEG Signal using Convolutional Neural Network,” *Journal of Computer Science*, vol. 18, no. 8, pp. 695–704, Aug. 2022, doi: 10.3844/jcssp.2022.695.704.

[22] M. Trăscău, M. Nan, and A. M. Florea, “Spatio-temporal features in action recognition using 3D skeletal joints,” *Sensors*, vol. 19, no. 2, Jan. 2019, doi: 10.3390/s19020423.

[23] F. Zhu, S. Wang, W. Liu, H. Zhu, M. Li, and X. Zou, “A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name,” *Frontiers in Psychiatry*, vol. 14, Jan. 2023, doi: 10.3389/fpsyt.2023.1039293.

[24] R. Kolaghassi, M. K. Al-Hares, G. Marcelli, and K. Sirlantzis, “Performance of deep learning models in forecasting gait trajectories of children with neurological disorders,” *Sensors*, vol. 22, no. 8, Apr. 2022, doi: 10.3390/s22082969.

[25] N. M. Rad *et al.*, “Deep learning for automatic stereotypical motor movement detection using wearable sensors in autism spectrum disorders,” *Signal Processing*, vol. 144, pp. 180–191, Mar. 2018, doi: 10.1016/j.sigpro.2017.10.011.

[26] M. Al-Jubouri *et al*., “Three-dimensional dataset combining gait and full body movement of children with autism spectrum disorder collected by Kinect v2 camera,” *Dryad Digital Repository: Davis*, 2020, doi: 10.5061/dryad.s7h44j150.

## BIOGRAPHIES OF AUTHORS

**Gulzat Ziyatbekova** Ph.D. in computer science, computer engineering and management, associate professor, a senior researcher at the RSE Institute of Information and Computational Technologies of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Associate Professor Al-Farabi Kazakh National University. She has co-authored more than 50 articles in journals and conference proceedings. Her professional interests are information systems, computer science, scientific research, and mathematical modeling. She can be contacted at email: ziyatbekova1@gmail.com.

**Svetlana Beglerova** is a candidate of technical sciences, acting associate professor of the Department of Information Systems, Taraz University named after M.Kh. Dulati. Currently, she is the head of the online learning department of the same university. She has more than 25 years of scientific and pedagogical experience, more than 70 scientific papers, including 6 articles in the Scopus database, 6 copyright certificates for programs and developments. Her professional interests include computer science, information systems, artificial intelligence, and machine learning. She can be contacted at email: nbaygabylov@gmail.com.

**Dinara Zhukenova** in 2014 graduated from the West Kazakhstan Agrarian-Technical University named after Zhangirkhan, speciality "Professional training". In 2016 she received a master's degree in ‘Information Systems’. Since 2021 till now she is a senior lecturer, master of the Department of Energy, Automation and Informatics of the West Kazakhstan Innovation and Technology University. She is author of more than 20 scientific articles and more than 10 educational and methodical works. Area of scientific publications: artificial intelligence, mobile applications, and information security. She can be contacted at email: clever_-92@mail.ru

**Alokhon Alikarieva** is an associate professor at the National University of Uzbekistan named after Mirzo Ulugbek, and a doctor of sociological sciences. She conducted a number of sociological studies as a project manager and participant in 4 grants from the Agency for Science and Technology. She has over 25 years of scientific and pedagogical experience, over 200 scientific papers, including 3 articles in the Scopus database, 3 textbooks, 3 monographs, 2 teaching aids, and 2 methodological manuals. Her professional interests include sociological sciences, economics, modeling, and management. She can be contacted at email: alohon77@ya.ru.

**Madi Akhmetzhanov** candidate of physical and mathematical sciences in the field of information technologies, artificial intelligence and data analysis, applied mathematical sciences and management in education, head of the department at Taraz University named after M.Kh.Dulaty. He is a co-author of more than 45 articles in journals and conference proceedings. His professional interests include artificial intelligence, big data, and mathematical modeling. He can be contacted at email: madiktz@gmail.com.

**Nuriddin Alikariev** is a professor at the National University of Uzbekistan named after Mirzo Ulugbek, Doctor of Economics. He conducted a number of sociological studies as a project manager and participant in 4 grants of the Agency for Science and Technology. He has over 30 years of scientific and pedagogical experience, over 300 scientific papers, including 3 articles in the Scopus database, 3 textbooks, 3 monographs, 2 teaching aids, 2 methodological manuals. Area of professional interests: sociological sciences, economics, modeling, forecasting, and management. He can be contacted at email: alinur1937@gmail.com.

**Quvvatali Rakhimov** head of the Department of Applied Mathematics and Computer Science at Fergana State University, Uzbekistan. He holds a Ph.D. in Engineering Sciences. His research interests include applied mathematics, computational modeling, and intelligent systems. He has published several papers in Scopus- and Web of Science-indexed journals and actively participates in international conferences and research collaborations. He can be contacted at email: quvvatali.rahimov@gmail.com.

**Yersultan Tulebayev** is a Candidate for the Doctor of Philosophy (Ph.D.) in Information Systems and a Senior Lecturer at Astana IT University. He is also a Bolashak International Scholar, who undertook an academic internship at the University of Illinois Urbana-Champaign (USA), specializing in Data Science and Information Sciences. His research focuses on the development of intelligent information systems for asset management, technological renewal of enterprises, and predictive maintenance using data-driven methods and machine learning. He can be contacted at email: yersultan.tulebayev@astanait.edu.kz.