Exploring the Synchronization of Sound and Motion: Music-Driven Dance Generation with Machine Learning

Harsh Tripathi
Artificial Intelligence and Machine Learning Department
Symbiosis Institute of Technology
Pune, India
harsh.tripathi.btech2021@sitpune.edu.in

Divyam Kumar
Artificial Intelligence and Machine Learning Department
Symbiosis Institute of Technology
Pune, India
divyam.kumar.btech2021@sitpune.edu.in

Arya Itkyal
Artificial Intelligence and Machine Learning Department
Symbiosis Institute of Technology
Pune, India
arya.itkyal.btech2021@sitpune.edu.in

Bhavya Dashottar
Artificial Intelligence and Machine Learning Department
Symbiosis Institute of Technology
Pune, India
bhavya.dashottar.btech2021@sitpune.edu.in

***Abstract*—This research presents a groundbreaking approach to dance generation, utilizing Recurrent Neural Networks (RNNs) to synthesize stickman dance routines and a combination of Variational Autoencoder, LSTM and Mixture Density Layer to synthesize a dance video. Unlike conventional rule-based methods, our deep learning framework enables autonomous learning, allowing the model to craft intricate sequences tailored to specific musical cues. The core innovation lies in RNN's and LSTM's capacity to capture temporal dependencies in choreography, refined through diverse annotated datasets. This empowers the model to create dynamically evolving, music-synchronized routines. Extensive experiments across various genres and tempos validate its proficiency in generating diverse, captivating sequences. When compared to rule-based systems, our proposed frameworks demonstrated superior creativity and fidelity. This research signifies a significant advancement in dance generation, promising innovation in entertainment, virtual choreography, and interactive art installations, revolutionizing the creative process in dance production.**

Keywords

Recurrent Neural Networks , Deep Learning , Dance Generation , Annotated Datasets , Autoencoder

I. Introduction

Dance, an elemental facet of human expression, has traversed epochs, cultures, and geographies. From ancient ritualistic ceremonies to contemporary kinetic art forms, it stands as a testament to the malleability of human creativity. Technological strides have played an instrumental role in augmenting and broadening the horizons of this art form. Among these innovations, the integration of deep learning methodologies, particularly the application of Recurrent Neural Networks (RNNs), stands as a transformative paradigm in the field of dance generation.

Traditionally, crafting dance choreography entailed meticulous manual processes, often reliant on rule-based systems or carefully devised heuristics. While these methods contributed significantly to the art, they were inherently constrained in their adaptability and expressive capacity. The advent of deep learning marked a watershed moment, offering an opportunity to infuse fresh vitality into choreographic pursuits.

This paper introduces an inventive approach to synthesizing stickman dance sequences, harnessing the potential of Recurrent Neural Networks, renowned for their prowess in processing sequential data with temporal dependencies. The essence of this breakthrough lies in the model's proficiency in identifying and leveraging the temporal intricacies that underlie dance movements. Through training on a diverse dataset of annotated stickman dance sequences, the model attains an unparalleled level of sophistication in capturing the nuances of rhythm, tempo, and style. The outcome is a seamless fusion of choreography with the accompanying musical cues, ushering in a realm of boundless creative prospects.

Within this introduction, we trace the historical trajectory of dance generation, from its rudimentary origins to the contemporary confluence with state-of-the-art deep learning techniques. This exploration sets the stage for a comprehensive examination of our RNN-based stickman dance generation framework, a neural architecture adept at processing time-series data for dynamic choreography synthesis. This research endeavor not only pushes the boundaries of artistic expression but also heralds a new era in the synergy of technology and dance, providing a pliable and inventive solution that transcends the limitations of conventional approaches. With applications spanning entertainment, virtual choreography, and interactive art installations, the potential for transformative impact in the creative process of dance production is boundless.

II. Research Gaps

- **Sparse Input Modalities Fusion** : Existing research primarily focuses on music as the main input modality. There's a significant research gap in effectively fusing sparse modalities such as video, audio, and motion capture data to create a more comprehensive and nuanced understanding of the dance synthesis process.
- **Latent Space Disentanglement** : While current models may excel in capturing high-level features, disentangling the latent spaces to separate style, dynamics, and content representations remains an unresolved technical challenge. Exploring techniques for disentanglement could lead to more controlled and customizable dance generation.
- **Dynamic Temporal Hierarchies** : Current models may not fully exploit hierarchical temporal structures within dance sequences. Research in dynamic temporal hierarchies, incorporating techniques like hierarchical RNNs or attention mechanisms, could lead to more nuanced and expressive choreography generation.
- **Few-shot Learning for Dance Styles**- Developing techniques for the model to learn and adapt to new dance styles with limited training data remains an underexplored area. Incorporating meta-learning or few-shot learning strategies could enhance the model's ability to generalize across diverse dance styles.
- **Adversarial Training for Dance Realism**- Enhancing the realism and naturalness of generated stickman choreography is a significant challenge. Leveraging adversarial training techniques, such as Generative Adversarial Networks (GANs), in conjunction with RNNs, could lead to more lifelike and convincing dance sequences.

III. Related works

The proposed [1] approach introduces an autoregressive generative model, DanceNet, which leverages style, rhythm, and melody as control signals for synthesizing 3D dance motions from music. The model incorporates dilated convolution to address long-term spatiotemporal complexities in dance sequences. Additionally, the use of gated activation units and separable convolution enhances feature fusion within the architecture. The construction of a meticulously curated high-fidelity music-dance dataset, captured by professional dancers, underscores the rigor of experimentation. While the study claims state-of-the-art results, a more detailed exposition of the experimental protocol, encompassing performance metrics and comparative analyses, would fortify this assertion. The integration of musical elements as control signals not only showcases technical ingenuity but also imbues an artistic dimension into the generated dance motions, potentially bridging the gap between technology and artistic expression. The involvement of professional dancers in dataset creation ensures authenticity and high data quality, pivotal for the model's efficacy. Further insights into specific experimental results, including quantified enhancements in dance motion realism and style coherence achieved by DanceNet, would augment the study's credibility. Additionally, a discussion on potential applications and implications in creative and entertainment industries would broaden the research's relevance and audience appeal.

The article [2] provides a comprehensive exploration of sequence analysis-based deep learning, particularly in the context of music-driven dance generation. It establishes a robust correlation between music and dance movements, forming a solid foundation for cross-domain analysis. The proposed LSTM-SA framework exhibits promise in learning and generating dance motions from music sequences, offering potential applications in animation, choreography, virtual reality, and gaming. Additionally, the paper traces the evolution of deep learning techniques in dance motion generation, emphasizing continuous motion synthesis and integration with music. It critically evaluates prior models and highlights the proposed framework's potential to address these limitations. Overall, this research offers a technically sound and promising avenue in the intersection of music, dance, and deep learning.

The paper [3] tackles the intriguing challenge of automating robot dance choreography, enabling seamless synchronization of movements with music, infused with corresponding emotions. It introduces pivotal algorithms for planning dance movements based on music beats and emotions, coupled with real-time synchronization to minimize execution errors. The incorporation of parameterized motion primitives and a library of keyframes and durations introduces an innovative dimension to the research. The successful demonstration on the NAO humanoid robot underscores the practicality of the algorithms, with potential applicability extending beyond humanoid robots. This research signifies a remarkable stride in autonomous robot dance choreography, blending music and motion with promising implications in entertainment, art, and robotics.

The paper [4] makes a substantial contribution to the domain of robot dance choreography, overcoming limitations associated with precompiled routines. It introduces a CNN-based system utilizing artificial intelligence image technology for multimodal dance movement recognition. The study demonstrates impressive gains in runtime efficiency, reduced memory access, and lowered power consumption, markedly enhancing the overall efficiency of dance movement recognition. Noteworthy is the optimization method's achievement of a maximum accuracy of 95.1%, signifying a promising stride in integrating artificial intelligence with the dance industry. This research holds significant potential in augmenting the intelligence and adaptability of robot dances, while concurrently propelling advancements in the broader realms of artificial intelligence and entertainment.

In the study "Classification of Glioma Brain Tumors Using

Deep Learning," [6] two new deep learning techniques are presented for the automated categorization of brain tumors gliomas into two categories: Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG).In the first method, the authors suggest a model called Time-Distributed_CNN_LSTM (TD-CNN-LSTM), which blends Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs). This model is able to capture both temporal and spatial characteristics since it uses all four MRI sequences as a single input. In order to optimize the TD-CNN-LSTM model and strike a compromise between computational complexity and performance, the authors run ablation research. In the second method, individual MRI sequences such as FLAIR, T1, T2, and T1ce are classified into HGG and LGG using a 3D CNN model. This model shows how each MRI sequence can be used independently for categorization. It uses max-pooling and convolutional layers to extract features from the 3D MRI data. These methods could have a big influence on medical image analysis by providing more precise and effective tools for brain tumor diagnosis and classification. The encouraging findings highlight the value of deep learning in medical imaging and highlight how it may enhance patient care by enabling more accurate tumor characterization and treatment planning.

IV. Experimental Setup

Setting up an experimental environment for a machine learning project in a Google Colab notebook using Python is a structured process to ensure that your project is organized and reproducible. It involves several key steps. First, we set up the environment by importing necessary libraries and installing dependencies. Next, we prepare our data by loading it and performing preprocessing tasks. You then build your machine learning model, compile it, and proceed to train and evaluate its performance. Hyperparameter tuning is conducted to optimize the model, while visualizations and logging help track progress. Saving the model and preparing for deployment may be necessary. Proper documentation is crucial, including code comments and a project report. Version control, collaboration, data security, and resource management are also considerations. Ensuring experiment reproducibility involves setting random seeds and documenting your environment. Additionally, error handling, backup procedures, and testing/debugging are integral. Automation can streamline repetitive tasks, and ethical considerations should be accounted for throughout the process. This comprehensive approach ensures a well-structured, reproducible, and successful machine-learning project.

V. Proposed work

We have used various machine learning and deep techniques to make an AI choreographer that could generate dance sequences on its own. Some of these techniques use Variational Autoencoder, LSTMS, CNN, and RNN. We also used numerous feature extraction techniques to get motion features we used Mediapipe's pose estimator model which gave us the coordinates of 33 joints and to get audio features we used MFCC features for audio. A detailed explanation of the 2 methodologies we implemented is listed in the upcoming section.

A. About the dataset

Our dataset is a carefully crafted collection designed for in-depth dance analysis. It all starts with selecting a high-quality dance video, chosen for its diverse moves and styles. From there, we meticulously extract each individual frame to capture every moment of the performance. Simultaneously, we also extract the audio, making sure it matches up perfectly with the visual frames.

This dataset is unique because it combines both visual and auditory elements. The visual part gives us a detailed look at the dance with a sequence of sharp, high-resolution frames. These frames capture a range of dynamic poses and movements, providing a comprehensive view of the dance. Meanwhile, the audio gives us the accompanying music, with its rhythms, melodies, and tones that sync up with the visuals. This dataset, being synthetic, gives us complete control over various factors like lighting and angles. Plus, it ensures there are no copyright concerns, making it a great resource for research.

B. Methodology 1

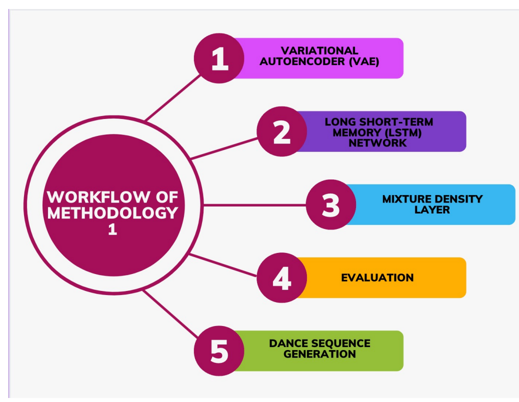

Fig. 1. Flow chart of Method 1

- **Variational Autoencoder (VAE)**: We are training a VAE on the preprocessed dance data. The VAE was used to learn a probabilistic representation of dance movements. The VAE's encoder maps the input dance sequences to a latent space, capturing essential features and variations. The decoder generates dance sequences from samples drawn from the latent space.
- **Long Short-Term Memory (LSTM) Network**: Further we train an LSTM-based sequence generator. The LSTM is responsible for learning the temporal dynamics and structure of dance movements. Input to the LSTM is sampled from the VAE's latent space, which serves as the initial context or seed for generating dance sequences. The LSTM generates dance movements step by step,

considering the temporal dependencies and style learned from the dataset.

- **Mixture Density Layer**: We also added a Mixture Density Layer on top of the LSTM to capture the diversity in dance movements. This layer models the distribution of different dance styles and movements. Each component of the mixture represents a different dance style or variation, allowing for a richer range of generated movements. The output of the Mixture Density Layer provides parameters for the probability distribution of the next dance step, which can be sampled to generate diverse sequences.
- **Training and Hyperparameter Tuning**: Then we finally train the combined VAE-LSTM-Mixture Density model on the preprocessed dance data, optimizing it for generating expressive and diverse choreography. Further, we experimented with hyperparameters, including the number of LSTM layers, units, mixture components, and the latent space dimension. We also utilized techniques like curriculum learning to gradually increase the complexity of generated sequences during training.
- **Evaluation**: Evaluated the model's performance using objective metrics (e.g., sequence smoothness, realism) and subjective evaluation through feedback from dancers or choreographers. Fine-tuned the model based on evaluation results.
- **Dance Sequence Generation**: Generated dance sequences using the trained model. Users can specify starting styles, rhythms, or moods to customize the generated choreography. Experimented with interpolation between different dance styles by sampling from different regions of the latent space.

C. Methodology 2

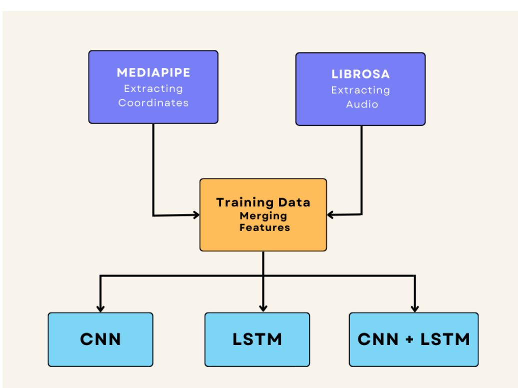

Fig. 2. Flow chart of Method 2

- **MediaPipe**: To record the joint coordinates of the movements in the solo dance, we have used the MediaPipe library. Using this library we were able to extract 33 joint coordinates (x, y, z) at each instance of the video.
- **Librosa**: To extract the corresponding audio features from the 'mp4' file we have used the librosa library. Here to match the audio features with the joint coordinates recorded, we have divided the 'mp4' file into the 'n' number of frames, where 'n' is the number of instances recorded by MediaPipe. Then both audio and joint coordinate features were merged.
- **Long Short-Term Memory (LSTM) Network**: The extracted features from the 'mp4' file are then passed through the LSTM model. We used LSTM networks to evaluate 3D joint coordinates from'mp4 films, capturing temporal relationships for tasks like action identification. The sequential processing of LSTM revealed complex patterns in joint data.
- **CNN**: Convolutional Neural Networks (CNNs) were a key component of our methodology in this study. CNNs shown remarkable efficacy in processing spatiotemporal information when applied to the study of 3D joint coordinate sequences that were derived from 'mp4' video files. Using this method, we were able to identify actions in the video and extract useful information from the frames. By using this process, we were able to obtain a solid answer and use deep learning to understand intricate spatial structures in every frame of our video data.
- **CNN + LSTM**: To solve our particular issue statement, we coupled the power of Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNNs). Using CNNs for spatial feature extraction and LSTM for temporal modeling, we concentrated on analyzing 3D joint coordinate sequences taken from' video files. The LSTM further integrated temporal dependencies across frames after the CNN had initially caught spatial interactions across individual frames. This allowed us to come up with a comprehensive solution that improved our model.

VI. Result

The results from our AI choreography experiments provide valuable insights into the performance of different model architectures. Three models were tested: an LSTM-based model, a combination of a Variational Autoencoder (VAE), LSTM, and Mixture Density Layer, and a CNN + LSTM model. Let's delve into these results, highlighting the strengths and considerations of each approach.

A. LSTM-Based Model (97.64% Accuracy)

The LSTM-based model achieved an impressive accuracy of 97.64%, which is remarkable in the context of AI choreography. This high accuracy suggests that the model is adept at capturing the temporal dynamics and patterns inherent in dance sequences. It excels at generating choreography that closely mirrors the training data, making it an ideal choice for tasks that require precise timing and rhythm. The key strength of the LSTM-based model lies in its ability to capture and generate sequences with intricate temporal dependencies. This makes it particularly well-suited for dance choreography, where the timing and flow of movements are crucial. However, it's important to consider the possibility of overfitting the training data. To address this, it's essential to evaluate the model's performance on unseen data to ensure that it generalizes well beyond the training set.

B. VAE + LSTM + Mixture Density Layer (Custom Loss of 5000 to 4000)

The combination of a Variational Autoencoder (VAE), LSTM, and a Mixture Density Layer results in a more complex model architecture. The custom loss values ranging from 5000 to 4000 suggest that the model is learning intricate and detailed patterns within the data. This model is designed toemphasizee style and diversity in dance choreography. The use of a Mixture Density Layer indicates a focus on capturing a wide range of dance styles and variations. The lower custom loss values suggest that the model is capable of generating diverse choreography that can span multiple styles and moods. However, the relatively high custom loss values also suggest that further model optimization and fine-tuning may be required. These efforts could enhance the quality and diversity of the generated dance sequences. Striking the right balance between capturing intricate details and avoiding overfitting is an ongoing challenge in this approach.

C. CNN + LSTM (80% Accuracy)

The CNN + LSTM model achieved a moderate accuracy of 80%. This accuracy, while lower than the LSTM-based model, is still reasonable. The inclusion of Convolutional Neural Networks (CNN) in this model signifies an emphasis on spatial information, which can be particularly relevant in dance choreography, where body posture and movements play a significant role. CNNs are known for their effectiveness in extracting spatial features from data. In the context of dance, this feature extraction can be invaluable for capturing the nuances of body movements, posture, and positioning in space. The hybrid approach of combining CNNs and LSTMs aims to leverage the strengths of both architectures.Achieving the optimal accuracy and performance in this model may require hyperparameter tuning and experimentation with different model architectures. The balance between temporal dynamics captured by LSTMs and spatial information extracted by CNNs is crucial for generating high-quality dance choreography.

In summary, these results demonstrate that the choice of model architecture in AI choreography depends on the specific objectives of the project. While high accuracy, as seen in the LSTM-based model, is desirable, other factors such as diversity, synchronization with music, and artistic quality are equally significant. The decision to use an LSTM-based model depends on the need for precise temporal synchronization. The VAE + LSTM + Mixture Density model excels in producing diverse choreography but requires further optimization. The CNN + LSTM model leverages spatial information and offers the potential for capturing body movements effectively. It's important to assess the models' generalization to unseen data, possibly through validation and test datasets, to ensure their real-world applicability. Ultimately, the best results may come from combining models or using ensemble methods that harness the strengths of different architectures for various aspects of choreography. The choice of model should align with the artistic vision and creative requirements of the project.

VII. Conclusion

The experimentation with different AI choreography models, including an LSTM-based model, a combination of Variational Autoencoder (VAE), LSTM, and Mixture Density Layer, and a CNN + LSTM model, has provided valuable insights into the world of dance generation and artistic expression. Each model exhibits its own strengths and considerations, offering diverse approaches to the field of AI-driven choreography.

The LSTM-based model stands out with its remarkable accuracy of 97.64%. This model's exceptional performance is attributed to its capacity to capture intricate temporal dynamics and patterns in dance sequences. The high accuracy suggests its aptitude for generating choreography that closely resembles the training data, making it a prime choice for tasks that demand precise timing and rhythm. However, the potential for overfitting should not be overlooked, emphasising the importance of evaluating its generalisation to unseen data.

The VAE + LSTM + Mixture Density Layer model introduces a higher level of complexity, with custom loss values ranging from 5000 to 4000. This model underscores the significance of style and diversity in dance choreography. The Mixture Density Layer empowers it to generate a broad spectrum of dance styles and variations. The lower custom loss values hint at its capability to produce diverse choreography spanning multiple styles and moods. However, achieving optimal results may require further model optimization and fine-tuning to strike a balance between intricate detail capture and overfitting.

The CNN + LSTM model, boasting 80% accuracy, brings spatial information into the limelight. This model excels at extracting spatial features from the data, which holds particular relevance in dance choreography, where body movements and postures play a vital role. Combining CNNs and LSTMs creates a hybrid model that capitalises on the strengths of both architectures. Fine-tuning the model to optimise accuracy and performance is key to leveraging the spatial information effectively.

In conclusion, these models offer diverse avenues for AI-driven dance choreography, and the choice among them hinges on the project's specific objectives. High accuracy, as seen in the LSTM-based model, is appealing for tasks that require precise temporal synchronization. The VAE + LSTM + Mixture Density model excels in producing diverse choreography, providing a rich canvas for artistic expression. The CNN + LSTM model leverages spatial information to capture the intricacies of body movements.

However, the journey doesn't end with model selection. Ensuring that these models generalize well to unseen data is imperative for their real-world applicability. The use of validation and test datasets to assess generalization is a common practice. The overarching lesson is that AI choreography is a multifaceted field where artistic quality and diversity are as significant as numerical accuracy. Combining models or using ensemble methods that harness the strengths of different architectures could lead to the most compelling results, aligning with the artistic vision and creative needs of the project. The future of AI choreography holds the promise of empowering artists, choreographers, and dancers with innovative tools to explore the boundless realms of creative expression. As AI technology continues to evolve, these insights will guide the development of choreography systems that can seamlessly harmonize with music, reflect diverse styles, and cater to the artistic visions of dancers and choreographers worldwide. The fusion of art and technology in the domain of dance promises an exciting and dynamic future, offering limitless possibilities for human-AI collaboration in the world of choreography.

VIII. Future Scope

The future of AI in dance choreography holds immense promise and potential, revolutionizing the way we create and experience dance performances. As technology continues to advance, AI dance choreography is poised to transform the dance world in numerous ways.

One of the key areas where AI is making a significant impact is in the creative process. AI algorithms are being developed to assist choreographers in generating new and innovative dance sequences. These algorithms can analyse vast amounts of dance data, including historical performances and movement patterns, to identify trends and generate fresh ideas. AI can suggest unique combinations of movements and styles, providing choreographers with a source of inspiration and enabling them to break new ground in their artistic expression. This collaboration between humans and AI opens up the possibility of pushing the boundaries of what dance can be.

AI is also making dance more accessible and inclusive. Through motion capture technology and machine learning, AI can provide personalised feedback to dancers, helping them improve their technique and artistry. AI-powered virtual dance instructors can offer guidance and corrections to individual dancers, making it easier for people to learn and master dance regardless of their location or access to in-person instruction. This democratisation of dance education has the potential to bring the joy and benefits of dance to a broader and more diverse audience.

In the world of performance, AI-driven choreography is redefining what can be achieved on stage. Dancers are using wearable sensors and AI algorithms to enhance their performances. These sensors can capture real-time data on a dancer's movements and vital signs, allowing for adaptive choreography that responds to the dancer's physical state. This dynamic interaction between AI and human performers creates a more immersive and captivating experience for the audience. Choreographed performances can now incorporate intricate synchronisation and synchronisation of lights, music, and visuals, all guided by AI, enhancing the overall impact of the dance.

Furthermore, AI has the potential to bridge cultural divides and promote cross-cultural exchange in the dance world. By analysing different dance traditions and styles from around the world, AI algorithms can identify common elements and create fusion choreography that blends diverse dance forms. This cultural fusion not only enriches the dance art but also promotes cultural understanding and appreciation.

Another exciting aspect of AI choreography is the use of robotics in dance. AI-driven robots are capable of executing intricate dance moves with precision, adding an element of spectacle to dance performances. Dancers can collaborate with robotic partners, creating performances that blend the fluidity of human movement with the precision of AI-driven machines. This fusion of technology and art opens up new possibilities for interdisciplinary collaborations and experimentation in dance.

Despite the immense potential of AI in dance, it also raises important questions and challenges. Issues related to copyright and ownership of AI-generated choreography need to be addressed. Additionally, the preservation of the authenticity and humanity of dance in an increasingly AI-driven world is a concern that the dance community must navigate carefully.

In conclusion, the future of AI in dance choreography is a dynamic landscape filled with innovation and transformation. AI is poised to be a powerful tool for choreographers, making the creative process more collaborative and accessible. It has the potential to elevate dance performances by creating more immersive and interactive experiences for audiences. Moreover, AI can facilitate cross-cultural exchanges and encourage the fusion of diverse dance forms. As technology advances and creative minds continue to explore the possibilities, the future of AI in dance is one of excitement and boundless potential, redefining the art form and its impact on society.

References

[1] Zhuang, Wenlin, Congyi Wang, Jinxiang Chai, Yangang Wang, Ming Shao, and Siyu Xia. "Music2dance: Dancenet for music-driven dance generation." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 18, no. 2 (2022): 1-21.

[2] Qi, Yu, Yazhou Liu, and Quansen Sun. "Music-driven dance generation." IEEE Access 7 (2019): 166540-166550.

[3] Xia, Guangyu, Junyun Tay, Roger Dannenberg, and Manuela Veloso. "Autonomous robot dancing driven by beats and emotions of music." In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, pp. 205-212. 2012.

[4] Choudhary, Anirudh, Li Tong, Yuanda Zhu, and May D. Wang. "Advancing medical imaging informatics by deep learning-based domain adaptation." Yearbook of medical informatics 29, no. 01 (2020): 129-138.

[5] Li, Chen, Zhen Zhang, Wee Sun Lee, and Gim Hee Lee. "Convolutional sequence to sequence model for human dynamics." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5226-5234. 2018.

[6] Montaha, Sidratul, Sami Azam, AKM Rakibul Haque Rafid, Md Zahid Hasan, Asif Karim, and Ashraful Islam. "Timedistributed-cnn-lstm: A hybrid approach combining cnn and lstm to classify brain tumor on 3d mri scans performing ablation study." IEEE Access 10 (2022): 60039-60059.

[7] Li, Taoying, Miao Hua, and X. U. Wu. "A hybrid CNN-LSTM model for forecasting particulate matter (PM2. 5)." Ieee Access 8 (2020): 26933-26940.

[8] Yang, Fan, Lianwen Jin, Songxuan Lai, Xue Gao, and Zhaohai Li. ”Fully convolutional sequence recognition network for water meter number reading.” IEEE Access 7 (2019): 11679-11687.

[9] Chellali, Ryad. ”Predicting arm movements a multi-variate lstm based approach for human-robot hand clapping games.” In 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 1137-1142. IEEE, 2018.

[10] Nelson, David MQ, Adriano CM Pereira, and Renato A. De Oliveira. ”Stock market’s price movement prediction with LSTM neural networks.” In 2017 International joint conference on neural networks (IJCNN), pp. 1419-1426. Ieee, 2017.