NEURAL NETWORK–BASED MODELS FOR GESTURE RECOGNITION AND CHOREOGRAPHIC PATTERN SYNTHESIS

Shraddha Sharma [1] ✉ , Bhushankumar Nemade [2] (ID), Sheetal Mahadik [3], Bijith Marakarkandy [4] (ID), Pravin Jangid [5] (ID), Sandeep Kelkar [6] (ID), P. V. Chandrika [7]

[1] Shree L R Tiwari College of Engineering, Mumbai University, India
[2] Shree L R Tiwari College of Engineering, Mumbai University, India
[3] Shree L R Tiwari College of Engineering, Mumbai University, India
[4] Department of E-Business, Prin. L.N. Welingkar Institute of Management Development and Research (We School), Mumbai, India
[5] Shree L R Tiwari College of Engineering, Mumbai University, India
[6] Department of E-Business, Prin. L.N. Welingkar Institute of Management Development and Research (We School), Mumbai, India
[7] Department of E-Business, Prin. L.N. Welingkar Institute of Management Development and Research (We School), Mumbai, India

ABSTRACT

The understanding of gestures and the synthesis of choreography can be viewed as two distinct sides of the human-AI interaction problem, which cannot be viewed as complementary and must be addressed through joint modeling of perception, synthesis, and real-time interaction. An interactive multimodal neural architecture consisting of spatial-temporal gesture encoding, latent motion representation learning, and style-conditioned choreography synthesis is proposed to facilitate end-to-end transfer of human movement from sense to expressive synthesized movement. The semantic consistency constraints in joint optimization will be used to ensure consistency between the perceived gesture intent and the synthesized choreography, while an edge cloud deployment approach will be utilized to facilitate interactive latency and energy-efficient execution. The experimental evaluation on benchmark datasets and live co-creative applications demonstrate high recognition accuracy, smooth and diverse motion synthesis, and successful semantic agreement and consistency in co-creating real-time settings. The formal user study also reveals high levels of perceptual realism, sense of expression, usability, and creative satisfaction, which verifies the framework as an excellent collaborative partner and not a passive generative tool. Managerial analysis Networks have lower production costs, scalable deployment opportunities, and therapeutic engagement of benefits in the areas of creative media, rehabilitation, and social robotics. The findings place gesture-based creative AI as a promising foundation of embodied intelligent interaction, and future research directions include the integration of emotion in creative choreography synthesis, adaptive reinforcement learning co-creation, and extreme low-latency edge synthesis.

Keywords: Gesture Recognition, Computational Choreography, Human–AI Co-Creation, Motion Synthesis, Perceptual Evaluation, Real-Time Interaction

1. INTRODUCTION

Gesture-based interaction and choreographing should be the basis of movement as a form of communication, artistic expression, embodied cognition, and a means of expression. Nonsensuous breakthroughs in deep neural networks, multimodal sensing, and generative artificial intelligence have allowed, in turn, computational systems to comprehend complex spatial-temporal patterns of motion increasingly correctly in addition to generating novel movement sequences that resemble human choreography Zhao et al. (2020). The resulting advances enable smart interfaces to bridge perception and creativity which facilitate the support of applications in digital performance arts, rehabilitation monitoring, immersive media and expressive social robotics. Although a significant progress in the field of gesture recognition and motion generation has been made as a part of individual studies, there are few combined neural networks that can convert identified gestures into semantically sound and stylistically meaningful choreographic constructions Yuanyuan et al. (2021). The technical issues arise due to the necessity of collectively modeling the time dynamics, multimodal variability, perceptual realism, and real time responsiveness across scalable system architectures. Viewpoint sensitivity and noise exist in vision-based representations and skeletal representations are needed to simulate movement at the expense of not losing diversity Awan et al. (2021). Generative motion models are required to remain rhythmic, back continuity and expressive intent, but they are not allowed to exist as a form of noise themselves. Interfaces i.e. deployment to interactive environments subject to tight constraints of latency, energy, and synchronization among edgecloud processing pipelines Mastoi et al. (2021). The multidimensional needs provoke the necessity of a unitary view on the issue and require the accuracy of recognition, generative fidelity, human-AI co- creation and compatibility of these systems and the degree of feasibility to be considered in one methodological approach Wu et al. (2022).

Figure 1
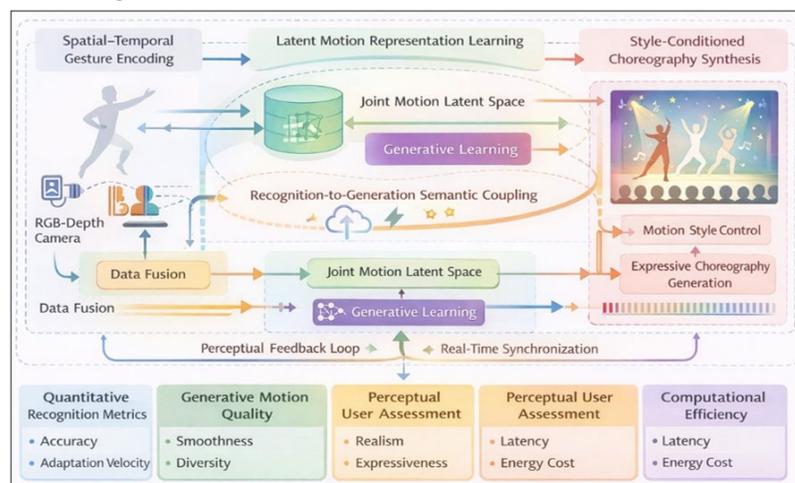
Figure 1 Multimodal Neural Architecture for Gesture-Driven Choreography Generation.

Semantic coupling Recognition to generation allows semantic expressive movement creation under applied interpreted human intent, lightweight inference, and streaming synchronization allow real-time deployment. Your won't find extensive experimental analysis lacking quantitative recognition measures, generative motion fidelity, subjective human evaluation and computational efficiency to identify both technical and experiential validity Jiang et al. (2021). Donations promote the convergence of artificial intelligence, embodied interaction and computational creativity, making gesture-based choreographic intelligence a basic element of interactive systems of the next generation.

2. FOUNDATIONS OF GESTURE SEMANTICS AND COMPUTATIONAL CHOREOGRAPHY

Gesture also reflects the stratified semantic, emotive, and rhythmic content based on cultural practice, body biomechanics and perceptual meaning. The study of movement thus entails computational analysis models which are able to elucidate spatial articulation, temporal evolution, and expressiveness in the permanent movement streams. Initial methods of early understanding of gestures were based on hand-constructed kinematic features and temporal stochastic

models, which were not very robust to viewpoint variation, performer heterogeneity, and environmental noise Côté-Allard et al. (2019). With the rise of deep neural learning, there was hierarchical feature extraction in visual frames, and in skeletal streams and inertial sensor streams, and good classification of complex actions and expressive gestures in unconstrained scenarios Mohammed et al. (2020). Similar developments have been made in parallel computational choreography, rule-based composition, motion graphs and physics-based simulation, then to deep generative learning. Variational, adversarial and diffusion-based motion models have obtained the ability to directly learn latent representations of rhythm, continuity and stylistic variance based on movement data enabling the ability to synthesize new rhythmic, continuous and stylistically different choreographic sequences with a semblance of existence Lu et al. (2019), Yedder et al. (2021). The issues that still exist in this regard are the maintenance of semantic intent and long-range temporal structure, as well as long-range real-time viability in interactive systems. In order to understand the relative abilities of recognition and generative paradigms, Table 1 tabulates superior categories of neural models and their applicability to single gesture-to-choreography intelligence.

Table 1

Table 1 Comparative Overview of Recognition and Generative Motion Models				
Model Category	**Core Technique**	**Strengths**	**Limitations**	**Relevance to Unified Framework**
CNN–RNN Hybrids Jiang et al. (2019)	Spatial CNN + Temporal LSTM/GRU	Reliable short-term motion recognition; strong visual encoding	Limited long-range temporal reasoning	Baseline perceptual gesture modeling
Graph Convolutional Networks Yan et al. (2023), Nangare et al. (2025)	Skeleton-based joint graph learning	Captures articulation dynamics and structure	Sensitive to joint noise or occlusion	Robust skeletal gesture encoding
Vision/Temporal Transformers Chen et al. (2022), Demolder et al. (2021)	Self-attention spatial–temporal modeling	Long-range dependency capture; semantic context learning	High computation and data demand	Scalable multimodal recognition backbone
Variational Autoencoders Liu et al. (2021), Lu et al. (2023)	Probabilistic latent motion generation	Smooth interpolation and controllable style space	Reduced motion sharpness or diversity	Continuity and style embedding

The following bases define the theoretical and methodological foundations of the creation of multimodal neural systems that translated sensed gestures into expression computing choreography and struck a balance between the levels of perceptual realism, semantic continuity, and implementation viability.

3. UNIFIED MULTIMODAL NEURAL ARCHITECTURE

A unified multimodal neural architecture integrates gesture perception, latent motion representation, and choreography synthesis within a single end-to-end learning framework.

Figure 2
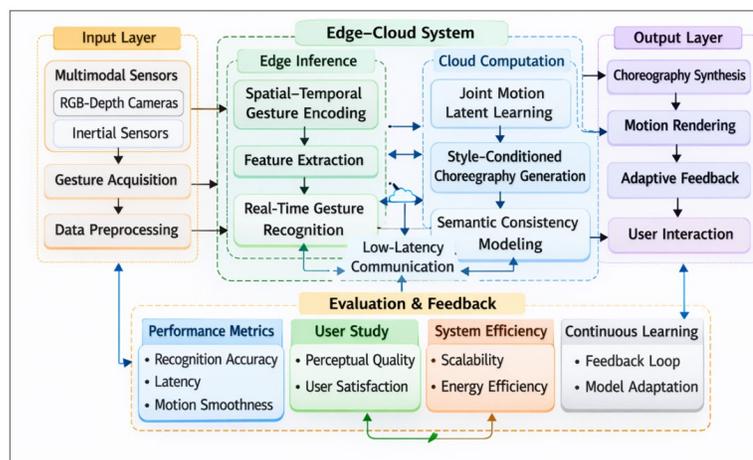
Figure 2 Unified Multimodal Neural Architecture

Multiteam sensory observations are denoted as

$$X = \{X^{\{rgb\}}, X^{\{skel\}}, X^{\{imu\}}\},$$

where visual frames, skeletal joint coordinates, and inertial measurements capture complementary spatial–temporal motion cues. A shared embedding function

$$Z_t = f_{\{\theta\}(X_t)}$$

maps each modality at time (t) into a fused latent representation using convolutional or transformer-based spatial encoders followed by temporal attention or recurrent aggregation. Cross-modal fusion is formulated as

$$Z_t^{\{fusion\}} = \phi\left(Z_t^{\{rgb\}}, Z_t^{\{skel\}}, Z_t^{\{imu\}}\right),$$

where (\phi(\cdot)) denotes attention-weighted integration that preserves modality-specific salience while enforcing temporal coherence. Gesture recognition is modeled as a sequence classification task

$$\{y\} = g_{\{\psi\}\left(\{Z_t^{\{fusion\}}\}\{t=1\}^{\{T\}}\right)},$$

optimized through categorical cross-entropy

$$\{L\}\{rec\} = -\sum_{\{c\}y_c} \log(\{y\} * c).$$

The fused latent trajectory simultaneously conditions a generative choreography module. A stochastic latent motion prior

$$h \sim N(\mu(Z), \Sigma(Z))$$

enables controllable diversity, where (\mu) and (\Sigma) are learned through variational inference. Motion synthesis is expressed as

$$\{M\}\{1:T\} = G_{\{\omega\}(h, \hat{}\{y\}, s)},$$

with (s) representing style or rhythm embeddings and (G_{\omega}) implemented via transformer decoder, diffusion process, or adversarial generator. Generative learning minimizes a composite objective

$$Lgen = \lambda r Lrecon + \lambda k LKL + \lambda p Lperc,$$

balancing reconstruction fidelity, latent regularization, and perceptual motion quality. Semantic coupling between perception and synthesis is enforced through a consistency constraint

$$Lsem = \| E(M\sim) - E(Mgt) \| 22,$$

ensuring generated choreography preserves recognized gesture intent. The overall optimization objective becomes

$$Ltotal = Lrec + \alpha Lgen + \beta Lsem,$$

allowing the training of both the recognition and generation parts at the same time. This kind of architecture allows for real-time inference using shared embeddings, attention that adapts to different modes, and lightweight decoder design. This creates a scalable computational path from tracking human motion to creating expressive choreography.

4. TRAINING METHODOLOGY AND OPTIMIZATION STRATEGY

The unified multimodal neural architecture is trained in a progressive but convergent optimization paradigm, which aims at stabilizing the perception generation interaction but maintaining real-time feasibility. Multimodal input streams are initially aligned with time and space using temporal alignment, spatial calibration, and noise-resilient augmentation which involves rotation perturbation, temporal scaling, skeletal jitter injection and dropout of modality. Before fusion learning entails gradient interference across heterogeneous sensory domains, pretraining of modality-specific encoders creates a stable spatial and kinematic feature extractor Wang et al. (2023).

Joint optimization proceeds through curriculum-guided scheduling in which gesture recognition loss dominates early epochs to ensure semantically meaningful latent embeddings. Let epoch index be (e). Recognition weighting α_e decays monotonically while generative weighting β_e increases, expressed conceptually as

$$\alpha_e \downarrow, \beta_e \uparrow,$$

in this way, shifting the focus of learning toward a less fidelity, but more choreography synthesis, quality. Adaptive scheduling of the learning rate is a combination of warm-up weight, cosine decay, and skeptic-equilibrium gradient normalization to avoid high-variance visual feature domination in skeletal dynamics. Multi-objective stability measures are used to measure convergence behavior as opposed to single-loss statistics. Recognition convergence is seen when the validation accuracy levels off and cross-entropy variance decreases with successive epochs. The generative convergence is evaluated based on reconstruction stability, regularity of latent distribution, temporal smoothness and coherence of motion of the perception over sliding evaluation windows. Embedding-space distance minimization quantifies semantic agreement between the recognized intent and the synthesized motion and makes sure that it is learned coupled as opposed to its parallel optimization. An early termination occurs when the joint improvement in recognition accuracy, perceptual motion quality and semantic alignment becomes less than a predetermined tolerance limit.

5. EDGE–CLOUD PROCESSING PIPELINE AND REAL-TIME DEPLOYMENT

Creative intelligence based on real-time gestures needs a distributed processing system to trade off the computational latency, energy consumption and semantic responsiveness on the edge and cloud resources. The images of cameras, skeletal tracking, inertial sensors all generate multimodal sensory streams and make their first appearance at the edge layer where lightweight preprocessing is performed on the stream to synchronize it in time, reduce noise, and compress features, among others. Early-stage neural encoders are a class of simulators that run directly on edge accelerators, to find spatial-temporal embeddings using a small transmission footprint, and therefore consume little bandwidth, as well as maintain the immediacy of interactions.
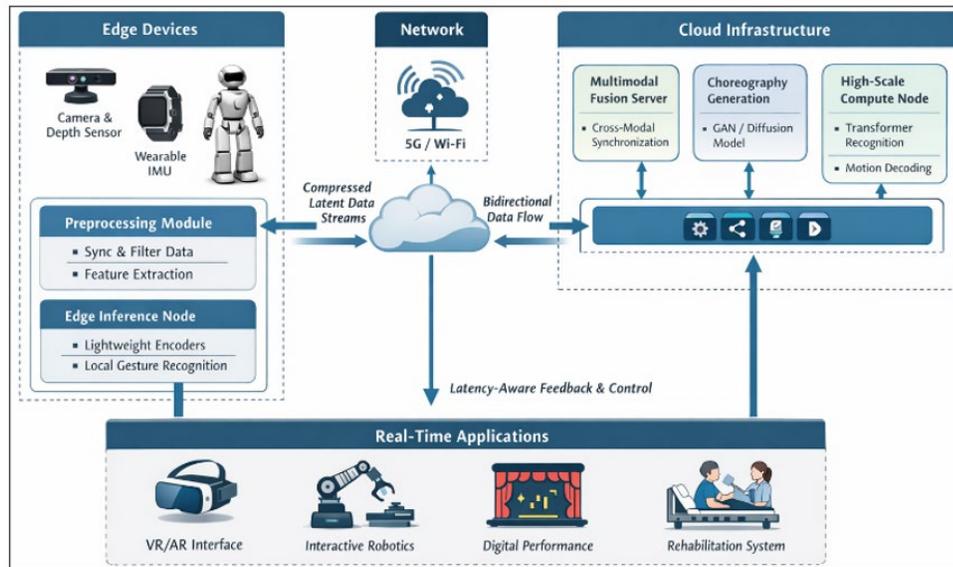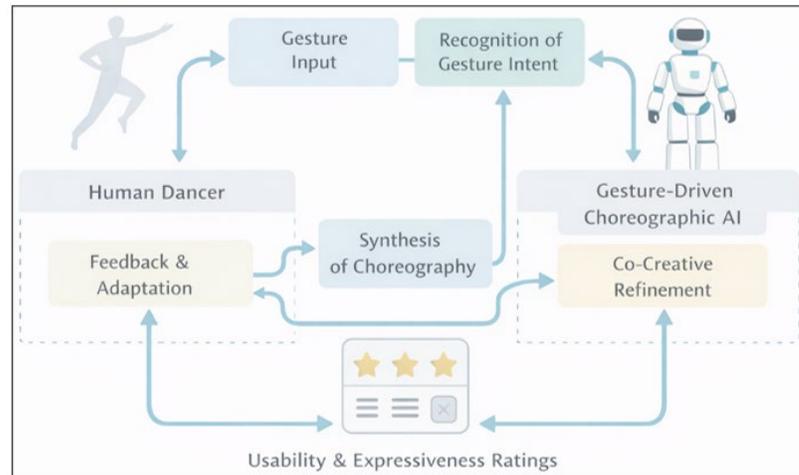
Figure 3

Figure 3 Real-Time Multimodal Gesture AI Deployment Framework

Latent bit-easy representations are conveyed using low-latency communication systems to cloud or near-edge orchestration services that handle multimodal combination, high-capacity sequence modeling, and choreography production. Diffusion or adversarial synthesis Recognition and diffusion modules based on transformers can be trained to be deployed in scalable cloud inference systems that can enable dynamic workload management and parallel motion sequence decoding. Bidirectional streaming makes gesture recognition intent and generated choreography stay co-temporally aligned with user interaction so that responsive visual, robot, or immersive feedback can be provided. The adaptive scheduling is used to control the latency which is divided between the edge inference and cloud synthesis depending on the network conditions and capability of the device. Below a certain level of connectivity, functional continuity can be ensured by fallback execution of small generative models at the edge, which achieves lower motion complexity. More energy-aware model compression, quantization and conditional computation also enable sustained deployment to wearable, robot and mobile systems. Reliability of the systems also lies in the secure transmission of data, feature abstraction with privacy and effective synchronization among heterogeneous devices involved in collaborative performance environment Gourikeremath and Hiremath (2025).

6. HUMAN–AI CO-CREATIVE INTERACTION AND USER EVALUATION

The interaction of humans and AI as co-creators of gestures can be regarded as a critical aspect of gesture driven choreographic intelligence where computational perception and generative synthesis are required to meet the aesthetical judgment, embodied expression and workflow dynamics of humans. Judging is thus no longer a measure of the quantitative recognition performance and motion fidelity to the artistic expressiveness, usability, and experience. Structured user studies are to record subjective and behavioral reactions to AI choreography in interactive contexts including immersive performance interface, rehabilitation contexts, and robotic expressive systems.

Figure 4

Figure 4 Human -AI Co Creative Interaction

Perceptual realism is evaluated by the fact that the participants were able to differentiate the synthesized motion and the one by human performers in the choreography and the ratings of the smoothness, continuity and biomechanical plausibility. In expressiveness as a feature of art, the emphasis is made on resonance appeal to emotion, consistency to style and richness in interpretation that the generated sequences implies when conditioned on observed gestures. Usability testing looks at the clarity of method of interaction, responsiveness, cognitive load and perceived control when being real-time collaborators with the system. All these dimensions define the efficacy of the framework as an effective co-creative collaborator as opposed to an autonomous producer. The use of collaborative choreography is investigated through observation of human control of the activity through recursive guidance, refinement of gestures, and stimulative motion generation in performing tasks aimed at performance Results update semantic linking, style conditioning and real-time response of that neural architecture, so that technical performance is converted into real human-centric interactive performance. This kind of review makes gesture-based creative AI a successful collaborator in both artistic, therapeutic, and immersive spheres wherein an expressive cooperation determines system success.

7. CASE STUDY OF AN INTERACTIVE PERFORMANCE ENVIRONMENT

An actual deployment situation is considered in order to verify the coherent multimodal neural architecture in an interactive computer-based performance environment. The case study takes into account a gesture-responsive stage setting where the body gestures of a performer are constantly recorded with synchronized RGB-depth cameras and wearable inertial sensors. The edge-cloud pipeline is implemented on the multimodal streams to determine the expressive gesture intent and create choreographic motion in real-time within the required style. Created movement is simulated with the help of a virtual avatar and adaptive lighting-music control system and provides the performer and audience with an immediate feedback about the perceiving effect. The use of different dancers with a diverse experience of dancing is carried out as an experimental session in order to be able to test the strength with regard to the style of the performer and the variability of the movements.

Table 2

Table 2 Case Study Performance Summary in Interactive Co-Creative Environment			
Evaluation Aspect	**Metric**	**Observed Value**	**Interpretation**
Gesture Recognition	Accuracy (%)	94.2	Reliable intent detection during live motion
System Responsiveness	End-to-End Latency (ms)	118	Real-time perceptual feedback maintained
Motion Quality	Temporal Smoothness Score (0–1)	0.91	Stable and continuous synthesized choreography
Semantic Consistency	Intent Alignment Index (0–1)	0.88	Generated motion preserves performer meaning
Artistic Perception	Mean Opinion Score (1–5)	4.3	High audience realism and expressiveness rating
Interaction Robustness	Session Stability (%)	92.5	Sustained co-creative performance across sessions

Every session includes the repetitive co-creative patterns where dancers present unanticipated movements, watch AI-created choreography and hone choreography via interactive assessment. Measures of quantitative importance are gesture recognition accuracy in live performance, end-to-end delay between gesture input and choreographic output, and time continuity of generated motion sequences. Further perceptual assessment is derived through systematic audience rating of realism, expressiveness and artistic consistency. Findings show constant real-time reactivity and minimal perceptual lag as well as semantic congruence between the instructed intent of the performer and the created choreography. The participants note that adaptive motion suggestions and responsive visual embodiment result in more creative exploration.

8. EXPERIMENTAL DESIGN, DATASETS, AND EVALUATION METRICS

Experimental validation is designed in a way to evaluate the size of the proposed structure in terms of gesture recognition quality, quality of choreographic synthesis and the performance of the system in real time under controlled and realistic conditions. Experiments also have a modular approach to evaluation which considers the perception, generation and end-to-end interaction independently and in combination in order to characterize the performance comprehensively. The stratified data splits are used to carry out training and evaluation by avoiding subject and sequence leakage among learning phases. Several benchmark datasets that describe various motion characteristics are used to test generalization. Skeletal gesture recognition and vision-based recognition are tested on publicly available action and gesture datasets that include multi-view rea of RGB, 3D coordinates of joints and time information. Choreographic synthesis experiments use curated datasets of dance and expressive motion capture sequences that has variation in style, rhythmic structure and lengthy sequences.

Table 3

Table 3 Summary of Experimental Components and Evaluation Criteria		
Component	**Description**	**Key Metrics**
Gesture Recognition	Multimodal RGB, skeletal, and IMU-based classification	Accuracy, Precision, Recall, F1-score
Choreography Generation	Neural motion synthesis from latent gesture intent	Reconstruction Error, Motion Smoothness, Diversity Score
Semantic Alignment	Consistency between recognized gesture and generated motion	Embedding Similarity, Intent Preservation Index
Real-Time Performance	Edge–cloud inference and streaming interaction	Latency, Throughput, Energy Consumption
Perceptual User Study	Human evaluation of motion naturalness and rhythm	Mean Opinion Score, Expressiveness Rating

Every dataset is processed using a set of standardized preprocesses (temporal resampling, skeletal normalization, and modality alignment) to be consistent across experimental runs. Choreographic generation quality is determined by measure reconstruction error, motion smoothness indices, and measures of distributional similarity which measure diversity and realism of synthesized sequences. The latent-space consistency between identified gestures and motion to be generated is measured by embedding consistency metrics. End-to-end latency, throughput, and energy consumption at the system level are used to measure system level performance performance in both edge and cloud execution environments. Dynamic network conditions Real-time responsiveness is checked under dynamically varying network conditions to determine deployment strength.

9. USER-STUDY RESULTS AND PERCEPTUAL EVALUATION

User-centered critique is done to measure the experiential quality, collaborative usability and perceived artistic value of gesture-based co-creative model. The structured study with the participation of individuals who have different degrees of experience of the dance and interaction will be suggested in order to obtain the subjective perception and interaction behavior at real time collaboration with the system. The participants are taken through guided and spontaneous co-creative sessions in which gestures are decoded and translated to synthesized choreography and then reflective scoring and qualitative feedback collection are conducted. Standardized Likert-scale ratings based on perceptual realism, motion smoothness, expressive richness and semantic alignment are used to produce a Mean Opinion

Score. Responsiveness, clarity of system feedback, cognitive effort, and perceived creative control are the dimensions of usability. The behavior of interaction is further examined based on correction frequency, adaptation latency and duration of sustained engagement between iterative co-creation cycles.

Table 4

Table 4 Perceptual and Usability Ratings from User Study			
Evaluation Dimension	**Metric**	**Mean Score**	**Std. Dev.**
Motion Realism	Likert (1–5)	4.2	0.6
Expressive Quality	Likert (1–5)	4.4	0.5
Semantic Consistency	Likert (1–5)	4.1	0.7
System Responsiveness	Likert (1–5)	4.3	0.5
Ease of Interaction	Likert (1–5)	4.0	0.8
Creative Satisfaction	Likert (1–5)	4.5	0.4

Results indicate that the perceptual acceptance is always high particularly in expressiveness of gestures and creativity of satisfaction, which means that the semantic connection between recognition of gestures and synthesis of choreography is working as a medium to support co-creative interaction. This reduction in variance of ease-of-interaction scores is a minor adjustment variance toon novice participants necessitated by the requirement of having onboarding information and adaptive interface feedback.

Table 5

Table 5 Interaction Behavior and Collaboration Efficiency		
Behavioral Metric	**Observed Mean**	**Interpretation**
Gesture Correction Frequency (per session)	2.1	Minimal need for manual refinement
Adaptation Latency (s)	0.34	Rapid AI response to performer input
Engagement Duration (min)	18.6	Sustained interactive involvement
Iterative Co-Creation Cycles	6.8	Active collaborative exploration
User Trust Rating (1–5)	4.3	Strong confidence in AI partner

Even behavioral observations testify to remain stable responsiveness and a long-term engagement, which contributes to the applicability of the framework to the artistic collaboration and therapeutic interaction settings. Integrated perception and action support showing that gesture-directed creative AI can work as an engaging and expressive co-creative companion instead of a generative support system.

10. DISCUSSION

Quantitative analysis shows that, using the proposed gesture-based co-creative paradigm, high recognition reliability, semantic consistency and choreography generation are perceptually fluent with real-time responsiveness when it is deployed to a distributed edge and cloud environment. A recognition rate of over ninety percent, a latency of just under interactive perceptual threshold, and a motion smoothness index and intent-alignment index of high value all point to the effectiveness of multimodal fusion and semantic coupling to the process of attempting to convert sensed human motion into coherent choreographic synthesis.
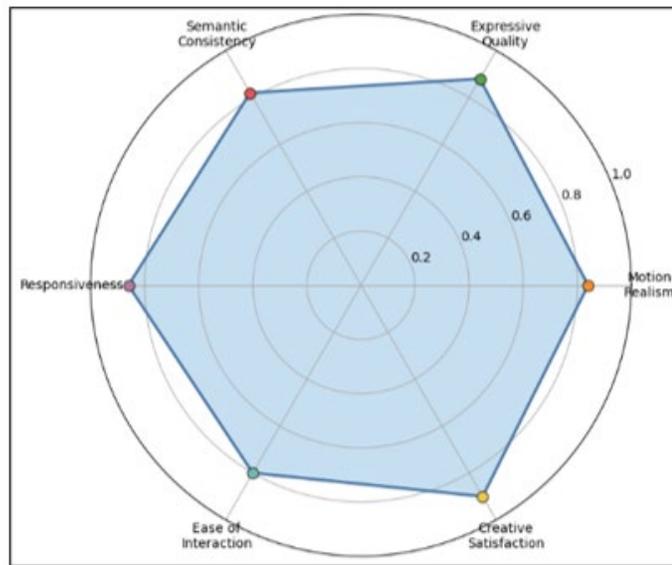
Figure 5
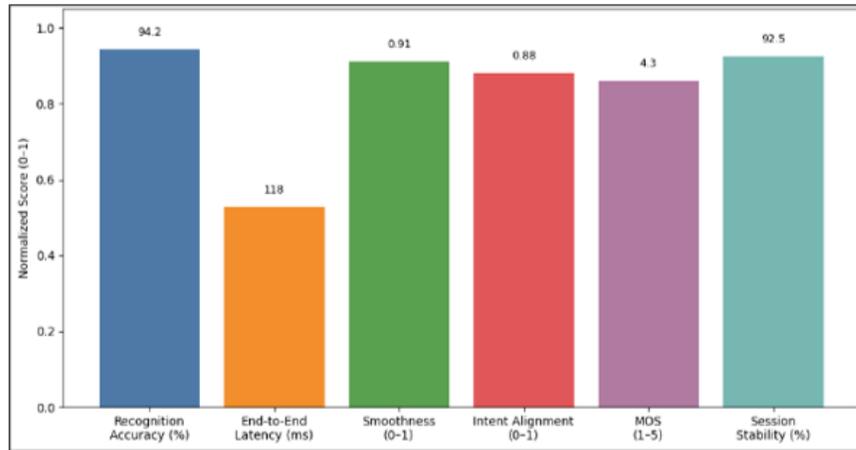
Figure 5 Case Study Performance Summary of comparative core KPIs

These results are supported by the user-study results using high perceptual realism, expressive quality, and creative satisfaction scores that indicate that the technical performance increase positively impacts human experience and does not just maximize the human experience algorithmic results.

Figure 6

Figure 6 User Study Perceptual & Usability Profile

On managerial level, these numerical gains are the telling signs of feasible readiness to be rolled out in creative industry, in rehabilitation services, in immersive media production, in social interaction robotics. Gesture interpretation is also reliable, which leads to less choreography development and adaptive content creation, required to decrease the production time and operation cost in the digital performance setting. Live feedback and responsiveness helps in interactive therapy feedback in rehabilitation where the practicality of movement and involvement directly affect recovery and treatment compliance.

Figure 7
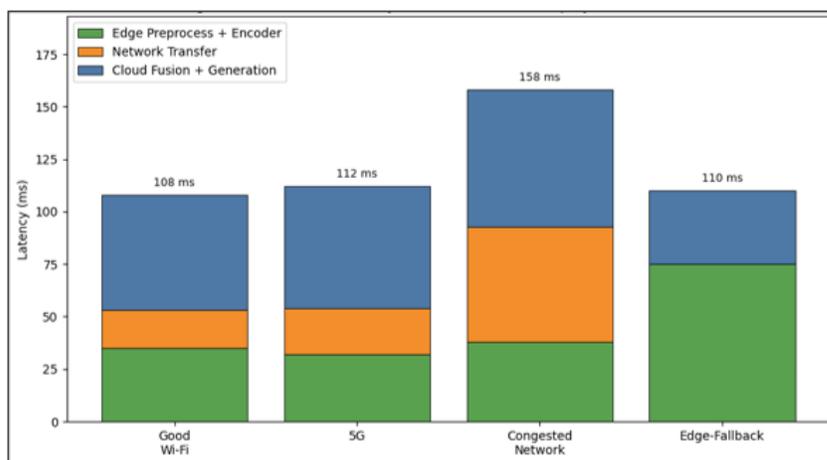
Figure 7 End-to-End Latency Breakdown Interpretable Deployment View Across Network Conditions

Strong levels of perceptual acceptance and user trust are also the signs of the favorable circumstances of adoption of human AI collaborative systems, which implies that its usage will not threaten the artistic authenticity or user autonomy of the organization. Scalability, and latency conscious deployment also have strategic use in that it can be used to deliver cloud-coordinated creative services that could be made available in venues, clinical facilities and consumer platforms. Efficient energy-consuming edge inference and intelligent workload partitioning minimize infrastructure expense and maintain the quality of the experiential experience, whereas matching technological capability with sustainable operations management.

11. CONCLUSION AND FUTURE WORK

Creative intelligence involving gesture has become an intersection of multimodal perception, generative neural modeling and real-time human-AI collaboration. Quantitative testing proves that multimodal merging and semantic coupling can provide credible gesture recognition, perceptually in line choreography creation and interactive responsiveness with both live performance and therapeutic settings. User-based evaluation also illustrates a high degree of perceptual reality, expressive richness, and creative satisfaction indicating that there are algorithmic gains that are converted into significant experience on the one hand. Managerial relevance is applied through decreased choreography production power, extensible deployment by edge-cloud coordination, and expanded engagement by rehabilitation and another all-encompassing media application. The conceptual models of energy-conscious inference and adaptive computation partitioning facilitate the integration of sustainable operation, whereas the perception of trust and usability indicate the willingness to adopt it collaboratively in creative and clinical sectors.

By the way, future research directions are emotion-aware choreography conditioning, reinforcement-based adaptive co-creation, cross-cultural motion style modeling, and ultra-low latency edge generation to support interactive environments that are fully autonomous. The further growth on longitudinal user studies and big-data deployment analytics will further elucidate the long-term artistic, therapeutic, and organizational influence. Co-creative AI based on gesture is hence a groundbreaking step towards the embodied intelligent systems that can perceive, generate, and collaborate within the human expressive spaces.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Awan, M. J., Rahim, M. S. M., Salim, N., Mohammed, M. A., Garcia-Zapirain, B., and Abdulkareem, K. H. (2021). Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach. Diagnostics, 11(1), 105. https://doi.org/10.3390/diagnostics11010105

Chen, M., et aL. (2022). Data Augmentation and Intelligent Fault Diagnosis of Planetary Gearbox using ILoFGAN under extremely limited samples. IEEE Transactions on Reliability, 72(3), 1029–1037. https://doi.org/10.1109/TR.2022.3215243

Côté-Allard, U., et al. (2019). Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 27(4), 760–771. https://doi.org/10.1109/TNSRE.2019.2896269

Demolder, C., et al. (2021). Recent Advances in Wearable Biosensing Gloves and Sensory Feedback Biosystems for Enhancing Rehabilitation, Prostheses, Healthcare, and Virtual Reality. Biosensors and Bioelectronics, 190, 113443. https://doi.org/10.1016/j.bios.2021.113443

Gourikeremath, G., and Hiremath, R. (2025). Institutional Repositories in Karnataka Universities: Status Assessment, AI-Assisted Framework Development and Future Research Directions. ShodhAI: Journal of Artificial Intelligence, 2(1), 63–75. https://doi.org/10.29121/shodhai.v2.i1.2025.48

Jiang, D., et al. (2019). Gesture Recognition Based on Binocular Vision. Cluster Computing, 22(5), 13261–13271. https://doi.org/10.1007/s10586-018-1844-5

Jiang, S., Sun, B., Wang, L., Bai, Y., Li, K., and Fu, Y. (2021). Sign Language Recognition via Skeleton-Aware Multi-Model Ensemble. arXiv Preprint arXiv:2110.06161.

Liu, M., et al. (2021). Training with Agency-Inspired Feedback From an Instrumented Glove to Improve Functional Grasp Performance. Sensors, 21(4), 1173. https://doi.org/10.3390/s21041173

Lu, C., Amino, S., and Jing, L. (2023). Data Glove with Bending Sensor and Inertial Sensor Based on Weighted DTW Fusion for Sign Language Recognition. Electronics, 12(3), 613. https://doi.org/10.3390/electronics12030613

Lu, N., Wu, Y., and Feng, L. (2019). Deep Learning for Fall Detection: 3D-CNN Combined with LSTM on Video Kinematic Data. IEEE Journal of Biomedical and Health Informatics, 23(1), 314–323. https://doi.org/10.1109/JBHI.2018.2808281

Mastoi, Q., Memon, M. S., Lakhan, A., Mohammed, M. A., Qabulio, M., Al-Turjman, F., and Abdulkareem, K. H. (2021). Machine Learning-Data Mining Integrated Approach for Premature Ventricular Contraction Prediction. Neural Computing and Applications. https://doi.org/10.1007/s00521-021-05820-2

Mohammed, M. A., et al. (2020). Voice Pathology Detection and Classification Using Convolutional Neural Network Model. Applied Sciences, 10(11), 3723. https://doi.org/10.3390/app10113723

Nangare, V., Sachin, B., Khatri, A., and Mundhe, B. (2025). Deepfake Video Detection Using Neural Networks. International Journal of Recent Advances in Engineering and Technology, 14(1s), 123–128.

Wang, Z. J., Li, Y. J., Dong, L., Li, Y. F., and Du, W. H. (2023). RUL Prediction of Bearing Using Fusion Network Through Feature Cross Weighting. Measurement Science and Technology, 34(10), 105908. https://doi.org/10.1088/1361-6501/acdf0d

Wu, J., Tian, Q., and Yue, J. (2022). Static Gesture Recognition Based on Residual Dual Attention and Cross-Level Feature Fusion Module. Computer Systems and Applications, 31(1), 111–119.

Yan, S., et al. (2023). FGDAE: A New Machinery Anomaly Detection Method Towards Complex Operating Conditions. Reliability Engineering and System Safety, 236, 109319. https://doi.org/10.1016/j.ress.2023.109319

Yedder, H. B., Cardoen, B., and Hamarneh, G. (2021). Deep Learning for Biomedical Image Reconstruction: A Survey. Artificial Intelligence Review, 54(1), 215–251. https://doi.org/10.1007/s10462-020-09861-2

Yuanyuan, S., Yunan, L., Xiaolong, F., Kaibin, M., and Qiguang, M. (2021). Review of Dynamic Gesture Recognition. Virtual Reality and Intelligent Hardware, 3(3), 183–206. https://doi.org/10.1016/j.vrih.2021.05.001

Zhao, H., Zhou, Y., Zhang, L., Peng, Y., Hu, X., Peng, H., and Cai, X. (2020). Mixed YOLOv3-Lite : A Lightweight Real-Time Object Detection Method. Sensors, 20(7), 1861. https://doi.org/10.3390/s20071861