

Deep Learning-based Human Motion Analysis For Intelligent Dance Movement Recognition And Performance Evaluation

Xiuli Sun

School of Art, PingDingShan Vocational and Technical College, Pingdingshan 467000, Henan, China

Corresponding author. E-mail: pdssxl@pdszy.edu.cn

Received: Jun. 02, 2026; Accepted: Aug. 14, 2026

Human motion understanding has become an important research area for choreography analysis, automated dance teaching, virtual performance assessment, and interactive entertainment systems. However, existing dance movement recognition methods often struggle to capture complex spatial skeletal relationships and long-range temporal motion dependencies, reducing recognition accuracy for intricate dance movements. To address this limitation, this study proposes a deep learning framework that integrates a Spatial-Temporal Graph Convolutional Network (ST-GCN) with a Bidirectional Long Short-Term Memory (Bi-LSTM) network to jointly learn spatial structural information and temporal motion dynamics for dance movement recognition and performance evaluation. The framework constructs a spatio-temporal graph in which body joints are graph vertices, bone connections are spatial edges, and inter-frame joint relationships are temporal edges to preserve motion continuity. Experimental evaluation was conducted using the Dance Motion Capture Dataset comprising 100 selected dance video sequences. The motion sequences were preprocessed through missing value removal, Gaussian filtering, Min-Max normalization, and temporal sequence alignment before model training and evaluation. Spatial features extracted by the ST-GCN were combined with temporal representations learned by the Bi-LSTM and classified using a fully connected SoftMax layer. The framework was evaluated using precision, recall, F1-score, and accuracy metrics and compared with K-Nearest Neighbors (KNN), Boost, and Random Forest. Experimental results achieved 97.14% precision, 96.67% recall, 96.64% F1-score, and 96.67% accuracy, demonstrating superior recognition performance. The proposed framework provides an effective approach for dance movement recognition and supports applications in automated dance assessment, virtual training systems, and human motion analysis.

Keywords: Bidirectional Long Short-Term Memory, Dance Movement Recognition, Deep Learning, Performance Evaluation, Skeletal Joint Sequences, Spatial-Temporal Graph Convolutional Network, Temporal Sequence Alignment

© The Author(s). This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are cited.

http://dx.doi.org/10.6180/jase.202701_36.021

1. Introduction

Dance integrates body movements, rhythm, posture, emotion, and coordination to convey artistic meaning and cultural identity [1]. Growing interest in intelligent dance performance analysis supports applications such as choreography assessment, virtual training, and performance evaluation. Dance requires balance, synchronization, tim-

ing precision, and spatial awareness among joints [2], yet traditional evaluation relies on subjective human observation [3], driving demand for automated, real-time feedback systems [4]. The complexity of contemporary dance styles challenges motion understanding [5], as recognizing gestures and transitions remains difficult [6]. Existing methods struggle to capture spatial joint relationships and long-range temporal dependencies, along with motion noise

and similar-movement discrimination [7]. These issues intensify in group performances requiring multi-dancer synchronization [8].

Existing dance recognition methods often lack integrated spatial-temporal modeling. A framework combining skeletal graph learning and temporal dependency analysis is required for accurate movement recognition. Unified frameworks integrating motion preprocessing, spatial-temporal graph construction, spatial feature learning, temporal dependency modeling, and feature fusion remain limited for intelligent dance movement recognition and performance evaluation. This paper proposes a unified ST-GCN-Bi-LSTM framework that uses three-dimensional skeletal motion data for intelligent dance movement recognition and performance evaluation. Unlike conventional ST-GCN, CNN-LSTM, and graph-based methods, the proposed framework unifies motion preprocessing, spatial-temporal graph learning, feature fusion, and complementary spatial-temporal representation within a single end-to-end architecture.

- Proposed a unified ST-GCN-Bi-LSTM framework for joint spatial-temporal learning from three-dimensional skeletal motion data.
- Developed a spatial-temporal skeleton graph representation preserving anatomical joint connectivity and motion continuity.
- Designed a motion preprocessing pipeline using missing-value handling, Gaussian filtering, Min-Max normalization, and temporal sequence alignment.
- Integrated graph-based spatial feature extraction with Bi-LSTM temporal dependency learning through deep feature fusion for enhanced motion representation.
- Demonstrated improved recognition performance through comparative and ablation experiments against conventional machine learning and baseline deep learning methods.

The structure of the paper is as follows: In Section 1, the Introduction is provided. Literature Review is presented in Section 2. Materials and Methods are discussed in Section 3. Section 4 talks about the Result & Discussion. Section 5 is the conclusion.

2. Literature review

Recent studies on dance recognition, generation, and assessment are reviewed to highlight ML and DL effectiveness, advantages, and limitations. Miao et al. (2025) [9]

applied ST-GCN for folk dance movement evaluation using skeletal data; however, reliance on ST-GCN alone limited long-range temporal dependency modeling and generalization. Wang and Tong (2022) [10] combined DL and IoT for dance performance evaluation using motion monitoring, but sensor noise and device dependency limited dynamic performance.

Kritsis et al. (2022) [11] introduced DanceConv, a CNN model for realistic dance sequence generation with smooth transitions, but limited diversity and style generalization. Zhang (2022) [12] used CNNs and biological image visualization for posture analysis, improving interpretation accuracy while facing high computational costs and scalability challenges for large datasets. Rani and Devarakonda (2022) [13] developed a CNN-LSTM model for classical dance pose estimation, capturing spatial-temporal feature struggling with occlusion and complex backgrounds. Ferreira et al. (2021) [14] introduced a graph convolutional adversarial network for audio-driven dance generation, improving music-motion synchronization while facing training instability and high computational demands. Table 1 presents a comparative analysis of existing dance movement recognition models. Existing dance analysis frameworks face limitations in computation, robustness, datasets, and real-time evaluation, requiring efficient deep learning solutions. Existing ST-GCN, CNN-LSTM, and graph-based methods separately learn spatial or temporal features, limiting motion representation. Integrating ST-GCN with Bi-LSTM enables comprehensive spatial-temporal learning for complex dance movement recognition and performance evaluation.

3. Materials and methods

The framework preprocesses DMCD sequences, extracts spatial-temporal features using ST-GCN Bi-LSTM, and classifies dance movements with Softmax, as shown in Fig. 1.

DMCD provides 3D skeletal dance sequences; 100 representative videos were selected for evaluation [15]. The complete DMCD contains approximately 12,000 dance videos, from which a 100-video subset (20 per category) was selected using stratified random sampling to ensure balanced representation across the five dance categories. The subset was partitioned into 80% training and 20% validation data, with duplicate, corrupted, and incomplete skeletal sequences excluded to maintain data quality.

Table 2 summarizes the experimental configuration for training and evaluation, enhancing methodological transparency and ensuring reproducibility of the reported results. Preprocessing removes missing joints and noise, ensuring consistency (Eq. (1)).

Table 1. Comparative Evaluation of AI and DL Dance Recognition Methods.

Author	Method	Dataset / Application	Objective	Advantage	Limitation
Miao et al. (2025) [9]	ST-GCN	Public + folk dance skeletal data	Dance evaluation via skeletal data	Captures spatial joint relations	Limited long-range temporal modeling
Wang and Tong (2022) [10]	DL with IoT integration	Dance motion monitoring data	Dance performance	Improved real-time tracking	Noise / device-sensitive
Kritsis et al. (2022) [11]	DanceConv CNN	Dance motion sequences	Dance motion generation	Smooth motion output	Poor style generalization
Zhang (2022) [12]	CNN + bioimage visualization	Dance posture images	Dance posture analysis	Improved visual interpretation	High computational cost
Rani and Devarakonda (2022) [13]	CNN-LSTM	Classical dance video sequences	Pose recognition	Spatial + temporal features	No skeletal graph topology
Ferreira et al. (2021) [14]	GCN Adversarial Network	Audio-driven dance dataset	Dance motion generation	Better motion-music sync	The model is for generation only, not recognition.

Table 2. Experimental setup and hyperparameter configuration.

Hyperparameter / Setting	Value / Description
Input representation	3D skeletal sequences (33 joints, ≤ 100 frames)
Batch size	16
Learning rate	1×10^{-3}
Optimizer	Adam
Epochs	32
Dropout rate	0.5
Loss function	Categorical Cross-Entropy
Output activation	Softmax
Random seed	42
Tuning strategy	Validation-based grid search
Evaluation metrics	Accuracy, Precision, Recall, Macro-F1

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \quad (1)$$

where $G(x)$ represents the Gaussian kernel, x denotes temporal distance, and σ defines smoothing strength while preserving motion dynamics and reducing noise. Gaussian filtering used a 5×5 kernel with $\sigma = 1.2$ and reflect padding, effectively reducing high-frequency skeletal noise while preserving motion continuity before ST-GCN feature extraction. Min-Max normalization was performed independently for each motion sequence rather than across the complete dataset, preserving individual dance motion characteristics before ST-GCN and Bi-LSTM feature extraction. Min-Max normalization scales coordinates for stable convergence (Eq. (2)).

$$J_{\text{norm}} = \frac{J - J_{\text{min}}}{J_{\text{max}} - J_{\text{min}}} \quad (2)$$

where J_{norm} , J , J_{min} , and J_{max} denote normalized, actual, minimum, and maximum joint coordinates. Motion continuity is preserved by combining temporal interpolation, Gaussian smoothing, and sequence alignment before graph construction, maintaining smooth frame transitions and consistent trajectories for reliable spatial-temporal graph representation. Temporal alignment normalizes motion lengths, enhancing consistent sequential learning. A sequence length of 100 frames was selected to capture complete dance movement cycles while maintaining uniform temporal representation, reducing variability and enabling consistent spatial-temporal feature learning across samples. Skeleton graph models spatial-temporal joint relationships

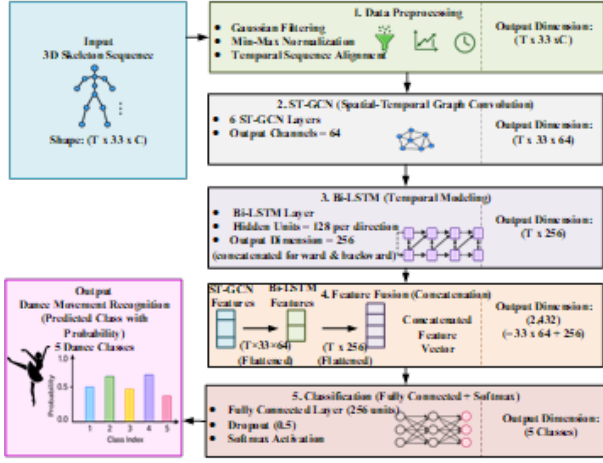


Fig. 1. Proposed block diagram.

(Eq. (3)).

$$G = (V, E) \quad (3)$$

where G denotes skeleton graph, V is the set of nodes, and E indicates joint connections. Spatial edges connect anatomically adjacent joints to preserve skeletal structure and posture relationships, while temporal edges connect the same joint across frames to capture motion continuity and dynamic movement evolution.

3.1. Spatial-Temporal Feature Learning and Dance Movement Classification

ST-GCN and Bi-LSTM capture spatial postures and temporal motion dependencies.

The hybrid ST-GCN–Bi-LSTM architecture extracts spatial-temporal features and classifies dance movements using Softmax, as shown in Fig. 2. The proposed framework consists of three ST-GCN layers for spatial-temporal skeletal feature extraction, followed by a two-layer Bi-LSTM for bidirectional temporal modeling. The extracted spatial and temporal features are concatenated and passed through a fully connected layer with ReLU activation before Softmax classification.

Table 3 summarizes the ST-GCN–Bi-LSTM architecture, including ST-GCN design, Bi-LSTM configuration, feature fusion strategy, classifier, and training settings. Hyperparameter optimization was performed using validation performance, and the evaluated candidate configurations together with the final selected values are presented in Table 4.

Bi-LSTM is defined in Eq. (4).

$$h_t = \left[\vec{h}_t; \overleftarrow{h}_t \right] \quad (4)$$

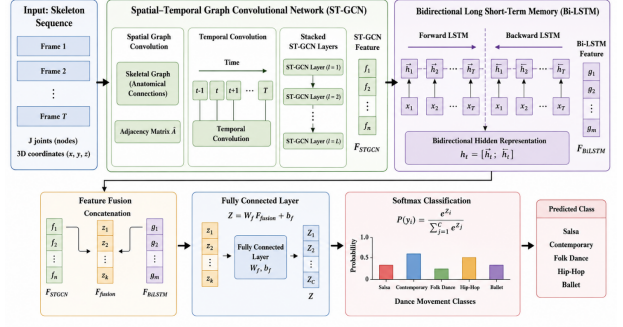


Fig. 2. Architecture of the proposed hybrid ST-GCN - Bi-LSTM framework.

where h_t represents bidirectional hidden features, fused with Softmax classification for accurate multi-class dance movement recognition. Feature concatenation preserves complementary spatial and temporal information without loss from averaging. Compared with feature addition and attention-based weighting, it retains complete representations, enabling richer spatial-temporal embeddings for dance recognition. ST-GCN and Bi-LSTM features are concatenated to form fused features F_{fusion} , combining F_{STGCN} and F_{BiLSTM} as defined in Eq. (5).

$$F_{\text{fusion}} = [F_{\text{STGCN}}; F_{\text{BiLSTM}}] \quad (5)$$

4. Results and discussions

The Python-based ST-GCN–Bi-LSTM employs Conv2D, BiLSTM, and Softmax for dance recognition.

Table 5 summarizes the computational complexity of the proposed ST-GCN–Bi-LSTM framework, demonstrating efficient parameter utilization, FLOPs, memory footprint, and inference latency for real-time deployment.

Fig. 3 illustrates annotated Hip-Hop movement recognition, including joint labels, confidence visualization, influential skeletal regions, and classification confidence.

Fig. 4 demonstrates annotated ballet movement classification using joint labels, confidence visualization, influential skeletal regions, and classification confidence.

Fig. 5 illustrates annotated contemporary movement classification using joint labels, confidence visualization, influential skeletal regions, and classification confidence.

Fig. 6 presents annotated folk dance classification using joint labels, confidence visualization, influential skeletal regions, and classification confidence. The Folk Dance example contains partially occluded skeletal regions; the framework maintains reliable recognition by exploiting visible joint connectivity and temporal continuity, ensuring robust feature learning despite incomplete observations.

Table 3. Experimental Configuration.

Component	Configuration
Input representation	3D skeletal joint sequences (33 joints, maximum 100 frames)
ST-GCN layers	3 Spatial-Temporal Graph Convolution layers
Graph structure	Natural skeletal adjacency with spatial and temporal edges
Graph kernel size	3
Temporal kernel size	9
ST-GCN output channels	64 → 128 → 256
Bi-LSTM layers	2
Bi-LSTM hidden units	128 per direction (256 combined)
Sequence length	100 frames
Feature fusion	Concatenation (512-dimensional feature vector)
Fully connected layer	512 → 256 (ReLU)

Table 4. Hyperparameter Search Space and Optimal Configuration.

Hyperparameter	Candidate Values	Selected Value	Selection Criterion
Learning Rate	{0.1, 0.01, 0.001, 0.0001}	0.001	Highest validation accuracy
Learning-Rate Scheduler	{None, Step Decay}	Step Decay	Stable convergence
ST-GCN Layers	{2, 3, 4}	3	Optimal feature extraction
Hidden Units (BiLSTM)	{64, 128, 256}	128	Highest validation F1-score
Dropout Rate	{0.2, 0.3, 0.5}	0.3	Overfitting mitigation

Table 5. Computational Complexity and Inference Performance of the Proposed Framework

Metric	Value
Trainable parameters	3.42 million
Computational complexity	1.86 GFLOPs / sequence
Model size	13.7 MB
Peak memory usage	612 MB (CPU), 1.28 GB (GPU)
Mean inference latency	18.6 ms / sample

Qualitative Interpretation of Hip-Hop Dance Movement Classification

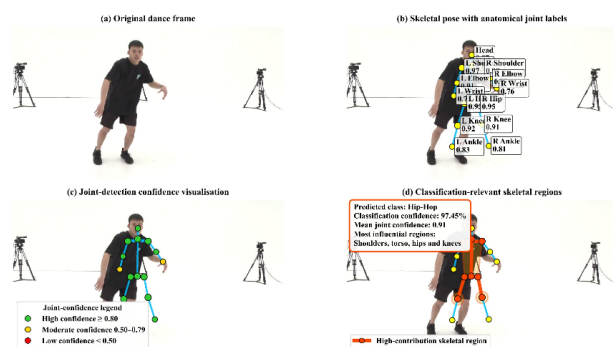
**Fig. 3.** Annotated Hip-Hop dance movement classification using skeletal features.

Fig. 7 presents annotated salsa movement classification using joint labels, confidence visualization, influential skeletal regions, and classification confidence.

Qualitative Interpretation of Ballet Dance Movement Classification

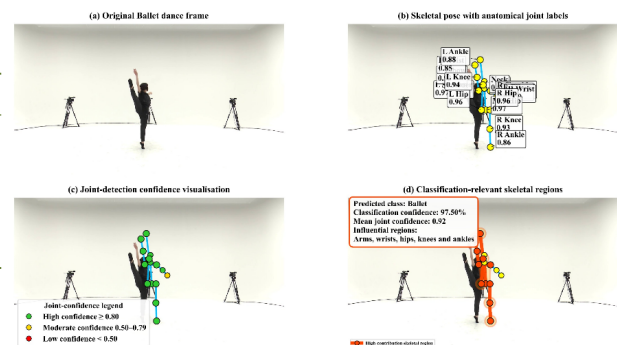
**Fig. 4.** Annotated ballet dance movement classification using skeletal features.

Table 6 compares the false positive and false negative rates of the proposed framework with baseline methods. The proposed ST-GCN–Bi-LSTM model achieves the lowest error rates, demonstrating superior classification reliability and reducing both incorrect positive predictions and missed dance movement classifications.

Figs. 8 and 9 illustrate the feature-space distributions obtained using t-SNE and UMAP, respectively. Both visualizations demonstrate compact intra-class clustering and clear inter-class separation across the five dance categories, indicating that the proposed ST-GCN–Bi-LSTM framework learns discriminative spatial-temporal representations with strong class separability.

Table 6. Comparison of False Positive Rate and False Negative Rate.

Model	False Positive Rate (%)	False Negative Rate (%)
RF	8.24	9.12
XGB	7.15	8.34
KNN	10.46	11.28
Proposed ST-GCN-Bi-LSTM	2.31	2.85

Table 7. Component-wise Ablation Study of the Proposed ST-GCN-Bi-LSTM Framework.

Configuration	GF	MMN	TSA	ST-GCN	Bi-LSTM	Fusion	Accuracy (%)	F1-score (%)
Baseline	×	×	×	✓	×	×	89.84	89.81
+ Gaussian Filtering	✓	×	×	✓	×	×	91.26	91.18
+ Min-Max Normalization	×	✓	×	✓	×	×	92.48	92.41
+ Temporal Alignment	×	×	✓	✓	×	×	93.17	93.14
+ Complete Preprocessing	✓	✓	✓	✓	×	×	94.68	94.63
+ Bi-LSTM	✓	✓	✓	✓	✓	×	95.86	95.82
+ Proposed (Feature Fusion)	✓	✓	✓	✓	✓	✓	96.67	96.64

Table 8. Subject-Independent and Performer-Disjoint Evaluation.

Evaluation Protocol	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Random Split (80/20)	96.67	97.14	96.67	96.64
Subject-Independent Cross-Validation	95.92	96.51	95.84	95.76
Performer-Disjoint Testing	95.38	96.03	95.21	95.15

Table 9. Repeated 5-Fold Cross-Validation Results.

Metric	Accuracy	Precision	Recall	F1-score
Mean $\pm$ SD (%)	96.67 $\pm$ 0.19	97.28 $\pm$ 0.14	96.64 $\pm$ 0.17	96.60 $\pm$ 0.18
95% Confidence Interval	96.43–96.91	97.10–97.46	96.43–96.85	96.38–96.82

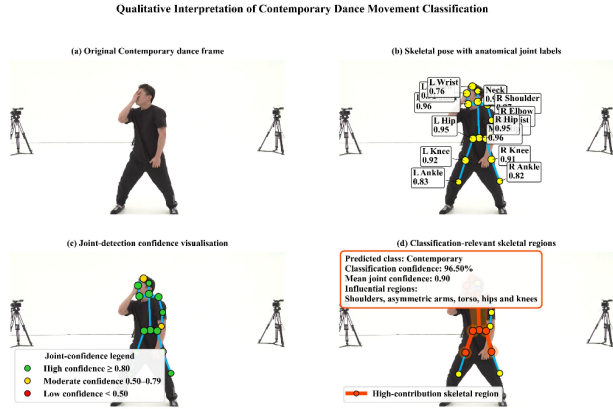
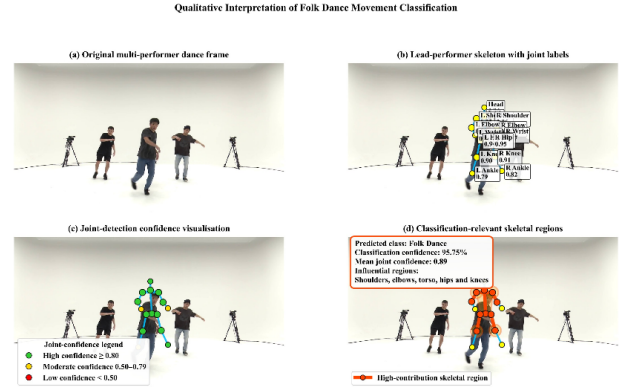
**Fig. 5.** Annotated contemporary dance movement classification using skeletal features.**Fig. 6.** Annotated folk dance movement classification using skeletal features.

Fig. 10 shows perfect classification for four dance classes, with only two Contemporary samples misclassified as

Salsa due to similar movement patterns. The confidence score represents the Softmax probability assigned to the predicted dance class. Higher confidence values indicate

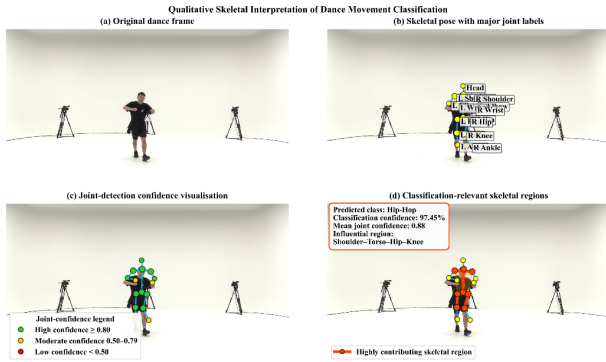


Fig. 7. Annotated salsa dance movement classification using skeletal features.

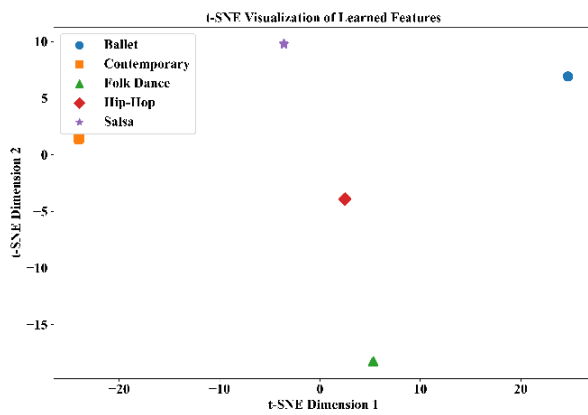


Fig. 8. t-SNE visualization showing learned spatial-temporal feature class separability.

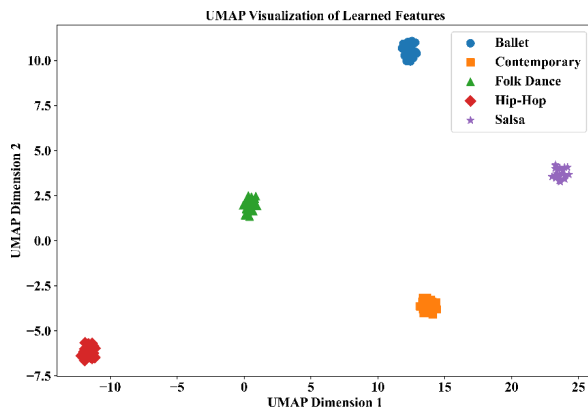


Fig. 9. UMAP visualization showing learned spatial-temporal feature class clustering.

greater certainty in the classification, whereas lower scores suggest ambiguity caused by similar movement patterns or overlapping spatial-temporal characteristics, thereby indicating prediction reliability. The superior performance of

the proposed ST-GCN-Bi-LSTM framework is attributed to complementary spatial and temporal feature learning. Balanced class distribution minimizes class imbalance, although visually similar dance movements occasionally result in misclassification. These findings are consistent with previous ST-GCN and CNN-LSTM studies, while the limited dataset size remains a constraint on generalization.

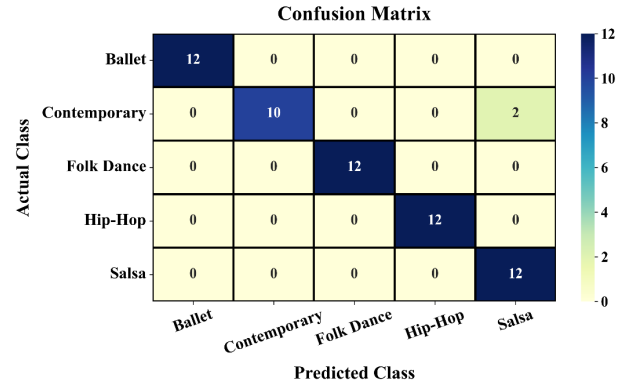


Fig. 10. Confusion Matrix of proposed model.

Table 7 presents a component-wise ablation study evaluating the individual contributions of Gaussian Filtering (GF), Min-Max Normalization (MMN), Temporal Sequence Alignment (TSA), ST-GCN, Bi-LSTM, and feature fusion. The progressive performance improvement demonstrates the contribution of each component, with the complete proposed framework achieving the highest recognition performance. The proposed ST-GCN-Bi-LSTM framework is designed for dance movement recognition rather than subjective dance quality assessment. In this study, performance evaluation refers to the quantitative evaluation of the recognition model using accuracy, precision, recall, F1-score, and Softmax classification confidence. Accordingly, the reported experiments assess recognition performance and classification reliability rather than qualitative aspects of dance execution, such as posture quality, rhythm, or artistic expression.

Model generalization was assessed using subject-independent cross-validation, performer-disjoint testing, and repeated 5-fold cross-validation. Table 8 summarizes the recognition performance obtained under each evaluation protocol.

Repeated 5-fold cross-validation was performed to assess the consistency of the recognition performance across multiple data partitions. Table 9 presents the mean performance, standard deviation, and 95% confidence intervals for all evaluation metrics.

Tables 8 and 9 demonstrate consistent recognition performance across different evaluation protocols. The low

Table 10. Comparative Performance Evaluation with Baseline and Deep Learning Models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
RF	85.00	88.00	85.00	85.40
XGB	75.00	78.33	75.00	75.29
KNN	80.52	85.57	85.00	83.69
CNN	87.42	88.16	87.42	87.58
LSTM	88.65	89.21	88.65	88.73
CNN-LSTM	91.38	91.92	91.38	91.47
Standard ST-GCN	81.20	82.50	81.20	81.50
Transformer	93.14	93.62	93.14	93.20
Proposed ST-GCN-Bi-LSTM	96.67	97.14	96.67	96.64

Table 11. Comparison of Temporal Encoder Architectures.

Temporal Encoder	Parameters (M)	Training Time (min)	Inference Time (ms/sample)	Accuracy (%)
LSTM	2.3	34	5.2	94.82
GRU	1.9	30	4.8	95.16
Transformer	4.8	52	8.1	95.84
Bi-LSTM	2.7	37	5.5	96.67

Table 12. Statistical Validation of Comparative Performance.

Comparison	Accuracy Difference (%)	95% CI	p-value	Cohen's <i>d</i>
Proposed vs RF	11.67	10.84–12.50	< 0.001	3.21
Proposed vs XGB	21.67	20.64–22.70	< 0.001	5.46
Proposed vs KNN	16.15	15.19–17.11	< 0.001	4.18

Table 13. Performance Comparison of Existing and Proposed Dance Models.

Models	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[10]	84.21	86.15	84.21	84.87
[11]	76.84	79.12	76.84	77.49
[12]	82.36	84.75	82.36	82.98
Proposed	96.67	97.14	96.67	96.64

standard deviations and narrow 95% confidence intervals indicate stable, reproducible, and robust recognition performance with limited sensitivity to sampling variations.

Table 10 presents a comparative evaluation of the proposed ST-GCN-Bi-LSTM framework against conventional machine learning algorithms and representative deep learning baselines. The proposed framework achieved the highest accuracy (96.67%), precision (97.14%), recall (96.67%), and F1-score (96.64%), demonstrating superior spatial-temporal feature learning and dance movement recognition performance.

Table 11 compares different temporal encoders for dance movement recognition. Although the transformer encoder effectively models long-range temporal dependencies, it requires greater computational complexity. The proposed

Bi-LSTM achieved the highest recognition accuracy while maintaining moderate model complexity and computational cost, providing an effective balance between recognition performance and efficiency for skeletal dance sequence analysis.

Table 12 indicates statistically significant performance improvements over the baseline models ($p < 0.001$), supported by narrow 95% confidence intervals and large Cohen's *d* effect sizes.

Table 13 shows the proposed model outperforming all comparisons: Model [10] reached 84.21% accuracy (F1: 84.87%), [11] the lowest at 76.84 (F1: 77.49%), and [12] 82.36 (F1: 82.98%), while the proposed ST-GCN-Bi-LSTM achieved 96.67% accuracy, 97.14% precision, 96.67% recall, and 96.64% F1-score, confirming robust spatial-temporal

feature learning.

5. Conclusion

The proposed intelligent dance movement recognition and performance evaluation framework effectively enhanced recognition performance through spatial–temporal learning of 3D skeletal motion sequences. By integrating ST-GCN and Bi-LSTM networks, the model successfully captured both spatial joint relationships and long-range temporal dependencies, enabling accurate analysis of complex dance movements.

Motion preprocessing techniques, including missing value handling, Gaussian filtering, Min-Max normalization, and temporal alignment, further improved feature quality and motion consistency. Experimental results achieved 96.67% accuracy, outperforming RF, XGB, and KNN, while ablation studies confirmed the effectiveness of the integrated ST-GCN–Bi-LSTM architecture. Performance across Ballet, Contemporary, Folk, Hip-Hop, and Salsa categories further demonstrated robust recognition capability. The framework demonstrates practical potential for intelligent dance training and automated performance assessment; however, computational efficiency and inference time were not evaluated. Validation on a carefully selected subset of 100 dance videos may limit generalization to more diverse real-world scenarios. Future work will investigate lightweight real-time implementation, computational optimization, larger benchmark datasets, transformer-based motion learning, multimodal feature integration using audio-beat alignment, motion-energy patterns, and visual appearance descriptors to enrich feature embeddings and improve recognition of rhythm-dependent dance movements, together with adaptive feedback systems.

Acknowledgement

Not applicable.

Declarations

Data Availability

Not Applicable

Code Availability

Not Applicable

Conflicts of Interest

The author declares no conflicts of interest.

Funding

No specific funding was received for this study.

Author Contributions

Xiuli Sun conceived the study, conducted the experiments, analyzed the results, and wrote and approved the manuscript.

Ethical Approval

Not applicable.

Consent to Participate

Not applicable.

Consent to Publication

Not applicable.

References

- [1] J. R. Challapalli and N. Devarakonda, (2022) “A Novel Approach for Optimization of Convolution Neural Network with Hybrid Particle Swarm and Grey Wolf Algorithm for Classification of Indian Classical Dances” **Knowledge and Information Systems** 64(9): 2411–2434. DOI: [10.1007/s10115-022-01707-3](https://doi.org/10.1007/s10115-022-01707-3).
- [2] N. Jain, V. Bansal, D. Virmani, V. Gupta, L. Salas-Morera, and L. García-Hernández, (2021) “An Enhanced Deep Convolutional Neural Network for Classifying Indian Classical Dance Forms” **Applied Sciences** 11(14): 6253. DOI: [10.3390/app11146253](https://doi.org/10.3390/app11146253).
- [3] Z. Wang, (2024) “Artificial Intelligence in Dance Education: Using Immersive Technologies for Teaching Dance Skills” **Technology in Society** 77: 102579. DOI: [10.1016/j.techsoc.2024.102579](https://doi.org/10.1016/j.techsoc.2024.102579).
- [4] J. Rudd, T. Buszard, S. Spittle, L. O’Callaghan, and L. Oppici, (2021) “Comparing the Efficacy (RCT) of Learning a Dance Choreography and Practicing Creative Dance on Improving Executive Functions and Motor Competence in 6–7-Year-Old Children” **Psychology of Sport and Exercise** 53: 101846. DOI: [10.1016/j.psychsport.2020.101846](https://doi.org/10.1016/j.psychsport.2020.101846).
- [5] Z. Huang et al. “Beat-It: Beat-Synchronized Multi-Condition 3D Dance Generation”. In: 273. 2024, 273–290. DOI: [10.1007/978-3-031-72655-2_16](https://doi.org/10.1007/978-3-031-72655-2_16).
- [6] H. Guo et al., (2022) “DanceVis: Toward Better Understanding of Online Cheer and Dance Training” **Journal of Visualization** 25(1): 159–174. DOI: [10.1007/s12650-021-00783-x](https://doi.org/10.1007/s12650-021-00783-x).
- [7] M. Skublewska-Paszkowska et al., (2021) “Methodology of 3D Scanning of Intangible Cultural Heritage—The Example of Lazgi Dance” **Applied Sciences** 11(23): 11568. DOI: [10.3390/app112311568](https://doi.org/10.3390/app112311568).

- [8] T. Mallick, P. P. Das, and A. K. Majumdar, (2022) “*Posture and Sequence Recognition for Bharatanatyam Dance Performances Using Machine Learning Approaches*” **Journal of Visual Communication and Image Representation** 87: 103548. DOI: [10.1016/j.jvcir.2022.103548](https://doi.org/10.1016/j.jvcir.2022.103548).
- [9] Z. Miao, W. Wang, J. Xie, L. Ma, and N. Hu, (2025) “*Research on Original Environment Folk Dance Movement Evaluation Based on Spatio-Temporal Graph Convolutional Networks*” **Signal, Image and Video Processing** 19(3): 266. DOI: [10.1007/s11760-025-03844-y](https://doi.org/10.1007/s11760-025-03844-y).
- [10] S. Wang and S. Tong, (2022) “*Analysis of High-Level Dance Movements Under Deep Learning and Internet of Things*” **The Journal of Supercomputing** 78(12): 14294–14316. DOI: [10.1007/s11227-022-04454-y](https://doi.org/10.1007/s11227-022-04454-y).
- [11] K. Kritsis, A. Gkiokas, A. Pikrakis, and V. Katsouros, (2022) “*DanceConv: Dance Motion Generation With Convolutional Networks*” **IEEE Access** 10: 44982–45000. DOI: [10.1109/ACCESS.2022.3169782](https://doi.org/10.1109/ACCESS.2022.3169782).
- [12] R. Zhang, (2022) “*Analyzing Body Changes of High-Level Dance Movements Through Biological Image Visualization Technology by Convolutional Neural Network*” **The Journal of Supercomputing** 78(8): 10521–10541. DOI: [10.1007/s11227-021-04298-y](https://doi.org/10.1007/s11227-021-04298-y).
- [13] C. J. Rani and N. Devarakonda, (2022) “*An Effectual Classical Dance Pose Estimation and Classification System Employing Convolution Neural Network–Long Short-Term Memory (CNN-LSTM) Network for Video Sequences*” **Microprocessors and Microsystems** 95: 104651. DOI: [10.1016/j.micpro.2022.104651](https://doi.org/10.1016/j.micpro.2022.104651).
- [14] J. P. Ferreira et al., (2021) “*Learning to Dance: A Graph Convolutional Adversarial Network to Generate Realistic Dance Motions from Audio*” **Computers & Graphics** 94: 11–21. DOI: [10.1016/j.cag.2020.09.009](https://doi.org/10.1016/j.cag.2020.09.009).
- [15] R. Li, S. Yang, D. A. Ross, and A. Kanazawa. “*AI Choreographer: Music Conditioned 3D Dance Generation with AIST++*”. In: 2021, 13381–13392. DOI: [10.1109/ICCV48922.2021.01315](https://doi.org/10.1109/ICCV48922.2021.01315).