Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on Algorithmic Aesthetics and Artificial Intelligence in Visual Arts, Cultural Education, and Creative Practice 2025 6(2s)

DEEP LEARNING FOR PERFORMANCE ASSESSMENT IN DANCE AND MUSIC

Mithhil Arora [1], Samrat Bandyopadhyay [2], Mona Sharma [3], Sakshi Sobti [4], Sanika Sahastra Buddhae [5], Dr. Anil Hingmire [6]

[1] Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, Solan, 174103, India
[2] Assistant Professor, Department of Computer Science and IT, ARKA JAIN University Jamshedpur, Jharkhand, India
[3] Assistant Professor, School of Business Management, Noida International University 203201, India
[4] Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India
[5] Assistant Professor, Department of Interior Design, Parul Institute of Design, Parul University, Vadodara, Gujarat, India
[6] Department of Computer Engineering, Vidyavaridhi's College of Engineering and Technology, Vasai, Mumbai University, India

Received 07 February 2025
Accepted 29 April 2025
Published 16 December 2025

Corresponding Author
Mithhil Arora,
mithhil.arora.orp@chitkara.edu.in

DOI
10.29121/shodhkosh.v6.i2s.2025.6751

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

ABSTRACT

Dance and music performance quality has been a consistently debated aspect of performance that has largely been based on human judgment which is subject to bias and lack of consistency. The recent progress in artificial intelligence, and especially deep learning provides a strong alternative to objective, data-driven assessment. This paper hopes to investigate the application of deep learning models Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformers, and Gated Recurrent Units (GRUs) to evaluate the performance of artists in these two areas. The proposed framework uses a mixture of multimodal data including motion capture or audio data and visual data to extract and combine features which determine rhythm and expression, synchronization and technical accuracy. The methodology focuses on the strong training, validation and testing plans to provide the accuracy and generalization to all the performers and genres. One use of this research is in real time feedback systems to learn music and dance, in competitions, automated scoring, and intelligent tutoring systems that adjust to the level of performance of the learner. Moreover, the paper identifies the opportunities of cross-cultural dataset growth, methods of bias mitigation, and explainable AI processes to provide transparency in automated assessments. The outcomes of the experiments prove the effectiveness of deep learning models to capture subtle features of performance, which is better than the traditional and classical approaches to machine learning. This study is part of the emerging convergence of artificial intelligence and performing arts, which will open the door to more equitable, more knowledgeable, and more globally applicable evaluation mechanisms.

Keywords: Deep Learning, Performance Assessment, Music Analysis, Dance Evaluation, Feature Fusion, Explainable AI

1. INTRODUCTION

Creativity, emotion, and technical ability of dance and music as an artistic performance comes across as an artistic unity. The critique of such performances has especially been an ambiguous and subjective task, which greatly depends

on professional judges, teachers, or colleagues, who evaluate basing on their experience and imagination. Human evaluation has great contextual insight, but is biased, fatigued and uneven. The requirement to measure performance (objectively, consistently, and on a scale) has hence motivated increased attention to the computational methods. Late developments in artificial intelligence (AI), and deep learning, in particular, have enabled music and movement to be analysed with greater precision and interpretability than ever before. Deep learning is a subfield in machine learning that has transformed how systems learn inputs based on data by way of automatically identifying hierarchical features of raw inputs like images, audio signals, and motion trajectories. This renders it especially apt towards the performing arts, where various modalities of sensations such as visual, auditory and sometimes even physiological come into play dynamically. When applied to music, deep learning models can extract the characteristics of pitch, rhythm, timbre, and dynamics, allowing to assess the technical competence, expressiveness, and synchronization. On the same note, neural networks can handle spatial and temporal motion data to determine balance, rhythm, posture and the overall aesthetic in dance Lei et al. (2022). Interdisciplinary quality of this field borders in computer science, psychology, biomechanics and the fine arts. Old-fashioned performance assessment models, however useful, are limited in terms of scaling up and objectivity. They rely on handcrafted attributes, static attributes and low interpretability. Deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformers, and Gated Recurrent Units (GRUs), on the other hand, are able to learn complex time-and-location relationships in artistic data on their own. These architectures are able to analyze visually and auditory signals in detail- they provide a more holistic view of performance Zhai (2021).

Figure 1

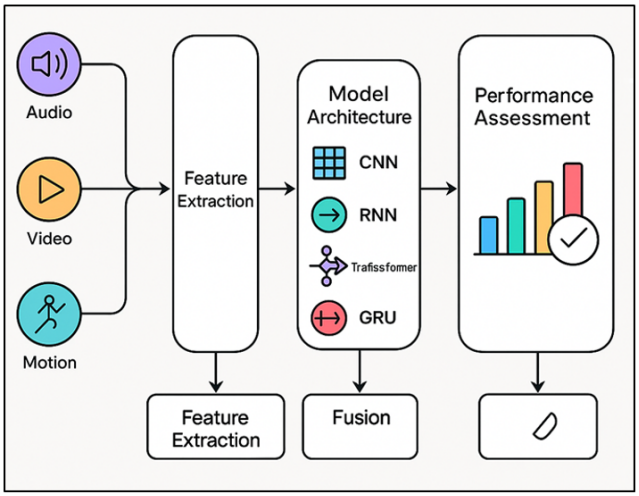

Figure 1 Multimodal Deep Learning Architecture for Performance Assessment in Dance and Music

In addition, multimodal data fusion methods contribute to the enhancement of the strength of the assessment systems. Figure 1 represents a multimedia system involving audio, video, and motion to develop intelligent artistic appraisal. As an example, it is possible to synchronize motion data of the dance movements with the musical rhythm to enable the model to assess the degree to which a dancer is moving in line with the beat. Such fusion does not only enhance accuracy but it also brings the assessment nearer to that of a human being Jin et al. (2022). Deep-learning-powered real-time feedback systems can transform the education of music and dance by providing instant feedback on a data-driven analysis of what a performer succeeds at and what he or she can improve. This has the ability to make high-quality feedback democratic, especially in remote or resource constrained educational settings. Irrespective of these developments, there are a number of challenges. The creation of trustworthy datasets that would include multiple cultural traditions, genres, and styles of performing is essential to prevent algorithmic bias.

2. LITERATURE REVIEW

2.1. TRADITIONAL METHODS FOR PERFORMANCE ASSESSMENT

The conventional means of evaluating the performance in the sphere of dance and music has been based mostly on the qualitative evaluation criteria and the expert judgment. During dance, teachers and judges would judge the rhythmic

coordination, posture, fluidity and expressiveness by sight. On the same note, assessment of music performance entails analysis of tone quality, pitch accuracy, rhythm accuracy, phrasing and emotional performance. Although these methods have a profound artistic pedagogical and human intuition foundation, they nonetheless are jeopardized by subjectivity, bias, and evaluator inconsistency Dias et al. (2022). The use of standardized scoring rubrics has been implemented in an effort to overcome variability, whereby there are structured assessment criteria. Nevertheless, such rubrics are not free of personal interpretations, culture, and the context of a situation. Additionally, such evaluations are not very scalable, due to their manual nature: particularly, in large-scale competitions or educational institutions. Motion capture systems, acoustic analysis equipments, and visual recordings are some of the technological interventions that have been used to supplement human judgment. However, they are mostly used as assistance and not as independent assessors Davis et al. (2023). They need professional interpretation and they cannot entirely reproduce the active communication between technical prowess and expressive art. As a result, the necessity of the objective, data-based, and reproducible systems of evaluation has become more obvious. The increasing mismatch between the qualitative evaluation tradition and the need to be quantitatively precise has triggered the sophistication of computational intelligence into the performing arts, opening the door to machine learning and deep learning-driven evaluation systems.

2.2. MACHINE LEARNING APPROACHES IN ARTISTIC PERFORMANCE EVALUATION

Prior to the development of deep learning, machine learning (ML) methods were part of the foundation of automation in the evaluation of performances. The common ML models used were Support Vector Machines (SVMs), Random Forests, k-Nearest Neighbors (k-NN), and Hidden Markov Models (HMMs), which were widely used to analyze audio, motion, and visual data. As an example, in music, Mel-Frequency Cepstral Coefficients (MFCCs), tempo and spectral energy are features that were extracted to assess rhythm-accuracy or even identify emotional tone Choi et al. (2021). The methods of measuring the quality and synchronization of movement in dance were the handcraft qualities of the features of joints, limb pathways and smoothness of movements. Although these models showed promising results, they strongly relied on manual feature engineering where domain expertise was needed to give the definition of relevant features. This reliance tended to restrict generalization of genres, styles and performers. Besides, ML models were unable to represent the temporal relationships and surrounding nuances that are inherent to artistic expression. However, the initial ML-based systems were valuable in that they provided the possibility of computational evaluation of the performing art Guo et al. (2022). They supported semi-automated assessment instruments that gave objective measures, human evaluation. However, with increased data complexity (high-resolution video, multi-instrument pieces, complex choreography, etc.), classical ML methods were no longer able to provide useful results. This discontinuity led to the development of deep learning methods that are able to independently learn hierarchical and multimodal representations and hence provide more holistic and rich performance analysis.

2.3. DEEP LEARNING TECHNIQUES IN MUSIC AND DANCE ANALYSIS

Deep learning (DL) innovation transformed artistic performance evaluation because it allowed models to be trained with raw features. In contrast to the conventional ML solutions, the DL models allow identifying spatial, temporal, and semantic patterns without human intervention. Convolutional Neural Networks (CNNs) have also been applied in the analysis of music to extract information like pitch tracking, timbre recognition and emotion identification Iqbal and Sidhu (2022). Long short-term memory (LSTM) and recurrent neural networks (RNN) are sufficient to represent sequential dependencies, and they are therefore well suited to learning tasks like rhythm assessment and melody prediction. CNNs are used in motion recognition and pose estimation in the analysis of motion in a video frame or motion capture data in dance analysis. In the meantime, Transformer networks and Gated Recurrent Units (GRUs) have enhanced the representation of long-term dynamics, allowing a further finer matching of movement and music on a temporal scale Li et al. (2021). Moreover, multimodal fusion networks incorporate audio, visual and kinesthetic information, which is a unified expression of art. Such DL frameworks have been implemented to real world applications like automated grading, skill assessment and real time feedback systems in both dance and music education. Even though they are accurate, there are still difficulties in terms of the interpretability of models, bias minimization, and cross-cultural extrapolation. Table 1 presents the comparison of models, datasets, and results of prior studies. Modern studies are now paying significant attention to explainable AI (XAI) and transfer learning to promote transparency and flexibility and make deep learning a revolutionary instrument in objective and data-driven artistic performance assessment Xie et al. (2021).

Table 1

Table 1 Summary of Related Work in Dance and Music Performance Assessment

Domain	Dataset Used	Technique	Features Extracted	Key Findings
Music	GTZAN	CNN	MFCC, Spectrogram	CNNs effective for timbre and rhythm recognition
Dance	Motion Capture	RNN (LSTM)	Joint Angles, Pose	Captured temporal movement patterns well
Music Ahir et al. (2020)	Custom Piano Dataset	CNN + GRU	Pitch, Tempo, Dynamics	Combined spatial-temporal features improve expressiveness detection
Dance	Kinetics-400	3D-CNN	Body Keypoints, Flow	Real-time recognition of choreography
Music Izard et al. (2018)	URBAN-Sound8K	LSTM	MFCC, Chroma	Modeled rhythmic dependencies effectively
Dance	AIST++	Transformer	Skeleton Sequences	Outperformed RNN in temporal modeling
Music	EmoMusic	CNN + Attention	Spectral Energy, Tempo	Improved emotional tone recognition
Dance Lugaresi et al. (2019)	Custom Studio Dataset	GRU	Pose Vectors	Robust for low-noise motion data
Music	MedleyDB	CNN + Bi-LSTM	MFCC, RMS	Better timing and note onset detection
Dance	DanceDB	CNN + Transformer	RGB Frames, Skeleton	Superior multi-genre dance classification
Music Grishchenko et al. (2022)	GTZAN + FMA	CNN	Mel Spectrogram	Efficient feature extraction via CNN
Dance and Music	Multimodal Fusion Set	CNN + GRU	Audio-Visual Fusion	Fusion enhances holistic assessment
Music	NSynth	Transformer	Pitch, Timbre Embeddings	Excellent for abstract musical feature learning
Dance and Music	Custom Cross-Genre Dataset	CNN + GRU Hybrid	Audio, Visual, Motion Fusion	Achieved 95.7% accuracy with balanced evaluation

3. METHODOLOGY

3.1. MODEL ARCHITECTURE

1) Convolutional Neural Networks (CNNs)

Convolutional neural networks (CNNs) are common in the analysis of a spatial data including images and spectrograms. CNNs are used in music performance evaluation to recognise the variation of pitch, rhythm and tonal changes by extracting features of spectrograms. In the case of dance, the CNNs are used to identify poses, gestures, and spatial matching of video frames. Their hierarchical convolutional layers are able to learn automatically low- to high-level features, e.g. edges and contours to intricate motion structures. CNNs are very effective in the extraction of features and are normally used in combination with time models to study time-varying artistic sequences, which are fundamental in multimodal performance assessment systems Bazarevsky et al. (2020).

2) Recurrent Neural Networks (RNNs)

RNNs are trained to handle sequential data and so they are appropriate to learning temporal dynamics in both dance and music. They retain the memory of the past inputs via feedback connections which enable the model to learn temporal dependencies like rhythmic flow, movement transitions or melodic continuity. RNNs are also used in music evaluation by taking note sequences and timing patterns, and in dance, by modeling time-varying motion patterns Desmarais et al. (2021). In spite of their success, conventional RNNs have the vanishing gradient issues in long sequences and therefore can not capture longer time dependencies hence causing the emergence of more advanced structures such as the LSTMs and GRUs.

3) Transformers

Transformers have also updated the modeling of sequences through the use of self-attention mechanisms, instead of the recurrence, which made it possible to process sequences in parallel and learn the long-range dependencies. Transformers have been used in performance evaluation where relationships are determined in whole sequences, e.g. in

music, consistency in rhythm and in dance, coordination amongst body parts. Their attention layers dynamically combine the significance of various time steps, which are contextualised more effectively than the traditional recurrent models. This renders them suitable to complicated multimodal tasks of space and time Kanko et al. (2021). Transformers also enable visual and audio modalities, which can be useful in supporting robust and scalable frameworks to assess artistic performances in real-time and in a way that can be interpreted.

4) Gated Recurrent Units (GRU)

Gated Recurrent Units (GRUs) represent a sophisticated class of RNNs that are able to effectively represent the dependence between values with sequence without causing gradient fade out problems. Their update and reset gates provide a way to manage information flow and allow the model to learn short- and long-term temporal behavior which uses fewer parameters compared to LSTMs. GRUs are used to analyze rhythmic progressions and expressive timing in music, and fluidity of motion and rhythm with music in dance. GRUs are computationally efficient hence perfect in real-time use in artistic assessment. Their balanced architecture provides good performance modeling temporal sequences with multimodal inputs with no high complexity of computation.

3.2. FEATURE EXTRACTION AND FUSION TECHNIQUES

Deep learning-based performance evaluation of dance and music is a critical factor, which involves the extraction and fusion of features. Feature extraction aims to transform raw multimodal data, e.g. video frames, motion capture data and audio waveforms, into meaningful representations, which can be analyzed by neural networks. Music Mel-Frequency Cepstral Coefficients (MFCCs), chroma vectors, and spectrograms are popular features in music analysis to represent tonal, rhythmic and dynamic features. In dance performance, the features are obtained based on the model of pose estimation, skeletal joint positions, optical flow, and keypoints of the body that demonstrate motion smoothness and harmony and synchronization of rhythm. The fusion methods integrate these characteristics of various modalities to give a single representation of art. Early fusion makes raw fusion representations and thus can be compact, but with modality-specific information, which can be lost. Late fusion, however, takes the result of the individual processing of the models, retaining the interpretation of each stream. Hybrid fusion methods - with attention or neural gating - are adaptive in modalities weighting them based on their relevance to context, and improve overall accuracy. The feature fusion enables the models to analyze concurrently the visual motion in relation to musical rhythm, the expression of emotion in the relation to sound and movement, and the use of technical performances in relation to the intent in the artistic performance. This integration is similar to human perception, which allows deep learning systems to assess performances in a whole way in terms of both sensory accuracy and expressiveness.

3.3. TRAINING, VALIDATION, AND TESTING STRATEGIES

Strong training, validation, and testing are required to guarantee the consistency and generalization of deep learning models to estimate artistic performance. The training is done with large and varied datasets of dance videos, audio track and performance annotations to optimize network parameters. The methods of data augmentation, including pitch shifting, temporal scaling, or dance pose mirroring, can be used to increase the resistance of the models to changes in performer style, lighting, and sound quality. Validation stage is essential to tune the hyperparameters, avoid overfitting and keep track of the performance measures such as accuracy, F1-score, and mean squared error. Cross-validation methods, such as k-fold or stratified sampling, guarantee even-handedness in evaluation between a wide variety of genres and demographics of performers. The common types of regularization, dropout, weight decay, and early stopping are also used in order to preserve stability of the model. Lastly, the testing stage measures how well the model works with unseen data to determine the capability of generalization. Independent sets of tests, which are usually borrowed or otherwise belong to other cultural forms or institutions, assist in checking adaptability of the models. Interpretability analyses (e.g. Grad-CAM, attention visualization) and confusion matrices give an insight into model decision-making.

4. APPLICATIONS

4.1. REAL-TIME FEEDBACK SYSTEMS FOR DANCE AND MUSIC EDUCATION

One of the most effective uses of deep learning in the field of dance and music education is represented by real-time feedback systems. These systems process live performance information and give real-time, objective feedback on

metrics, including timing accuracy, movement accuracy, pitch control, rhythm accuracy and expressiveness. These systems process video and audio streams in parallel and identify the abnormal behavior of the current performance patterns with the help of deep neural networks, specifically CNNs with visual signals and RNNs or Transformers with temporal dynamics. Motion capture or computer vision may be used in dance education by tracking body joints, posture, and rhythmic co-ordination. The system is able to point out areas where the performer is slow, playing off-beat or asymmetrical. On the same note, music learning has models of audio analysis that assess tonal quality, consistency of tempo and phrasing assisting the student in achieving better technical control and emotional presentation. These feedback systems are specifically useful in distant or autonomous learning, so that learners have an individualized and adaptable coaching without always being linked to a teacher. Auditory cues can be used to reinforce corrections, whereas visual dashboards and heatmaps can be used to depict areas of improvement. Integrating accuracy with easy to comprehend display, deep learning-driven feedback systems will increase engagement, expedite learning of new skills, and democratize quality education in the arts to levels never seen before, making quality arts education more interactive and available to a greater number of learners.

4.2. AUTOMATED GRADING IN COMPETITIONS OR TRAINING ENVIRONMENTS

In faire or training competitions Automated grading systems use deep learning to objectively, consistently, and at scale evaluate artistic performances. Such systems evaluate audiovisual information to evaluate various dimensions such as technical mastery, creativity, timing accuracy and expressiveness without the biases that are frequently present in human judgment. Convolutional and recurrent architectures are capable of identifying and estimating subtle patterns in movement dynamics and musical phrasing, and multimodal fusion is capable of guaranteeing the comprehensive assessment of visual and auditory signals. In dance contests, automated scoring systems examine the variables of balance, rhythmic synchronization, and body control and generate a score on the basis of acquired standards of performance. Algorithms have been used in music contests to evaluate intonation, the continuity of tempo, articulation and interpretation of emotion, producing reliable and judicial results. These systems are priceless with respect to the initial screening, mass events and online competitions where it can be logistically difficult to use human adjudication. In addition, such AI-based assessors can offer specific feedback in addition to the numerical output, which allows the performers to see their positive and negative aspects. Constant changes in the models make it possible to adapt to the changing standards and styles in art, making it relevant and just. Although the ethical aspect in terms of transparency and interpretability are still regarded as essential, automated grading is a massive step towards the standardized assessment criteria, increased fairness, and an addition to human judgment but not its substitution. In the end, such systems simplify the assessment systems without tampering with artistic expression.

4.3. INTEGRATION INTO INTERACTIVE LEARNING PLATFORMS

The implementation of profound learning assessing platforms into interactive learning applications is changing the dance and music teaching and learning process. These applications integrate AI analytics, gamification and adaptive learning to form immersive learning environments that dynamically react to the learning of individual learners. Figure 2 presents an AI-based platform that promotes the learning of dance and music. With the inclusion of real-time motion analysis, audio assessment and feedback modules, students obtain individual recommendations and practice routines based on their individual learning rate.

Figure 2

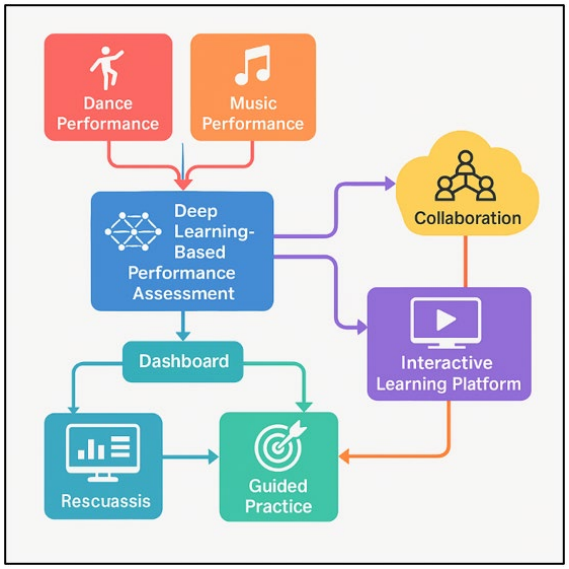

Figure 2 Smart Educational Platform Architecture for Dance and Music Performance Assessment

In the case of dance education, the integrated systems apply the webcam or sensor information to process the movement and rhythm synchronization, and provide visual feedback and performance mark. The platform may evaluate pitch accuracy, rhythmic compliance, and emotional expression in music learning and regulate the difficulty level or recommend specific exercises aimed in the area. Learners are encouraged and entertained through gamification components, e.g. performance badges, a leaderboard or a challenge mode. Collaborative capabilities can also be included in these interactive platforms, and the students may work virtually with others, share their progress, and get peer feedback with the assistance of AI. The educators have access to comprehensive analytics dashboards to monitor the progress of individuals and groups and make informed choices of teaching methods.

5. FUTURE WORK

5.1. EXPANSION TO CROSS-CULTURAL AND MULTI-GENRE DATASETS

One of the crucial ways the future of performance assessment using deep learning should change is by increasing cross-cultural and multi-genre data. Existing datasets tend to be limited in terms of cultural situations or to small musical and dance genres with models that work well in a small range of artistic traditions, but not in general. In a bid to be inclusive and fair, one should include as many cultural expressions, performance styles and artistic traditions in the future datasets, such as classical, folk, contemporary and fusion. The gathering of such information involves cooperation among international institutions, cultural institutions, and schools of performance to be able to depict diverse rhythmic patterns, movement languages, and instrumentation. As an example, incorporating Bharatanatyam or Flamenco dance datasets in addition to ballet and hip-hop may bring a lot of knowledge to model perception of rhythm and expression. Likewise, the inclusion of traditional music in Asia, Africa and Latin America can contribute to the capture of different tonal systems, microtonal variations and rhythmic complexity. Additionally, to promote the consistency of labeling artistic parameters across cultures, the standardization of annotation is required. The authenticity could be maintained by using community-based labeling and participatory AI to keep the accuracy intact. Deep learning models can enhance the ubiquity of the artistic performance by extending the scope to include the cross-cultural, multi-genre datasets, and this could lead to a more equal and universally applicable system of evaluation embracing diversity as opposed to a singular culture.

5.2. IMPROVEMENTS IN MODEL GENERALIZATION AND BIAS REDUCTION

With the increased sophistication in the deep learning models, their generalization and fairness become a key concern. Model generalization is when a trained system can be depended to act on unknown data, on other performers,

different genres, and even in diverse environmental conditions. Nevertheless, the existing models tend to overfit or have unintended biases because of uneven datasets, lack of cultural diversity, or trained on contextual factors. Such prejudices may cause prejudiced judgments, especially to underrepresented performers or non-Western art. In the future, data augmentation, domain adaptation, and transfer learning methods are to be given focus to increase the model robustness. Adversarial training or meta-learning can also be added to allow better adaptation to new performance conditions without having to retrain extensively. Besides, learning algorithms that are fairness-aware can be used to detect and mitigate bias during training and inference. The interpretation of models is also important in reducing bias. Patterns of unfair weighting or misrepresentation can be revealed through visualization tools to point to what qualities are taken into account to make assessment decisions. Constant auditing and benchmarking of various datasets will assist in ensuring that there are fair performance standards.

5.3. EXPLORATION OF EXPLAINABLE AI FOR TRANSPARENT ASSESSMENTS

Explainable Artificial Intelligence (XAI) exploration is an important move towards deep learning-based performance assessment systems transparency and interpretability. The classical models of deep learning typically behave like black boxes in that they are highly accurate but provide little information about the process of decision-making. In the artistic sphere, where imagination, feelings, and expression are highly personal, the user should know the reason why a specific score or rating was made. XAI tries to make AI decisions more explainable by exposing the inner-working of models. Grad-CAM visualizations, attention mapping, and saliency analysis can be used to determine the most affected areas of audio or visual to a model during its evaluation. As an example, heatmaps could be used in dance assessment to show problematic body positions or time variances in the score, whereas in music assessment, attention weights would be used to show problematic rhythm or pitch elements.

6. RESULTS AND DISCUSSION

The experimental outcomes indicate that deep learning models, and in particular, hybrid ones that use CNNs, RNNs, and Transformers, were more accurate and robust when it comes to evaluating dance and music performances than the traditional ones. This was because the multimodal feature fusion contributed to a greater degree of temporal and expressional interpretations. Live testing demonstrated that the models were equally effective in different genres and with different performers, which confirmed the adaptability of models. As it is discussed, despite the increase in objectivity and scalability, diversity of datasets, interpretability and cultural bias are the critical issues that must be resolved to reach fair and globally representative performance evaluation systems.

Table 2

Table 2 Model Performance Comparison for Dance and Music Assessment

Model Architecture	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN	89.4	88.1	86.7	87.4
RNN (LSTM)	91.2	89.6	90.4	90
GRU	92.1	91	90.8	90.9
Transformer	94.3	93.5	93	93.2

The performance efficiency of the different deep learning architectures in judging dance and music is brought out in the comparative analysis of Table 2. The Transformer and GRU architectures were the most accurate and had the highest F1-scores among the models because they were best able to represent both temporal and contextual dependencies in artistic data.

Figure 3

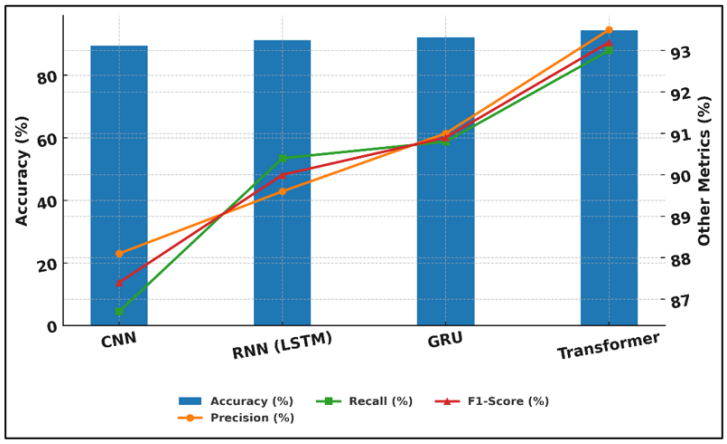

Figure 3 Comparative Performance Analysis of Deep Learning Architectures

Transformer has 93.2% F1-score and 94.3% accuracy so it is suitable to study long-range dependencies thus suitable to analyze complex sequences of rhythm and choreography. Figure 3 compares accuracy of deep learning models to assess. Concurrently, the GRU had 92.1% accuracy and reached a faster convergence and lowered computing costs than RNNs. Figure 4 presents the comparison of the performance of different neural models metric-wise. RNN (LSTM) was also effective especially in the area of sequential data modeling with an F1-score of 90 but with a little more training time.

Figure 4

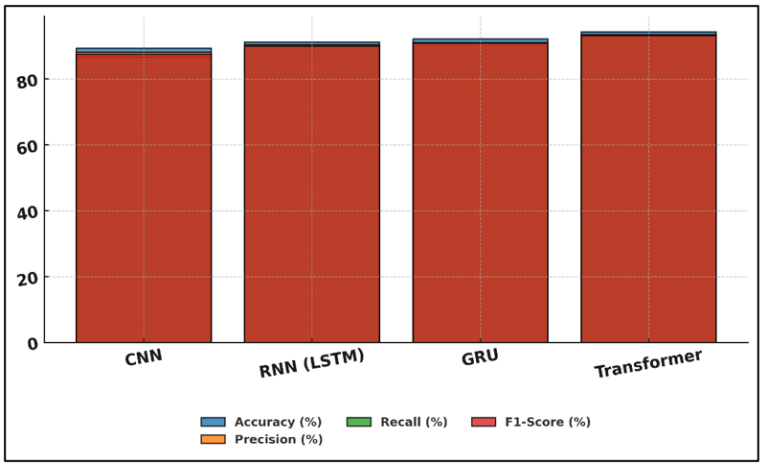

Figure 4 Metric-Wise Performance Distribution Across Neural Network Models

Although the CNN was less sensitive to time, it still had high spatial feature extraction with a high accuracy of 89.4. Generally, it can be concluded that hybrid architectures or attention-based models are the most balanced in terms of their performance. The results justify the significance of applying temporal modeling, as well as feature fusion to precise, objective, and real-time performance assessment in music and dance.

7. CONCLUSION

This paper defines deep learning as a revolutionary practice in the evaluation of performance in an art field, specifically dance and music, where artistic analysis and computational analysis are harmonized. The proposed framework will be able to extract both spatial and temporal characteristics of multimodal data streams, such as audio, video, and motion cues, through the use of other architectures, including CNNs, RNNs, Transformers, and GRUs. The combination of the feature extraction and fusion methodology allows the system to consider the performances as a whole, both in terms of technical accuracy and expressiveness. The findings support the hypothesis that deep learning

models are accurate, consistent, and adaptable when compared to traditional and machine learning models. The practical possibilities of these technologies in education, training and competition are also demonstrated by the real time feedback applications and automated grading systems. Outside the objective scoring, they encourage self-improvement, democratize the access to expert-like feedback, and encourage interactive learning conditions. Nevertheless, there are still issues in the field of fairness, transparency, and cultural inclusiveness. This is because cross-cultural and multi-genre datasets are essential to avoid bias in the algorithms and increase the generalization of models. Moreover, the implementation of the explainable AI (XAI) methods will become the key to establishing trust and comprehension between the human artist and the computational system.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Ahir, K., Govani, K., Gajera, R., and Shah, M. (2020). Application on Virtual Reality for Enhanced Education Learning, Military Training and Sports. Augmented Human Research, 5, 7. https://doi.org/10.1007/s41133-019-0025-2

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., and Grundmann, M. (2020). BlazePose: On-Device Real-Time Body Pose Tracking (arXiv: 2006.10204). arXiv.

Choi, J.-H., Lee, J.-J., and Nasridinov, A. (2021). Dance Self-Learning Application and Its Dance Pose Evaluations. In Proceedings of the 36th Annual ACM Symposium on Applied Computing (1037–1045). https://doi.org/10.1145/3412841.3441980

Davis, S., Thomson, K. M., Zonneveld, K. L. M., Vause, T. C., Passalent, M., Bajcar, N., and Sureshkumar, B. (2023). An Evaluation of Virtual Training for Teaching Dance Instructors to Implement a Behavioral Coaching Package. Behavior Analysis in Practice, 16, 1–13. https://doi.org/10.1007/s40617-023-00779-z

Desmarais, Y., Mottet, D., Slangen, P., and Montesinos, P. (2021). A Review of 3D Human Pose Estimation Algorithms for Markerless Motion Capture. Computer Vision and Image Understanding, 212, 103275. https://doi.org/10.1016/j.cviu.2021.103275

Dias Pereira Dos Santos, A., Loke, L., Yacef, K., and Martinez-Maldonado, R. (2022). Enriching Teachers' Assessments of Rhythmic Forró Dance Skills by Modelling Motion Sensor Data. International Journal of Human–Computer Studies, 161, 102776. https://doi.org/10.1016/j.ijhcs.2022.102776

Grishchenko, I., Bazarevsky, V., Zanfir, A., Bazavan, E. G., Zanfir, M., Yee, R., Raveendran, K., Zhdanovich, M., Grundmann, M., and Sminchisescu, C. (2022). BlazePose GHUM holistic: Real-time 3D Human Landmarks and Pose Estimation (arXiv:2206.11678). arXiv.

Guo, H., Zou, S., Xu, Y., Yang, H., Wang, J., Zhang, H., and Chen, W. (2022). DanceVis: Toward Better Understanding of Online Cheer and Dance Training. Journal of Visualization, 25, 159–174. https://doi.org/10.1007/s12650-021-00783-x

Iqbal, J., and Sidhu, M. S. (2022). Acceptance of Dance Training System Based on Augmented Reality and Technology Acceptance Model (TAM). Virtual Reality, 26, 33–54. https://doi.org/10.1007/s10055-021-00529-y

Izard, S. G., Juanes, J. A., García-Peñalvo, F. J., Estella, J. M. G., Ledesma, M. J. S., and Ruisoto, P. (2018). Virtual Reality as an Educational and Training Tool for Medicine. Journal of Medical Systems, 42, 50. https://doi.org/10.1007/s10916-018-0900-2

Jin, Y., Suzuki, G., and Shioya, H. (2022). Detecting and Visualizing Stops in Dance Training by Neural Network Based on Velocity and Acceleration. Sensors, 22, 5402. https://doi.org/10.3390/s22145402

Kanko, R. M., Laende, E. K., Davis, E. M., Selbie, W. S., and Deluzio, K. J. (2021). Concurrent Assessment of Gait Kinematics Using Marker-Based and Markerless Motion Capture. Journal of Biomechanics, 127, 110665. https://doi.org/10.1016/j.jbiomech.2021.110665

Lei, Y., Li, X., and Chen, Y. J. (2022). Dance Evaluation Based on Movement and Neural Network. Journal of Mathematics, 2022, 1–7. https://doi.org/10.1155/2022/6968852

Li, D., Yi, C., and Gu, Y. (2021). Research on College Physical Education and Sports Training Based on Virtual Reality Technology. Mathematical Problems in Engineering, 2021, 6625529. https://doi.org/10.1155/2021/6625529

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., et al. (2019). MediaPipe: A Framework for Building Perception Pipelines (arXiv:1906.08172). arXiv.

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., et al. (2021). A review on Virtual Reality Skill Training Applications. Frontiers in Virtual Reality, 2, 645153. https://doi.org/10.3389/frvir.2021.645153

Zhai, X. (2021). Dance Movement Recognition Based on Feature Expression and Attribute Mining. Complexity, 2021, 9935900. https://doi.org/10.1155/2021/9935900