Received 20 May 2026, accepted 11 June 2026, date of publication 15 June 2026, date of current version 25 June 2026.
Digital Object Identifier 10.1109/ACCESS.2026.3704046

RESEARCH ARTICLE

A Hybrid Multi-Branch Deep Learning Model for Autism Spectrum Disorder Detection Using Spatiotemporal Gait Analysis

AYAGOZ MUKHANOVA[1], AIZAT AMIRBAY[1], ZHAZIRA TASZHUREKOVA[2], ABDUGANI KHOLBEKOV[3], KHILOLA KHAYDAROVA[4], SABOKHAT KALANOVA[3], GULZIRA ABDIKERIMOVA[1], AIDANA AKHMETOVA[1], ROHAYA LATIP[5], AND MAKBAL BAIBULOVA[1]

[1]Department of Information Systems, L. N. Gumilyov Eurasian National University, Astana 010000, Kazakhstan
[2]Department of Applied Informatics and Programming, M. Kh. Dulaty Taraz University, Taraz 080000, Kazakhstan
[3]Department of Sociology, National University of Uzbekistan named after Mirzo Ulugbek, Tashkent 100174, Uzbekistan
[4]Samarkand State University named after Sharof Rashidov, Samarkand 140163, Uzbekistan
[5]Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Seri Kembangan 43400, Malaysia

Corresponding authors: Aizat Amirbay (amirbay.aiz@gmail.com) and Zhazira Taszhurekova (zhk.taszhurekova@dulaty.kz)

This work was supported by the Committee of Science of the Ministry of Science and Higher Education of Republic of Kazakhstan through Research Project, ''Development of Artificial Intelligence Models and Algorithms for Analyzing Social Signals in Children with Autism'' under Grant AP26105045.

ABSTRACT Early and objective diagnosis of autism spectrum disorders (ASD) remains a pressing issue in modern medical and cognitive informatics, where motor behavior analysis is considered a promising source of digital biomarkers. In particular, kinematic gait characteristics obtained with contactless sensors enable the identification of subtle motor impairments that may not be apparent on visual assessment. In this paper, we propose a hybrid multi-branch deep learning model, HMB-TSC, for the automatic classification of children with ASD based on spatiotemporal gait data extracted from the Kinect v2 camera. An open dataset containing gait recordings of children with ASD and typically developing (TD) peers was used for the experiments. Data preprocessing included outlier filtering, feature normalization, feature expansion, and time-series segmentation using a sliding-window method. The HMB-TSC architecture combines three complementary branches: a Transformer for modeling global temporal dependencies, a CNN-BiLSTM for extracting local spatiotemporal patterns, and a CNN-BiGRU for efficient sequential analysis. The performance evaluation was conducted using the holdout, 5-fold, and 5×3 nested validation protocols. In a reproducible released run, HMB-TSC achieved Accuracy = 0.850, AUROC = 0.948, and F1 = 0.880 in holdout validation. Under 5-fold and 5×3 nested cross-validation, the model achieved Accuracy = 0.850 $\pm$ 0.035, AUROC = 0.940 $\pm$ 0.029, and F1 = 0.845 $\pm$ 0.036. Compared with the baseline CNN, LSTM, and Transformer models, the proposed model demonstrated strong and stable overall performance across validation protocols, especially in terms of AUROC and balanced F1-score.

INDEX TERMS Autism spectrum disorder, gait analysis, hybrid deep learning, kinematic features, multi-branch architecture, screening, spatiotemporal data.

I. INTRODUCTION

Autism spectrum disorders (ASD) are common neurodevelopmental disorders characterized not only by deficits in social communication but also by persistent motor and sensorimotor abnormalities that often manifest early in development. Despite advances in clinical practice, early and objective identification of ASD remains a global scientific and medical challenge, as standard diagnostic protocols, based primarily on clinical observations and parent

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

questionnaires, are subjective, labor-intensive, and have limited scalability [1]. These limitations have spurred growing interest in computational biomarkers derived from behavioral signals, such as eye movements, facial expressions, posture, and body kinematics, which enable inexpensive, noninvasive, and scalable screening methods, particularly in children with limited verbal communication [2], [3]. Recent reviews further emphasize the clinical relevance of objective behavioral biomarkers for early ASD screening [9] while also highlighting the limitations of subjective diagnostic procedures and the growing role of computational motor biomarkers in neurodevelopmental disorders [24].

Among digital phenotyping approaches, particular attention in recent years has been paid to the analysis of gait kinematics, represented as three-dimensional joint trajectories. Neurobiological and biomechanical studies show that children with ASD exhibit characteristic gait disturbances, including reduced stride length, decreased symmetry, increased coordination variability, and atypical interarticular synchronization [4], [5], [6]. The use of modern non-contact sensor technologies, such as depth cameras and RGB-based pose estimation algorithms, enables the recording of multidimensional gait parameters in a marker-free and noninvasive manner, expanding the scope of motion analysis beyond specialized laboratories [7]. Despite growing interest in this area, a unified approach to computational gait analysis in ASD is lacking in the literature. Methods based on manually constructed spatiotemporal features are often insufficient to describe the nonlinear and multiscale structure of human movement [8]. Deep learning architectures, including convolutional neural networks, recurrent models (LSTM, GRU), graph networks, 3D-CNN, and Transformer approaches, demonstrate improved classification performance but yield inconsistent results regarding the most informative temporal representations [9], [10]. Furthermore, it has been shown that model quality depends significantly on data collection conditions, preprocessing steps, and validation protocols, which negatively affect reproducibility and limit the clinical applicability of existing solutions [7], [11]. Recent methodological reviews in biomedical informatics further emphasize that reproducibility, robustness, explainability, and standardized validation protocols remain critical challenges for the clinical translation of AI-driven ASD diagnostic systems [34]. Most existing models focus either on local temporal patterns extracted by convolutional filters or on global temporal dependencies modeled by attention mechanisms. Recurrent networks are capable of capturing sequential dynamics but struggle with high-dimensional, multi-joint trajectories. Graph models, in turn, are sensitive to the given skeletal topology and generalize poorly to changes in sensor configuration or recording conditions [12], [13], [14], [15], [16], [17], [18], [19]. Modern hybrid architectures demonstrate improved performance, while existing approaches typically fail to integrate global temporal relationships, local spatiotemporal representations, and directed sequential dependencies within a single framework [20], [21], [22], [23], [24], [25]. Recent advances in deep learning have further demonstrated the effectiveness of data-driven approaches for analyzing complex behavioral and neurological patterns in various neurodevelopmental and neurodegenerative disorders. In the case of ASD, recent studies have explored multimodal deep learning frameworks capable of capturing subtle behavioral and motor abnormalities from high-dimensional data [26]. Similarly, explainable artificial intelligence (XAI) methods have been successfully applied to Parkinson's disease detection, providing not only accurate classification but also clinically interpretable data on movement-related biomarkers [27]. In the field of emotional and behavioral analysis, LSTM-based architectures have demonstrated a high ability to model temporal dependencies in affective signals, enabling improved recognition of emotional states and behavioral patterns [28]. Furthermore, GRU-based models have shown promising results in the analysis of neurodegenerative conditions such as Alzheimer's disease, effectively capturing the temporal progression of cognitive and physiological signals [29]. Taken together, these studies highlight the growing importance of hybrid and sequence-based deep learning architectures for modeling complex temporal dynamics in biomedical and behavioral data, further stimulating the development of the proposed multi-component framework for the detection of ASD.

In this paper, we propose a hybrid multi-branch time series classifier, HMB-TSC, which combines three complementary modeling paradigms: Transformer for capturing global temporal dependencies, CNN-BiLSTM for encoding local spatiotemporal patterns, and CNN–BiGRU for directed sequential analysis. The goal of this study is to develop and experimentally evaluate an architecture that provides a multi-scale representation of gait kinematics in children with ASD and typical development. To achieve this goal, we address the challenges of generating a cleaned and segmented gait dataset, developing a hybrid architecture, conducting rigorous quality assessment using holdout sampling, k-fold cross-validation, and nested cross-validation, and analyzing the contributions of individual model branches. The scientific novelty of this work lies in integrating Transformer, CNN-BiLSTM, and CNN-BiGRU branches within a unified framework for gait-based ASD classification. The proposed architecture captures complementary temporal and spatiotemporal patterns and was evaluated using holdout, k-fold, and nested cross-validation protocols. This design provides a robust multi-scale representation of gait dynamics for automated ASD screening.

II. METHODS

A. DATA PREPARATION

To address ambiguities related to sample expansion and potential data leakage, the Datasets and Preprocessing section was clarified as follows. This study used the open dataset "Three Dimensional Dataset Combining Gait and

Full Body Movement of Children with Autism Spectrum Disorders Collected by Kinect v2 Camera" [15], which includes raw gait recordings from 100 children: 50 children with ASD and 50 TD children (https://datadryad.org/dataset/doi:10.5061/dryad.s7h44j150).

To improve the robustness of the proposed framework to variability in gait patterns, time-series augmentation was applied based on the protocol described by Al-Jubouri, Ali, and Rajihy [15]. Augmentation was performed only after subject-level splitting of the dataset. Synthetic variants were generated exclusively for the training subsets within each evaluation split, while validation and test samples remained completely unaugmented. This strategy prevented data leakage and eliminated the possibility of participant-derived variants appearing simultaneously in both training and evaluation subsets. The implemented augmentation pipeline included seven transformations: coordinate scaling, temporal warping using B-spline interpolation, Gaussian noise injection, horizontal mirroring, random rotation, sequential scaling with noise injection, and sequential temporal warping with noise injection. All reported evaluation metrics were computed exclusively on original non-augmented recordings.

Crucially, augmentation was not used as a preliminary step before the overall dataset partitioning. In all experiments, partitioning by child identifier was performed first, and then augmentation was applied only to the training dataset. Accordingly, synthetic instances were never generated for test observations and were not included in the evaluation set. This processing procedure was identical for both the hold-out evaluation and the 5-fold and nested cross-validations. For cross-validation, group-wise 5-fold cross-validation was used, with the participant identifier serving as the grouping feature. This means that all records for the same child, including all augmented variants derived from them, could only be found in one subset of the current split. In other words, if a specific participant's original trajectory fell into the test fold, then none of its augmented versions could end up in the train fold, and vice versa. Formally, for each split, the condition of no intersection between participant identifiers in the training and test sets is enforced (1):

$$|\, ID_{train} \cap ID_{test} \,| = 0 \tag{1}$$

where ID_train and ID_test denote the sets of participant identifiers in the training and test subsets, respectively. This protocol ensured complete separation between participants across all evaluation splits and prevented data leakage. Augmentation was applied exclusively to the training data, whereas all validation and test metrics were calculated using only original recordings. Consequently, the reported performance reflects the model's ability to generalize to previously unseen participants.

The Kinect v2 sensor was selected due to its low cost, portability, and marker-free acquisition capabilities, making it suitable for practical ASD screening scenarios. Previous studies report average joint localization errors of approximately 13–15 mm compared with Vicon systems, which were further mitigated in this study through normalization, temporal smoothing, relative kinematic features, and augmentation-based regularization.

To improve the diversity of the training data, seven augmentation strategies were applied, including magnitude scaling, time warping, jittering, horizontal mirroring, random rotation, and two composite transformations. Augmentation was performed exclusively on the training subsets after subject-level partitioning, while validation and test samples remained unaugmented. The augmentation methods and corresponding hyperparameters are summarized in Table 1.

TABLE 1. Summary of data augmentation methods and parameter settings for gait time-series data.

No.	Method	Description	Hyperparameters
1	Magnitude Scaling (MS)	Global scaling of joint coordinates	Scaling factor ∈ [0.9, 1.1]
2	Time Warping (TW)	Nonlinear temporal deformation using B-splines	σ = 0.2, 3 control points
3	Jittering (JT)	Gaussian noise injection	μ = 0, σ = 0.01 m
4	Horizontal Mirroring (HM)	Reflection across sagittal plane	X-coordinate reflection
5	Random Rotation (RR)	Rotation around vertical axis	θ ∈ [−5°, 5°]
6	Scaling + JT	Sequential application of MS followed by JT	α ∈ [0.9, 1.1]; μ = 0; σ = 0.01 m
7	TW + JT	Sequential application of TW followed by JT	σ = 0.2; 3 control points; μ = 0; σ = 0.01 m

A summary of demographic and trajectory-related characteristics is reported in Table 2.

TABLE 2. Demographic and gait-related characteristics of the study cohort.

Group	Sex	Mean Age (years)	Mean Trajectory Length (m)
TD	26 boys 24 girls	boys: 7.8 girls: 5.4	boys: 3.98 girls: 3.49
ASD	42 boys 8 girls	boys: 8.19 girls: 8.66	boys: 1.299 girls: 1.308

For subsequent processing and model development, only the XLSX-format files were used. Each file contains a time-series representation of two-dimensional joint coordinates recorded during a single walking sequence. The data are organized in tabular form, where rows correspond to consecutive temporal frames and columns represent joint-specific coordinate values.

As summarized in Table 3, a single XLSX file corresponds to one complete gait cycle and consists of 37 frames acquired at a uniform temporal resolution of 33 ms, which is consistent with the standard acquisition rate of the Kinect v2 sensor. Temporal alignment is ensured through the Time_ms column, which provides an explicit timestamp for each frame and allows precise synchronization of the time series.

The total dimensionality of a single sample comprises 51 features, including the temporal index and 25 pairs of (X, Y) coordinates associated with anatomically meaningful joints spanning the upper limbs, lower limbs, trunk, and head. To reduce noise and avoid redundancy, only the planar (X, Y) components of the joint trajectories were used. Due to the lateral position of the Kinect v2 sensor, the Z-axis primarily reflects forward motion and is highly correlated with temporal progression rather than intra-cycle gait dynamics. Furthermore, depth measurements exhibit higher error (≈2-5 mm RMS), which can reduce the signal-to-noise ratio. Previous studies show that key gait characteristics relevant for ASD classification are effectively captured by planar components, while our preliminary experiments showed that including the Z-coordinate yields only minor and statistically insignificant improvements. Therefore, restricting the representation to (X, Y) improves robustness and computational efficiency without sacrificing discriminatory performance. Therefore, each file represents a separate instance of a two-dimensional kinematic gait trajectory, forming the basis for subsequent feature extraction and model training.

TABLE 3. Structure of a single gait-cycle file.

No.	Feature	Description
1	Time_ms	Timestamp (ms)
2	MidSpine_X/Y	Mid-spine coordinates
3	Neck_X/Y	Neck coordinates
4	Head_X/Y	Head coordinates
5	SpineBase_X/Y	Base of spine
6	SpineShoulder_X/Y	Upper spine
7	ShoulderLeft_X/Y,ShoulderRight_X/Y	Shoulder joints
8	ElbowLeft_X/Y, ElbowRight_X/Y	Elbow joints
9	WristLeft_X/Y, WristRight_X/Y	Wrist joints
10	HandLeft_X/Y, HandRight_X/Y	Hands
11	HandTipLeft_X/Y,HandTipRight_X/Y	Fingertips
12	ThumbLeft_X/Y, ThumbRight_X/Y	Thumbs
13	HipLeft_X/Y, HipRight_X/Y	Hip joints
14	KneeLeft_X/Y, KneeRight_X/Y	Knees
15	AnkleLeft_X/Y, AnkleRight_X/Y	Ankles
16	FootLeft_X/Y, FootRight_X/Y	Feet

As a preprocessing step, a data cleaning procedure was applied to eliminate artifacts caused by tracking losses and inaccurate joint localization. For each temporal frame, the physiological plausibility of joint coordinates was examined, and values falling outside realistic anatomical ranges were treated as sensor-induced errors and excluded from further analysis. Outlier filtering was performed at the frame level. For each feature, outliers were identified using Tukey's interquartile range (IQR) test. Specifically, values outside the interval [Q1 - 1.5 IQR, Q3 + 1.5 IQR] were considered anomalous. If even one joint coordinate in a given frame violated this condition, the entire frame was considered a tracking artifact and removed. This criterion was chosen due to its robustness to non-Gaussian distributions and suitability for biomechanical data acquired with the Kinect. This conservative filtering strategy resulted in the removal of less than 2% of the total observations, while preserving the original statistical and demographic characteristics of the dataset and maintaining the overall structure of informative features. The effectiveness of the cleaning procedure was quantitatively evaluated using mutual information (MI) between individual features and the binary class label. The ten most informative variables before and after filtering are reported in Table 4. The results show only negligible changes in MI values ($\Delta MI \leq 0.005$), indicating that the filtering procedure preserved the relative importance of the most informative gait features. Notably, vertical coordinates of the upper torso and shoulder region remained the strongest contributors to ASD-TD discrimination, with *ShoulderRight_y* retaining the highest relevance (MI ≈ 0.38).

TABLE 4. Top-10 features by mutual information with the class label before and after outlier filtering.

Feature	MI (original)	MI (filtered)
ShoulderRight_y	0.383847	0.381441
MidSpine_y	0.374316	0.369399
Neck_y	0.371170	0.369508
SpineShoulder_y	0.370991	0.370866
ShoulderLeft_y	0.364928	0.361637
ElbowRight_y	0.349527	0.348560
WristRight_y	0.329146	0.324886
ElbowLeft_y	0.321191	0.317790
HandTipRight_y	0.306640	0.304373
HandRight_y	0.305684	0.303232

To further assess the impact of filtering on the global dataset structure, descriptive statistics were compared before and after outlier removal. As summarized in Table 5, the filtering procedure removed only 55 records (1.84% of the dataset) while preserving statistically significant differences in age and trajectory length between the ASD and TD groups. These findings indicate that the preprocessing step improved data quality without substantially altering the overall dataset structure.

TABLE 5. Summary comparison of the dataset before and after outlier filtering.

Metric	Before filtering	After filtering
Rows removed	–	55 (1.84%)
t-test (Age)	p = 0.0004; ASD = 8.19; TD = 7.20	p = 0.0004; ASD = 8.62; TD = 6.66
t-test (Trajectory length, m)	p < 0.0001; ASD = 1.31; TD = 3.99	p < 0.0001; ASD = 1.32; TD = 3.75

After outlier removal, joint coordinates were normalized to reduce inter-individual variability caused by differences in body size and recording conditions. Additional statistical and kinematic descriptors were computed, and Euclidean norms were used to obtain a compact representation of gait dynamics. Each gait trajectory was temporally normalized to a fixed length of 256 time steps, resulting in a standardized sequence of size 256 × 25. Since each XLSX file represented a complete gait cycle, a single non-overlapping sliding window ($w = 256$, $s = 256$) was applied to each recording. Data augmentation was performed exclusively on the training subsets, whereas all validation and test metrics were calculated using only original recordings. The resulting dataset provided a homogeneous and reproducible set of time-series samples for training and evaluating the baseline and proposed HMB-TSC models. To prevent data leakage, all data partitioning was performed at the subject level. Augmented samples derived from the same participant were assigned exclusively to a single fold, ensuring complete separation between training and test data. Data augmentation was applied only to the training subsets, while validation and test samples remained unchanged. Consequently, the reported performance reflects the model's ability to generalize to previously unseen participants. The gender distribution of participants before filtering is shown in Figure 1.

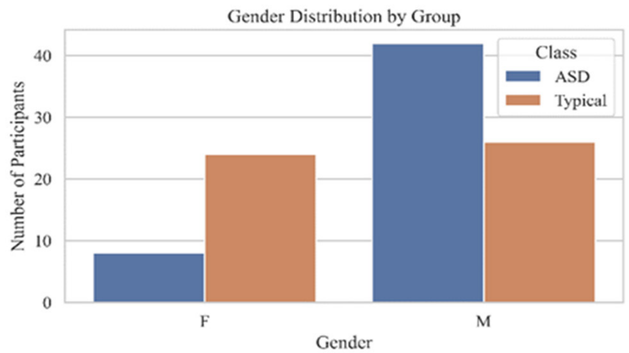


FIGURE 1. Gender distribution of participants before outlier filtering.

The group of individuals with ASD exhibits a significant male predominance, while the gender balance remains roughly balanced in the TD group. This pattern persists after pre-processing, confirming the absence of demographic bias caused by the filtering procedure. The gender distribution after applying the outlier filtering procedure is presented in Figure 2.

Removing a small number of records (1.84%) does not affect the overall composition of the classes: the ASD group remains predominantly male, while the typically developing group maintains a roughly balanced structure, indicating the absence of demographic bias in the filtering process.The age distribution before and after preprocessing is presented in Figures 3 and 4.

The ASD cohort is generally concentrated in middle childhood, while the typically developing group exhibits a bias toward younger ages. Importantly, the overall distribution patterns remain stable after filtering, confirming that the preprocessing step preserves the age characteristics of the dataset.

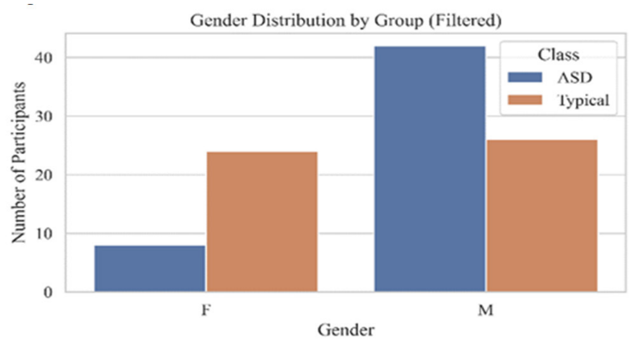


FIGURE 2. Gender distribution of participants after outlier filtering.

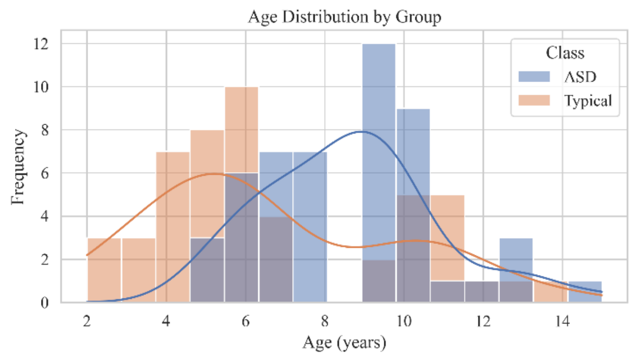


FIGURE 3. Age distribution of participants before outlier filtering.

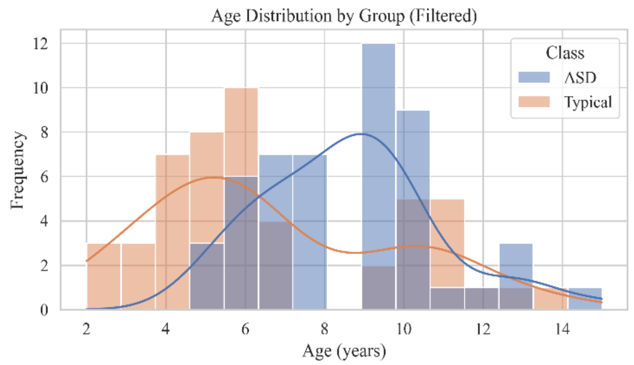


FIGURE 4. Age distribution of participants after outlier filtering.

Figures 5 and 6 present a comparison of age distributions as boxplots. The ASD group exhibits a higher median age, while both groups exhibit moderate dispersion and partial overlap. Importantly, the distribution characteristics remain stable after outlier removal, indicating that preprocessing eliminates extreme values without altering the underlying demographic structure.

Since the age and gender distributions are already visualized in Figures 1-6, detailed text descriptions have been

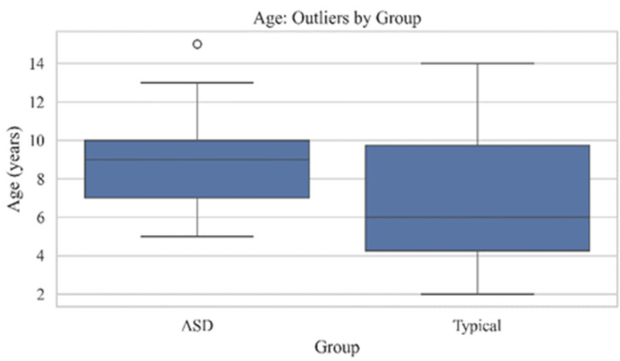


FIGURE 5. Age outlier distribution by group before outlier filtering.

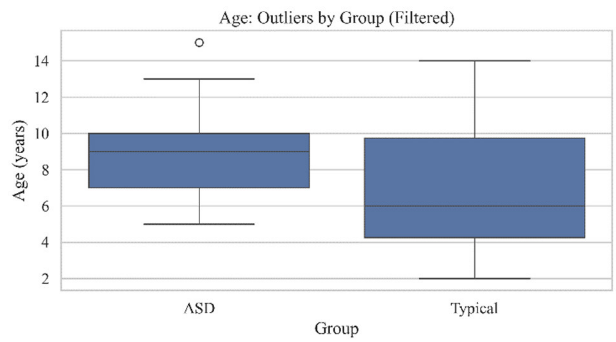


FIGURE 6. Age outlier distribution by group after outlier filtering.

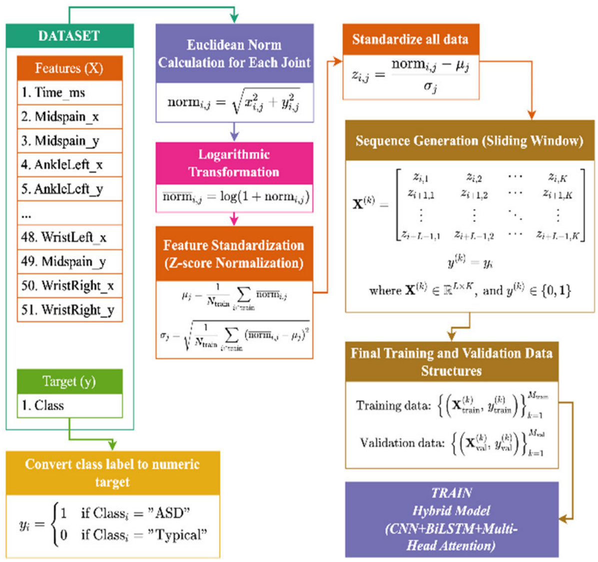


FIGURE 7. Algorithmic pipeline for feature extraction and time-series construction from joint coordinate data.

intentionally kept to a minimum. Overall, the demographic structure remains unchanged after preprocessing: the ASD group exhibits a higher proportion of male participants, while the typically developing group remains approximately balanced. These characteristics should be considered as contextual factors when interpreting model performance and assessing its generalizability.

B. FEATURE ENGINEERING AND TIME-SERIES CONSTRUCTION

An informative feature space was constructed, and temporal sequences were organized based on the cleaned and normalized joint coordinates obtained in the previous preprocessing step. The objective of this stage is to transform raw kinematic measurements into a compact, structured representation that preserves the spatial and temporal characteristics of gait dynamics while remaining suitable for deep learning-based sequence modeling. The complete feature extraction and time-series construction pipeline is illustrated in Figure 7, which summarizes the sequence of transformations applied to the original joint coordinate data and their conversion into fixed-length time-series inputs used for model training.

1) STRUCTURE OF THE INPUT GAIT DATASET

Let the original gait dataset be defined as shown in (2)

$$D = \{(x_i, y_i)\}_{i=1}^{N} \tag{2}$$

where $\mathrm{x}_i \in \mathbb{R}^{T\times J\times 2}$ denotes the raw joint-coordinate sequence of the i-th gait trial, T is the number of time steps, $J = 25$ is the number of tracked body joints, and each joint is represented by two planar coordinates (x, y). The corresponding class label y_i is defined as a binary variable according to (3):

$$y \in \{0, 1\} \tag{3}$$

where class 1 corresponds to children with ASD and class 0 denotes TD subjects.

2) EUCLIDEAN MAGNITUDE TRANSFORMATION

To reduce the dimensionality of the coordinate representation while preserving essential spatial information, each joint coordinate pair is transformed into a scalar magnitude using the Euclidean distance from the origin. This operation is defined as shown in (4):

$$r_{t,j}^{(i)} = \sqrt{\left(x_{t,j}^{(i)}\right)^2 + \left(y_{t,j}^{(i)}\right)^2} \tag{4}$$

where $r_{t,j}^{(i)}$ is the scalar amplitude of the position of the j-th joint at time t. Next, a logarithmic transformation is applied to stabilize the variance. $\left(x_{t,j}^{(i)} \mid y_{t,j}^{(i)}\right)$ - pairs of plane coordinates for each joint $j = 1, \ldots, J$.

3) LOGARITHMIC TRANSFORMATION

To mitigate the influence of extreme values and stabilize variance across joints with different motion amplitudes, a logarithmic transformation is applied to the Euclidean norms. The transformed feature is computed as defined in (5):

$$\tilde{r}_{t,j}^{(i)} = log(1 + r_{t,j}^{(i)}) \tag{5}$$

where the additive constant ensures numerical stability and avoids undefined values. This transformation compresses large magnitudes and improves the robustness of the subsequent normalization step.

4) FEATURE STANDARDIZATION

To ensure that all features contribute equally during model training, Z-score normalization is applied to each transformed joint feature. The standardized feature is computed according to (6):

$$z_{t,j}^{(i)} = \frac{\tilde{r}_{t,j}^{(i)} - \mu_j^{train}}{\sigma_j^{train}} \tag{6}$$

where μ_j^{train} and σ_j^{train} denote the mean and standard deviation of feature j, estimated exclusively from the training subset. This step aligns all features to a common scale and prevents dominance of joints with larger numerical ranges.

5) SLIDING WINDOW SEQUENCE CONSTRUCTION

The standardized joint sequences were segmented into fixed-length temporal windows using a sliding window approach. Each generated sequence is represented as shown (7):

$$X^{(k)} \in \mathbb{R}^{L \times J} \tag{7}$$

where L denotes the window length and J the number of joint-based features. A stride parameter S is used to control overlap between adjacent windows. Each window inherits a binary class label (8):

$$y^{(k)} = y_i \tag{8}$$

This segmentation strategy preserves temporal continuity and enables the learning of both local motion patterns and longer-range dependencies.

6) FINAL TRAINING AND VALIDATION STRUCTURES

The resulting dataset is partitioned into training and validation subsets as follows (9):

$$D_{train} = \left\{(X^{(k)}, y^{(k)})\right\}_{k=1}^{N_{train}} \}, D_{val} = \left\{(X^{(k)}, y^{(k)})\right\}_{k=1}^{N_{val}} \} \tag{9}$$

These structured representations serve as direct inputs to the baseline and hybrid deep learning architectures described in the following section. The proposed feature engineering and temporal structuring pipeline ensures data consistency, robustness to sensor artifacts, and effective preservation of discriminative spatiotemporal gait patterns relevant to ASD classification.

To quantify the effect of preprocessing on feature distributions, a comparative statistical analysis was performed on the joint-coordinate variables before and after filtering and transformation. The analysis focused on features exhibiting the most pronounced distributional changes, enabling an objective assessment of noise suppression while verifying that physiologically meaningful motion patterns were preserved. Table 6 presents the standard deviation (Std), maximum values (Max), and coefficient of variation (CV) for the selected joint features, where (B) and (A) denote the values before and after filtering, respectively. These metrics provide additional information about the dispersion, behavior of extreme values, and relative variability. The results demonstrate a significant reduction in variability for all features. For example, the standard deviation of Head_x and Neck_x decreases from approximately 1580 to 70, while CV values decrease by more than an order of magnitude (e.g., from 1.884 to 0.100 for AnkleLeft_x), indicating significantly improved data stability. Furthermore, maximum values are reduced from unrealistically high values (exceeding 25,000 units) to physiologically plausible ranges of approximately 1,000 units, confirming the effective removal of sensor-induced artifacts.

TABLE 6. Statistical characteristics of joint-coordinate features affected by outlier filtering.

Feature	Std (B)	Std (A)	Max (B)	Max (A)	CV (B)	CV (A)
MidSpine_x	425.94	71.32	23706.20	1076.55	0.551	0.093
AnkleLeft_x	1573.58	74.76	29777.25	1110.79	1.884	0.100
Head_x	1580.94	71.27	27554.92	1069.07	1.824	0.093
Neck_x	1459.65	70.10	29549.83	1061.56	1.726	0.092
ElbowLeft_y	645.95	62.96	726.20	726.20	1.223	0.115
ShoulderLeft_y	1111.95	56.82	644.78	644.78	2.530	0.115

The most pronounced changes are observed in the features describing the spatial position of the head and limbs, confirming the physiologically based nature of the filtering procedure. A significant reduction in the standard deviation and coefficient of variation indicates effective noise suppression and improved data homogeneity. These results confirm the validity of the adopted data cleaning strategy and justify the use of filtered features for subsequent deep learning-based modeling.

C. BASELINE DEEP LEARNING MODELS FOR GAIT TIME-SERIES CLASSIFICATION

To provide a consistent benchmark for the proposed HMB-TSC architecture, three representative deep learning baselines were implemented: LSTM, 1D-CNN, and Transformer. These models represent recurrent, convolutional, and attention-based paradigms for time-series learning, respectively. All architectures were trained using the same input representation of size 256 × 25 obtained from the preprocessing pipeline described in Section II-B.

1) LSTM BASELINE

The architecture of the baseline LSTM model is presented in Figure 8. The network takes an input tensor of shape 256 × 25 and processes it through a single LSTM layer with 32 hidden units, which encodes temporal dependencies

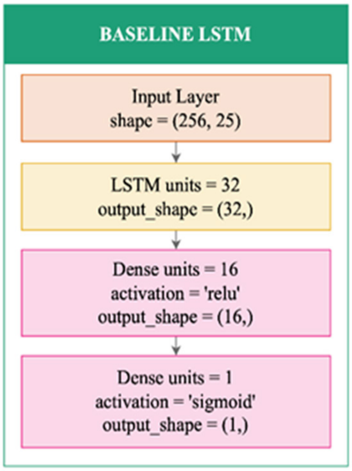


FIGURE 8. Architecture of the baseline single-branch LSTM model for binary gait classification.

across the gait cycle. The resulting representation is fed into a fully connected layer with 16 neurons and ReLU activation to introduce nonlinear feature transformation. Finally, a sigmoid-activated output neuron produces the probability of class membership for binary classification.

2) CNN BASELINE

The 1D-CNN architecture shown in Figure 9 comprises two convolutional layers (32 and 64 filters) separated by max pooling. Global average pooling and a dense layer with ReLU activation are used to generate the final representation before binary classification with a sigmoid output neuron.

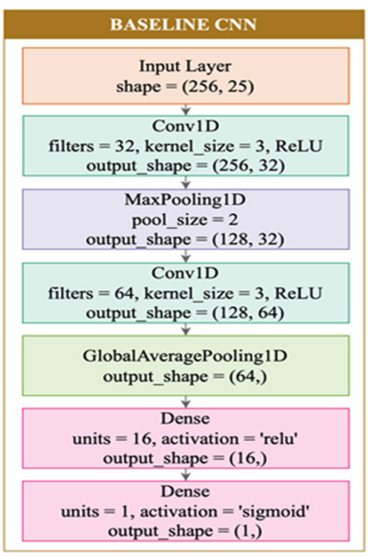


FIGURE 9. Architecture of the baseline 1D-CNN model for gait time-series classification.

3) TRANSFORMER BASELINE

The Transformer baseline model is shown in Figure 10. The input sequence is first projected into a latent embedding space of dimension $d_{model} = 64$ using a dense layer. Positional encoding is added to preserve temporal ordering. The embedded sequence is then processed by a Transformer encoder block consisting of a multi-head self-attention layer with four attention heads, residual connections, and layer normalization. A position-wise feedforward network with hidden dimension 128 is applied to refine the learned representation. After global average pooling along the temporal axis, the resulting 64-dimensional vector is passed through a dense layer with 16 neurons and ReLU activation, followed by a sigmoid output neuron. This model enables explicit learning of long-range temporal relationships within gait sequences.

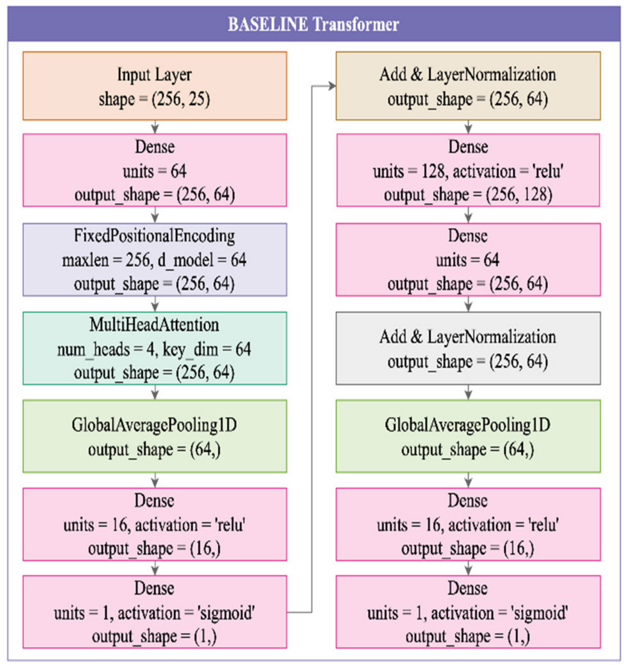


FIGURE 10. Architecture of the baseline Transformer model for binary classification of gait time-series data.

Together, these baseline models provide representative recurrent, convolutional, and attention-based benchmarks, enabling a systematic evaluation of the proposed HMB-TSC architecture for ASD gait classification.

D. ARCHITECTURE OF THE HMB-TSC MODEL

While baseline architectures are effective in capturing specific characteristics of gait time-series signals, their isolated use often limits the representation of heterogeneous and multi-scale motor patterns observed in children with ASD. To overcome this limitation, this study proposes a hybrid multi-branch architecture, referred to as HMB-TSC, which integrates global temporal modeling, local spatiotemporal feature extraction, and bidirectional sequential learning within a unified framework. The overall architecture of the proposed model is illustrated in Figure 11. Input gait segments are represented as (10):

$$X \in \mathbb{R}^{N,256,N_{feat}} \tag{10}$$

where N denotes the batch size, 256 is the number of time steps, and N_{feat}) denotes the number of joint-based features extracted after preprocessing.

The first branch is based on a Transformer encoder and is intended to model long-range temporal dependencies across the gait cycle. The input sequence is first mapped into an internal embedding space using a fully connected layer with 32 units ($d_{model} = 32$). Positional encoding is then added to preserve temporal ordering. The embedded sequence is subsequently processed by four stacked Transformer encoder blocks, each consisting of a multi-head self-attention module with four attention heads, followed by dropout, residual connections, and layer normalization. A feedforward subnetwork further refines the learned representation, and global average pooling is applied to produce a compact 32-dimensional feature vector.

The second branch combines convolutional processing with bidirectional LSTM modeling. A one-dimensional convolutional layer with 16 filters and a kernel size of three is applied to extract localized motion patterns and short-term kinematic variations. Temporal downsampling is performed using max pooling, and the resulting feature maps are passed to a bidirectional LSTM layer with 12 hidden units. This component encodes sequential dependencies in both forward and backward directions, which is important for capturing symmetric and cyclic gait characteristics. Layer normalization and dropout are employed to improve training stability and mitigate overfitting.

The third branch follows a similar structure, but replaces the BiLSTM layer with a bidirectional GRU. This design provides an alternative recurrent representation with reduced computational complexity while maintaining the ability to capture temporal dynamics. The CNN-BiGRU pathway,

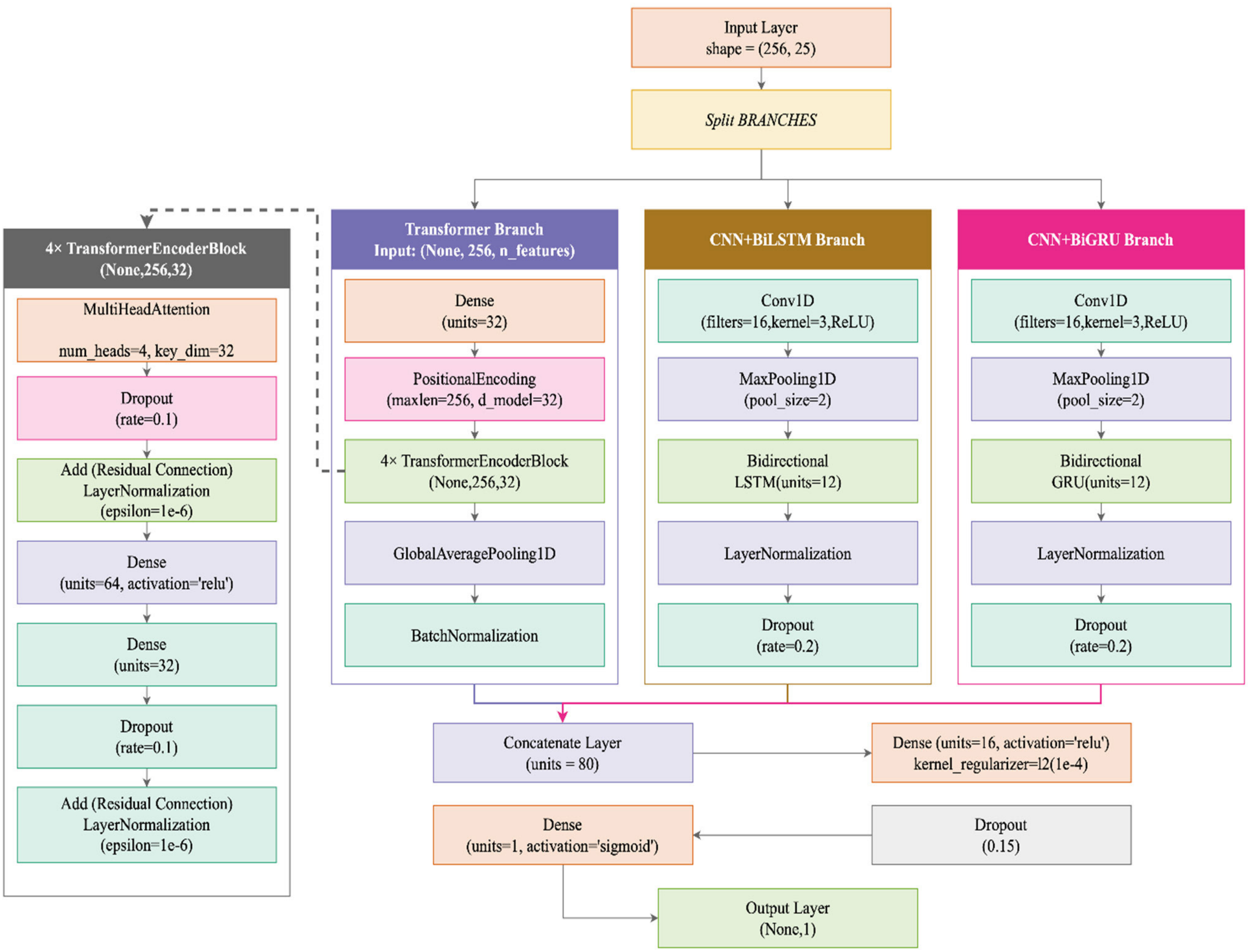


FIGURE 11. Architecture of the proposed HMB-TSC multi-branch model for ASD gait time-series classification.

therefore, serves as a complementary temporal abstraction mechanism, strengthening the overall architecture's robustness.

The results of the three branches are combined into a single vector representation that combines global temporal relationships, local spatiotemporal patterns, and bidirectional sequential information. The fused representation is processed by a fully connected layer with ReLU activation, L2 regularization, and dropout to reduce redundancy and improve generalization. Finally, a sigmoid-activated output neuron produces the probability of ASD class membership. By combining attention-based global modeling with convolutional-recurrent feature learning, the proposed HMB-TSC architecture enables a multi-level representation of gait dynamics. It improves discrimination performance under limited data conditions. This synergistic design improves resilience to sensor noise, enhances discriminative feature learning, and supports effective generalization under limited data conditions, making it well-suited for objective gait-based ASD classification.

To ensure reproducibility and a transparent description of the training procedure, all models were trained using a common set of hyperparameters and optimization strategies. The chosen configuration was determined through empirical evaluation and a search for key parameters, including the learning rate, regularization strength, and training stability constraints. Particular attention was paid to balancing convergence stability and generalization ability, especially under limited data conditions. To reduce overfitting and ensure stable training dynamics, optimization strategies such as adaptive learning rate scheduling, Dropout regularization, and gradient clamping were used. The final set of hyperparameters and training settings adopted for all experiments is presented in Table 7.

To ensure full reproducibility, all baseline models and the proposed HMB-TSC architecture were trained using the same optimization configuration. Specifically, the models were implemented using the same fixed-length input representation of size 256×25 and optimized using the Adam algorithm. The initial learning rate was set to 1×10^{-4}, the batch size was fixed at 32, and training was conducted for 50 epochs for all architectures. Since the problem was formulated as a binary classification with an output neuron activated by a sigmoid function, the binary cross-entropy loss function was used consistently across all experiments. This unified training configuration was adopted to ensure fair comparisons between the LSTM, CNN, Transformer, and HMB-TSC models and to guarantee consistency of experimental conditions across all validation protocols. To promote transparency and facilitate independent verification of the results, all source code associated with this study has been made publicly available. The repository includes the complete implementation of the data preprocessing pipeline, baseline models (LSTM, CNN, Transformer), the proposed HMB-TSC architecture, ablation experiments, and scripts for figure and table generation. The codebase is accessible via the following GitHub repository:

TABLE 7. Hyperparameter configuration and training setup for the HMB-TSC architecture.

Parameter	Value	Rationale
Optimizer	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)	Provides stable convergence for sequences of moderate length
Learning rate	1×10^{-4} with cosine decay schedule ($T_0 = 20$ epochs)	Selected via grid search in the range 1e−3 to 1e−5; lowest validation loss achieved at 1e−4
Batch size	32 sequences ($\approx$ 2.4M parameters per model)	Balances gradient estimation quality and GPU memory usage (8 GB)
Dropout	0.2 after each Conv block; 0.1 in Transformer encoder; 0.3 before final MLP head	Empirically reduces overfitting without degrading convergence
L2 regularization (weight decay)	1×10^{-5}	Prevents weight norm explosion for long sequences
Gradient clipping	$\|g\|_2 \leq 5$	Mitigates gradient spikes during early training stages
Epochs / Early stopping	50 epochs; patience = 10 defined, early stopping disabled	Sufficient for convergence across all folds
Framework	PyTorch v2.2, CUDA 11.8	Ensures consistent computation on CPU and GPU

https://github.com/amirbayaiz-svg/A-Hybrid-Multi-Branch-Deep-Learning-Model-for-Autism-Spectrum-Disorder-Detection-/tree/main

The repository is accompanied by comprehensive documentation describing the experimental workflow, model configurations, and evaluation procedures. Environment configuration files and dependency specifications are provided to support reproducibility across different computational platforms. All experiments can be reproduced by following the instructions outlined in the repository's README file. Where permitted, links to the original dataset sources, along with scripts for data loading and preprocessing, are also included.

III. RESULTS

A. CORRELATION ANALYSIS OF JOINT COORDINATES

To quantitatively assess the relationship between gait kinematic features and class membership, a correlation analysis was conducted on key joint coordinates. The primary objective of this analysis was to identify biomechanical patterns potentially discriminative between children with ASD and TD peers, as well as to examine the robustness of these relationships with respect to the applied data-cleaning procedures. Pearson correlation coefficients were computed between joint coordinate features and the binary class label for both the original dataset and the dataset after outlier removal. Figure 12 shows the correlation matrix between the key features of the joint coordinates and the binary class label calculated on the original dataset before outlier removal, and Figure 13 shows the corresponding relationships after extreme value filtering.

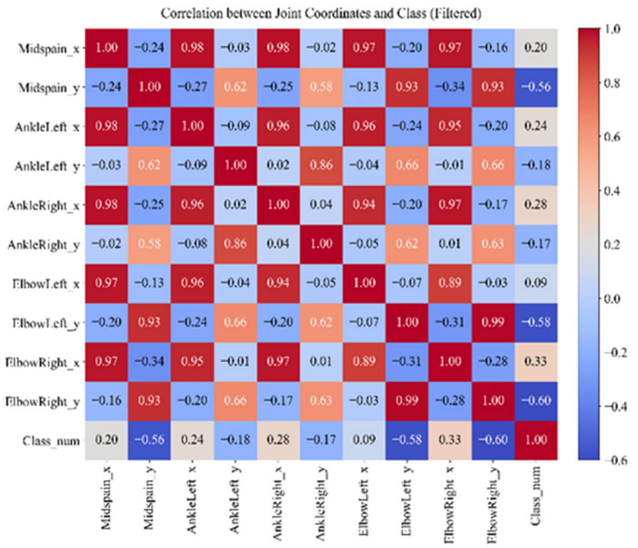


FIGURE 12. Correlation matrix before outlier removal.

In the original correlation matrix in Figure 12, the strongest negative correlations with the class label are observed for vertical coordinates of the upper body, including MidSpine_y, ElbowLeft_y, and ElbowRight_y ($\rho \approx -0.57$ to -0.61). These features also exhibit strong mutual correlations, reflecting coordinated variations in the posture and movement of the shoulder girdle and upper trunk. In contrast, horizontal coordinates of the lower limbs (AnkleLeft_x, AnkleRight_x) and the central body axis (MidSpine_x) show weak positive correlations with the class label, indicating less pronounced group differences in horizontal body displacement.

After outlier removal in Figure 13, the dominant role of vertical upper-body coordinates in class discrimination is preserved, while the corresponding correlation coefficients exhibit only minor changes. For instance, the correlation between MidSpine_y and the class label decreases marginally from -0.57 to -0.56, and for ElbowRight_y from -0.61 to -0.60, indicating the stability of these features and their high diagnostic relevance. At the same time, an increase in positive correlations is observed for the horizontal ankle coordinates (AnkleLeft_x: $0.06 \rightarrow 0.24$; AnkleRight_x: $0.09 \rightarrow 0.28$). This effect is likely attributable to the removal of sensor-related noise and a more accurate representation of foot placement, particularly in the TD group. Overall, a comparison of the correlation matrices before and after outlier removal confirms that the data cleaning procedure does not distort the interpretable biomechanical relationships. On the contrary, it stabilizes and partially strengthens the class-related correlations for the selected features. These results support the validity of using filtered joint coordinate data for subsequent feature engineering

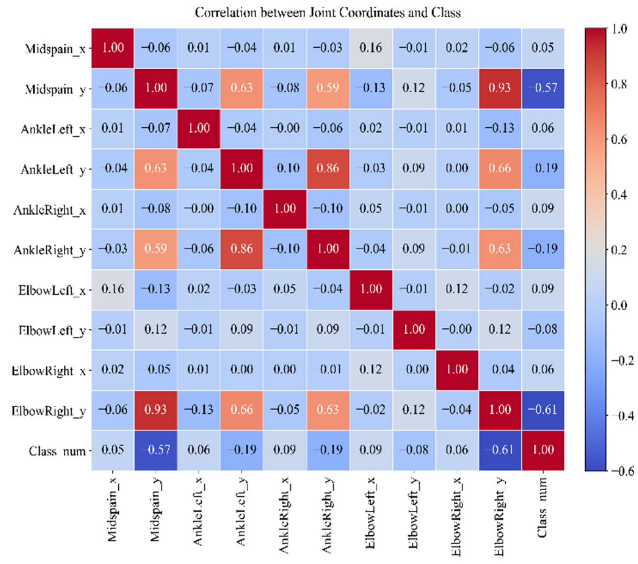


FIGURE 13. Correlation matrix after outlier removal.

and deep learning-based classification, while preserving the interpretability of the biomechanical gait characteristics. Importantly, the identified correlation patterns are consistent with previously described gait impairments in children with ASD, including impaired postural control, decreased upper body stability, and altered interarticular coordination [4], [5], [6]. In particular, the dominant role of vertical upper body features is consistent with studies highlighting atypical trunk control and increased movement variability in individuals with ASD. This agreement confirms that the extracted features reflect physiologically relevant motor characteristics rather than pre-processing artifacts, thereby strengthening the clinical relevance of the proposed approach.

B. QUANTITATIVE EVALUATION OF CLASSIFICATION PERFORMANCE

This subsection presents a quantitative analysis of the training results obtained for the baseline models and the proposed hybrid HMB-TSC architecture in the task of binary classification of gait patterns in children with ASD and TD peers. The nested 5×3 cross-validation consists of a 5-fold outer GroupKFold procedure used to evaluate the generalization ability of the model and a 3-fold inner GroupKFold procedure applied within each outer fold region to tune hyperparameters. To ensure strict separation between subjects and prevent data leakage, a GroupKFold strategy based on participant IDs was used. This evaluation protocol provides a robust assessment of classification accuracy, robustness, and generalization ability. Furthermore, the training dynamics of the hybrid architecture were analyzed in selected training epochs to study convergence behavior and the formation of stable internal representations. The validation results obtained at the 50th training epoch for all considered models are summarized in Table 8, providing a direct comparative assessment of their classification performance.

TABLE 8. Validation metrics of baseline and hybrid models at epoch 50.

Model	Val Loss	Val Acc	Val AUC	Val Precision	Val Recall	Val F1	Val Spec
HMB-TSC	0.3524	0.850	0.948	0.846	0.917	0.880	0.750
LSTM	0.6962	0.400	0.438	0.000	0.000	0.000	1.000
CNN	0.6768	0.650	0.740	1.000	0.417	0.588	1.000
Transformer	0.6022	0.650	0.906	1.000	0.417	0.588	1.000

The proposed hybrid model demonstrated the best overall results across all assessed indicators (Table 8), indicating its high discriminatory power and balanced classification performance. The combination of high recall and precision suggests that the model effectively identifies patterns associated with ASD while maintaining a controlled false-positive rate. From a clinical perspective, high recall is particularly important as it reduces the risk of missing ASD cases in early screening. In a holdout validation experiment, the HMB-TSC model demonstrated the best overall results, achieving an accuracy of 0.850, an AUROC of 0.948, and an F1 score of 0.880. The baseline Transformer model demonstrated a competitive AUROC value (0.906), indicating its ability to model global temporal dependencies; however, its threshold metrics, including recall and F1 score, remained significantly lower than those of the proposed hybrid architecture. The baseline CNN models, especially the LSTM, demonstrated weaker overall performance, with the LSTM failing to identify positive samples in the validation experiment reliably. These results indicate that using only one modeling paradigm, whether recurrent or convolutional, is insufficient for reliable gait-based ASD classification. In contrast, the hybrid multi-branch architecture benefits from combining additional temporal and spatiotemporal representations, resulting in a more balanced discriminative ability across all evaluation metrics [30].

The CNN-based model demonstrates moderate performance, reflecting the limited ability of purely convolutional filters to capture long-term temporal dependencies in gait dynamics. This suggests that local pattern extraction alone is insufficient to model the full temporal structure of gait cycles. These results suggest a weaker balance between precision and recall relative to recurrent and hybrid approaches, reflecting the limited ability of purely convolutional filters to capture long-range temporal dependencies in gait dynamics. This behavior is consistent with studies on convolutional sequence modeling, where CNN-based architectures rely on limited receptive fields and may be less effective at identifying long-term temporal dependencies [11]. The transformer-based model demonstrates high efficiency in modeling long-term time dependencies, although its performance remains sensitive to the data representation method and limited sample size. This suggests that, despite their theoretical advantages, transformer-based approaches may require large datasets or pre-training strategies to achieve consistent performance in gait analysis tasks. This observation is consistent with results from time-series studies, where Transformer models may be sensitive to how the data is presented and to limited sample sizes, especially in time-series applications. [31]. Recent research in gait recognition also shows that transformer-based architectures achieve optimal performance when trained on large or self-trained datasets [32]. Overall, these results demonstrate the advantage of the proposed hybrid architecture, which integrates convolutional, recurrent, and attention-based mechanisms, enabling more robust learning and a more balanced classification performance compared with the baseline models. This observation is consistent with previous work showing that hybrid CNN-BiLSTM architectures achieve high performance in modeling complex gait dynamics in noisy environments [33]. Importantly, these results suggest that the observed performance improvement is due not only to increased model capabilities but also to the complementary integration of multiple temporal modeling paradigms. To further investigate the contribution of each component to the proposed HMB-TSC architecture, an ablation study was conducted. The primary goal of this analysis was to assess whether the observed performance improvements were due to complementary interactions between model branches or simply to an increase in the number of parameters. Ablation experiments were designed by systematically removing individual branches from the full architecture and independently evaluating configurations with each branch removed. This approach allows direct assessment of the role of each modeling paradigm, including Transformers, CNN-BiLSTM, and CNN-BiGRU. All configurations were evaluated using a 5-fold cross-validation protocol. To statistically evaluate the contribution of each architectural component, paired two-tailed t-tests were conducted on identical folds to compare the full HMB-TSC model with each ablation variant. The quantitative results of the ablation analysis are presented in Table 9.

The full HMB-TSC model was used as the reference configuration; therefore, p-values are reported only for comparisons with the corresponding ablation variants. As shown in Table 9, most pairwise differences did not reach statistical significance at the predefined threshold of $\alpha = 0.05$. A statistically significant reduction in AUROC was observed only for the Transformer-only configuration ($p = 0.0247$), indicating that the complete hybrid architecture provides a more robust representation than the attention-based branch alone. For the remaining variants, the observed differences in Accuracy, Macro-F1, and AUROC were not statistically significant, suggesting that the individual branches capture partially overlapping information at the current dataset size. These findings indicate that the primary benefit of the proposed HMB-TSC framework lies in the complementary integration of convolutional, recurrent, and attention-based mechanisms, which improves overall robustness and stability rather than

TABLE 9. Ablation analysis of HMB-TSC variants with paired t-test p-values across 5-fold cross-validation.

Model	Parameters	Accuracy (%), mean ± std	Macro-F1 (%), mean ± std	AUROC, mean ± std	p Accuracy	p Macro-F1	p AUROC
Full HMB-TSC	6818113	85.00 ± 3.54	84.98 ± 3.54	0.940 ± 0.029	—	—	—
Without Transformer branch	5509441	81.00 ± 12.94	80.90 ± 13.10	0.934 ± 0.071	0.5122	0.5094	0.7753
Without CNN-BiLSTM branch	3968577	89.00 ± 6.52	88.99 ± 6.52	0.940 ± 0.035	0.2420	0.2408	1.0000
Without CNN-BiGRU branch	4166209	83.00 ± 7.58	82.86 ± 7.75	0.940 ± 0.024	0.4766	0.4689	1.0000
Transformer only	1316673	77.00 ± 7.58	76.86 ± 7.51	0.900 ± 0.046	0.1202	0.1148	0.0247
CNN-BiLSTM only	2857537	85.00 ± 9.35	84.84 ± 9.46	0.932 ± 0.070	1.0000	0.9756	0.7075
CNN-BiGRU only	2659905	88.00 ± 5.70	87.99 ± 5.70	0.946 ± 0.048	0.4263	0.4239	0.7160

producing uniformly significant gains across all evaluation metrics. While Table 9 focuses on the relative contribution of individual architectural components, Table 10 summarizes the overall performance of all models under the same 5-fold cross-validation protocol.

TABLE 10. Results of 5-fold cross-validation (mean ± std) for all models.

Model	Loss	Accuracy	AUC	Precision	Recall	F1	Specificity
HMB-TSC	0.288 ± 0.064	0.850 ± 0.035	0.940 ± 0.029	0.873 ± 0.041	0.820 ± 0.045	0.845 ± 0.036	0.880 ± 0.045
LSTM	0.692 ± 0.001	0.550 ± 0.094	0.646 ± 0.155	0.454 ± 0.310	0.380 ± 0.356	0.383 ± 0.294	0.720 ± 0.311
CNN	0.658 ± 0.009	0.700 ± 0.094	0.774 ± 0.124	0.890 ± 0.152	0.480 ± 0.205	0.597 ± 0.165	0.920 ± 0.130
Transformer	0.628 ± 0.017	0.650 ± 0.079	0.734 ± 0.090	0.746 ± 0.107	0.460 ± 0.134	0.561 ± 0.121	0.840 ± 0.089

In the 5-fold cross-validation protocol, HMB-TSC demonstrated the best overall performance among the compared architectures, achieving accuracy = 0.850 ± 0.035, AUROC = 0.940 ± 0.029, and F1 = 0.845 ± 0.036. The proposed model showed a more balanced tradeoff between precision and recall than the baseline methods, indicating improved discriminative ability when separating subject-level data. In contrast, the baseline LSTM model demonstrated substantially lower performance and greater variability, suggesting that recurrent modeling alone is insufficient to represent complex gait dynamics fully. The baseline CNN model achieved moderate stability but lower classification performance, reflecting the limitations of relying solely on local convolutional features. The Transformer model demonstrated comparatively higher AUROC values. However, its lower recall and F1-scores indicate reduced robustness under limited data conditions. While standard cross-validation provides an estimate of performance stability, it does not fully decouple hyperparameter tuning from model evaluation. To obtain a more conservative and unbiased estimate of generalization performance, a 5 × 3 nested cross-validation was employed. The corresponding results are presented in Table 11.

TABLE 11. Results of nested 5 × 3 cross-validation (mean ± std).

Model	Loss	Accuracy	AUC	Precision	Recall	F1	Specificity
HMB-TSC	0.288 ± 0.064	0.850 ± 0.035	0.940 ± 0.029	0.873 ± 0.041	0.820 ± 0.045	0.845 ± 0.036	0.880 ± 0.045
LSTM	0.692 ± 0.001	0.550 ± 0.094	0.646 ± 0.155	0.454 ± 0.310	0.380 ± 0.356	0.383 ± 0.294	0.720 ± 0.311
CNN	0.658 ± 0.009	0.700 ± 0.094	0.774 ± 0.124	0.890 ± 0.152	0.480 ± 0.205	0.597 ± 0.165	0.920 ± 0.130
Transformer	0.628 ± 0.017	0.650 ± 0.079	0.734 ± 0.090	0.746 ± 0.107	0.460 ± 0.134	0.561 ± 0.121	0.840 ± 0.089

The results of the 5 × 3 nested cross-validation are presented in Table 11. In the 5 × 3 nested validation, the HMB-TSC model demonstrated the best overall results among the compared architectures, achieving accuracy = 0.850 ± 0.035, AUROC = 0.940 ± 0.029, and F1 = 0.845 ± 0.036. The proposed model also demonstrated the lowest mean loss and a balanced tradeoff between precision and recall, indicating robust discrimination capability across subject-level splits. In contrast, the baseline LSTM model demonstrated significantly lower performance and greater variability, suggesting its limited ability to model complex gait dynamics within a single recurrent network. The baseline CNN model achieved moderate results, with relatively high specificity but lower recall and F1 scores, suggesting reduced sensitivity to positive ASD cases. The Transformer model demonstrated competitive AUROC values, indicating the usefulness of attention-based temporal modeling; however, lower recall and F1 scores indicate reduced robustness under limited data. Overall, the nested validation results support the effectiveness of combining complementary temporal and spatiotemporal representations within the proposed hybrid multi-branch architecture. The similarity between the standard 5-fold and nested cross-validation results indicates stable hyperparameter selection and low optimization variance across folds. This consistency suggests that the proposed architecture exhibits robust generalization behavior and that the reported performance is not strongly dependent on a specific validation partition.

To assess the statistical significance of the observed performance differences, a comparative analysis of the proposed HMB-TSC model with the baseline architectures (LSTM, CNN, and Transformer) was conducted using both 5-fold cross-validation and 5 × 3 nested cross-validation. Accuracy and AUROC values obtained across all folds were assessed using a two-tailed paired t-test. Additionally, McNemar's test was applied to aggregated fold predictions as an additional nonparametric evaluation based on the contingency statistics of TP, FP, FN, and TN. All statistical analyses were performed using SciPy (version 1.11), and p-values were adjusted for multiple comparisons using the Bonferroni correction. The results show that HMB-TSC achieved statistically significant improvements over the baseline LSTM model across all evaluation parameters and metrics. Comparisons with the baseline CNN and Transformer models also revealed significant differences across several evaluation metrics, particularly in AUROC and McNemar's test. Although some paired t-test results for the CNN approached the significance threshold, the overall statistical analysis confirms the improved discriminatory ability of the proposed hybrid architecture. Detailed statistical comparison results are presented in Table 12.

TABLE 12. Statistical comparison of HMB-TSC and baseline models under cross-validation protocols.

Model Pair	Accuracy (paired t-test, p-value)	AUROC (paired t-test, p-value)	McNemar's test (p-value)
5-fold CV			
HMB-TSC vs LSTM	0.0020	0.0109	5.30e-06
HMB-TSC vs CNN	0.0231	0.0234	0.0107
HMB-TSC vs Transformer	0.0135	0.0023	0.0003
Nested 5 × 3 CV			
HMB-TSC vs LSTM	0.0020	0.0109	5.30e-06
HMB-TSC vs CNN	0.0231	0.0234	0.0107
HMB-TSC vs Transformer	0.0135	0.0023	0.0003

To further confirm the generalizability of the results to previously unobserved participants, an additional subject-level evaluation protocol was applied using leave-10%-subjects-out cross-validation. In this protocol, all samples from each participant were strictly assigned to either the training or test set in each iteration, ensuring complete separation at the subject level. Each iteration included 90 training subjects and 10 test subjects, with a balanced class distribution (5 with ASD and 5 with typical development). Quantitative results are presented in Table 13, including both mean and worst-case performance measures across all iterations. The HMB-TSC model achieved an average accuracy of 0.880 ± 0.140, an AUROC of 0.920 ± 0.118, and an F1 score of 0.875 ± 0.145. Importantly, the worst-case scores were analyzed separately, revealing accuracy = 0.600, AUROC = 0.680, and F1 = 0.600. These results indicate that, although the model exhibits high average generalization ability, there is variability in performance across subject subdivisions due to interindividual differences. Unlike standard k-fold cross-validation, this protocol directly assesses the model's generalization ability to previously unseen participants, rather than to new segments of known individuals.

TABLE 13. Leave-10%-subjects validation: mean and worst-case performance of HMB-TSC.

Model	Accuracy mean ±std	Accuracy worst	AUROC mean ±std	AUROC worst	F1 mean ±std	F1 worst	Specificity mean ±std	Specificity worst
HMB-TSC	0.880 ± 0.140	0.600	0.920 ± 0.118	0.680	0.875 ± 0.145	0.600	0.900 ± 0.141	0.600

To further confirm the absence of data leakage at the individual subject level, a subject-level split audit was conducted, and the results are presented in Table 14.

TABLE 14. Subject-level split audit across evaluation protocols.

Protocol	Folds	Train subjects	Test subjects	Max overlap	Train TD/ASD	Test TD/ASD
Holdout	1	80	20	0	42/38	8/12
5-fold outer	5	80	20	0	40/40	10/10
Nested outer	5	80	20	0	40/40	10/10
Leave-10%-subjects	10	90	10	0	45/45	5/5

The analysis confirms the absence of subject overlap between the training and test sets across all assessment protocols. This ensures that the reported results reflect true generalization, rather than the memorization of subject-specific patterns. To assess the proposed model's robustness to measurement noise, a sensitivity analysis was conducted by perturbing the joint coordinates with uniform noise of ±13 mm and ±25 mm, corresponding to typical Kinect sensor error ranges (Table 15).

TABLE 15. Sensitivity analysis of HMB-TSC under Kinect coordinate noise.

Noise (mm)	AUROC, mean ± std	Accuracy, mean ± std	F1, mean ± std	Mean AUROC drop	Worst AUROC drop
0	0.940 ± 0.029	0.850 ± 0.035	0.845 ± 0.036	–	–
±13	0.940 ± 0.029	0.810 ± 0.065	0.835 ± 0.043	0.0000	0.0000
±25	0.934 ± 0.055	0.740 ± 0.139	0.801 ± 0.079	0.0060	0.0500

The results presented in Table 15 show that the HMB-TSC model maintains a stable area under the ROC curve (AUROC) of 0.940 ± 0.029 with ±13 mm noise, without noticeable degradation in ranking performance. With a stronger perturbation (±25 mm), the AUROC decreases slightly to 0.934 ± 0.055, corresponding to an average drop of 0.006 and a worst-case drop of 0.050 across all folds. These results demonstrate that the model maintains high discriminatory performance even with significant coordinate noise. Although the AUROC remains largely unchanged, accuracy and F1 scores exhibit greater sensitivity to noise, especially at perturbations of ±25 mm. This behavior is expected, as coordinate noise can bias individual samples relative to the threshold without significantly altering the overall ranking structure. Therefore, the HMB-TSC model can be considered robust to typical Kinect measurement noise in terms of AUROC, while threshold-based classification metrics may require recalibration in higher-noise environments.

To avoid ambiguity, Table 8 presents the results obtained using a single holdout validation split, Table 10 presents the average results of the 5-fold cross-validation, and Table 11 presents the average results of the 5 × 3 nested cross-validation. This distinction clearly distinguishes the single-sample assessment from the more robust estimates of generalizability obtained through cross-validation protocols. In addition to aggregate performance metrics, the convergence behavior of the proposed hybrid architecture was analyzed across selected training epochs. The evolution of validation metrics is summarized in Table 16.

In addition to overall performance metrics, the training dynamics of the proposed HMB-TSC model were further examined to assess convergence and validation stability during the optimization process. At the initial stage (epoch 1), the model demonstrated high validation loss (0.7359), low Accuracy (0.400), and zero recall, reflecting the lack of stable learned representations early in the optimization process. By epoch 10, the validation loss had significantly decreased to 0.3803, while Accuracy and AUROC increased

TABLE 16. Validation metric evolution of the HMB-TSC model across selected epochs.

Epoch	Val Loss	Val Acc	Val AUC	Val Precision	Val Recall
1	0.7359	0.400	0.844	0.000	0.000
10	0.3803	0.850	0.917	0.909	0.833
20	0.4017	0.850	0.917	0.909	0.833
30	0.4046	0.850	0.906	0.909	0.833
50	0.4102	0.800	0.969	0.900	0.750

to 0.850 and 0.917, respectively, indicating rapid acquisition of discriminative spatiotemporal patterns. Stable performance was maintained at epochs 20 and 30, where the model retained high Accuracy (0.850) and a good balance of precision and recall. At the final epoch (epoch 50), the model achieved the highest AUROC value (0.969), indicating strong discrimination between the ASD and TD classes. However, the simultaneous decrease in precision (0.800) and recall (0.750) suggests that the threshold-dependent and ranking-based metrics evolved differently during the later stages of optimization. These observations indicate that the proposed HMB-TSC architecture converges to a stable operating mode while maintaining robust discrimination across training epochs. Overall, the results presented in Tables 8–16 confirm the effectiveness of combining transformer-based temporal modeling, convolutional feature extraction, and bidirectional recurrent learning for gait-based ASD classification.

IV. DISCUSSION

These results demonstrate that spatiotemporal gait patterns derived from Kinect kinematic data provide a meaningful basis for ASD classification. In the evaluated validation protocols, the proposed HMB-TSC architecture demonstrated the best overall performance among the compared baseline models, particularly in terms of AUROC and the balanced F1-score. The observed improvements suggest that the combination of additional temporal and spatiotemporal representations facilitates more efficient modeling of heterogeneous gait dynamics than single-branch architectures.

From a methodological perspective, the effectiveness of HMB-TSC stems from its hybrid, multi-branch structure. The Transformer branch captures long-term temporal dependencies, convolutional layers extract localized spatiotemporal patterns, and bidirectional recurrent units model sequential gait dynamics. The integration of these additional representations enables a more comprehensive characterization of gait behavior than architectures based solely on convolutional, recurrent, or attentional processing mechanisms. Subject-level validation results show that the proposed architecture maintains relatively stable discriminatory performance across different participant partitions, although some variability remains observable between groups. This variability is expected given the limited dataset size and the heterogeneous nature of ASD-related motor behavior.

Furthermore, the analysis highlights the influence of demographic factors on model evaluation. The observed gender imbalance in the ASD group and differences in age distribution between participants with ASD and typically developing participants may influence gait characteristics, as motor behavior changes with age and biological development. Therefore, some of the classification performance may reflect demographic variability rather than solely ASD-specific motor characteristics, highlighting the importance of stratified validation and demographic-based modeling in future studies.

Convergence analysis further demonstrated that the proposed hybrid architecture provides stable optimization behavior during training. Validation metrics improved significantly in the early stages and remained relatively stable in later stages of optimization, while AUROC values remained consistently high. Sensitivity analysis using simulated Kinect coordinate perturbations further demonstrated that the model maintains stable discrimination performance under moderate noise, although threshold-based metrics degrade gradually with stronger perturbations. These results indicate that the proposed approach is sufficiently robust to realistic sensor-related variability.

Despite these promising results, several limitations should be considered. The evaluation was conducted using a single Kinect-based dataset, which may limit external generalizability. These limitations are consistent with recent methodological reviews in AI-based ASD and ADHD diagnostics, where insufficient reproducibility, limited external validation, and the lack of standardized assessment protocols remain key barriers to clinical application [34]. Furthermore, demographic imbalances and age-related variability may impact the learned gait representations. Another limitation is reliance on Kinect coordinates, which remain sensitive to data-collection conditions and sensor noise. Therefore, future research should focus on validation using larger, more diverse cohorts, on multimodal behavioral integration, and on the development of interpretable models capable of providing clinically meaningful explanations.

Overall, the obtained results confirm the utility of hybrid multi-branch architectures for gait-based ASD screening and demonstrate the importance of rigorous subject-level validation and reproducible assessment protocols. To enhance transparency and reproducibility, the full source code and experimental processing pipeline have been published for independent verification of the proposed approach.

V. CONCLUSION

In this study, we proposed a hybrid multi-branch architecture, HMB-TSC, for classifying autism spectrum disorders (ASD) using spatiotemporal gait features extracted from Kinect v2 recordings. The proposed framework integrates transformer-based temporal modeling with CNN-BiLSTM and CNN-BiGRU branches to capture additional gait representations at different time scales. Experimental evaluation showed that HMB-TSC demonstrated the best overall performance among the compared baseline architectures in the validation protocols with holdout, 5-fold, and nested 5×3 sampling. In a reproducible release run, the model achieved Accuracy $= 0.850$, AUROC $= 0.948$, and F1 $=$ 0.880 on holdout validation. Under 5-fold and 5×3 nested cross-validation, the model achieved Accuracy $= 0.850 \pm 0.035$, AUROC $= 0.940 \pm 0.029$, and F1 $= 0.845 \pm 0.036$. These results confirm the effectiveness and stability of the proposed hybrid architecture for gait-based ASD screening.

The proposed approach provides a noninvasive, scalable framework for objective analysis of motor behavior using markerless motion capture data. However, the study remains limited by the use of a single Kinect-based dataset and by demographic variability across the evaluated groups. Future research will focus on validation using larger, more diverse datasets, on multimodal behavioral integration, and on the development of interpretable ASD screening systems.

REFERENCES

[1] C. Okoye, C. M. Obialo-Ibeawuchi, O. A. Obajeun, S. Sarwar, C. Tawfik, M. S. Waleed, A. U. Wasim, I. Mohamoud, A. Y. Afolayan, and R. N. Mbaezue, "Early diagnosis of autism spectrum disorder: A review and analysis of the risks and benefits," *Cureus*, vol. 15, no. 7, Aug. 2023, Art. no. e43226, doi: 10.7759/cureus.43226.

[2] F. Mohammadi, H. Shahrokhi, A. Asadzadeh, S. Pirmoradi, A. Moghtader, and P. Rezaei-Hachesu, "Artificial intelligence in autism spectrum disorder diagnosis: A scoping review of face, voice, and text analysis methods," *Health Sci. Rep.*, vol. 8, no. 11, Nov. 2025, Art. no. e71476, doi: 10.1002/hsr2.71476.

[3] H. Rakotomanana and G. Rouhafzay, "A scoping review of AI-based approaches for detecting autism traits using voice and behavioral data," *Bioengineering*, vol. 12, no. 11, p. 1136, Oct. 2025, doi: 10.3390/bioengineering12111136.

[4] U. J. Ganai, A. Ratne, B. Bhushan, and K. S. Venkatesh, "Early detection of autism spectrum disorder: Gait deviations and machine learning," *Sci. Rep.*, vol. 15, no. 1, Jan. 2025, Art. no. 873, doi: 10.1038/s41598-025-85348-w.

[5] A. Pradhan, V. Chester, and K. Padhiar, "Classification of autism and control gait in children using multisegment foot kinematic features," *Bioengineering*, vol. 9, no. 10, p. 552, Oct. 2022, doi: 10.3390/bioengineering9100552.

[6] J. A. G. Lum, K. Shandley, N. Albein-Urios, M. Kirkovski, N. Papadopoulos, R. B. Wilson, P. G. Enticott, and N. J. Rinehart, "Meta-Analysis reveals gait anomalies in autism," *Autism Res.*, vol. 14, no. 4, pp. 733–747, Apr. 2021, doi: 10.1002/aur.2443.

[7] W. W. T. Lam, Y. M. Tang, and K. N. K. Fong, "A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation," *J. NeuroEngineering Rehabil.*, vol. 20, no. 1, May 2023, Art. no. 57, doi: 10.1186/s12984-023-01186-9.

[8] R. A. J. de Belen, T. Bednarz, A. Sowmya, and D. Del Favero, "Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019," *Translational Psychiatry*, vol. 10, no. 1, Sep. 2020, Art. no. 333, doi: 10.1038/s41398-020-01015-w.

[9] S. Pandya, S. Jain, and J. Verma, "A comprehensive analysis towards exploring the promises of AI-related approaches in autism research," *Comput. Biol. Med.*, vol. 168, Jan. 2024, Art. no. 107801, doi: 10.1016/j.compbiomed.2023.107801.

[10] A. Amirbay, N. Baigabylov, A. Mukhanova, K. Mukhambetova, E. Zaitov, R. Burganova, K. Khusanova, and F. Akhmedova, "Autism detection using facial and motor analysis using machine learning," *Bull. Electr. Eng. Informat.*, vol. 14, no. 5, pp. 3985–4000, Oct. 2025, doi: 10.11591/eei.v14i5.10319.

[11] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[12] F. Colonnese, F. Di Luzio, A. Rosato, and M. Panella, "Bimodal feature analysis with deep learning for autism spectrum disorder detection," *Int. J. Neural Syst.*, vol. 34, no. 2, Feb. 2024, Art. no. 2450005, doi: 10.1142/s0129065724500059.

[13] F. Colonnese, F. D. Luzio, A. Rosato, and M. Panella, "Enhancing autism detection through gaze analysis using eye tracking sensors and data attribution with distillation in deep neural networks," *Sensors*, vol. 24, no. 23, p. 7792, 2024, doi: 10.3390/s24237792.

[14] J. Shin, A. S. M. Miah, M. Kakizaki, N. Hassan, and Y. Tomioka, "Autism spectrum disorder detection using skeleton-based body movement analysis via dual-stream deep learning," *Electronics*, vol. 14, no. 11, p. 2231, May 2025, doi: 10.3390/electronics14112231.

[15] M. Al-Jubouri, "Three-dimensional dataset combining gait and full body movement of children with autism spectrum disorder collected by Kinect v2 camera," Tech. Rep., 2020, doi: 10.5061/dryad.s7h44j150.

[16] H. Tian, X. Ma, H. Wu, and Y. Li, "Skeleton-based abnormal gait recognition with spatio-temporal attention enhanced gait-structural graph convolutional networks," *Neurocomputing*, vol. 473, pp. 116–126, Feb. 2022, doi: 10.1016/j.neucom.2021.12.004.

[17] J. Kim, H. Seo, M. T. Naseem, and C.-S. Lee, "Pathological-gait recognition using spatiotemporal graph convolutional networks and attention model," *Sensors*, vol. 22, no. 13, p. 4863, Jun. 2022, doi: 10.3390/s22134863.

[18] Z. Yin, Y. Jiang, J. Zheng, and H. Yu, "STJA-GCN: A multi-branch spatial–temporal joint attention graph convolutional network for abnormal gait recognition," *Appl. Sci.*, vol. 13, no. 7, p. 4205, Mar. 2023, doi: 10.3390/app13074205.

[19] Y. Wang, T. Song, Y. Yang, and Z. Hong, "Research on multi-scale spatio-temporal graph convolutional human behavior recognition method incorporating multi-granularity features," *Sensors*, vol. 24, no. 23, p. 7595, Nov. 2024, doi: 10.3390/s24237595.

[20] X. Han, Y. Cui, X. Chen, Y. Lu, and W. Hu, "Spatio-temporal dynamic attention graph convolutional network based on skeleton gesture recognition," *Electronics*, vol. 13, no. 18, p. 3733, Sep. 2024, doi: 10.3390/electronics13183733.

[21] D. K. Ghosh, A. Chakrabarty, H. Moon, and M. J. Piran, "A spatio-temporal graph convolutional network model for Internet of Medical Things (IoMT)," *Sensors*, vol. 22, no. 21, p. 8438, Nov. 2022, doi: 10.3390/s22218438.

[22] Y. Jing, P. Qin, X. Fan, W. Qiang, Z. Wencheng, W. Sun, F. Tian, and D. Wang, "Deep learning–assisted gait parameter assessment for neurodegenerative diseases: Model development and validation," *J. Med. Internet Res.*, vol. 25, Jul. 2023, Art. no. e46427, doi: 10.2196/46427.

[23] A. Franco, M. Russo, M. Amboni, A. M. Ponsiglione, F. Di Filippo, M. Romano, F. Amato, and C. Ricciardi, "The role of deep learning and gait analysis in Parkinson's disease: A systematic review," *Sensors*, vol. 24, no. 18, 2024, Art. no. 5957, doi: 10.3390/s24185957.

[24] A. Sepas-Moghaddam and A. Etemad, "Deep gait recognition: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 264–284, Jan. 2023, doi: 10.1109/TPAMI.2022.3151865.

[25] M. Khaliluzzaman, A. Uddin, K. Deb, and M. J. Hasan, "Person recognition based on deep gait: A survey," *Sensors*, vol. 23, no. 10, p. 4875, May 2023, doi: 10.3390/s23104875.

[26] H. Lu, J. Chen, Z. Zhang, R. Liu, R. Zeng, and X. Hu, "Emotion recognition from skeleton data: A comprehensive survey," 2025, *arXiv:2507.18026*.

[27] C. Bunterngchit, L. H. Baniata, and A. Rasool, "NeuroXAI: Explainable deep learning for EEG-based detection of Alzheimer's and Parkinson's diseases," in *Proc. IEEE 25th Int. Conf. Bioinf. Bioengineering (BIBE)*, Athens, Greece, Nov. 2025, pp. 612–619, doi: 10.1109/BIBE66822.2025.00107.

[28] C.-W. Chang, C.-Y. Chang, and Y.-Y. Lin, "A hybrid CNN and LSTM-based deep learning model for abnormal behavior detection," *Multimedia Tools Appl.*, vol. 81, no. 9, pp. 11825–11843, Apr. 2022, doi: 10.1007/s11042-021-11887-9.

[29] S. M. Mahim, M. S. Ali, M. O. Hasan, A. A. N. Nafi, A. Sadat, S. A. Hasan, B. Shareef, M. M. Ahsan, M. K. Islam, M. S. Miah, and M.-B. Niu, "Unlocking the potential of XAI for improved Alzheimer's disease detection and classification using a ViT-GRU model," *IEEE Access*, vol. 12, pp. 8390–8412, 2024, doi: 10.1109/ACCESS.2024.3351809.

[30] 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, p. 9101, Nov. 2023, doi: 10.3390/s23229101.

[31] A. Zeng, M. Chen, L. Zhang, and Q. Xu, "Are transformers effective for time series forecasting?" in *Proc. AAAI Conf. Artif. Intell.*, Jun. 2023, vol. 37, no. 9, pp. 11121–11128, doi: 10.1609/aaai.v37i9.26317.

[32] A. Cosma, A. Catruna, and E. Radoi, "Exploring self-supervised vision transformers for gait recognition in the wild," *Sensors*, vol. 23, no. 5, p. 2680, Mar. 2023, doi: 10.3390/s23052680.

[33] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TimesNet: Temporal 2D-variation modeling for general time series analysis," 2022, *arXiv:2210.02186*.

[34] A. Rasool, F. Ahmad, C. Bunterngchit, and S. Aslam, "Challenges in translating AI-driven ASD/ADHD diagnosis: A methodological systematic review," *Int. J. Med. Informat.*, vol. 214, Jul. 2026, Art. no. 106417, doi: 10.1016/j.ijmedinf.2026.106417.

AYAGOZ MUKHANOVA received the Ph.D. degree in information systems from L. N. Gumilyov Eurasian National University, Kazakhstan, in 2014. Currently, she is an Associate Professor with the Department of Information Systems, L. N. Gumilyov Eurasian National University. Her research interests include image processing, computer vision, satellite imagery, artificial intelligence, and machine learning.

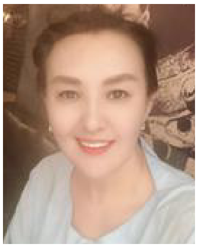
AIZAT AMIRBAY received the master's degree in information systems from L. N. Gumilyov Eurasian National University, Astana, Kazakhstan, in 2019, where she is currently pursuing the Ph.D. degree. Her research interests include artificial intelligence, spatiotemporal data analysis, and computer vision.

ZHAZIRA TASZHUREKOVA completed her Ph.D. dissertation in the specialty of geoecology at M. Kh. Dulaty Taraz State University, Taraz, Kazakhstan. She has more than 40 publications, one monograph, and five articles in the Scopus database. Her research interests include information systems development, information retrieval, and machine learning.

ABDUGANI KHOLBEKOV received the Doctor of Science degree in sociology. He is currently a Professor with the National University of Uzbekistan named after Mirzo Ulugbek and a Professor with Tashkent State University of Economics. He has authored more than 178 papers monography in journals and conference proceedings. His professional research interests include sociology and transformation processes of modern society.

KHILOLA KHAYDAROVA received the Ph.D. degree in philosophical sciences in sociological sciences. She is currently an Assistant with the Department of Sociology, Samarkand State University named after Sharof Rashidov. Her professional research interests include the sociology of management, environmental sociology, urban sociology, sociology of everyday life, ecology, urban ecology, and environmental education of youth.

SABOKHAT KALANOVA received the Ph.D. degree in sociological sciences. She is currently an Associate Professor with the Department of Sociology, Faculty of Social Sciences, National University of Uzbekistan named after Mirzo Ulugbek. Her professional interests include scientific research in sociology, ethnosociology, and social psychology.



GULZIRA ABDIKERIMOVA received the Ph.D. degree in information systems from L. N. Gumilyov Eurasian National University, Kazakhstan, in 2020. Currently, she is an Associate Professor with the Department of Information Systems, L. N. Gumilyov Eurasian National University. Her research interests include image processing, computer vision, artificial intelligence, and machine learning.

AIDANA AKHMETOVA received the master's degree in computer science and engineering from Novosibirsk State University, Russia, in 2013, and the Ph.D. degree in information systems from L. N. Gumilyov Eurasian National University, Kazakhstan, in 2019. Her research interests include information systems, data analysis, and database technologies.

ROHAYA LATIP received the Ph.D. degree in distributed database systems from Universiti Putra Malaysia, Malaysia, in 2009. Currently, she is an Associate Professor with the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia. Her research interests include distributed systems, cloud computing, workflow scheduling, and resource management.

MAKBAL BAIBULOVA received the degree in information systems from Taraz State University named after M. Kh. Dulati, in 2010, and the master's degree in computer science, in 2012. Her research interests include information and computer technologies and systems, and artificial intelligence.