Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Emotion-Aware AI and Digital
Transformation of Visual Culture 2025 6(5s)

DEEP LEARNING MODELS FOR CHOREOGRAPHY GENERATION

Afroj Alam [1], Sadhana Sargam [2], Jyoti Rani [3], Pavas Saini [4], Amol Bhilare [5], S. Balakrishnan [6]

[1] Assistant Professor, Department of Computer Science and Engineering, Presidency University, Bangalore, Karnataka, India
[2] Assistant Professor, School of Business Management, Noida International University, Uttar Pradesh, India
[3] Assistant Professor, Department of Fashion Design, Parul Institute of Design, Parul University, Vadodara, Gujarat, India
[4] Centre of Research Impact and Outcome, Chitkara University, Rajpura 140417, Punjab, India
[5] Department of Computer Engineering, Vishwakarma Institute of Technology, Pune 411037, Maharashtra, India
[6] Professor and Head, Department of Computer Science and Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation (DU), Tamil Nadu, India

Received 15 May 2025
Accepted 19 August 2025
Published 28 December 2025

Corresponding Author
Afroj Alam,
afroj.alam@presidencyuniversity.in

DOI
10.29121/shodhkosh.v6.i5s.2025.6908

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Due to the rapid development of deep learning, the new opportunities of computational production of human movement have been opened, especially in the field of dance choreography. The paper discusses deep learning choreography generators that combine movement information, music framework, and time in order to create expressive and sensible sequences of dances. Conventional choreography models tend to have a handmade regulation or professionalized composition where flexibility and creative variety is restricted. Conversely, deep learning methods that are data driven can directly learn complex spatio-temporal patterns using big datasets of motion and video. The suggested framework uses pose representations, video frames and rhythmic information based on music to simulate the inherent relationship between motion and sound. Recurrent neural networks as LSTM and GRU models are used to learn long-lasting temporal dependencies in dance sequences whereas transformer-based models are used to improve global context awareness and sequence coherence. Also, generative adversarial networks, diffusion-based networks are explored to achieve motion synthesis which provides smooth transitions, stylistic variability and a sense of realistic continuity in movement. A modular system architecture is structured in such a way that it can allow multimodal inputs, convolutional feature extraction and temporal sequence generation. The evaluation of the experimental results is carried out on standard choreography and motion datasets and performance is measured by quantitative evaluation measures including mean absolute error, Fréchet Inception Distance, and a smoothness index specific to the movement evaluation.

Keywords: Deep Learning, Choreography Generation, Motion Synthesis, Lstm and Transformers, Generative Adversarial Networks, Diffusion Models

1. INTRODUCTION

Dance is a human essential that is an amalgamation of physical activities, rhythm, emotion, and cultural identity. In classic and folk traditions as well as in modern and experimental ones, choreography is the systematic language, by means of which movement is arranged in space and time. The process of choreography design is a creative endeavor, and it is a complicated matter, which involves profound knowledge of body mechanics, musical harmony, aesthetic

values and expression. Conventionally, the art of choreography used to be based on experience and intuition of human choreographers, who routinely create, develop and pass on movement patterns through practice and performance. Though such humanistic approach is still the main focus of dance, it is time consuming, personal and hard to scale and analyse in a systematic way. As digital media and motion capture technologies become more and more accessible, and large repositories of dance videos emerge, computational techniques on modelling and generating dance movements have become of high interest. Initial computational choreography systems were mostly rule-based, representing predefined movement grammars, symbolic representations or biomechanical restrictions Pleshakova et al. (2024). Though these systems offered formalized portrayals of dance, there was no flexibility and they were not able to embrace the depth of variability and fluidity of time and style of human movement. These restrictions inspired the transition to data-driven approaches that were able to learn choreography patterns straight out of observed performances. Deep learning has become an effective paradigm of modelling high-dimensional, sequential, and multimodal data, so it is especially appropriate to choreography generation Hou (2024). Human motion may be modelled either as time-varying sequences of pose, skeletal joint paths or visual attributes derived by video frames and all these shows strong interaction between time and nonlinear multi-directional interactions.

Recurrent deep neural networks, in particular, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models have been shown to be effective at modelling long-range temporal variations in sequential data. More recently, architectures based on transformers have gone further to support sequence modelling with self-attention mechanisms to learn global interactions in a sequence of motion and enhance coherence and long-term structure of generated choreography. Simultaneously, generative modelling types, including Generative Adversarial Networks (GANs) and diffusion models, have demonstrated impressive performance in synthetic data generation of realistic and variety in the fields of visual, audio, and motion data Hong (2024). These models in combination with choreography generation allow creating continuous and smooth movements that maintain a style of consistency and physical reasonableness. In particular, diffusion-based methods provide better stability and better control over motion transitions, dealing with the typical problems of jitter, discontinuities, unnatural poses, and other problems in previous generative systems. The other dimension of choreography generation that is crucially important is the incorporation of music and rhythm. Dance is closely connected with musical framework, speed and mood. The current deep learning systems permit multimodal learning, i.e. musical instruments like beats, tempo, spectral patterns can be synchronized with motion patterns to produce dance sequences that are rhythmically aligned and expressively significant Wu et al. (2023). Such multimodal approach creates new opportunities to adaptive choreography which is dynamically responsive to the musical styles and performance contexts. The results of research in the creation of deep learning models that generate choreography have serious implications outside the artistic sphere.

2. LITERATURE REVIEW

2.1. TRADITIONAL AND RULE-BASED CHOREOGRAPHY SYSTEMS

The first computational methods of choreography generation were mostly rule-based, and tried to formalise dance by means of fixed structures and symbolic representations. These systems were inspired by the known systems of dance notation like Labanotation and Benesh Movement Notation, which represent body positions, direction, and timing as symbolic rules Croitoru et al. (2023). Translation of such notations into computational grammars was a goal of researchers to capture the choreographic knowledge in a form that can be recognized and interpreted. Choreography engines with rules were commonly based on manually constructed libraries of movement, biomechanical restrictions and transition rules to produce practicable dance sequences Tay et al. (2023). The physical plausibility and style consistency within a small domain was achieved by this method, and it was applicable in educational demonstrations and restricted performance simulations. Nevertheless, the conventional systems were not very flexible and creative. The choreography produced was very reliant on previous assumptions and was not very diverse as rules were manually designed by experts Sun et al. (2023). Such systems could not generalize across styles of dances, musical variations and performer specific nuances. Besides, expressive attributes of emotion, improvisation, and light differences in time could not be achieved through rule-based models, which form the core of human dance Copet et al. (2023).

2.2. MOTION CAPTURE AND MOVEMENT DATASETS FOR DANCE ANALYSIS

The development of the choreography modeling has been strongly associated with the existence of the motion capture and movement datasets that give quantitative depiction of the human dance. Motion capture (MoCap) systems capture the exact 3D joint movements with the use of optical markers, inertial sensors, or depth cameras and provide a more detailed study of body dynamics. Initial data used to track human movement in general ways like walking or running, but more recently there has been an increase in the types of dances tracked, including ballet, contemporary, hip-hop, folk and social dances Chen et al. (2021). These data sets would measure the changes in time, stylistic features, and artist specific features that are vital in the study of choreography. Besides MoCap data, video-based large dance datasets have become popular (because of its availability and diversification of cultures). Skeletal information can now be remotely estimated in RGB videos by pose estimation algorithms, and it is now possible to create datasets without any special capture equipment Zeng (2025). Nevertheless, the problem of dataset preparation, such as noise in pose extraction, missing joints, occlosures, and inconsistency across different recording conditions persists Zhang and Zhang (2022).

2.3. DEEP LEARNING IN GENERATIVE ARTS AND PERFORMANCE MODELING

Deep learning has played an important role in generative arts and performance modeling where machines are able to learn creative patterns, which are generally complex, based on data. AI in visual arts applications include style transfer, image generation, and stylization, whereas in music, recurrent networks, transformers, and other models have shown results in composition and improvisation Chen et al. (2024). Such improvements have naturally been applied to the performance arts where movement and expression is dealt with as a high-dimensional and sequential signal. The first deep learning models to be used in motion generation were recurrent neural networks, and specifically, LSTM and GRU, which were very effective at representing temporal relationships in a sequence of dances and gestures. Transformer-based architectures have been becoming increasingly popular in recent years as they can represent long-range dependencies with self-attention models Lauriola et al. (2022). Table 1 is a summary of deep learning methods in automated choreography generation. Such models have demonstrated enhanced coherence and the global structure in the generated performance as compared to the old-fashioned recurrent methods. The Generative Adversarial networks also helped in bringing in the paradigms of adversarial training that promote realism and authentic style in the generated movements.

Table 1

Table 1 Related Work on Deep Learning Models for Choreography Generation

Study Focus	Data Type Used	Key Contribution	Limitations
Rule-based Dance Generation	Symbolic poses	Formalized choreographic rules	Low creativity, poor scalability
MoCap-Based Dance Synthesis	MoCap skeletons	Early probabilistic motion modeling	Style rigidity
RNN Dance Motion Modeling	MoCap	Captured temporal dependencies	Long-sequence drift
Music-to-Dance Mapping	Audio + MoCap	Beat-synchronized motion	Limited diversity
Video-Based Dance Learning	RGB video	Pose learning from videos	Pose noise
GAN-Based Dance Generation Lund et al. (2023).	Pose sequences	Improved realism	Training instability
Style-Conditioned Dance	MoCap	Style-aware choreography	Mode collapse
Transformer for Motion	MoCap	Global sequence coherence	High computation
Cross-Modal Dance Synthesis	Audio + Pose	Strong music–motion coupling	Data intensive
Diffusion Motion Models	Pose	Smooth motion transitions	Slow inference
Dance Style Transfer	Pose + Music	Style transfer across dances	Style leakage
Multimodal Dance Generation	Video + Audio	Rich multimodal fusion	Complex training
Real-Time Dance Avatars	Pose streams	Interactive performance	Limited realism

Hybrid Transformer–Diffusion	Pose + MIDI	High realism & smoothness	Computational cost

3. METHODOLOGY

3.1. DATASET PREPARATION (POSE DATA, VIDEO FRAMES, MIDI BEATS)

The choreography generation methodology starts with an extensive dataset preparation procedure that receives various data modalities necessary in choreography generation. Pose data are the fundamental representation of human motion, which come in either by motion capture system or pose estimation on videos of dances. These data are time-indexed joint coordinates giving the kinematic structure of the performer. Raw video frames are simultaneously captured to maintain visual context, space relations and other style clues like body positioning, arm length and body posture. Video data also make it possible to check the accuracy of poses and extend this knowledge to appearance-based features in the future. MIDI beat information is mined out of other music tracks to integrate music structure. MIDI data give direct indications of tempo, beat positions, rhythm and note intensity, enabling them to be programmed with movement and music in perfect sync. In case no MIDI files are found, audio signals are manipulated to recognize beats and rhythmic patterns and transmitted into MIDI-like time indicators. The sequence of dances is divided into synchronized windows at the temporal coordinates of pose path, video frame, and beat marking. There is also metadata like style of dance, range of tempo, and identity of the performer to facilitate conditional generation and control of style. This multi-mod data preparation makes sure that the learning models not only view the physical dynamics of motion but also the rhythmic alignment in movement to music creating a strong basis of expressive choreography synthesis.

3.2. PREPROCESSING: NORMALIZATION, POSE SKELETON EXTRACTION, SEQUENCE ALIGNMENT

Preprocessing is important in converting the raw multimodal data to structured inputs that can be processed using the deep learning models. Pose skeleton pose estimation Pose skeleton pose estimation is the stage of transforming raw motion capture signals or approximated joint positions into conventional skeleton models. Hierarchies are computed jointly based on anatomical constraints, such that all the samples share similar connectivity. Missing/Noisy background on the values of the joints, which is normal in the process of video-based pose estimation is addressed with interpolation and temporal smoothing, which minimise jitter and discontinuities. It is then normalized in order to enhance the model stability and generalization. The process of spatial normalization is with respect to a reference joint, (say the hip or torso) and normalizes them to compensate performer height variation and variation in camera distance. Temporal normalization brings about the same frame rate in sequences by resampling the motion data to the fixed number of frames/s. The beat sequences in music are also scaled to a usual tempo grid allowing it to be regularly associated with motion frames. Sequence alignment involves the combination of movement and music in time-synchronized streams.

3.3. MODEL DESIGN: RECURRENT (LSTM, GRU) AND TRANSFORMER-BASED NETWORKS

The model design is aimed at the complex temporal dependencies and multimodal interaction of the choreography of a dance. The ability to capture contextual information over a long duration of time means that recurrent neural networks, specifically Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are used as a baseline sequence model. These processing models encode and decode frame by frame, and for such processes to be learned, temporal relationships between one movement and the next, one beat of the music and the next, and so on, are required.

The explicit memory cells and gating mechanisms allow LSTM networks to be well modelled to work with the long dance phrases with anything resembling the performance with their more efficient computation mechanism that GRUs provide. Figure 1 illustrates recurrent and transformer architecture which allows choreography generation. Transformer-based architectures are added to the framework in order to address the weaknesses of recurrent models in working with very long sequences. To learn global dependencies across the motion sequences of a body part, transformers employ self-attention mechanisms to mimic the recurrent motifs, structural symmetry and long-range coordination of body parts. Multimodal embeddings are pose features that were enriched with beat and tempo information and it is jointly processed using stacked attention layers. Positional encoding is used in order to maintain time in the sequence.

Figure 1

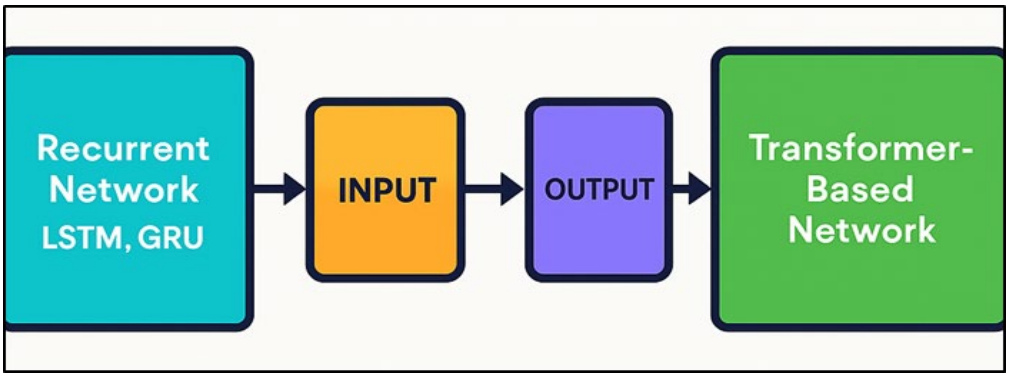

Figure 1 Architecture of Recurrent and Transformer-Based Models for Choreography Generation

4. SYSTEM ARCHITECTURE

4.1. INPUT MODULE: MUSIC, RHYTHM, AND MOVEMENT DATA STREAMS

The input module is the cornerstone of the choreography generation system as it combines various data sources that constitute auditory and physical features of dancing. The input of music can be in the form of raw audio signals or symbolic input in the form of MIDI files which are coded with tempo, beat placement, and rhythmic strength. Based on these inputs, spectral energy, tempo curves and beat onsets are obtained as low-level features that represent the musical structure of the movement dynamics. Streams of rhythm are focused on the time and thus generated choreography is synchronized to musical accents and phrasing. Movement data streams are pose-based skeletal representations, which are generated as a result of motion capture or video-based pose estimation. The positions of joints, angles and relative limb orientations are represented in each frame and create a time-ordered record that captures human motion. Music and movement streams are synchronized in time with a common timestamp structure in order to facilitate multimodal learning. Conditioning variables may include optional metadata (e.g. dance style, emotional tone, performance constraints, etc.) to allow choreography generation to be controlled.

4.2. FEATURE EXTRACTION USING CNNS AND TEMPORAL ENCODERS

The layer of feature extraction converts raw multimodal inputs into informative representations, which are small and can be used in sequence modeling. Convolutional Neural Networks (CNNs) are used to run visual data, such as video frames or heatmap of the position of humans in the space. These networks fly-capture spatial relationships including body posture, limb positioning and inter-joint association which are vital in comprehending movement organization. In case of pose-only input, skeletal connectivity and joint correlations can be effective with the help of graph-based or 1D convolutional layers. The musical features are analyzed using special encoders which derive rhythmic and harmonic patterns out of audio spectrograms or MIDI representations. These encoders maintain time continuity as well as dimensionality reduction that facilitates the effective fusion of multimodal information. The mechanisms of feature fusion integrate a representation of encoded music and movement by concatenation, by attention-based weighting or cross-modal transformers.

4.3. SEQUENCE GENERATION VIA LSTM/TRANSFORMER LAYERS

The sequence generation component aims at generating timely coherent and expressive choreography based on latent feature embeddings. The recurrent architectures that are applied to capture frame to frame dependencies and gradual change in movement are recurrent architectures like Long Short-Term Memory (LSTM) networks. LSTMs can learn the dynamics of contemporary poses during motion and rhythmic patterns and predict the future progression of the dance, generating an entirely continuous sequence of posts, by remembering internal states of memory. Figure 2 depicts the LSTM Transformer architecture in sequence generation in choreography synthesis. Information flow is controlled by gated mechanisms by decreasing sudden changes and providing biomechanical plausibility. Addition of transformer-based layers is done to improve global sequence modelling.

Figure 2

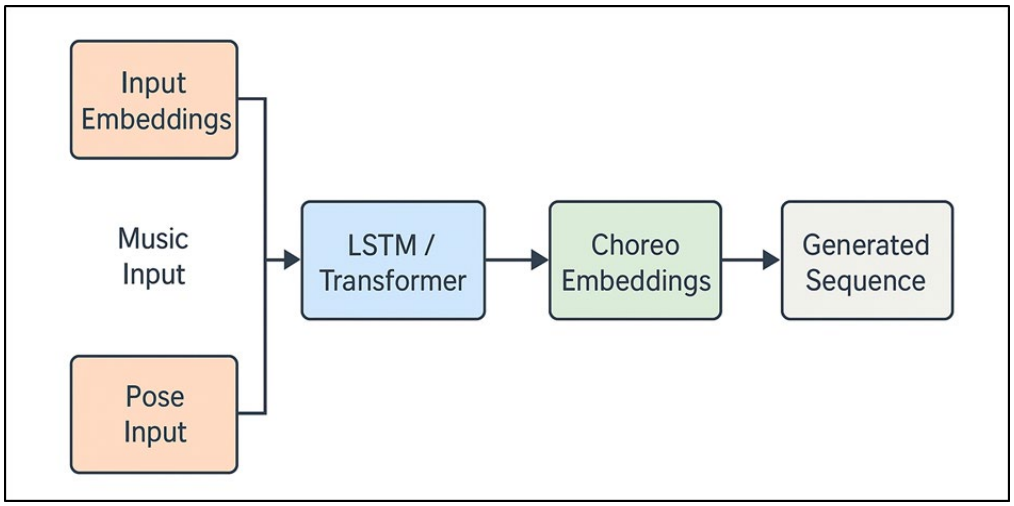

Figure 2 LSTM–Transformer–Based Sequence Generation Architecture for Choreography Synthesis

Transformers can compare the relationships between all the time steps simultaneously by means of self-attention and thus discover the long-range dependencies, repetitive motifs, and structural patterns in choreography. Such an ability is specifically relevant to the creation of long dance patterns and their alignment with the same theme and musical orientation. The positional encodings maintain the time series, whereas the multi-head attention enables the model to pay attention to various elements of motion and rhythm at the same time.

5. APPLICATIONS AND IMPLICATIONS

5.1. AI-ASSISTED DANCE EDUCATION AND TRAINING SYSTEMS

Deep learning generated choreography can be used to revolutionize dance education and training through the provision of intelligent, adaptable and accessible learning devices. Systems based on AI have the potential to produce personalized practice routines, depending on the skill set of the learner, style of dance, and dance tempo, therefore, allowing them to engage in individual training plans. Students have visual references of posture, timing and transitioning of movements by creating skeletal animation with images or virtual avatars. Repetitive practice may also be aided through such systems that provide variations of the same sequence allowing learners to become flexible, to memorize and to range. AI-assisted tools are helpful to instructors in terms of automatic content generation and analytical feedback. Teaching content can be generated choreography where rhythm alignment is demonstrated, spatial patterns, or a difference in style between the different forms. AI systems can be used together with motion capture or camera-guided tracking so that the movement of the student can be compared to generated or expert reference sequences and offered quantitative feedback regarding truthfulness, smoothness, and time. This objective measure is a supplement of the conventional qualitative training and assists in identifying points of improvement.

5.2. VIRTUAL PERFORMANCE DESIGN AND STAGE SIMULATION

The application of AI-created choreography is crucial to the design of virtual performance and simulation of the stage, especially with regard to digital and hybrid performance. Deep learning models allow generating dynamic dance sequences the visualization of which can be performed in the form of a virtual avatar, motion graphics or holograph projection. These functions assist choreographers and stage designers to explore movement patterns, spatial structures and timing without necessarily having to physically practice them. Tests in virtual stage simulation enable designers to experiment with choreographies in varying lighting and camera angles and stage layouts to make planning more efficient and exploration of creative possibilities. Live performances Artificial intelligence can be used to generate choreography in real-time, with a musical backdrop and interactive systems to make adaptable performances that change tempo, respond to user input, or environmental changes. This brings new prospects of participatory and immersive dance experiences. In computer-generated performances, including online performances and metaverse events, AI

choreography allows the production of the visually interesting one on a scale, without the need to involve people in the large-scale human motion capture sessions.

5.3. CROSS-DOMAIN USE IN ANIMATION, FILM, AND VR ENVIRONMENTS

In addition to dance-related uses, choreography generation using deep learning has wide application in the fields of animation, film, and virtual reality (VR). The production of lifelike and expressive movement of characters is a constant challenge in animation and development of games. Choreography models based on AI offer automated motion generation, which eliminates the use of manual keyframing, or the costly motion capture, with visual realism at any rate of production speed. Dance sequences generated can be scaled to alternate character models, styles, and narrative settings and help to create content at scale. Artificial intelligence-generated choreography may be useful in film and digital media, helping pre-visualization in movies and animation films, where the director and animators can experiment with the storytelling through movements prior to full production. Sophisticated crowd scenes, background dancers, or staged performances can be created at high efficiency to increase the flexibility of creativity. In VR and augmented reality (AR), AI choreography helps provide immersion through the generation of real-life-like avatars that behave in a natural manner to the music and the user's actions. It will be useful especially on virtual concerts, interactive exhibitions and social VR platforms.

6. EXPERIMENTAL SETUP AND RESULTS

6.1. DATASET DESCRIPTION AND SPLIT RATIOS

Experimental assessment is performed on a curated set of choreography videos in the shape of synchronized dance videos, pose-based skeletal sequences and related music tracks with beat annotations. It has a variety of dance styles, tempos, and variations of the performers to achieve diversity and generalization. The pose sequences are separated at a fixed frame rate and musical features are synchronized with the movement data. To train and evaluate the models, the dataset is split by standard split strategy 70 percent of the sequences will be used to create the training set, 15 percent to create the validation set, and 15 percent to create the testing set. Such a divide guarantees enough data to learn the model and at the same time provides an ability to evaluate the performance without any bias. All the subsets maintain cross-style representation to prevent any distributional bias.

Table 2

Table 2 Dataset Composition and Split Ratios

Dance Style	Total Sequences	Avg. Duration (s)	Training (70%)	Validation (15%)
Ballet	420	18.5	294	63
Contemporary	510	16.2	357	77
Hip-Hop	380	14.8	266	57
Folk/Traditional	290	20.1	203	44
Jazz	310	15.6	217	47

Table 2 shows the structure of choreography data and its distribution on training and validation subsets of the various styles of dances. The collection of data shows equal stylistic diversity including classical, contemporary, urban, and traditional styles of dance. The total generated sequence of dance in each style is distributed as indicated in Figure 3. Contemporary dance adds the most sequences (510) as it has a broad stylistic range and can be studied to learn how to move in a flexible manner.

Figure 3

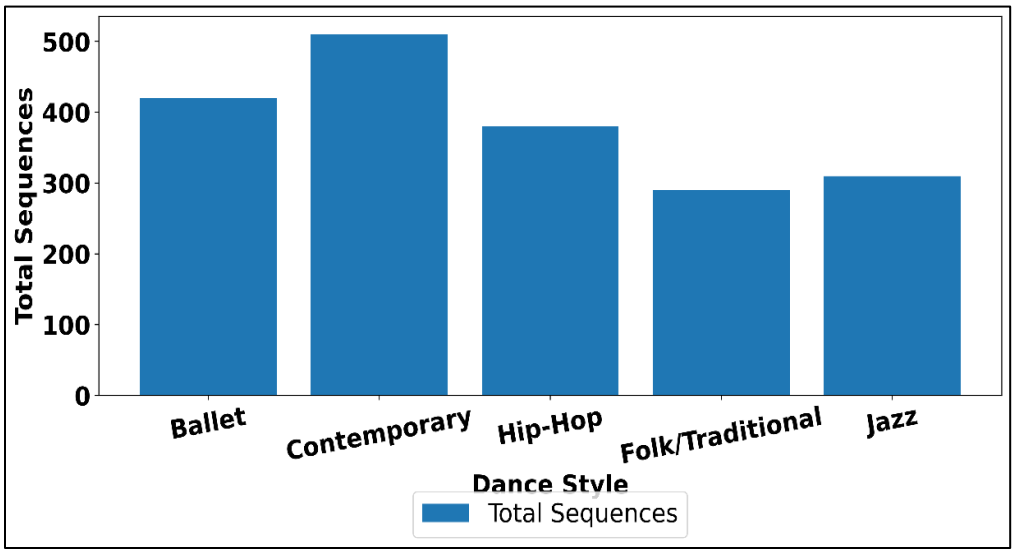

Figure 3 Distribution of Total Dance Sequences Across Styles

Ballet and jazz offer orderly movement vocabularies with moderate number of sequences which facilitates the modelling of controlled and rhythmically accurate choreography. The Hip-hop sequences highlight high tempo and change-of-direction and can therefore be useful in assessing the temporal sensitivity of deep learning models. Although folk and traditional dances are the least numerous, they have the longest average (20.1 seconds), and they embrace extensive rhythmic cycles and culturally rich movement patterns. The bar visualization of the metrics of dance style datasets is compared as seen in Figure 4.

Figure 4

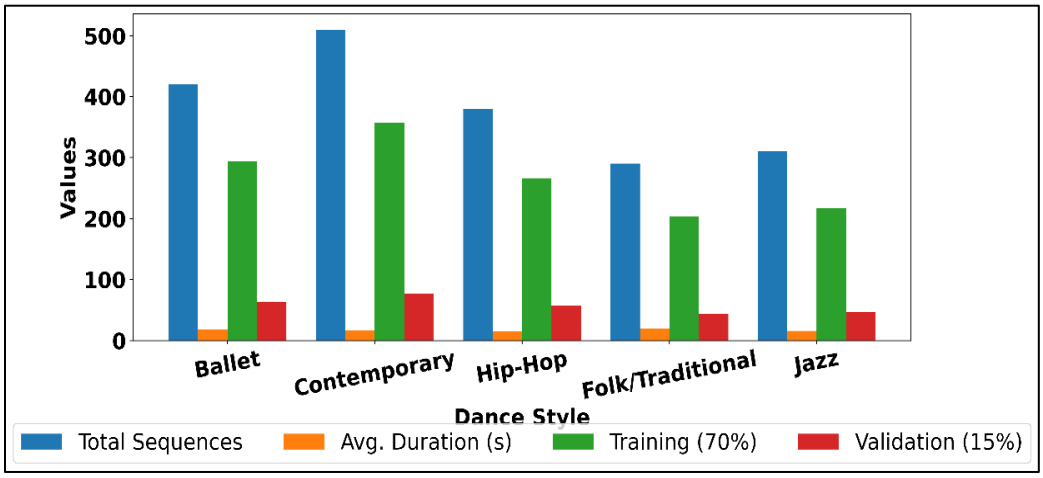

Figure 4 Comparative Bar Visualization of Dance Style Metrics for Dataset Preparation

The 70% training/15 percent validation split is constant across all styles so that there is consistency in exposure to learning and the performance can also be tuned on representative validation data. It offers a strong generalization capability by avoiding style-specific bias and offers a justifiable comparative appraisal of choreography generation models in various spheres of movement.

6.2. QUANTITATIVE EVALUATION METRICS (MAE, FID, SMOOTHNESS INDEX)

Quantitative measures of model performance are applied to measure the accuracy, realism, and motion continuity. Mean Absolute Error (MAE) is used to measure the mean error between generated and ground-truth joint positions,

which is a measure of the accuracy of pose prediction. In motion features, Fréchet Inception Distance (FID) is used to measure the similarity of the distributions of real and generated dance sequences, which is perceptual realism and diversity. Temporal coherence is specifically measured with the help of a Smoothness Index, which is calculated using frame-to-frame change in joint velocities and accelerations, and punishing jitter and sudden transitions. The decreasing values of MAE and FID with increasing values of smoothness suggest better choreography generation quality in balance between accuracy, realism, and continuity of expressiveness.

Table 3

Table 3 Quantitative Performance Comparison of Choreography Generation Models

Model Type	MAE (↓)	FID (↓)	Smoothness Index (%)	Rhythm Alignment (%)
LSTM	0.084	42.6	86.3	82.1
GRU	0.079	39.8	87.5	83.6
Transformer	0.062	31.4	91.2	89.7
LSTM + GAN	0.068	34.9	92.4	88.1

Table 3 provides a comparison of the performance of various deep learning models in choreography generation based on the metrics of accuracy, realism, smoothness and rhythm alignment. The most basic LSTM model has the highest values of MAE and FID, which makes it less accurate in poses and less realistic, but still with a good smoothness thanks to the sequential memory structure. Figure 5 presents the comparison of the MAE of sequential and hybrid prediction models.

Figure 5

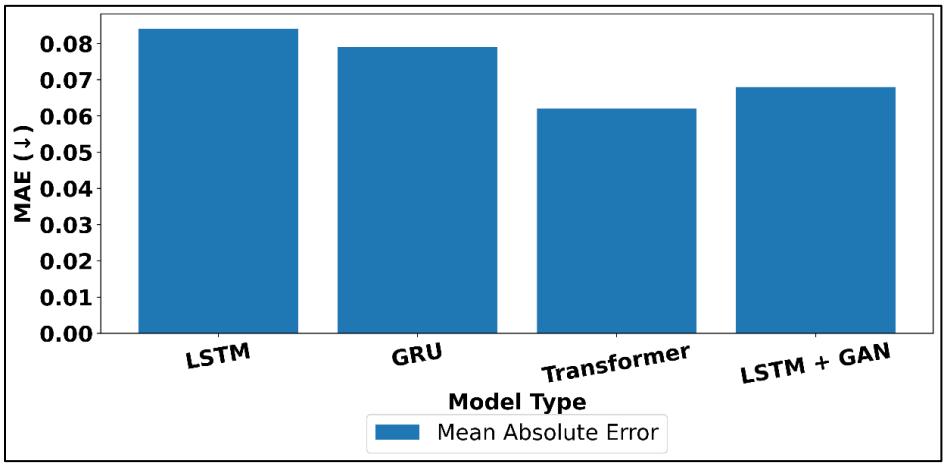

Figure 5 MAE Comparison Across Sequential and Hybrid Prediction Models

GRU model has slight improvements over LSTM in all metrics, which is a valid indication of its more efficient gating process and better response to the time dependencies. Figure 6 displays the comparison of FID, smoothness, rhythm alignment of architectures using area plot. Transformer based models show a significant performance improvement with a much lower MAE and FID value and an increased score of Smoothness Index and Rhythm Alignment.

Figure 6

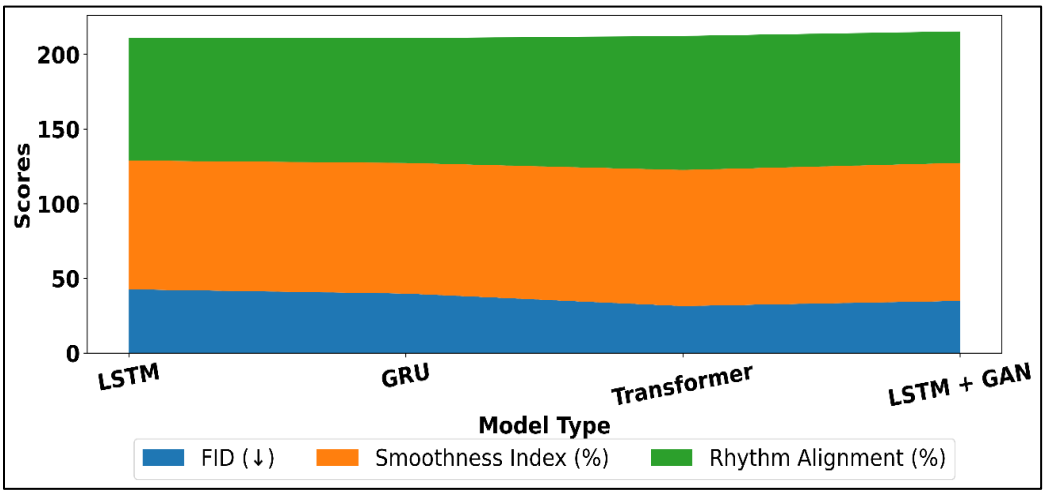

Figure 6 Area Plot of FID, Smoothness, and Rhythm Alignment Across Sequence Modelling Architectures

This brings out the success of self-attention in the attendance of long-range dependencies and international choreographic framework. The hybrid LSTM + GAN model goes further to increase the quality of perceptual, as indicated by the improvement of smoothness and rhythm alignment, because of the adversarial training, which promotes realistic motion transition. All in all, the findings suggest that there is a broad evolution of recurrent to attention-based and hybrid generative architectures, and more advanced models are able to generate a more fluid, accurate, and musically synchronized sequence of choreography.

7. CONCLUSION

This paper has provided extensive research on the deep learning models in choreography generation and the ways the model can revolutionize the process of creating, analyzing, and sharing dance. The proposed framework shows how complicated connections between movement and music can be effectively modeled with the help of the current techniques of artificial intelligence by employing multimodal data sources, including pose-based skeletal representations, video frames, and musical rhythm cues. With the shift towards the data-driven deep learning models of the dance sequence creation, the problem of traditional rule-based choreography systems is replaced with the idea of flexibility, colouring stylistic diversity, and expressiveness in the generated dance sequences. Recurrent architectures, such as LSTM models and GRU models, were found to be effective to express local temporal continuity and motion transition, whereas transformer-based networks to be more effective in expressing global sequence coherence and long-range dependencies. The use of modern generative methods also increased the realism and the fluidity, and it solved most of the popular issues like motion stuttering and unnatural transitions of poses. Quantitative measures (MAE, FID, Smoothness Index) of experimental evaluation showed that models based on deep learning could produce choreography, which was rhythmically and visually convincing across a wide range of dance styles. In addition to the technical performance, the consequences of this study are applicable in various fields of application. AI-assisted choreography systems can be used in the field of dance education and training to facilitate personalized learning, objective feedback, and remote training. Like in performance design, the virtual stage simulation and adaptive choreography introduces novel creative possibilities.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Chen, K., Tan, Z., Lei, J., Zhang, S. H., Guo, Y. C., Zhang, W., and Hu, S. M. (2021). ChoreoMaster: Choreography-Oriented Music-Driven Dance Synthesis. ACM Transactions on Graphics, 40, 1–13. https://doi.org/10.1145/3450626.3459849

Chen, Y., Wang, H., Yu, K., and Zhou, R. (2024). Artificial Intelligence Methods in Natural Language Processing: A Comprehensive Review. Highlights in Science, Engineering and Technology, 85, 545–550.

Copet, J., Kreuk, F., Gat, I., Remez, T., Kant, D., Synnaeve, G., Adi, Y., and Défossez, A. (2023). Simple and Controllable Music Generation. arXiv.

Croitoru, F. A., Hondru, V., Ionescu, R. T., and Shah, M. (2023). Diffusion Models in Vision: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45, 10850–10869. https://doi.org/10.1109/TPAMI.2023.3288679

Hong, C. (2024). Application of Virtual Digital People in the Inheritance and Development of Intangible Cultural Heritage. People's Forum, 6, 103–105.

Hou, C. (2024). Artificial Intelligence Technology Drives Intelligent Transformation of Music Education. Applied Mathematics and Nonlinear Sciences, 9, 21–23. https://doi.org/10.2478/amns-2024-0021

Lauriola, I., Lavelli, A., and Aiolli, F. (2022). An Introduction to Deep Learning in Natural Language Processing: Models, Techniques, and Tools. Neurocomputing, 470, 443–456. https://doi.org/10.1016/j.neucom.2021.10.040

Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., and Wang, Z. (2023). ChatGPT and a New Academic Reality: Artificial Intelligence-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing. Journal of the Association for Information Science and Technology, 74, 570–581. https://doi.org/10.1002/asi.24750

Pleshakova, E., Osipov, A., Gataullin, S., Gataullin, T., and Vasilakos, A. (2024). Next Gen Cybersecurity Paradigm Towards Artificial General Intelligence: Russian Market Challenges and Future Global Technological Trends. Journal of Computer Virology and Hacking Techniques, 7, Article 23.

Sun, Y., Dong, L., Huang, S., Ma, S., Xia, Y., Xue, J., Wang, J., and Wei, F. (2023). Retentive Network: A Successor to Transformer for Large Language Models. arXiv.

Tay, Y., Dehghani, M., Bahri, D., and Metzler, D. (2023). Efficient Transformers: A Survey. ACM Computing Surveys, 55, 16–21. https://doi.org/10.1145/3530811

Wu, J., Gan, W., Chen, Z., Wan, S., and Lin, H. (2023). AI-Generated Content (AIGC): A Survey. arXiv.

Zeng, D. (2025). AI-Powered Choreography Using a Multilayer Perceptron Model for Music-Driven Dance Generation. Informatica, 49, 137–148.

Zhang, L., and Zhang, L. (2022). Artificial Intelligence for Remote Sensing Data Analysis: A Review of Challenges and Opportunities. IEEE Geoscience and Remote Sensing Magazine, 10, 270–294. https://doi.org/10.1109/MGRS.2021.3132935