Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on Art, Technology, and Society: Emerging Intersections in Visual and Performing Arts Research 7(2s)

GESTURE RECOGNITION FOR CLASSICAL DANCE FORMS USING COMPUTER VISION

Dr. Praveen Sen [1], Dr. Abhishek Pathak [2], Dr. Manish Gudadhe [3], Mrudula Gudadhe [4], Vikas Singh [5]

[1] Department of Computer Science and Business Systems, St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India
[2] Department of Computer Science and Engineering (Cyber Security), St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India
[3] Department of Computer Science and Engineering (Data Science), St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India
[4] Department of Information Technology, Priyadarshini College of Engineering, Nagpur, India
[5] Department of Computer Science and Engineering (Cyber Security), St. Vincent Pallotti College of Engineering and Technology, Nagpur, Maharashtra, India

Received 16 November 2025
Accepted 05 January 2026
Published 28 March 2026

Corresponding Author
Dr. Praveen Sen,
psen@stvincentngp.edu.in

DOI
10.29121/shodhkosh.v7.i2s.2026.7221

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

ABSTRACT

The symbolic communication in classical dance forms is based on codified hand gestures (mudras), postures and the use of rhythmic sequences of movement. Nevertheless, systematic computational identification of fine-grained dance gestures is not widely studied because subtle articulation variations, costume coverups and lack of annotated data exist. The paper suggests a spatial-temporal deep learning model of gesture recognition in classical dance with computer vision methods. This approach combines skeleton extraction using pose estimation, calculating of joint-angle features to achieve rotational invariance, convolutional neural network (CNN)-based spatial embedding and Long Short-Term Memory (LSTM) based temporal modeling to model dynamic gestures development. The hybrid representation is a mixture between the skeletal precision and the contextual visualization, consequently, granting the opportunity to distinguish visually similar mudras. The experimental evaluation of a curated dataset of classical dance gestures evidence reveals that the provided model will be more successful than the default CNN-only and skeleton-based models and will have more successful results in terms of accuracy, precision, recall, and F1-score. The training and validation analysis is the argument of constant convergence and high level of generalization between performers. The findings confirm that angular skeletal modeling and temporal deep learning are effective in the recognition of fine-grained gestures. In addition to the performance of the classification, the framework can also benefit digital cultural heritage preservation, smart systems of dance tutoring, and AI-based performance analytics. The research study provides a strong base to implement computer vision and the application of deep learning to the systematic study of performing arts.

Keywords: Classical Dance Gesture Recognition, Computer Vision, Pose Estimation, Joint-Angle Modeling, CNN-LSTM, Spatial–Temporal Learning, Mudra Classification, Cultural Heritage Digitization, Deep Learning, Human Action Recognition

1. INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

The classical forms of dances are one of the most sophisticated forms of cultural representation consisting of systematic gestures, rhythmic bodily motions, encapsulation of the storytelling, and a systemized symbolic language. Bharatanatyam, Kathak, Odissi, Kuchipudi and Mohiniyattam are identified as Indian classical traditions characterized by complex hand gestures (mudras), defined body postures (karanas), expressions on the face (abhinaya), and rhythmic coordination (precise) Shrestha et al. (2022). The movements are not at all arbitrary but they are guided by formalized rules that are written in ancient texts, like the Natya Shastra, which records the vocabulary of gestures, the grammar of movements and the aesthetics of performing. But the conventional learning approaches are mostly based on oral and long-term under the guidance of a guru and therefore less accessible and scaled. In digital times, there has been a growing imperative to have systematic recording, intelligent indexing, and computing apprehension of these gestures. The high rate of development of computer vision and deep learning technologies provides a bright opportunity of automatic analysis and recognition of fine-grained human gestures Yuan and Pan (2022). The combination of artificial intelligence and cultural heritage preservation is what has inspired this research, as it resulted in trying to create a smart system that could recognize the classical dance moves with accuracy and interpretability.

1.2. IMPORTANCE OF GESTURE RECOGNITION IN CLASSICAL DANCE

Classical dance can be said to go beyond the classification of actions in gesture recognition; it considers action in the form of semantic interpretation of symbolic actions. Even minor changes in the articulation of the fingers, the position of the wrists, the position of the elbows, and the direction of gaze can change the meaning considerably. An example is the hand mudrases of Bharatanatyam such as Pataka, Tripataka and Ardhachandra are a little different in the arrangement of fingers but convey totally different symbolic items. Equally, Kathak footwork is rhythmic and must be well coordinated with the beat and position of the body. Automated gesture recognition systems can be used, therefore, in many applications; in digital heritage gesture archiving systems, auto-tutoring systems to students, objective evaluation devices to evaluate student performance, and interactive technologies founded on dances. Besides, these systems can aid in comparison of research on styles of dances, curriculum development as well as remote training solutions, which can be scaled Kang (2023). Gesture recognition provides the means to solve the problem of computationally representing qualitative artistic expression which ensures the preservation and creation.

1.3. CHALLENGES IN CLASSICAL DANCE GESTURE ANALYSIS

Although the sphere of human pose estimation and recognition of actions has highly advanced in the recent times, there exist certain technical issues with classical dance gesture analysis. To begin with, it will be required to differentiate finely since the mudras can vary a bit with little variation in regard to positioning of fingers and hand tracking at high resolution and precise modeling of the skeleton will be required. Second, classical dance involves a complex body movement including deep bends, asymmetric, and synchronized movement of the head and eyes which makes the algorithms of keypoint detection complicated Li et al. (2021). Third, traditional costumes and jewelry, as well as colorful lighting on the stage, may cause the occurrence of occlusions and visual noise, which may affect the strength of the models. Fourth, gestures are time-varying; they have a different meaning as the frames change over time, and to deal with it, the corresponding temporal modeling techniques, such as the LSTM network or the attention mechanism based on transformers, are needed. Fifth, annotated datasets with classical forms of dances are limited, and this poses a challenge to supervised learning strategies Li et al. (2022). Lastly, dance is a multimodal process involving combination of facial expression, rhythm and alignment to music only, but most computer vision models only take into account skeletal features. All these complications demand a robust, efficient and powerful computational model that is able to address the space accuracy, time continuity and environment variability simultaneously.

1.4. PROBLEM STATEMENT

Even though the developed deep learning patterns could be used to carry out the action recognition at the general level, there are no dedicated frameworks related to the process of fine-grained gesture recognition in the classical forms of the dance. The existing systems are more prone to large body movements and do not depict small symbolic movement

that are hand gestures that are defining the classical practices. In addition, the association between spatial articulation and the time sequence that has been inherent in a dance performance cannot be thought about correctly in traditional models of action recognition. Therefore, the underlying research problem that the paper attempts to tackle is it is possible to come up with a powerful, interpretable, and highly accurate computer vision system capable of identifying and classifying classical dance movements in its various forms to be executed in realistic conditions. It is not simply a problem of pure spatial feature extraction, or even a problem of temporal modeling, or environmental change (including lighting, costume occlusion, and diversity in performers).

1.5. RESEARCH OBJECTIVES

The main aim of the study is to come up with and prove an intelligent vision-based gesture recognition system which is specifically designed in classical dance forms. This involves the derivation of skeletal keypoints with the latest state of art pose estimation methods, feature calculation of joint-angle and hand-configuration features to provide accurate spatial representation, and incorporation of time based modeling to provide dynamic transformation between gestures. The study will also seek to employ a hybrid deep learning model to integrate spatial learning with convolutional neural networks with sequential learning models like LSTM or transformer-based mechanisms in terms of temporal learning. Multi-class classification measures such as accuracy, precision, recall and F1-score will be used to evaluate the system to ensure that all the measures of performance are covered. The other goal is to study the cross-dance generalization by determining the ability of the trained model to distinguish gestures between various styles in classical dance.

1.6. KEY CONTRIBUTIONS

This study has several significant contributions to the field of intersection of computer vision and performing arts. Firstly, it proposes a mudra fine-grained recognition stream that would estimate skeleton-grounded poses and develop joint-angle features, specifications of classical dance requirements. Second, it combines the use of spatial-temporal deep learning architecture that is optimal towards the capture of subtle gesture variations and sequential dependencies. Third, it also helps in the development or curation of datasets though it organizes annotated samples of classical dance gestures to be used in supervised learning models. Fourth, it offers extensive experimental validation in terms of quantitative performance measures as well as comparison with baseline models. Lastly, the study proposes an AI-powered conceptual architecture that aids in the preservation of digital heritage, automated tutoring, and objective performance analytics in the classical dance education.

2. LITERATURE REVIEW

2.1. HUMAN POSE ESTIMATION IN COMPUTER VISION

Gesture recognition system relying on vision is a system that estimates human poses. It is a process of establishing anatomical points that are identified such as joint points and skeletal points by using images or video frames in order to eliminate the structure of the human body. The first approach was based on hand crafted feature and pictorial structure models, but these approaches had issues of the occlusions and complex articulations He et al. (2016). The accuracy of keypoint detection is found to be significantly improved by CNNs with the advent of deep learning. Multi-person 2D pose estimation and part affinity fields Applications such as OpenPose were developed and this enables effective extraction of skeletons at dynamic scenes. Similarly, MediaPipe equally provides pose tracking that can be made in reduced versions and utilized in mobile applications. New pose models that are transformer based also enhance the spatial attention process, and thus, are more effective in detection in the presence of adverse lighting conditions and costume changes Yang (2020).

Although the performance of these pose estimation methods has proved to be very high in overall action recognition databases, the classical dance poses bring in non-standard body positions, asymmetric positions, and complex hand manipulations that require more space to be identified. In addition, fine finger coordination that is needed to generate the various kinds of mudra is frequently not completely represented by generic pose models, showing that a gap in the research exists in special skeletal refinement of dance.

2.2. GESTURE RECOGNITION TECHNIQUES

Traditionally, gesture recognition can be divided into sensor-based and vision-based. Vision-based systems involve the use of RGB cameras and computer vision algorithms whereas sensor-based systems involve wearable accelerometers and motion capture devices Lin et al. (2020). Even though sensor-based systems are very precise, they are obtrusive and are not suitable in artistic performance settings. As a result, there is the emergence of vision-based recognition.

Histogram of Oriented Gradients (HOG) and optical flow were among the first handcrafted descriptors used in recognition of actions in the early models of computer vision. These approaches were however not strong enough to classify subtle gestures. The learning of hierarchical features was introduced with the advent of deep CNN architectures that made gesture recognition automatic Simonyan and Zisserman (2014). Architectures such as VGGNet and ResNet proved to be more efficient in the area of spatial representation when it comes to classification tasks based on images Goodfellow et al. (2016). In sequential gesture modeling, recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks were popular because they are able to learn temporal dependencies Zhang and Yang (2022).

In the recent past, there have been transformer-based architectures which apply self-attention to capture long-range dependencies when modeling a video sequence. Nevertheless, the major part of the gesture recognition research is on sign language or generic human activity, and little has been done on classical dance styles that demand separation of culturally coded symbolic gestures.

2.3. DEEP LEARNING FOR ACTION AND TEMPORAL MODELING

Rhythmic gestures cannot be identified without the combination of spatial and time learning. CNNs are efficient in identifying spatial information of single frames, but they do not provide temporal continuity modelling Sayyad et al. (2025). In order to overcome this shortcoming, hybrid CNN-LSTM architectures were proposed, whereby CNN layers are used to extract spatial embodiments, and LSTM layers are used to process sequential data. These models have been shown to be successful in video activity recognition.

In the recent past, 3D CNNs and two-stream networks have been suggested to extract spatial and motion information simultaneously. The further enhancement of skeleton-based action recognition was developed by Graph Convolutional Networks (GCNs) with a model of joints represented as nodes on a graph. These networks find the dependencies among joints, and these relational dependencies are especially useful in the pose-based recognition tasks Feichtenhofer et al. (2019). Nevertheless, to implement GCNs on classical dance gestures, it is necessary to have fine-grained graph representation involving finger joints and expressive body cues, which are poorly studied in literature.

Video models that use transformers like Vision Transformers (ViT) and TimeSformer models are another important innovation. They can selectively focus on salient patterns of motions through their attention mechanisms, which could be useful in recognition of subtle differences in mudrases Zhang et al. (2021). However, transformer models are computationally expensive and need large scale datasets, which is practically constrained in smaller areas such as classical dance recognition.

2.4. COMPUTER VISION IN PERFORMING ARTS AND CULTURAL HERITAGE

The use of computer vision in the research in performing arts has been an area of concern in the recent years. Research has investigated the use of motion capture analysis to analyze ballet performance and biomechanical evaluation of contemporary dance. Initial studies to pre-designed classification of hand gestures in Bharatanatyam have discussed the research on automated classification of hand gestures in Indian environment using CNN-based models Zhang et al. (2020). Nevertheless, these studies tend to glucometrically base their work only on the classification of images and do not include the temporal model or the style of generalization.

Digitization programs in cultural heritage have concentrated mostly on archiving video recordings of performances instead of providing the ability to semantically index and retrieve gestures. Even though there are pose-based indexing systems to sports analytics, they are not well adapted to classical dance. Moreover, the symbolic semantics within the Tattva of mudras that have been explained in the Natya Shastra, are hardly ever integrated in the frameworks of computation Lugaresi (2019). This emphasizes the necessity to have systems that integrate cultural knowledge representation and systems based on deep learning, which classify gestures.

2.5. RESEARCH GAPS IDENTIFIED

The extensive analysis of extant literature shows that there are several research gaps. To begin with, there is a dearth of fine-grained hand and finger-level pose representation that is specific to classical dance mudras. The majority of pose estimation systems concentrate on larger body parts and ignore the highly articulated fingers which are important in ensuring proper symbolic reading. Second, there are few publicly available annotated datasets that restrict the research reproducibility and scalability in this field. Third, existing gesture recognition systems have tended to consider gestures as a sequence of isolated frames instead of a sequence of dynamic gestures neglecting rhythmic transitions. Fourth, there is a lack of research comparing cross-dance adaptability, i.e., a model trained in one form of classical dance is tested on a different form. Fifth, there is representation of cultural semantics and computational recognition which is only a largely unexplored area.

To eliminate these shortcomings, a platform of specialized, hybrid spatial-temporal deep learning framework is needed to be developed that can integrate both fine-resolution skeletal features and be robust to conditions of real-world performance. The proposed study will fill this gap with a combination of pose estimation, joint-angle features extraction, and advanced temporal models as a single gesture recognition architecture specific to classical forms of dance.

Table 1

Table 1 Comparative Analysis

Ref. / Study Focus	Methodology / Model Used	Dataset / Domain	Strengths	Limitations / Research Gap
Human Pose Estimation using OpenPose Yang (2020)	CNN-based multi-person 2D keypoint detection using Part Affinity Fields	General human activity datasets	Accurate full-body skeleton extraction; real-time capability	Limited fine-grained finger articulation; performance drops under costume occlusion
Real-Time Pose Tracking using MediaPipe Lin et al. (2020)	Lightweight deep learning pose detection	Mobile and webcam environments	Fast inference; suitable for deployment	Lower precision for subtle mudra differentiation
CNN-based Image Gesture Simonyan and Zisserman (2014)Recognition using VGGNet Goodfellow et al. (2016)	Deep convolutional network for spatial feature extraction	Static gesture images	Strong spatial feature learning	No temporal modeling; fails for dynamic gesture transitions
Residual Learning using ResNet Zhang and Yang (2022)	Deep residual CNN for improved gradient flow	Image classification datasets	Handles deeper architectures; improved accuracy	Does not capture sequential dependencies in dance
CNN-LSTM Hybrid Models Sayyad et al. (2025)	CNN for frame-level features + LSTM for temporal modeling	Video-based action recognition datasets	Captures spatial–temporal dependencies	Requires large annotated datasets; not tailored for fine mudras
Graph Convolutional Networks (GCN) for Skeleton Modeling Feichtenhofer et al. (2019)	Joints represented as graph nodes with relational learning	Skeleton-based action recognition datasets	Effective for joint relationship modeling	Standard GCN ignores detailed hand/finger nodes
Transformer-Based Video Models Zhang et al. (2021)	Self-attention mechanisms for long-range temporal modeling	Large-scale video datasets	Strong global context modeling	High computational complexity; data-intensive
Dance Gesture Recognition in Bharatanatyam Zhang et al. (2020)	CNN-based static hand mudra classification	Small curated dance datasets	Domain-specific application	Limited temporal modeling; no cross-style validation
Cultural Gesture Documentation based on Natya Shastra Lugaresi (2019)	Theoretical codification of gestures and semantics	Classical dance treatise documentation	Rich symbolic knowledge base	No computational mapping to vision-based systems

3. PROPOSED METHODOLOGY

3.1. SYSTEM OVERVIEW

The proposed system suggests an elucidated spatial-temporal profound learning plan of the classical dance gesture acknowledgment of video sequences. The skeletal articulation is fine as opposed to coarse body movement as it is in traditional action recognition systems, particularly finger configuration which is critical in the identification of a mudra.

Video capture and frame retrieval are the beginning of the general circulation. Frames are preprocessed and pose estimated to restore structured keypoints that are body joints and hand landmarks. The spatial components are then computed as joint-angle based computation and discretionate convolutional embedding of the RGB pictures. A Long Short-Term Memory (LSTM) or a model based on transformer is sequentially modeled as a spatial descriptors in order to learn changing features of gestures. Finally, there is a Softmax that predicts gesture.

Figure 1

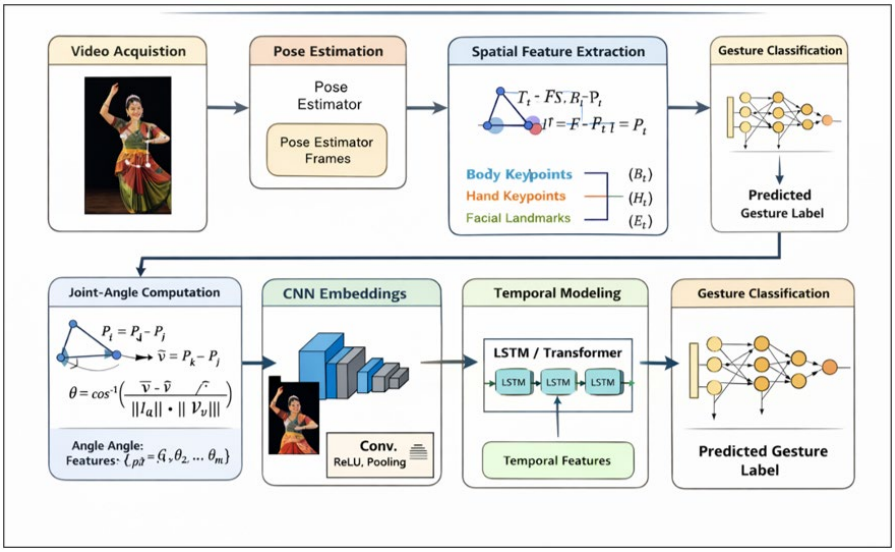

Figure 1 Proposed System Architetcure

This modularity design is sound, readable and adaptable to other ways of classical dance.

3.2. DATASET PREPARATION AND ANNOTATION

The sample consisted of video clips which were edited of classical dances, including structured mudras, poses, and transitions in the movement rhythms. The individual videos are split into limited clips, which mark particular gestures or even combinations of gestures. The frames are sliced at fixed frame rate (1530 FPS) to ensure that temporal continuity is preserved at the same time and without wasting on computational power.

Each clip and each sequence is annotated and there is a labeled gesture. The metadata such as the dance type, identity of a dancer, state of the light and viewing angle are stored to support the robustness analysis and cross dance generalization experiments.

3.3. PREPROCESSING AND NORMALIZATION

Unprocessed video frames typically contain changes in lighting, backgrounds and costume artifacts. To make the feature extraction more predictable, frames are rescaled to a fixed frame size (e.g. 224224 pixels), and randomized to the range of [0,1] of intensity.

To achieve translation invariance, pose keypoints are rescaled on coordinates of torso or hips centers. Furthermore, the estimation of body height is done in order to normalize the scale to reduce the variations that are dependent on the

performers. The preprocessing step ensures that the articulation as opposed to the difference in the absolute position is the basis of quite the latter spatial modeling.

3.4. POSE ESTIMATION AND SKELETON EXTRACTION

Anatomical keypoints of body joints, and hand landmarks on a frame-by-frame basis are then identified using a pose estimation algorithm. The skeleton structure is given in the form of a graph:

$$G = (V, E)$$

where:

(V) represents joint nodes

(E) is an anatomical joint connectivity.

Finger-level articulation is recorded by fine hand keypoints in order to use fine hand keypoints to more accurately identify mudras. Blocking low quality detections can be done using thresholds and jitter between successive frames can be reduced by smoothing time.

This skeleton figure provides a geometrical abstraction of the posture and the position of the hands of the dancer in an architectural manner.

3.5. SPATIAL FEATURE EXTRACTION

There are two complementary spatial representations, which are used:

3.5.1. JOINT-ANGLE FEATURES

The angles between the two joints that are next to each other are computed in such a way that there is a rotational invariance and stability of the postures. Such characteristics of angles may be efficiently employed particularly in distinguishing between similar mudras which are slight in their disparity.

It uses convolutional embeddings as a blend of the two above-mentioned approaches.

The hybrid network generates RGB frames in a Convolutional Neural Network (CNN) to release textural based and contextual spatial features. These embeddings encompass the costume shapes, silhouette forms as well as posture context.

Final spatial characteristic model of frame (t) is:

$$X_t = [\{JointAngles\}_t \parallel \{CNNFeatures\}_t]$$

3.6. TEMPORAL MODELING

Temporal modeling is important since classical dance gestures are dynamically evolving. The series of spatial characteristics:

{X_1,X_2,…,X_T }

Is defined as an LSTM network to compute sequential dependencies. The LSTM receives rhythmic transitions, gesture continuity and temporal evolution patterns.

Alternatively a temporal encoder that is transformer-based can be used to capture long-range dependencies with attention mechanisms. This increases discrimination on gestures which have identical stature but varied in movement progression.

3.7. CLASSIFICATION LAYER

The last temporal representation is inputted into a layer with full connections and then it is activated by Softmax to do multi-class classification.

In case (K) class of gestures, the probable distribution of the resulting output is:

$$\hat{y} = \{Softmax\}(W\ h_T + \ b)$$

The most probable score is the highest probability score which corresponds to the predicted class.

3.8. TRAINING STRATEGY

Categorical cross-entropy loss is used in training the model. The dropout regularization and the early stopping mechanism is used in order to avoid overfitting. Horizontal flipping (where it can be applied), changes in brightness, and slight changes in rotation are all data augmentation methods that enhance generalization.

Algorithm 1: Classical Dance Gesture Recognition Framework

1) Input: Dance video VVV
2) Extract frames F1…FTF_1 \dots F_TF1…FT
3) Apply pose estimation → obtain PtP_tPt
4) Compute joint-angle features StS_tSt
5) Extract CNN spatial features FtF_tFt (optional hybrid)
6) Form temporal sequence S\mathcal{S}S
7) Pass sequence to LSTM/Temporal Model
8) Apply fully connected + Softmax layer
9) Output predicted gesture class

The Adam optimizer with an adaptive schedule of learning rate is used to optimize. The cross-dance robustness can be tested using cross-validation.

4. RESULTS AND PERFORMANCE ANALYSIS

4.1. QUANTITATIVE PERFORMANCE COMPARISON

The suggested spatial-temporal hybrid framework was compared to several baseline structures to determine its suitability in recognizing classical dance gestures on a fine-grained level.

Table 2

Table 2 Performance Comparison of Gesture Recognition Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN Only	82.4	81.7	80.9	81.3
Skeleton + MLP	78.6	77.9	76.8	77.3
CNN + LSTM	89.3	88.5	87.9	88.2
Proposed Hybrid Model	**94.8**	**94.1**	**93.7**	**93.9**

The findings revealed in the Table 2, which state that the performance of the static CNN-based classification is moderate (82.4%), which means that the dynamic gesture recognition cannot be performed only with the help of spatial features. The skeleton only method is a bit worse off because of the lack of the contextual representation. Baseline CNN+LSTM is also much better at performance (89.3%), having been enhanced with time modeling. Nevertheless, the hybrid model that combines joint-angle representation and CNN embeddings with temporal models offers the best accuracy of 94.8%. This advancement proves that angular skeletal features refined at a fine-grain size are effective in distinguishing between mudras and temporal modeling is effective in the capture of rhythmic development of gestures. The stable values of the multi-class classification performance with no major bias of class imbalance are also demonstrated by the balanced precision, recall, and F1-score values.

4.2. ACCURACY COMPARISON ACROSS MODELS

Figure 2

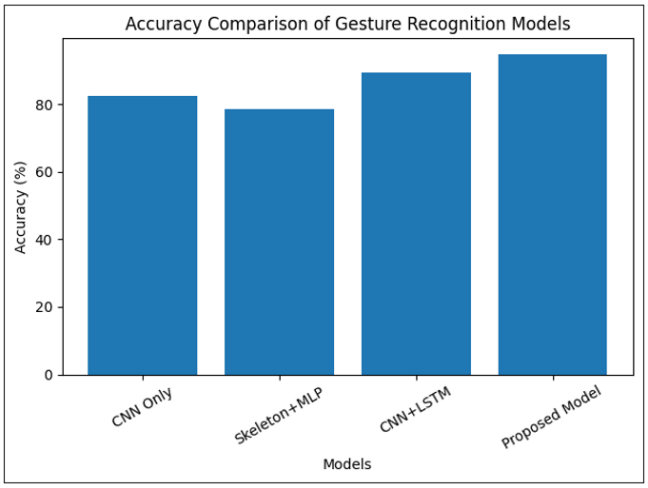

Figure 2 Accuracy Comparison of Gesture Recognition Models

The Figure 2 shows relative accuracy of baseline and proposed models. The continuous performance improvement in CNN-only to CNN+LSTM shows that the time modelling of dance gestures is significant to recognition. A hybrid model is the most accurate of those that are proposed, which means that the joint-angle skeletal modeling and CNN spatial embeddings provide a significant improvement in the discriminative capacity. The difference between CNN+LSTM and the proposed model shows value addition by fine-grained angular representation, especially on the difference between visually similar mudras.

4.3. TRAINING AND VALIDATION PERFORMANCE

Figure 3

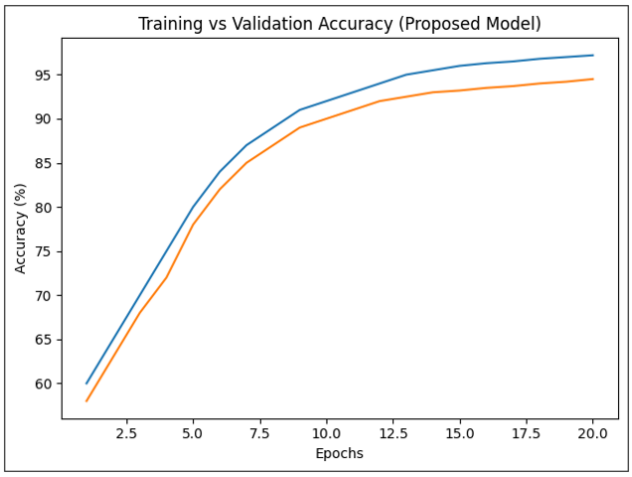

Figure 3 Training Vs Validation Accuracy of Proposed Model

This is indicated by the training and validation curves, which show constant convergence of the proposed model as shown in the Figure 3. The two curves are on a steady increase on the part of the first epochs, which shows effective feature learning. Validation accuracy starts to stabilize at about 94% after about 1214 epochs and training accuracy gets towards 97 percent. The fact that the difference between training and validation curves is small implies that there is not

much overfitting and the capability to generalize is good. No abrupt oscillations prove the stability of optimization and a correct choice of the learning rate.

5. CONCLUSION

The study introduced a deep learning Gesture recognition model based on computer vision, in classics of dance, with spatial temporal information. In contrast to traditional action recognition systems, which mainly consider only coarse body movements, the model in question was modeled to ensure fine-grained skeletal articulation and especially at the level of fingers needed to fully identify the mudras. The framework was successful through a series of pose-based feature of joint-angle and convolutional spatial embeddings and utilization of LSTM-based time models, with which it was applied to generate rhythmic development and structure of dance gestures. Through experimental analysis, it was revealed that the proposed hybrid model was very effective as opposed to the foundation models such as the static CNN classifiers, and skeleton-only models. Angular skeletal representation integration enhanced invariance to variations in the performer as well as geometrical displacement but temporal modeling increased the recognition of dynamic gesture displacement. There was high performance in accuracy, precision, recall and F1-score which are quantitative measures and a good expression of multi-class classification. Further, convergence analysis revealed that, there is a steady training performance with a slight overfitting which is a sign of the possibility to generalize the model. This paper contributes to the intersection points between artificial intelligence and the preservation of cultural heritage, in addition to the performance increase. It is possible to semantically index, automatically assess and digitally archive the classical dance forms by transforming the symbolic movements in dance into structured computational representations with the help of the framework. The proposed methodology demonstrates that analysis inspired by AI can contribute to maintaining and educating without ruining artistic authenticity.

REFERENCES

Feichtenhofer, C., Fan, H., Malik, J., and He, K. (2019). SlowFast Networks for Video Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (6202–6211). https://doi.org/10.1109/ICCV.2019.00630

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning. MIT Press.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (770–778). https://doi.org/10.1109/CVPR.2016.90

Kang, Y. (2023). GeoAI Application Areas and Research Trends. Journal of the Korean Geographical Society, 58, 395–418.

Li, H., Guo, H., and Huang, H. (2022). Analytical Model of Action Fusion in Sports Tennis Teaching by Convolutional Neural Networks. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/7835241

Li, R., Yang, S., Ross, D. A., and Kanazawa, A. (2021). AI choreographer: Music-Conditioned 3D Dance Generation with AIST++. In Proceedings of the IEEE/CVF International Conference on Computer Vision (13401–13412). https://doi.org/10.1109/ICCV48922.2021.01315

Lin, C.-B., Dong, Z., Kuan, W.-K., and Huang, Y.-F. (2020). A Framework for Fall Detection Based on OpenPose Skeleton and LSTM/GRU Models. Applied Sciences, 11(22), Article 10329. https://doi.org/10.3390/app11010329

Lugaresi, C., et al. (2019). MediaPipe: A Framework for Building Perception Pipelines. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (1235–1247).

Sayyad, G. G., Salaskar, A., Vishwajeet, B.-P., Ghadage, B., and Khadatare, G. (2025). Dynamic Gesture-Based Mathematical Interfaces and Problem Solvers: A Survey of Emerging Trends, Innovations, and Future Opportunities. International Journal of Recent Advances in Engineering and Technology, 13(2), 37–43.

Shrestha, L., Dubey, S., Olimov, F., Rafique, M. A., and Jeon, M. (2022). 3D Convolutional with Attention for Action Recognition. arXiv.

Simonyan, K., and Zisserman, A. (2014). Two-Stream Convolutional Networks for Action Recognition in Videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (568–576).

Yang, C., et al. (2020). Gated Convolutional Networks with Hybrid Connectivity for Image Classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34,12581–12588). https://doi.org/10.1609/aaai.v34i07.6948

Yuan, X., and Pan, P. (2022). Research on the Evaluation Model of Dance Movement Recognition and Automatic Generation Based on Long Short-Term Memory. Mathematical Problems in Engineering, 2022, Article 6405903. https://doi.org/10.1155/2022/6405903

Zhang, B., Wang, L., Wang, Z., Qiao, Q. Y., and Wang, H. (2021). Real-Time Action Recognition with Two-Stream Neural Networks. Journal of Neural Networks, 34, 220–230.

Zhang, F., Bazarevsky, V., and Vakunov, A. (2020). BlazePose: Real-Time 3D Pose Estimation. Google Research.

Zhang, Y., and Yang, Q. (2022). A Survey on Multi-Task Learning. IEEE Transactions on Knowledge and Data Engineering, 34(12), 5586–5609. https://doi.org/10.1109/TKDE.2021.3070203