ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20241218002203
Original scientific paper

Development of a Performance Evaluation System in Turkish Folk Dance Using Deep Learning-Based Pose Estimation

Erdem BÜYÜKGÖKOĞLAN, Sinan UĞUZ*

Abstract: Folk dances are integral cultural expressions that reflect a society's heritage, values, and historical development. In the context of education, folk dances promote belonging, discipline, and creativity. This study explores an innovative approach to teaching and evaluating traditional Turkish folk dance, through the integration of deep learning, time-series analysis, and classical methods. The system developed in this study enables objective performance assessment by students and provides teachers with valuable feedback for monitoring progress. The performances of one teacher and nine students, each performing five distinct dance figures, were captured via webcam recordings, and skeletal data were analysed using Mediapipe and YOLO. To evaluate performance, classical methods such as Euclidean distance and Cosine similarity were employed in conjunction with time-series techniques like TLCC and DTW, as well as deep learning models such as LSTM and Siamese Networks. Among these methods, the LSTM model emerged as the most effective, achieving an average score of 68,43 and an MSE of 56,11. The DTW method followed, achieving an average score of 60,64 and an MSE of 139,32. Overall, the integration of technology, particularly deep learning, into traditional dance education presents a transformative opportunity for enhancing performance evaluation in folk dance.

Keywords: dance evaluation, DTW, folk dance, LSTM, pose estimation, siamese networks, TLCC

1 INTRODUCTION

Traditional folk dances serve as a vital medium for preserving and transmitting a society's cultural identity and heritage. Embedded within these dances are the values, beliefs, and historical narratives of communities, making them essential components of intangible cultural heritage. These dances not only embody artistic expression but also function as living archives that connect generations, reinforcing social cohesion and communal belonging [1]. The ritualistic and celebratory aspects of folk dance reflect the lived experiences of people, commemorating historical events, agricultural cycles, and social customs [2]. Folk dance performances have traditionally been evaluated through subjective assessments conducted by expert juries. These evaluations are based on standardized scoring criteria, including movement accuracy, rhythmic precision, balance, body control, energy, and aesthetic quality. However, such assessments are inherently susceptible to bias and inconsistency, thereby presenting significant challenges to the standardization and long-term preservation of folk dance traditions [3]. The need for objective and systematic evaluation methods is crucial in ensuring the integrity and continuity of folk dance as an art form, particularly in the context of evolving pedagogical approaches.

Moreover, there is a notable lack of effective pedagogical approaches that foster the development of students' dance skills and enable objective self-assessment during the learning process. This gap poses significant challenges for dance education and can negatively influence students' motivation. Traditional dance instruction heavily relies on resemblance, imitation, repetition, and reflection. However, common limitations of conventional training include restricted and often cost-prohibitive access to expert instructors and reliance on mirrors for feedback, which only provide instantaneous visual input without quantitatively assessing movement accuracy [4, 5].

Early research in dance evaluation using pose estimation has primarily focused on the automated evaluation of dance performances and the provision of visual feedback through Kinect-based skeletal tracking. While Kinect technology shows considerable potential, its adoption has been constrained by limitations such as the high cost of sensors and associated software [6, 7].

Recently, deep learning has emerged as a faster and more cost-effective alternative to sensors and motion capture technologies for pose estimation tasks, such as joint point identification and angle computation. The data obtained through pose estimation enables the calculation of a dancer's position, velocity, joint angles, and acceleration, facilitating skill and performance evaluation. This capability has driven the growing popularity of deep learning in this field [8].

This study aims to design a deep learning-based system to evaluate Turkish folk dance performances. The proposed system analyses the movements of student dancers using pose estimation methods and scores their performances by comparing them to those of instructor dancers. This approach seeks to provide an objective and automated evaluation of dancers' performances without the need for an instructor's presence.

2 RELATED WORKS

Recent advancements in dance education and evaluation through human pose recognition demonstrate the increasing relevance and applicability of this field. The research conducted in this domain exhibits significant methodological differences across several key criteria, as outlined in Tab. 1. It is evident that pose data undergo a series of preprocessing steps, key point detection is performed using a variety of pose extraction algorithms, and distinct methodologies are employed for the evaluation of dance performance. These differences highlight the diversity of approaches in the literature and the ongoing exploration of advanced techniques aimed at enhancing the effectiveness of dance training and assessment.

An examination of Tab. 1 reveals that OpenPose, PoseNet, AlphaPose, Mediapipe, and YOLO are among the most commonly used pose estimation algorithms in the

literature. These algorithms have gained significant recognition for their advanced capabilities in detecting and tracking human keypoints with high accuracy. They have become essential tools for analyzing movement and performance across various disciplines, including dance education and assessment [9-12].

Significant challenges in pose estimation for dance evaluation include occlusion (where one joint obstructs the view of another), dancer synchronization, variations in body size, camera angle differences, and the influence of irrelevant points on performance assessment. To address the problem of occlusion, techniques such as point imputation using directograms or interpolation have been explored to enhance the accuracy and reliability of pose detection and performance evaluation [10, 12].

Techniques including the Ant Colony Optimization algorithm and Dynamic Time Warping (DTW) have been employed for dancer synchronization, aligning movements across multiple performers despite timing or sequencing differences [2, 11, 13]. To address variations in body size and camera angle differences, affine transformations are frequently utilized. These transformations help normalize the data, ensuring consistency in pose estimation by compensating for perspective and scaling differences, thereby improving movement analysis accuracy under diverse conditions [9, 11].

For performance evaluation, Cosine Similarity and Euclidean Distance are the most prominent techniques. These methods quantify the similarity between pose data, enabling precise movement alignment and contributing to the objective assessment of dance performance. Additionally, advanced machine learning methods, including Hidden Markov Models (HMM) and Support Vector Regression (SVR), have been employed alongside deep learning approaches such as Long Short-Term Memory (LSTM) networks. These techniques capture temporal dependencies and complex movement patterns, offering improved accuracy in dynamic performance evaluation [10].

In this study, affine transformations were applied as a preprocessing step to enhance the accuracy and consistency of pose data. MediaPipe and YOLO algorithms were chosen for 2D pose estimation, as they operate on the CPU and offer a balance of speed and accuracy. For performance evaluation, a comprehensive approach was employed, incorporating Euclidean distance, cosine similarity, DTW, Time Lagged Cross Correlation (TLCC), LSTM networks, and Siamese networks. This multifaceted evaluation strategy enabled a rigorous and precise assessment of the dancers' performances.

Table 1 Related Works

Works	Year	Dance	Pre-processing	Pose Estimation	Evaluation
Lee vd. [9]	2020	K-Pop	Affine	Kinect	Object Keypoint Similarity Angle Similarity
Zhou vd. [10]	2021	K-Pop	Direktogram	AlphaPose	LSTM SVR
Tang vd. [11]	2022	Modern	Similarity Matching Ant Colony	MMPose VideoPose3D	Euclidean Distance Cosine Similarity
Kang vd. [12]	2023	Modern		Openpose	Cosine Similarity
Sheng vd. [13]	2023	Classic	SSA	OpenPose	DTW
Ding & Wang [14]	2023	Modern		Mediapipe, YOLO v7	Cosine Similarity
Our Works	2024	Folk Dance	Affine	Mediapipe YOLO v8	Euclidean Distance, Cosine Similarity, DTW, TLCC, LSTM, Siamese Networks

2.1 Research Gaps and Contributions

Traditional dance education, primarily based on observation and imitation, often results in subjective assessments and insufficient monitoring of student progress. There is a critical necessity for technological advancements in this field to provide more objective evaluation methods and to facilitate more effective tracking of students' development [4, 15].

Traditional dance performances are often evaluated based on personal criteria, emphasizing the necessity of a fair and consistent evaluation tool. Such a tool would ensure more objective and standardized assessments, reducing subjectivity and enhancing the reliability of performance evaluations [3].

Although there have been studies on movement analysis and evaluation of dance performances, research specifically focused on folk dances is significantly underexplored. Further investigation in this area could enhance the quality of dance education and contribute to the preservation of cultural heritage. Additionally, research on utilizing Long Short-Term Memory (LSTM) networks, Siamese Networks, and other deep learning methods to analyze dance movements remains limited, presenting an opportunity to develop more accurate and comprehensive evaluation systems for dance performance.

This study presents a novel system that employs various deep learning techniques, including Long Short-Term Memory (LSTM) networks and Siamese Networks, for the instruction and evaluation of traditional folk dance. By digitally adapting the mirror technique widely practiced in dance education, the system facilitates self-assessment for novice dancers. The developed system provides objective and detailed feedback, which in turn supports the advancement of students' skills.

The application of deep learning and time series analysis methods to traditional dance evaluation provides groundbreaking insights in the field. In particular, the effectiveness of the LSTM model underscores the potential of these advanced methodologies in enhancing the precision and depth of dance performance assessment.

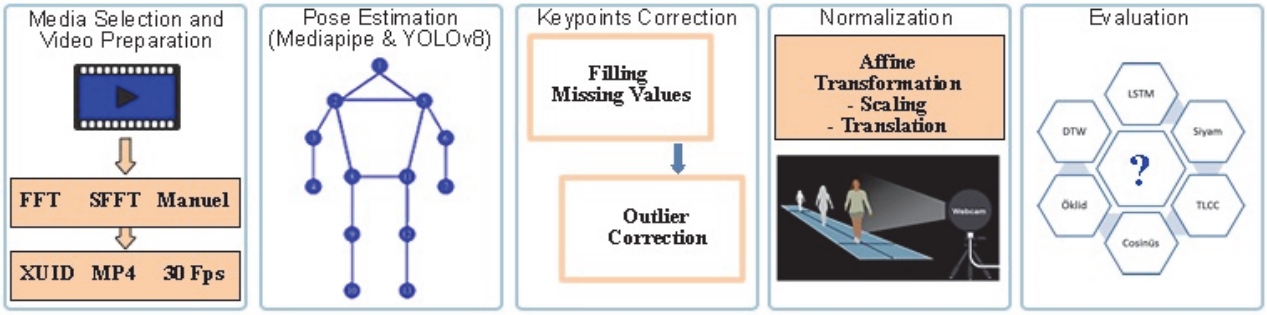


Figure 1 Dance Performance Evaluation Pipeline

3 MATERIALS AND METHODS

A detailed overview of the processes implemented in this study is illustrated in Fig 1. The subsequent subheadings provide an in-depth discussion and detailed analysis of these processes.

3.1 Media Selection and Video Preparation

The initial phase of the study involves the processing and integration of both student and teacher videos into the system. In cases where the student requires the automatic synchronization of the video's start and end times with the accompanying music, it is necessary to upload the music track in MP3 format.

In this scenario, the Short-Time Fourier Transform (STFT) is employed to automatically detect the beginning and end points of the music within the video. The corresponding sections are then extracted and saved into a new video file. Alternatively, users have the option to manually trim the video through the application's interface (Fig. 2).

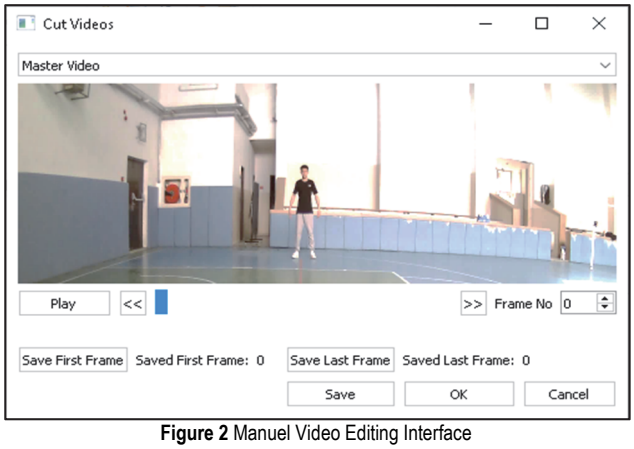


Figure 2 Manuel Video Editing Interface

The application provides functionality for users to adjust the frames per second (fps) rate of the video. Ensuring that two videos have the same fps value is essential, particularly for time synchronization, movement analysis, and performance evaluation. Discrepancies in fps may hinder the comparison of simultaneous movements and result in timing errors [16]. Additionally, reducing the fps rate helps reduce computational costs [17]. In this section, video segments defined by user-specified start and end frame numbers are recorded in MP4 format using the XVID video codec and the OpenCV library, with a user-defined fps value.

3.2 Pose Estimation

At this stage, pose extraction is applied to the finalized video files. In this study, MediaPipe and YOLO v8 were selected for pose estimation due to their computational efficiency and suitability for real-time applications [18].

Mediapipe is a deep learning and computer vision-based framework developed by Google, designed for detecting human skeletal posture. It identifies 33 distinct keypoints on the human body with high accuracy and speed. Mediapipe processes data at a rate of 44 fps on a CPU, making it an efficient tool for real-time pose estimation and movement analysis. Its ability to operate with minimal computational resources while maintaining high performance makes it particularly useful for applications such as motion tracking, human-computer interaction, and performance evaluation in dynamic environments [19].

YOLO is a popular deep learning architecture specialized in object detection and classification. The YOLOv8s-pose model utilized in this study processes a single frame in approximately 233.2 ms. In YOLOv8 pose models, there are 17 key points that represent different parts of the human body. Notably, points located in the facial region, such as the eyes and ears, are not relevant to dance performance evaluation. Their inclusion in the analysis can introduce irrelevant data, which can compromise the accuracy of the assessment, as they do not contribute to the analysis of body movements and posture during the performance [20].

Both YOLO and Mediapipe models, particularly those that detect facial features such as eyes and ears, may negatively impact performance evaluation in dance studies. Since these points are not directly related to dance movements or posture, their inclusion in the analysis can introduce irrelevant data, which can compromise the accuracy of the assessment. The focus should be on key body joints that are more pertinent to movement and posture evaluation for dance performance [10, 14].

In this study, based on similar works in the literature, many facial keypoints (except for the nose) were omitted from the analysis. Instead, 13 keypoints critical for dance performance (Tab. 2) were selected for analysis. When determining these points, those that were common to YOLO and Mediapipe were prioritized. This approach ensures that the most relevant and consistent keypoints across both methods are used, enhancing the accuracy and reliability of the dance performance evaluation.

3.3 Keypoint Correction

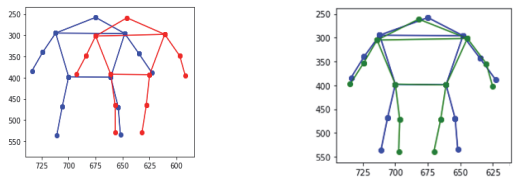
Although the accuracy of human pose algorithms has improved over time, challenges persist, particularly in occlusion situations [21]. To minimize the impact of such errors, erroneous or missing data points were corrected using a forward-backward imputation method. Forward-backward imputation is an effective technique for filling missing values in time series data. This method calculates the average of the nearest valid observations before and after the missing data point. The resulting value is used to fill the gap, ensuring consistency with surrounding observations. This technique maintains consistency with consecutive observations and effectively preserves the trend and structure of the data [22].

Table 2 Revised Keypoints

Mediapipe Keypoint ID	YOLO Keypoint ID	Body Part
0	0	Nose
11	5	Left Shoulder
12	6	Right Shoulder
13	7	Left Elbow
14	8	Right Elbow
15	9	Left Wrist
16	10	Right Wrist
23	11	Left Hip
24	12	Right Hip
25	13	Left Knee
26	14	Right Knee
27	15	Left Ankle
28	16	Right Ankle

3.4 Normalization

In pose estimation methods used to evaluate students' performances, variations in students' physical characteristics and camera distances introduce significant challenges. These factors impact the accuracy and consistency of keypoint detection, potentially leading to discrepancies in performance assessment [23]. Therefore, to standardize the body proportions of the teacher and the student, an Affine Transformation was applied. Scaling was conducted based on the distance between the nose and the midpoint of the feet. The body width was normalized using the distance between the right and left hips. Subsequently, a translation transformation was applied. The student's hip coordinates were aligned with the teacher's hip coordinates to synchronize the central points of the two dancers (Fig. 3). This approach was designed to enhance the consistency of the evaluations.



a) Scaling b) Translation

Figure 3 Affine Transformations

3.5 Evaluation

In this study, performance scores were calculated through six distinct methods: LSTM, Siamese Network, DTW, TLCC, Euclidean distance, and cosine similarity. While the Euclidean distance and cosine similarity methods were computed on a frame-by-frame basis, the other methods were applied to the entire video data. In video-based dance performance evaluation, time series analysis methods were utilized to analyze the temporal flow of the dancers' movements and their alignment with the choreography. These analyses aim to identify the strengths and weaknesses of the performances, evaluate choreography compliance, and provide feedback [24].

3.5.1 Euclidean Distance

For each keypoint, the Euclidean distances between the student's and the teacher's poses were computed and aggregated to obtain a total error value. This error value represents the difference between the two poses. Euclidean distance $d(p,q)$ is defined as shown in Eq. 1.

$$d(p,q) = \sqrt{\left((p1-q1)^2 + (p2-q2)^2\right)} \tag{1}$$

where, p and q represent the coordinates of the two key points being compared [25]. In each frame, the calculation was performed separately for each of the 13 key points, and the average was taken.

3.5.2 Cosine Similarity

Cosine similarity quantifies the similarity between poses by calculating the cosine of the angle between the vectors representing the key points. Due to its scale-invariance and directionality, it ensures robust comparisons even in poses captured from different camera angles or distances. This feature offers a significant advantage in comparing skeleton data of varying sizes. The cosine similarity is calculated using the following formula:

$$\cos(\theta) = (A \cdot B) \Big/ \left(\left|\left(|A|\right)\right|\left|\left(|B|\right)\right|\right) \tag{2}$$

where, A and B are the two vectors representing the poses, $(A \cdot B)$ is the dot product of the two vectors, $\|A\|$, $\|B\|$ are the magnitudes of the vectors. The dot product (·\cdot·) is a scalar value calculated by multiplying and summing corresponding elements of the two vectors (e.g., the *x*-coordinates of the left elbow in both vectors) [25, 26].

3.5.3 Time Lagged Cross Correlation

TLCC is a statistical method applied to analyze how changes in one time series affect changes in another time series and the time delay at which this influence occurs. TLCC is particularly useful for understanding the temporal dynamics between two time series. In the context of dance education, TLCC can be used to evaluate the timing and movement accuracy by measuring how delayed a student follows the instructor's movements and how similarly specific movements are performed [27].

3.5.4 Dinamic Time Wrapping

The DTW algorithm is designed to handle timing discrepancies arising from individual speed differences in dance performances by comparing the movements of the instructor and the student. This algorithm optimally aligns pose sequences from both performances, minimizing the distance between poses at different time steps [4]. As a result, the synchronization of performances and the accuracy of the student's mimicry can be quantitatively assessed. The DTW distance, typically calculated using a metric like Euclidean distance, is inversely proportional to the similarity between the sequences [24].

3.5.5 Long Short Time Memory

LSTM is a type of recurrent neural network widely employed in deep learning applications. LSTMs are particularly effective in handling time series data and modeling long-term dependencies [28]. In this study, LSTM networks process pose data extracted from the dance videos of the instructor and the student to score the student's performance. By learning the temporal flow of movements and choreography, LSTM is capable of evaluating the overall quality of the student's performance [29]. Twelve LSTM models with different parameters were developed and trained. The parameters used for these training sessions are presented in Tab 3. Among these, the model LSTM_11 achieved the lowest error rate. The architecture of the LSTM model that achieved the best results is illustrated in Fig 4.

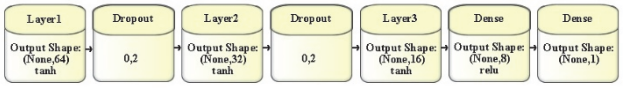


Figure 4 Structure of the Most Successful LSTM Network

Table 3 LSTM Models' Paremeters

Model	Layer1	Dropout1	Layer2	Dropout2	Layer3	Dense	Epoch	Optimizer	Activation
LSTM_1	128	0,2	64	0,2	32	16	10	adamax	relu
LSTM_2	128	0,2	64	0,2	32	16	10	sgd	relu
LSTM_3	128	0,3	64	0,2	32	16	10	sgd	relu
LSTM_4	128	0,1	64	0,2	32	16	10	sgd	relu
LSTM_5	52	0,2	26	0,2	13	8	10	adam	relu
LSTM_6	64	0,2	32	0,2	16	8	10	adam	tanh
LSTM_7	64	0,2	32	0,2	16	8	10	adamax	tanh
LSTM_8	64	0,2	32	0,2	16	8	10	adamax	tanh
LSTM_9	64	0,2	32	0,2	16	8	10	adam	relu
LSTM_10	64	0,2	32	0,2	16	8	10	adam	tanh
LSTM_11*	**64**	**0,2**	**32**	**0,2**	**16**	**8**	**10**	**sgd**	**tanh**
LSTM_12	64	0,2	32	0,2	16	8	20	adam	tanh

3.6.6 Siamese Network

Siamese networks represent an image matching technique developed by Bromley and colleagues in the early 1990s [30]. These networks offer robust solutions, particularly in similarity learning problems. Siamese networks consist of two or more identical sub-networks sharing the same structure. These sub-networks are trained using a shared loss function. The most distinctive feature of Siamese networks is their ability to quantify the similarity or dissimilarity between two input examples. This capability makes Siamese networks an indispensable tool for pose comparison [31]. In this study, eight different Siamese models were created to measure the similarity in motion data, as shown in Tab 4.

Table 4 Siammese Networks' Parameters

Model	Subnet	LSTM Unit	Dense Unit	Optimizer
Siamese_1	LSTM	32	16	Adam
Siamese_2	LSTM	32	16	RMSprop
Siamese_3	LSTM	32	8	Adam
Siamese_4	LSTM	32	8	RMSprop
Siamese_5	LSTM	64	16	Adam
Siamese_6*	**LSTM**	**64**	**16**	**RMSprop**
Siamese_7	LSTM	64	8	Adam
Siamese_8	LSTM	64	8	RMSprop

Each model was tested using different configurations of LSTM neurons, dense layer neurons, and dropout rates. The models were trained with a learning rate of 0,001; and similarity scores between teacher and student movements were obtained. After training, the best-performing model was identified as Siamese_6. The structure of the best-performing Siamese network model created in this study is illustrated in Fig. 5.

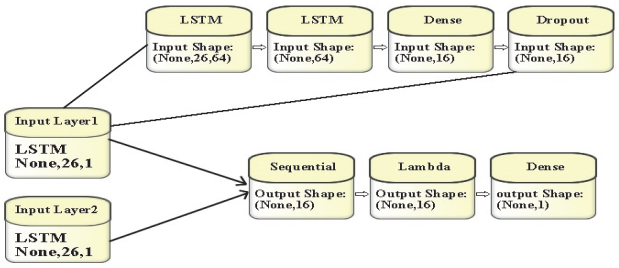


Figure 5 The Structure of the Most Successful Siamese Network

3.7 Graphical User Interface

In this study, a GUI consisting of five different screens was developed to allow students to thoroughly analyze their dance performances. Through the Pose Comparison Screen shown in Fig. 6, the student can monitor their synchronized performance with the teacher. At the bottom of the form, the score calculated for the relevant frame and the average score are displayed. Additionally, the user can change the evaluation method using the menu located in the lower left section. Another crucial screen in the application is the evaluation screen, dedicated to scoring and evaluation (Fig. 7). This screen features a digital

referee evaluation system. It presents an evaluation list covering all the methods used in this study. The user selects the expert video and their own video for comparison and also selects the pose estimation method.

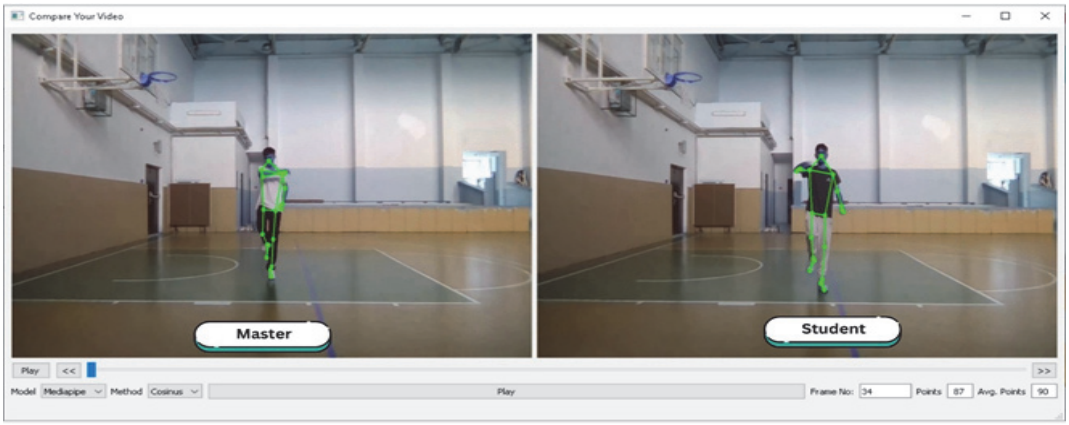


Figure 6 The Pose Comparison Screen

Subsequently, the evaluation methods (TLCC, DTW, Euclidean, Cosine, Siamese Networks, and LSTM) are then chosen for scoring. Using these selections, the system performs the necessary calculations and generates a score ranging from 0 to 100, displaying the detailed results. Both frame-based averages and video-based averages are independently calculated.

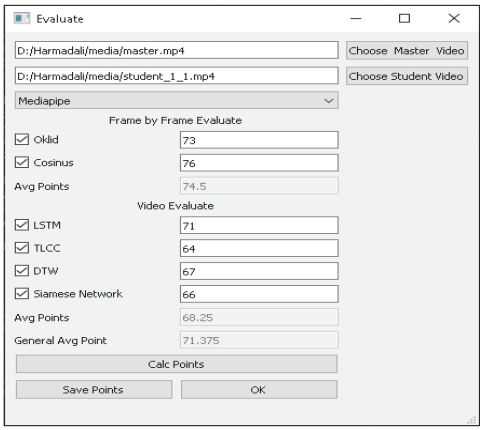


Figure 7 Evaluation Menu

To enable the use of this section in dance competitions for performance evaluations, a save button was implemented for storing performance scores. This feature allows users to monitor their progress over time by saving their performances. Teachers, in particular, can utilize this feature to closely monitor their students' development.

4 RESULTS AND DISCUSSION

The system developed in this study evaluates dancers' poses, movements, and overall choreography while providing constructive feedback. The primary objective of the system is to enhance dance practice, offer personalized guidance, and accelerate skill development. The performances of six distinct evaluation metrics were analyzed by comparing them with expert evaluations. The analysis relied on data collected from 9 students performing 5 different dance movements. The similarity in the averages between YOLO and Mediapipe suggests that both methods effectively predict the positions of keypoints. However, it has been observed that Mediapipe offers a speed advantage, as demonstrated in the tutorials [20, 32]. When comparing the model averages with expert ratings, it is evident that the model averages closely approximate the expert evaluations. Fig. 8 presents the average scores of all models as percentages. The LSTM model, with an average score of 68,43; demonstrates the closest alignment with the expert average of 66,67.

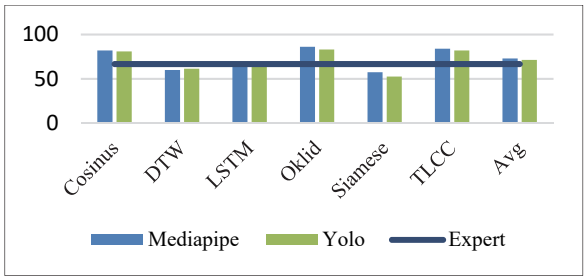


Figure 8 Average Points by Pose Estimation Algorithms

Euclidean distance achieved the highest average score, while Siamese Networks had the lowest (Fig. 8). LSTM models demonstrated consistent performance across both pose estimation methods. The best-performing LSTM model, LSTM_11, with the lowest MSE value of 56,11; was selected for the application. Similarly, the best Siamese model, Siamese_6, which achieved the lowest MSE value, was integrated into the system.

Table 5 LSTM and Siamese Network Models MSE Scores

LSTM Models	MSE		Siamese Models	MSE
LSTM_1	573		Siamese_1	225
LSTM_2	56,13		Siamese_2	239
LSTM_3	56,25		Siamese_3	251
LSTM_4	56,70		Siamese_4	267
LSTM_5	700		Siamese_5	251
LSTM_6	736		**Siamese_6**	**224**
LSTM_7	528		Siamese_7	301
LSTM_8	527		Siamese_8	271
LSTM_9	714			
LSTM_10	725			
LSTM_11	**56,11**			
LSTM_12	706			

Table 6 MSE Scores of all evaluation methods

Method	MSE	MAE	PCC
Cosine Similarity	290,98	8,0	0,60
DTW	139,32	8,5	0,58
LSTM	56,11	5,2	0,85
Euclidean Distance	432,68	7,8	0,62
Siamese Networks	224,08	5,5	0,82
TLCC	355,74	7,1	0,65

The similarity in the averages between YOLO and Mediapipe suggests that both methods effectively predict the positions of key points. However, it has been observed that Mediapipe offers a speed advantage, as demonstrated in the tutorials [20, 32]. ANOVA analysis was conducted to evaluate the overall agreement between different scoring methods and expert ratings and helps determine whether there is a statistically significant difference among the scoring methods.

The results indicated a significant difference between the various scoring methods ($F(5,94) = 12,5$, $p < 0,001$). According to the Tukey HSD post-hoc test results, the LSTM and Siamese Networks methods demonstrated significantly higher performance than others.

As presented in Tab. 6, the performance of various dance movement evaluation methods is compared using Mean Squared Error (*MSE*), Mean Absolute Error (*MAE*), and Pearson Correlation Coefficient (*PCC*) metrics. LSTM method, with the lowest *MSE* (56,11) and *MAE* (5,2), most accurately approximates expert ratings. This finding suggests that LSTM predictions are the most consistent with expert evaluations, exhibiting the least error. Additionally, the high correlation coefficients of LSTM ($PCC = 0,85$) and Siamese Networks ($PCC = 0,82$) indicate that these methods produce results significantly aligned with expert assessments. Conversely, the Euclidean Distance method, which has the highest *MSE* (432,68), deviates the most from expert ratings, demonstrating the lowest accuracy. Furthermore, the low correlation coefficient of the DTW method ($PCC = 0,58$) suggests that it produces results less aligned with expert ratings.

When the model averages are compared with the expert ratings, it can be concluded that the model averages effectively approximate the expert ratings. Although significant similarities exist in the performance of the Siamese models, it is evident that their success rates are relatively low. This suggests that modifying the overall structure of the network, rather than changing individual parameters, may yield better results. Analyzing the LSTM Models results further reveals that models with 64 LSTM neurons generally have higher MSE values compared to those with 32 neurons. Increasing the number of LSTM neurons does not always improve performance, while reducing the number of dense layer neurons often results in higher MSE values. Similarly, increasing the number of epochs (as in the LSTM_12 model) does not consistently enhance performance. Models using the 'sgd' optimizer, however, generally exhibit lower MSE values.

As a result, LSTM is the most successful method because it can correctly analyze the temporal change and contextual relationships of movements. Although the Siamese Network, which has similar features, has a high correlation, it is predicted that the success rate will increase with the changes to be made in the general structure of the network. TLCC, DTW and Cosine Similarity performed at a medium level because although they correctly analyzed the movements in certain directions, they could not capture the changes over time well enough. Euclidean Distance is the least successful method because it focuses only on the simple distance between two points and ignores the dynamic structure of the movement.

5 CONCLUSION

The aim of this study was to develop a deep learning-based system capable of evaluating Turkish folk dance performances. The proposed system analyzes the movements of student dancers using pose estimation methods, compares their performance to that of an instructor, and provides a numerical score. This approach was designed to provide an objective and automated evaluation of dance performances without the need for an instructor. In this study, pose data extracted using Mediapipe and YOLOv8 were analyzed with various methods to assess dance performances.

Among the evaluation methods used in the study, Euclidean distance fell short of expectations due to the dynamic nature of dance and its sensitivity to outliers. However, given its simplicity and computational efficiency (making it a common choice in methods like DTW and TLCC), its performance could be improved in future studies by employing alternative transformation and normalization techniques or by incorporating bone or planar distances.

In this study, similarity calculations were performed using unsupervised learning with LSTM and Siamese Networks. Future research could incorporate expert scores directly into model training to train more tailored and accurate models. Beyond the individual success of methods, combining multiple models could improve accuracy and robustness in addressing complex problems like dance evaluation. Researchers may implement similar applications tailored to different dance styles, expanding the applicability and practical utility of such systems.

6 REFERENCES

[1] Georgios, L. (2017). The transformation of traditional dance from its first to its second existence: The effectiveness of music-movement education and creative dance in the preservation of our cultural heritage. *Journal of Education and Training Studies, 6*(1), 104. https://doi.org/10.11114/jets.v6i1.2879

[2] Kim, Y. & Kim, D. (2018). Real-time dance evaluation by markerless human pose estimation. *Multimedia Tools and Applications, 77*(23), 31199-31220. https://doi.org/10.1007/s11042-018-6068-4

[3] Lee, S. & Kim, M. (2020). Analysis of inter-rater agreement of Latin American and modern dance sport. *Korean Journal of Sport Science, 31*(4), 830-839. https://doi.org/10.24985/kjss.2020.31.4.830

[4] Diehl, K. (2016). The mirror and ballet training: Do you know how much the mirror's presence is really affecting you? *Journal of Dance Education, 16*(2), 67-70. https://doi.org/10.1080/15290824.2015.1110854

[5] Trajkova, M. & Cafaro, F. (2018). Takes Tutu to ballet: Designing visual and verbal feedback for augmented mirrors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2*(1), 1-30. https://doi.org/10.1145/3191770

[6] Alexiadis, D. S., Kelly, P., Daras, P., O'Connor, N. E., Boubekeur, T., & Moussa, M. B. (2011). Evaluating a dancer's performance using Kinect-based skeleton tracking. *Proceedings of the 19th ACM International Conference on Multimedia*, 659-662. Scottsdale, Arizona, USA: ACM. https://doi.org/10.1145/2072298.2072412

[7] Laraba, S. & Tilmanne, J. (2016). Dance performance evaluation using hidden Markov models. *Computer Animation and Virtual Worlds, 27*(3-4), 321-329. https://doi.org/10.1002/cav.1715

[8] Haberkamp, L. D., Garcia, M. C., & Bazett-Jones, D. M. (2022). Validity of an artificial intelligence, human pose estimation model for measuring single-leg squat kinematics. *Journal of Biomechanics, 144*, 111333. https://doi.org/10.1016/j.jbiomech.2022.111333

[9] Lee, J. J., Choi, J. H., Chuluunsaikhan, T., & Nasridinov, A. (2020). Pose evaluation for dance learning application using joint position and angular similarity. *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, 67-70. Virtual Event, Mexico: ACM. https://doi.org/10.1145/3410530.3414402

[10] Zhou, Z., Xu, A., & Yatani, K. (2021). SyncUp: Vision-based practice support for synchronized dancing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5*(3), 1-25. https://doi.org/10.1145/3478120

[11] Tang, H., Luo, Y., & Yang, J. (2022). Research on dance movement evaluation method based on deep learning posture estimation. *2022 2nd International Conference on Big Data Engineering and Education (BDEE)*, 173-177. Chengdu, China: IEEE. https://doi.org/10.1109/BDEE55929.2022.00036

[12] Kang, J., Chaewon, K., Jeewoo, Y., Houggeun, J., Taihu, L., Hyunmi, M., Minsam, K., & Jinyoung, H. (2023). Dancing on the inside: A qualitative study on online dance learning with teacher-AI cooperation. *Education and Information Technologies, 28*(9), 12111-12141. https://doi.org/10.1007/s10639-023-11649-0

[13] Sheng, B., Chen, X., Zhang, X., Tao, J., & Qu, H. (2023). Research on dance movement recognition and assessment through human pose estimation. *2023 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 1-6. Koh Samui, Thailand: IEEE. https://doi.org/10.1109/ROBIO58561.2023.10354805

[14] Ding, Y. & Wang, S. (2023). AIDanceFriend: An intelligent mobile application to automate the dance rating using artificial intelligence and computer vision. In *Computer Science, Engineering and Applications*, 41-50. Academy and Industry Research Collaboration Center (AIRCC). https://doi.org/10.5121/csit.2023.130505

[15] Xu, S. (2023). Development of online dance teaching system based on computer technology. Z. Zhan, B. Zou, & W. Yeoh (Eds.). *Proceedings of the 2022 3rd International Conference on Big Data and Informatization Education*, 692-699. Dordrecht: Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-034-3_71

[16] Mehta, D., Srinath, S., Oleksandr, S., Helge, R., Mohammad, S., Hans-Peter, S., Weipeng, X., Dan, C., & Christian, T. (2017). VNect: Real-time 3D human pose estimation with a single RGB camera. *arXiv.* https://doi.org/10.1145/3072959.3073596

[17] Pavllo, D., Feichtenhofer, C., Grangier, D., & Auli, M. (2018). 3D human pose estimation in video with temporal convolutions and semi-supervised training. *arXiv.* https://doi.org/10.1109/CVPR.2019.00794

[18] Ben Gamra, M. & Akhloufi, M. A. (2021). A review of deep learning techniques for 2D and 3D human pose estimation. *Image and Vision Computing, 114*, 104282. https://doi.org/10.1016/j.imavis.2021.104282

[19] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device real-time body pose tracking. *arXiv.*

[20] Ultralytics. (2024). *Pose estimation*. Retrieved June 15, 2024.

[21] Chen, H., Feng, R., Wu, S., Xu, H., Zhou, F., & Liu, Z. (2023). 2D human pose estimation: A survey. *Multimedia Systems, 29*(5), 3115-3138. https://doi.org/10.1007/s00530-022-01019-0

[22] Ahn, H., Sun, K., & Kim, K. P. (2022). Comparison of missing data imputation methods in time series forecasting. *Computers, Materials & Continua, 70*(1), 767-779. https://doi.org/10.32604/cmc.2022.019369

[23] Labinghisa, B., & Lee, D. M. (2021). A pose estimation scheme based on distance scaling algorithm in real-time environment. *Multimedia Tools and Applications, 80*(26-27), 34359-34367. https://doi.org/10.1007/s11042-021-11027-3

[24] Kato, M., Goncharenko, I., & Gu, Y. (2023). Investigation of posture similarity metrics for online dance learning support. *2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, 229-230. PingTung, Taiwan: IEEE. https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226933

[25] Srikaewsiew, T., Khianchainat, K., Tharatipyakul, A., Pongnumkul, S., & Kanjanawattana, S. (2022). A comparison of the instructor-trainee dance dataset using cosine similarity, Euclidean distance, and angular difference. *2022 26th International Computer Science and Engineering Conference (ICSEC)*, 235-240. Sakon Nakhon, Thailand: IEEE. https://doi.org/10.1109/ICSEC56337.2022.10049368

[26] Uğuz, S. (2019). *Makine öğrenmesi teorik yönleri ve Python uygulamaları ile bir yapay zekâ ekolü* (2nd ed.). Ankara: Nobel Yayıncılık.

[27] Brockwell, P. J. & Davis, R. A. (2016). *Introduction to time series and forecasting* (3rd ed.). *Springer texts in statistics.* Switzerland: Springer.

[28] Uguz, S. & Buyukgokoglan, E. (2022). A hybrid CNN-LSTM model for traffic accident frequency forecasting during the tourist season. *Tehnički Vjesnik, 29*(6). https://doi.org/10.17559/TV-20220225141756

[29] Olah, C. (2015). Understanding LSTM networks. Retrieved May 19, 2024. http://colah.github.io/posts/2015-08-Understanding-LSTMs

[30] Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a "siamese" time delay neural network. *Advances in Neural Information Processing Systems, 6*, 737-744.

[31] Tao, B., Huang, L., Zhao, H., Li, G., & Tong, X. (2021). A time sequence images matching method based on the Siamese network. *Sensors, 21*(17), 5900. https://doi.org/10.3390/s21175900

[32] Lugaresi, C., et al. (2019). MediaPipe: A framework for building perception pipelines. *arXiv*. Retrieved May 18, 2024.

Contact information:

Erdem BÜYÜKGÖKOĞLAN
Isparta University of Applied Sciences,
Department of Computer Engineering,
Isparta, Turkey
E-mail: erdembgo@gmail.com

Sinan UĞUZ, Associate Professor
(Corresponding author)
Isparta University of Applied Sciences,
Department of Computer Engineering,
Isparta, Turkey
E-mail: sinanuguz@isparta.edu.tr