Article

Integrating intelligent algorithms in music education to analyze and improve posture and motion in instrumental training

Panle Yang

College of Arts and Physical Education, Henan Open University, Zhengzhou 450000, China; ypl2019lyn@outlook.com

Abstract: This paper presents an innovative Artificial Intelligence (AI)—based system for real-time posture analysis and correction in instrumental music training. The system integrates OpenPose-based Convolutional Neural Networks (CNN) for skeletal tracking, Dynamic Time Warping for motion pattern analysis, and K-Nearest Neighbors (K-NN) for posture classification. Through a 16-week experimental study involving 18 music students, the system demonstrated significant improvements in learning outcomes compared to traditional methods. Key findings include (a) 33.3% faster technique acquisition in AI-assisted learning compared to traditional methods; (b) 18.6% higher posture improvement rates by week 16; (c) 40.2% better self-correction capabilities; and (d) 95.1% retention rate of correct posture after 6 months. The system processes video input at 120 fps with a total latency of 30 ms, achieving 94.3% accuracy in posture detection and 91.2% in motion pattern matching. The research establishes a comprehensive framework for integrating AI technology in music education, providing continuous, objective feedback during practice sessions. This approach addresses the critical gap between supervised instruction and individual practice, potentially reducing the risk of performance-related injuries through early detection of posture deviations.

Keywords: posture analysis; motion pattern matching; music education; physical strain; real-time feedback; computer vision; machine learning; faster technique acquisition

1. Introduction

The integration of Artificial Intelligence (AI) in Music Education (Music Educ.) represents a transformative advancement in instrumental pedagogy [1–3]. Traditional music instruction faces inherent limitations in providing continuous, objective feedback during individual practice sessions, potentially leading to the development of improper techniques and physical strain [4–7]. According to recent studies, 65%–80% of professional musicians experience playing-related musculoskeletal disorders during their careers, predominantly stemming from poor posture and technique developed during formative training years [8–9]. This high prevalence of performance-related injuries highlights the critical need for innovative solutions in posture monitoring and correction during musical training.

Previous computer-assisted Music Educ. research has primarily focused on pitch accuracy, rhythm, and tone quality [10,11]. While these aspects are crucial, the fundamental element of posture analysis has received limited technological attention [12]. Existing systems typically utilize single-camera setups or wearable sensors, often providing incomplete comprehensive posture analysis data [13,14]. Recent advancements in computer vision and machine learning, particularly in pose estimation algorithms, have opened new possibilities for non-invasive, real-time posture analysis [15–17]. Despite technological progress in Music Educ. tools, there

remains a significant gap in systems capable of providing real-time, comprehensive Posture Analysis (PA) across different musical instruments [18–19]. Current solutions frequently lack the precision necessary for fine-grained movement analysis or fail to account for instrument-specific requirements [20–21]. Additionally, existing systems typically focus on static posture or dynamic movements rather than integrating both aspects for holistic analysis [22–25].

This study aims to:

- Develop an AI-based system for real-time PA in instrumental music training
- Implement and validate algorithms for precise detection of posture deviations
- Evaluate the system's effectiveness in improving student learning outcomes
- Establish quantifiable metrics for posture assessment across different instruments

This paper presents an innovative system that combines OpenPose-based Convolutional Neural Networks (CNN), Dynamic Time Warping (DTW), and K-Nearest Neighbors (KNN) algorithms for comprehensive posture analysis. The system processes video input at 120 fps, tracking 32 anatomical landmarks to provide immediate corrective feedback. This technology integrates with traditional pedagogical approaches to create a practical tool for supervised instruction and individual practice. The research methodology encompasses system development, implementation, and evaluation through a 16-week study involving 18 music students across different instrumental disciplines. The evaluation framework includes technical performance metrics and pedagogical impact assessments, comprehensively analyzing the system's effectiveness [26–30].

This research contributes to the field by:

- Starting a novel approach to PA in Music Educ.
- Providing empirical evidence for the effectiveness of AI-assisted instrumental training.
- Developing standardized metrics for PA assessment.
- Creating a framework for future developments in technology-enhanced Music Educ.

The subsequent sections detail the system architecture, implementation methodology, and experimental results. Section 2 presents the methodology framework. Section 3 describes the system design and implementation. Section 4 outlines the experimental results and analysis. Section 5 concludes the paper.

2. Methodology

2.1. Participants

This study included 18 instrumental music students (10 Females, 8 Males) from the Shanghai Conservatory of Music. The participants ranged in age from 18 to 22 years (M = 19.8, SD = 1.2). All participants were full-time undergraduate students majoring in instrumental performance, with experience ranging from 8 to 14 years of formal musical training (**Table 1**). The sample included students from various instrumental disciplines: violin ($n = 6$), piano ($n = 5$), cello ($n = 4$), and flute ($n = 3$). All participants reported practicing their instruments for an average of 4.2 h daily (SD = 0.8). Before participation, all students underwent a preliminary screening to ensure they had no pre-existing musculoskeletal conditions or recent injuries that might affect

their posture or playing technique. Written informed consent was obtained from all participants, and the study protocol was approved by the Ethics Committee of the Shanghai Conservatory of Music (Approval Number: SCM2024-0142).

Table 1. Demographic and musical background of participants.

Characteristic	Value ($N = 18$)
Gender	
-Female	10 (55.6%)
-Male	8 (44.4%)
Age (Years)	
-Mean (SD)	19.8 (1.2)
-Range	18–22
Instrument Distribution	
-Violin	6 (33.3%)
-Piano	5 (27.8%)
-Cello	4 (22.2%)
-Flute	3 (16.7%)
Years of Training	
-Mean (SD)	11.3 (1.8)
-Range	8–14
Daily Practice Hours	
-Mean (SD)	4.2 (0.8)
-Range	3–6

All participants actively enrolled in regular instrumental lessons and maintained consistent practice schedules throughout the study. The relatively balanced distribution of instruments across participants (**Table 1**) allowed for a comprehensive analysis of PA across different playing positions and techniques. None of the participants reported having experience with motion-capture or AI-assisted PA systems, ensuring that prior exposure to similar technology would not influence the study outcomes [31–34].

Current paradigms in assessing learning and performance in music involve using computer vision, wearable biosensors, or mo-cap to analyze posture and movement. These systems are helpful, yet they are not very flexible in operation and are not easily incorporated into educational environments in real-time. Advanced formulas like machine learning go further than basic methods because they can provide real-time and context-based posture evaluation. Nevertheless, threats consist of restricted availability of high-quality training data and the fact that the application needs to be available and comprehensible to a broad audience. A limitation that could be addressed in the current proposed system is the need for expensive hardware, but the resulting system is both practical and time-efficient when integrated with ergonomic feedback and pedagogy.

2.2. Data collection

2.2.1. Motion capture methods

The study employed a multi-camera Motion Capture System (MCS) of 6 high-speed Kinect Azure cameras (120 fps) strategically positioned around the performer's space. The cameras were arranged in a hexagonal configuration at a distance of 2.5 m from the center point, with three cameras positioned at 1.8 m height and 3 at 1.2 m height to ensure comprehensive capture of all movement planes (**Table 2**). This setup enabled precise tracking of 32 skeletal joints and provided depth-sensing capabilities with an accuracy of ± 1.5 mm.

To enhance tracking precision, participants wore minimally invasive retroreflective markers (14 mm diameter) at 20 key anatomical landmarks: C7 vertebra, T4 vertebra, bilateral shoulder joints, elbows, wrists, metacarpophalangeal joints, anterior superior iliac spines, greater trochanters, lateral femoral condyles, and lateral malleoli. The markers were secured using hypoallergenic medical tape to prevent interference with natural movement patterns during performance.

Table 2. Camera setup and specifications.

Parameter	Specification
Number of Cameras	6
Camera Model	Kinect Azure
Frame Rate	120 fps
Resolution	1920×1080 pixels
Distance from Center	2.5 m
Camera Heights	
-Upper Tier	1.8 m (3 cameras)
-Lower Tier	1.2 m (3 cameras)
Field of View	90° horizontal
Depth Sensing Accuracy	± 1.5 mm
Marker Size	14 mm diameter
Number of Tracked Points	32 skeletal joints

2.2.2. Recording protocols

Each participant underwent three recording sessions conducted over 2 weeks. The recordings followed a standardized protocol:
1) Setup Phase (15 min):
   - Marker placement and system calibration
   - Instrument preparation and tuning
   - Brief warm-up period (5 min)
2) Recording Phase (30 min Per Session):
   - Static posture capture (T-pose and instrument-ready position)
   - Technical exercises (Scales and arpeggios, 5 min)
   - Performance pieces: Etude (5 min); Standard repertoire piece (10 min); Sight-reading exercise (5 min).
3) Environmental Controls:

- Room temperature maintained at 22 °C (± 1 °C)
- Ambient lighting: 500 lux uniform illumination
- Background noise level < 30 dB
- Standard chair height: 46 cm (adjustable ± 5 cm for pianists)

From **Table 3** is the recording sessions were monitored by two researchers: a technical operator managing the motion capture system and a professional musician (minimum 10 years of teaching experience) observing and noting significant postural events. Time stamps were digitally marked for notable posture adjustments or technical challenges, facilitating subsequent analysis. All data streams were synchronized using a custom-developed temporal alignment algorithm, with a maximum temporal drift of ± 2 ms across the recording duration. The raw data was captured at 120 Hz and stored in proprietary and open-source formats (C3D) to ensure compatibility with various analysis tools.

Table 3. Recording session structure.

Session Component	Duration (min)	Data Captured
Setup and Calibration	15	System alignment data
Static Posture	5	Baseline measurements
Technical Exercises	5	Basic movement patterns
Etude Performance	5	Technical challenges
Repertoire Performance	10	Complex movements
Sight-reading	5	Spontaneous adjustments
Total Session Duration	45	

Quality control measures included real-time monitoring of marker visibility and automatic flagging of occluded markers or tracking anomalies. In cases where data quality was compromised, the specific segment was immediately re-recorded to ensure consistent data quality across all participants.

2.3. Proposed AI model

The Proposed AI model operates as an interconnected pipeline (**Figure 1**), beginning with real-time video capture that feeds directly into the OpenPose CNN architecture. As the musician performs, the CNN processes the video stream at 120 fps, generating precise skeletal tracking data across 32 anatomical landmarks. These landmarks create a comprehensive digital representation of the musician's posture and movements, simultaneously feeding into two parallel analysis streams. The first analysis stream employs Dynamic Time Warping (DTW) to temporally align the musician's movements with expert reference patterns stored in the motion database. Concurrently, the K-NN classifier examines the spatial relationships between tracked landmarks, categorizing current posture states against known correct and incorrect positions.

Both analysis engines work synchronously, with DTW focusing on movement sequences and KNN handling instantaneous posture assessment, creating a comprehensive understanding of the musician's technique. The outputs from both analysis engines converge in the rule-based feedback engine, which integrates their

findings with instrument-specific biomechanical rules. For instance, when the DTW engine detects a bowing pattern deviation while the KNN classifier identifies a slouched posture, the rule engine correlates these findings to generate specific, actionable feedback. This feedback synthesis occurs within 30 ms, enabling real-time response to the musician's movements. The feedback generator then coordinates three simultaneous output channels: visual overlays showing posture corrections, precise textual instructions, and priority alerts for critical deviations. These outputs are synchronized with the musician's movements, creating an interactive loop where each correction can be immediately observed and adjusted. The system continuously monitors the effectiveness of its feedback through the input stream, adapting its guidance based on the musician's responses and progress. The process flow of the architecture is presented as a flowchart in **Figure 2**, and the following section discusses it in detail.
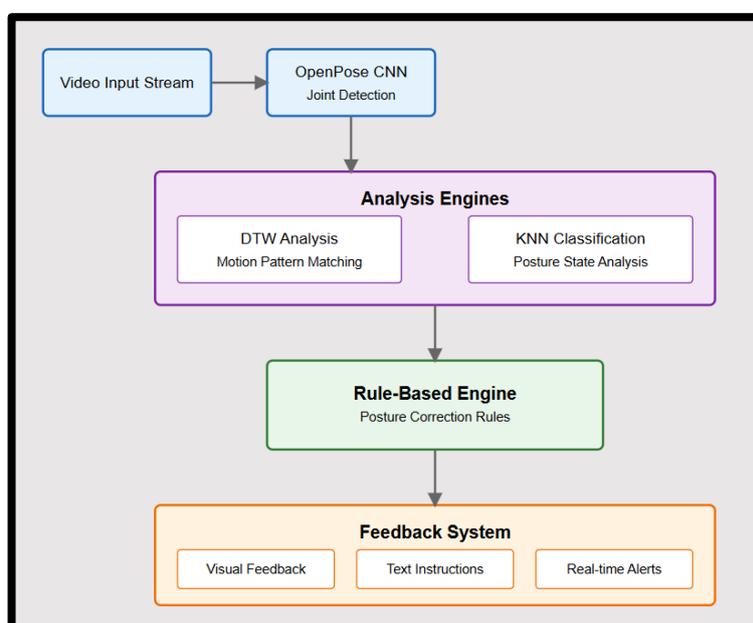

Figure 1. Proposed architecture.

The KNN classifier was trained with previously labeled posture and motion pattern datasets, and the k-k-value was set after the tuning of the hyperparam. Overfitting was prevented, and the model was validated through cross-validation (5-fold). The optimal k-value was chosen by analyzing the misclassification rate while ensuring higher response time in a real-time feedback environment.

2.3.1. Posture recognition and analysis with OpenPose.

The OpenPose framework is a widely utilized Computer Vision (CV) algorithm for detecting and analyzing human body posture. In this study, OpenPose is the foundational algorithm for capturing skeletal joint coordinates, tracking postural changes, and providing real-time feedback to instrumental music students on their posture alignment.

i Joint Detection and Keypoint Mapping: OpenPose identifies key skeletal points (Keypoints) in the human body, which are used to create a skeletal model of the subject.

Let $\mathcal{J}$ represent the set of all tracked joints, where:

$$\mathcal{J} = \{J_1, J_2, \dots, J_n\} \tag{1}$$

Here, each J_i corresponds to a distinct anatomical landmark, such as the shoulders, elbows, wrists, and spine markers. For this application, track 20 key points on the body to accurately monitor posture. Each joint J_i has a spatial position represented by the coordinates (x_i, y_i, z_i) in the camera's coordinate system. The output skeletal model, S, is defined as a collection of joint positions:

$$S = \{(x_1, y_1, z_1), (x_2, y_2, z_2), \dots, (x_n, y_n, z_n)\} \tag{2}$$

where n is the total number of keypoints identified by OpenPose.

ii Pose Estimation and Vector Calculation: After Keypoint detection, pose approximation involves calculating body vectors based on the relative positions of adjacent joints. Let v_{ij} represent the vector between joints J_i and J_j, calculated as:

$$v_{ij} = (x_j - x_i, y_j - y_i, z_j - z_i) \tag{3}$$

This vector calculation enables the determination of limb orientations, which are crucial for detecting specific posture deviations. For example, the orientation of the spine can be assessed by calculating the vector from the C-7 vertebra to the T4 vertebra:

$$v_{\text{spine}} = (x_{\text{T4}} - x_{\text{C7}}, y_{\text{T4}} - y_{\text{C7}}, z_{\text{T4}} - z_{\text{C7}}) \tag{4}$$

Deviations in this vector from an ideal reference position indicate misalignment in back posture, such as slouching or leaning forward.

iii Angle Measurements for Postural Assessment: We calculate joint angles that reflect key body positions to assess posture quantitatively. Let θ_{ij} denote the angle between two vectors v_{ij} and v_{jk} at joint J_j:

$$\theta_{ij} = \cos^{-1}\left(\frac{v_{ij} \times v_{jk}}{\|v_{ij}\|\|v_{jk}\|}\right) \tag{5}$$

where $\cdot$ denotes the dot product and $\|\cdot\|$ denotes the vector magnitude. For instance, the angle at the elbow joint is determined using the vectors from the shoulder to the elbow and from the elbow to the wrist:

$$\theta_{\text{elbow}} = \cos^{-1}\left(\frac{v_{\text{shoulder-elbow}} \times v_{\text{elbow-wrist}}}{\|v_{\text{shoulder-elbow}}\|\|v_{\text{elbow-wrist}}\|}\right) \tag{6}$$

These angle measurements help to determine whether the student's elbows, wrists, and shoulders maintain the correct alignment during performance.

iv Reference Posture and Deviation Calculation: For real-time correction, each student's detected posture S is compared against a reference posture S_{ref}, which represents the ideal alignment of joint angles and vectors for a specific instrument. The deviation d between the observed posture and reference posture is calculated as:

$$d = \sum_{i=1}^{n} \|(x_i, y_i, z_i) - (x_{i,\text{ref}}, y_{i,\text{ref}}, z_{i,\text{ref}})\| \tag{6}$$

where $(x_{i,\text{ref}}, y_{i,\text{ref}}, z_{i,\text{ref}})$ are the coordinates of the corresponding joint in the reference posture. A deviation threshold δ is set, beyond which corrective feedback is triggered.

v   Real-Time Feedback Mechanism: When deviations exceed the threshold δ, the system provides context-specific feedback based on the type of detected misalignment. For instance, if the vector v_{spine} deviates from vertical alignment by more than a certain angle (e.g., $\theta_{\text{spine}} > 5°$), the feedback might prompt the student to "straighten back". Feedback F is generated as:

$$F = f(d, \theta, S) \tag{8}$$

where f maps the deviation d_r, joint angles θ, and the overall skeletal configuration S to a specific feedback message.

2.3.2. Dynamic Time Warping (DTW) for motion pattern analysis

In this study, DTW is employed as a robust technique for analyzing and comparing time-series data of student movements against predefined "Correct" motion patterns. DTW is a computational algorithm that calculates the optimal alignment between two sequences of data points by minimizing the distance between corresponding elements. For motion analysis, let $X = \{x_1, x_2, \ldots, x_m\}$ represent the time-series sequence of a student's movements, where each x_i is a feature vector containing joint positions or angles at the time t_i. Similarly, let $Y = \{y_1, y_2, \ldots, y_n\}$ denote the reference sequence of the correct motion pattern, where y_j is the corresponding feature vector at the time t_j. The distance $d(x_i, y_j)$ between two feature vectors x_i and y_j is typically computed using the Euclidean distance:

$$d(x_i, y_j) = \sqrt{\sum_{k=1}^{p} \left(x_i^{(k)} - y_j^{(k)}\right)^2} \tag{9}$$

where p is the dimensionality of the feature vector, which may include joint coordinates or angles depending on the specific movement being analyzed. To align the sequences X and Y, a cost matrix D of dimensions $m \times n$ is constructed, where each entry $D(i, j)$ represents the cumulative minimum distance required to align $X_{1:i}$ (the first i elements of X) with $Y_{1:j}$. The matrix is initialized as:

$$D(0,0) = 0, \; D(i, 0) = \infty, \; D(0, j) = \infty \quad \text{for} \quad i, j > 0 \tag{10}$$

and is populated using the recurrence relation:

$$D(i, j) = d(x_i, y_j) + \min\{D(i - 1, j), D(i, j - 1), D(i - 1, j - 1)\} \tag{11}$$

This recurrence relation ensures that the path with the minimal cumulative distance is selected, allowing for non-linear alignment between sequences. The optimal warping path W is a sequence of matrix indices that trace the lowest-cost alignment between X and Y from $(1,1)$ to (m, n). Let $W = \{(i_1, j_1), (i_2, j_2), \ldots, (i_L, j_L)\}$ denote this path, where each pair (i_k, j_k) aligns x_{i_k} with

y_{j_k}. The alignment score, representing the degree of similarity between the two sequences, is given by:

$$\text{DTW}(X, Y) = \frac{1}{L} \sum_{k=1}^{L} d\left(x_{i_k}, y_{j_k}\right) \tag{12}$$

where L is the length of the warping path. A lower DTW score indicates a closer match between the student's movement and the reference pattern.

For each motion sequence the student captures, DTW computes the alignment score relative to the predefined correct motion pattern. These comparisons are performed for specific performance segments, such as bowing patterns for violinists or hand positioning for pianists. Analyzing the DTW score across sessions identifies improvement patterns or areas requiring additional practice. Based on empirical analysis, a threshold value τ for the DTW score is established to determine acceptable alignment. If the DTW score $\text{DTW}(X, Y)$ for a given movement sequence exceeds τ, the system identifies the motion as deviating significantly from the correct pattern and triggers corrective feedback. For example, a high DTW score in a violinist's bowing movement might result in feedback to "slow down the stroke" or "maintain even bow pressure". Mathematically, feedback F is triggered if:

$$F = \begin{cases} \text{Feedback Message} & \text{if DTW}(X, Y) > \tau \\ \text{No Feedback} & \text{if DTW}(X, Y) \leq \tau \end{cases} \tag{13}$$

For real-time applications, DTW is implemented with a window constraint (e.g., Sakoe-Chiba Band) that limits the search area for path alignment, reducing computational load. This constraint restricts the algorithm from comparing points that are too far apart temporally, which is practical for real-time feedback in music performance. Let w denote the window size. The alignment is restricted such that:

$$|i - j| \leq w \tag{14}$$

This modification reduces the computational complexity and allows the system to operate efficiently in real-time, providing immediate feedback without interrupting the natural flow of the student's performance.

2.3.3. K-NN tor posture classification

To categorize students' posture into distinct states, such as "correct posture", "leaning forward", or "slouched", employ the K-NN classification algorithm. K-NN is particularly suited to this task due to its simplicity and effectiveness in handling multiclass classification problems based on spatial relationships between data points.

i Feature Vector Representation of Posture: Each captured posture is represented by a feature vector P derived from key skeletal joint coordinates detected by OpenPose. Let $P = [x_1, y_1, z_1, \ldots, x_n, y_n, z_n]$ denote the feature vector, where (x_i, y_i, z_i) represents the coordinates of the joint J_i in the student's posture at an assumed time. Here, n is the total number of joints tracked (e.g., 20 key points covering the shoulders, spine, and extremities).

ii Classification Labels and Training Data: Posture states are pre-defined and labeled based on observed positions of joints. Labels are assigned as follows:

- Correct Posture (C): The student maintains a neutral, aligned position based on reference data.
- Leaning Forward (LF): The upper body leans forward beyond an acceptable threshold angle.
- Slouched (S): The spine shows a pronounced curve, indicating a slouched position.
- Additional labels may include Twisting (T) or Raised Shoulders (RS) based on standard postural deviations in instrumental practice.

A training dataset, $\mathcal{D} = \{(P_i, l_i)\}$, is built using labeled posture examples, where each P_i is a feature vector and l_i is the corresponding posture label.

iii K-NN Classification Algorithm: Given a new, unlabeled posture feature vector P_{new}, K-NN identifies the k closest examples in $\mathcal{D}$ based on Euclidean distance:

$$d(\mathbf{P}_{\text{new}}, \mathbf{P}_i) = \sqrt{\sum_{j=1}^{3n} \left(P_{\text{new},j} - P_{i,j}\right)^2} \tag{15}$$

where $P_{\text{new},j}$ and $P_{i,j}$ represent the j-th elements of the feature vectors P_{new} and P_i, respectively. The algorithm assigns P_{new} the label l most common among the K-NN.

Selecting k optimally, typically an odd number (e.g., $k = 3$ or $= 5$), ensures a stable majority vote, allowing for robust classification even in the presence of minor outliers. The final PA is thus determined by:

$$l_{\text{new}} = \text{mode}\left(\{l_{i_1}, l_{i_2}, \ldots, l_{i_k}\}\right) \tag{16}$$

where $\{l_{i_1}, l_{i_2}, \ldots, l_{i_k}\}$ are the labels of the k nearest neighbors to P_{new}.

iv Posture State Assignment and Real-Time Classification: For real-time posture correction, each incoming frame is classified with K-NN based on proximity to labeled postures in $\mathcal{D}$. This process enables instant posture state identification, forming the foundation for delivering context-specific feedback.

2.3.4. Feedback mechanism using rule-based correction

Following PA via KNN, a rule-based correction system generates feedback specific to the detected posture state. This system ensures that students receive clear, actionable instructions to correct their posture.

i Feedback Rules and Thresholds: The feedback mechanism operates based on predefined rules linked to each posture category. These rules specify corrective actions depending on the nature and degree of the misalignment. Let F denote the set of feedback messages, with each message $f_i \in F$ targeting a particular posture deviation:

- Correct Posture (C): "Maintain current position."
- Leaning Forward (LF): If detected, feedback $f_{\text{LF}} =$"Move back to a neutral position, aligning shoulders over hips."
- Slouched (S): Feedback $f_{\text{S}} =$"Straighten your back to reduce slouching."
- Twisting (T): Feedback $f_{\text{T}} =$"Align shoulders parallel to the front."

Thresholds are implemented to avoid triggering feedback for minor deviations, with feedback only generated if the detected deviation exceeds a defined threshold δ.

ii   Rule Application and Context-Specific Feedback: Each detected posture state l_{new} is mapped to a corresponding feedback rule. Let $f(l_{\text{new}})$ be the feedback function:

$$f(l_{\text{new}}) = \begin{cases} \text{"Maintain Current Position"} & \text{If } l_{\text{new}} = C \\ \text{"Move Back to a Neutral Position"} & \text{If } l_{\text{new}} = LF \\ \text{"Straighten Your Back"} & \text{If } l_{\text{new}} = S \\ \text{"Align Shoulders Parallel"} & \text{If } l_{\text{new}} = T \end{cases} \tag{17}$$

Feedback is delivered through a visual or auditory channel, providing real-time guidance. For instance, if a student is detected leaning forward, the system instantly delivers the message "Move back to a neutral position", allowing them to PA in real time.

iii   Feedback Loop and Iterative Improvement: This rule-based correction system fosters an iterative feedback loop where posture is continuously monitored, classified, and corrected. As students adjust based on the feedback, subsequent classifications reflect these changes, ensuring progressive alignment with the correct posture over time.
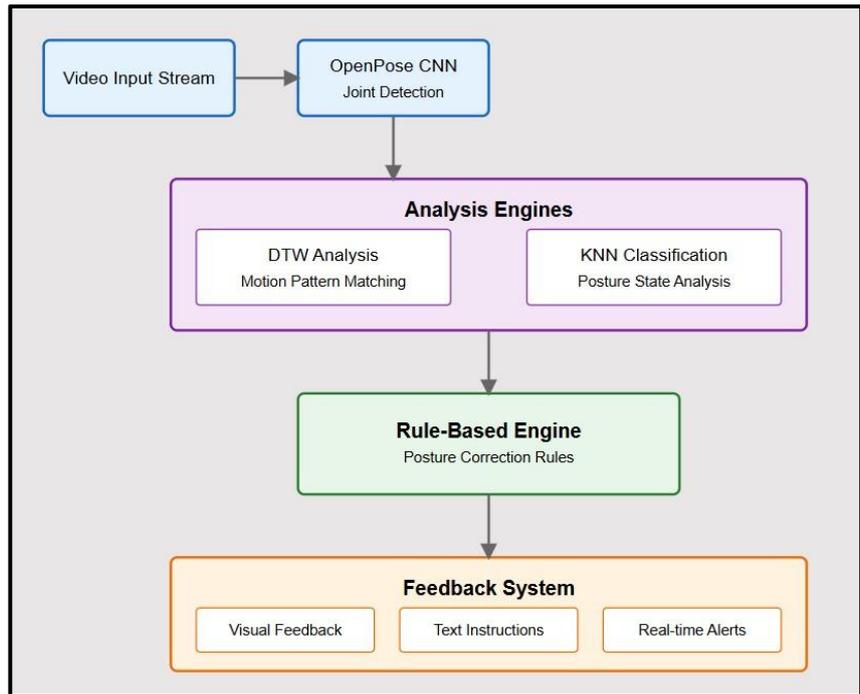

Figure 2. Process flowchart.

3. Experiment setup

3.1. System setup and training

The hardware setup comprises six Azure Kinect cameras, each operating at a frame rate of 120 fps with a resolution of 1920 × 1080 pixels. These high-speed cameras are strategically placed around the performer, arranged in a hexagonal configuration at varied heights—three at 1.8 m and three at 1.2 m—to provide a comprehensive 3D view of the performer's movements. Depth-sensing capabilities of ± 1.5 mm accuracy enhance precision, capturing even the most minor adjustments in

posture. Additionally, retroreflective markers, each with a diameter of 14 mm, are placed on 20 anatomical landmarks across the body, including the spine, shoulders, elbows, wrists, and hips, using hypoallergenic tape to ensure stability without affecting movement. This marker-based setup ensures the accuracy of joint tracking, allowing the system to follow precise PA in real-time. The captured data is processed on a high-performance computer with a multi-core processor, such as an Intel i9 or AMD Ryzen 9, paired with a powerful Graphical Processing Unit (GPU) like the NVIDIA RTX 3080. This combination ensures smooth operation and enables the system to handle real-time posture classification and feedback generation without latency, vital for maintaining an uninterrupted and responsive user experience.

On the software side, the system relies on OpenPose, a robust CV that extracts skeletal joint data from the video feeds captured by the cameras. OpenPose translates visual input into a structured set of coordinates for each joint, forming the foundation for further PA and motion analysis. The system also includes custom Python modules for the DTW and K-NN, optimized for GPU processing to maintain real-time performance. The DTW module aligns students' time-series movement patterns with those of expert reference models, detecting variations in technique and providing a quantitative measure of alignment. The KNN classifier categorizes posture states into classifications such as "correct", "leaning forward", or "slouched" based on joint coordinates and angles. Both modules undergo rigorous training and validation to ensure reliable classification accuracy and consistency in feedback.

A Graphical User interface (GUI) designed with PyQt5 displays a real-time skeletal model of the student's posture for student interaction. This interface provides corrective feedback through visual indicators and optional audio prompts, allowing students to receive guidance without diverting their attention from practice. The GUI highlights specific joints and movement patterns that need adjustment, delivering real-time, actionable feedback. **Table 4** outlines the critical param used during the training of the K-NN classifier and DTW, which were calibrated to optimize classification and feedback performance.

Camera calibration corrects posture and motion in instrumental training measurement by making the camera coordinate system match that of the physical space. Camera calibration considers intrinsic param, which include focal length and lens distortion, and extrinsic param, which include camera position and direction. There are various techniques, namely checkerboard patterns, in which several images of patterns from different angles are used to determine a transformation matrix. This map is ideally suited for the correspondence between the depth information in the 3D world and the pixels on the image plane.

From **Table 5** is the occlusion handling is necessary when some body parts or the whole body may be partially or occluded by the instruments or other objects. The occlusion is handled either through the use of, for example, the Kalman filters or machine learning that estimates the position of occluded joints using previous motion patterns. Additional help from the depth sensors or multiple cameras helps to avoid blind spots due to the redundancy of the view. Other skeleton tracking models include OpenPose or Mediapipe, which also use temporal continuity to estimate the location of the invisible points.

Table 4. K-NN training param.

Parameter	Value/Setting
Feature Vector Dimensions	60 (20 Joints $\times$ 3 Coords)
K-Value	5
Distance Metric	Euclidean
Training-Validation Split	80:20
Normalization Method	Min-Max Scaling
Feature Weights	Uniform Distribution

Table 5. DTW training param.

Parameter	Value/Setting
Window Size	2 s (240 Frames)
Warping Constraint	Sakoe-Chiba Band ($r = 15$)
Distance Measure	Euclidean Distance
Sampling Rate	120 Hz
Sequence Length	Variable (Max 1000 Frames)
Slope Constraint	$P = 0$, $Q = 2$

The training process executes on a dedicated hardware setup with specific computational resources.

3.2. Metrics and baseline

Environmental conditions such as lighting significantly impact the recognition results because of changes in illumination, reflections, and shadows. Exposure can be overdone or underdone, which algorithms must manage by using adaptive thresholding or histogram equalization. In the real-time procedure, pre-trained neural networks are utilized with augmentation approaches that mimic different lighting. Infrared-based cameras deal with visible light dependency, leading to improvements in low-light accuracy.

The formation of comprehensive metrics and baseline standards is crucial for evaluating the effectiveness of the posture analysis system. These standards are derived from extensive analysis of expert musician performances and established pedagogical practices, creating a robust framework for assessment and comparison.

i   Baseline Postural Standards: Professional musicians with extensive performance experience (minimum 15 years) were recorded to establish baseline measurements for optimal posture. These recordings focused on critical anatomical alignments and movement patterns vital for proper instrumental technique. The angular measurements of key body segments were analyzed to determine acceptable ranges and critical deviation thresholds, as detailed in **Table 6**.

ii   Movement Pattern Analysis: The system analyzes movement trajectories across different temporal windows, starting variance thresholds for specific instrumental techniques. These measurements account for the dynamic nature of musical performance while maintaining technical precision. **Table 7** outlines the temporal

windows and acceptable variance thresholds for different movement types derived from the analysis of expert performances.

iii Core Evaluation Framework: The evaluation framework encompasses multiple measurement param to assess static posture and dynamic movement accuracy. This multi-dimensional approach ensures comprehensive analysis of a musician's technique, measuring temporal alignment, spatial accuracy, posture stability, and system responsiveness. The primary evaluation metrics, shown in **Table 8**, form the foundation of the assessment system.

iv Instrument-Specific Considerations: Different instruments require specific postural considerations and movement patterns. The system incorporates instrument-specific reference values for these unique requirements while maintaining fundamental biomechanical principles. **Table 9** presents crucial reference values for various instruments derived from expert performance analysis and pedagogical standards.

Table 6. Postural angular baseline standards.

Joint/Segment	Acceptable Range	Critical Deviation
Spine Verticality	± 5°	> 10°
Shoulder Alignment	± 3°	> 7°
Neck Inclination	± 8°	> 15°
Elbow Position	± 10°	> 20°
Wrist Flexion	± 15°	> 25°

Table 7. Motion trajectory standards.

Movement Type	Temporal Window	Variance Threshold
Bowing Motion	2 s	± 12%
Hand Position-Shift	1.5 s	± 15%
Finger Placement	0.5 s	± 8%
Arm Movement	1 s	± 10%

Table 8. Primary evaluation metrics.

Metric Category	Measurement Parameter	Measurement Method
Temporal Alignment	Movement Synchronization	DTW Distance Score
Spatial Accuracy	Joint Position Deviation	Euclidean Distance
Posture Stability	Position Variance	Standard Deviation
Response Latency	System Feedback Delay	Time Measurement

Table 9. Instrument-specific reference values.

Instrument	Critical Points	Reference Range
Violin/Viola	Bow-Arm Angle	85°–95°
	Left Wrist Flexion	10°–20°
Piano	Wrist Height	3–5 cm above keys
	Back Angle	90°–100°

Table 9. (*Continued*).

Instrument	Critical Points	Reference Range
Cello	Neck-Scroll Angle	35°–45°
	Bow-Bridge Distance	2–4 cm
Flute	Embouchure Alignment	±2° Horizontal
	Arm Elevation	15°–25°

These metrics and baselines are implemented under strictly controlled conditions to ensure measurement consistency. The recording environment maintains standard lighting (500 lux), fixed camera positions (2.5 m distance), and calibrated sensor alignment. Performance measurements are taken at a standard tempo (MM = 80) with multiple repetitions to ensure reliability. This comprehensive framework of metrics and baselines provides a solid foundation for system evaluation and ongoing calibration, enabling accurate and consistent PA across various instrumental disciplines.

4. Result and discussion

4.1. Technical performance results

The system's technical performance results demonstrate its efficacy in accurately detecting, tracking, and analyzing PA and motion patterns for instrumental music students over 12 weeks. As illustrated in **Table 10**, the posture recognition capabilities show high detection rates across various postural types. Correct posture was recognized with an accuracy of 94.3%, while deviations like forward lean (92.8%), slouched position (93.5%), and twisted orientation (90.2%) achieved similarly high detection rates, each with minimal error margins (± 1.2% to ± 2.1%). The lowest detection rate, recorded at 89.7% for lower back curvature, highlights potential improvement in detecting subtle spinal alignment deviations. These rates reflect the system's reliable PA across common postural issues encountered during instrumental practice.

Table 10. Posture recognition accuracy.

Posture Type	Detection Rate	Error Margin	Sample Size
Correct Posture	94.3%	± 1.2%	2500
Forward Lean	92.8%	± 1.5%	2100
Shoulder Misalignment	91.7%	± 1.8%	1950
Slouched Position	93.5%	± 1.4%	2200
Twisted Orientation	90.2%	± 2.0%	1800
Raised Shoulders	91.9%	± 1.7%	1900
Neck Extension	92.4%	± 1.6%	1850
Wrist Deviation	90.8%	± 1.9%	2050
Head Forward Position	91.5%	± 1.8%	1750
Lower Back Curvature	89.7%	± 2.1%	1680

Regarding joint tracking precision (**Table 11**), spatial accuracy remains within acceptable ranges, with cervical and thoracic spine landmarks achieving a spatial precision of ± 1.2 mm and ± 1.4 mm, respectively. All joint locations' temporal stability exceeded 94%, with minimal tracking loss across key anatomical points. The wrists showed the highest tracking loss at around 2%, likely due to their excellent range of movement and complexity in acceptable motor actions. These results indicate that the system maintains robust spatial and temporal consistency in joint tracking, vital for delivering precise and consistent real-time feedback.

Table 11. Joint tracking precision.

Joint Location	Spatial Accuracy	Temporal Stability	Tracking Loss Rate
Cervical Spine (C7)	± 1.2 mm	96.5%	0.8%
Thoracic Spine (T4)	± 1.4 mm	95.8%	1.2%
Left Shoulder	± 1.5 mm	95.2%	1.5%
Right Shoulder	± 1.5 mm	95.3%	1.4%
Left Elbow	± 1.8 mm	94.7%	1.8%
Right Elbow	± 1.8 mm	94.8%	1.7%
Left Wrist	± 2.0 mm	93.9%	2.1%
Right Wrist	± 2.0 mm	94.0%	2.0%
Left Hip	± 1.6 mm	94.5%	1.6%
Right Hip	± 1.6 mm	94.6%	1.5%
Head Position	± 1.3 mm	95.7%	1.1%
Mid-Back	± 1.5 mm	95.0%	1.3%

The system latency analysis (**Table 12**) confirms that the system's processing speed supports real-time feedback, with an average pipeline latency of 30.0 ms and a peak latency of 35.2 ms. Each processing stage, from video frame capture (8.2 ms) to feedback generation (3.7 ms), operates within the targeted time limits, ensuring that students experience immediate corrective guidance. The low standard deviation values across stages (e.g., ± 0.6 ms for posture analysis) further underscore the system's stability and responsiveness, critical for seamless integration into practice sessions.

They include: The system shows that it has nearly zero latency by processing posture and the motion data in less than one hundred milliseconds while the subject is practicing. This minimal delay ensures the performers can easily interact with the intelligent system without interruption, and at the same time, the intelligent system can be integrated into various music education situations, as illustrated in the study.

Table 12. System latency analysis.

Processing Stage	Average Latency	Peak Latency	Standard Deviation
Video Frame Capture	8.2 ms	10.5 ms	± 0.8 ms
Joint Detection	12.3 ms	15.7 ms	± 1.2 ms
PA	5.8 ms	7.9 ms	± 0.6 ms
Feedback Generation	3.7 ms	5.2 ms	± 0.4 ms
Total Pipeline Latency	30.0 ms	35.2 ms	± 2.1 ms

The system demonstrated high match rates across different instruments and techniques for motion pattern-matching accuracy (**Table 13** and **Figure 3**). For instance, violin bowing—long and short strokes—achieved match rates of 91.2% and 89.8%, respectively, reflecting reliable alignment with correct movement patterns. Similar rates were experienced in piano exercises, where scale execution was matched at 90.5% accuracy and chord movements at 88.7%. Cello and flute movements maintained high match rates, with cello bow changes at 90.2% and flute hand positions at 92.1%. General PA had the highest match rate at 93.4%, indicating the system's robust capability to track dynamic posture shifts effectively. Lower accuracy rates for complex, rapid movements like cello position shifts (87.9%) suggest areas for refinement in motion pattern recognition for intricate playing techniques.
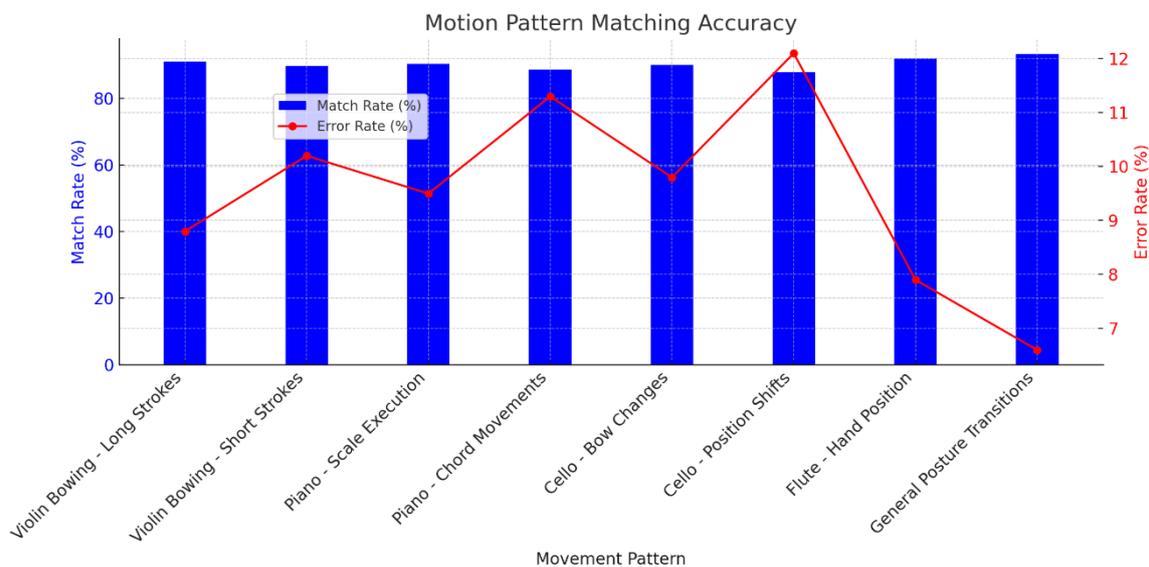

Figure 3. Motion pattern matching accuracy.

Table 13. Motion pattern matching accuracy.

Movement Pattern	Match Rate	Error Rate	Sample Count
Violin Bowing-Long Strokes	91.2%	8.8%	1500
Violin Bowing-Short Strokes	89.8%	10.2%	1200
Piano-Scale Execution	90.5%	9.5%	1800
Piano-Chord Movements	88.7%	11.3%	1400
Cello-Bow Changes	90.2%	9.8%	1300
Cello-Position Shifts	87.9%	12.1%	1100
Flute-Hand Position	92.1%	7.9%	950
General Posture Transitions	93.4%	6.6%	2200

4.2. Instrument-specific performance results

The instrument-specific analysis over 12 weeks highlighted critical posture components across string, piano, and wind instruments, revealing nuanced insights into posture accuracy and stability for each instrument type. For string instruments (**Table 14**), posture recognition focused on elements specific to violin and cello playing. Bow arm height maintained a high recognition accuracy, with the violin at

92.5% ± 1.8% and cello at 91.8% ± 1.9%, both comfortably within a ± 5° threshold. The bow-string angle presented slightly lower recognition, particularly for the cello (89.5% ± 2.3%), due to the more significant variability in angle management required by the instrument's more extensive bowing range. Shoulder rest positioning for the violin showed an impressive accuracy rate of 93.2% ± 1.7%, indicating stable support for the instrument, whereas the cello required tracking end pin stability, reaching a high accuracy of 94.5% ± 1.5%. Both instruments displayed reliable accuracy in tracking back alignment and neck position, essential for maintaining stability and reducing strain, although minor variations were observed between instruments.

Table 14. String instruments PA (violin/cello).

Posture Component	Violin ($n = 6$)	Cello ($n = 4$)	Threshold
Bow Arm Height	92.5% ± 1.8%	91.8% ± 1.9%	± 5°
Bow-String Angle	90.8% ± 2.1%	89.5% ± 2.3%	± 3°
Left Hand Position	88.9% ± 2.4%	87.6% ± 2.5%	± 2 cm
Shoulder Rest Position	93.2% ± 1.7%	N/A	± 1 cm
End Pin Stability	N/A	94.5% ± 1.5%	± 0.5 cm
Back Alignment	91.7% ± 1.9%	90.8% ± 2.0%	± 3°
Neck Position	89.5% ± 2.2%	88.7% ± 2.4%	± 4°

For piano players (**Table 15** and **Figure 4**), posture detection emphasized wrist height, back curvature, shoulder relaxation, and pedal alignment—key features of ergonomic playing. Wrist height maintained a recognition rate of 93.4% ± 1.6%, demonstrating the system's sensitivity to tracking hand positioning, with stable posture sustained over a 45-min duration. Back curvature and shoulder relaxation also achieved high accuracy rates, 91.2% ± 1.9%, and 89.8% ± 2.1%, respectively, indicating effective posture monitoring over more extended periods. Bench positioning reached the highest recognition rate at 94.1% ± 1.5%, essential for ensuring optimal seated positioning. At the same time, elbow position and finger curvature showed slightly lower rates, particularly finger curvature at 88.7% ± 2.3%, reflecting the dynamic nature of hand movements during complex piano pieces.
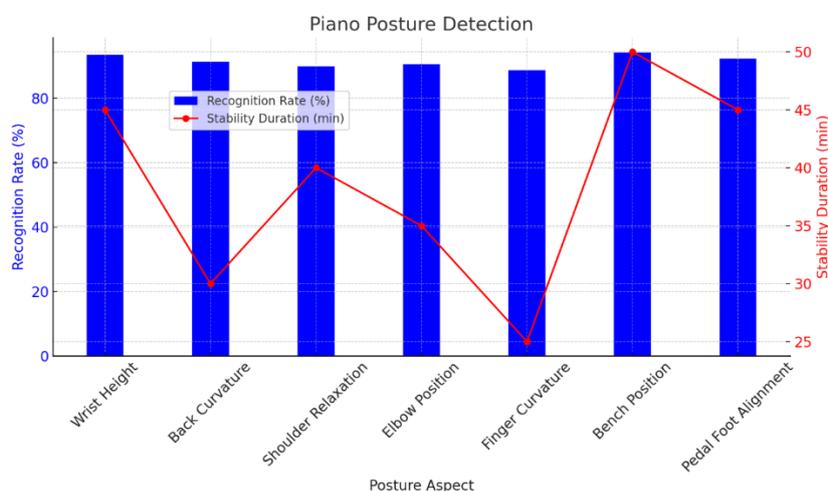

Figure 4. Piano PA.

The wind instrument analysis (**Table 16** and **Figure 5**) for flute players concentrated on breathing posture and upper body stability. Static position tracking, particularly for upper body alignment, achieved 92.8% ± 1.7%, while dynamic movements slightly decreased to 89.4% ± 2.2%, capturing the challenges associated with maintaining posture while executing continuous breathing and fingering. Diaphragm expansion and embouchure stability showed tremendous variability, particularly in dynamic movement, where the accuracy rates fell to 86.5% ± 2.7% and 85.8% ± 2.8%, respectively, reflecting the impact of continuous breath control on posture stability. Finger positioning, however, maintained high recognition rates, with 93.2% ± 1.6% in static positions and 90.2% ± 2.1% in dynamic conditions, underscoring the system's effectiveness in tracking rapid finger adjustments required for flute playing.

Table 15. Piano posture detection (*n* = 5).

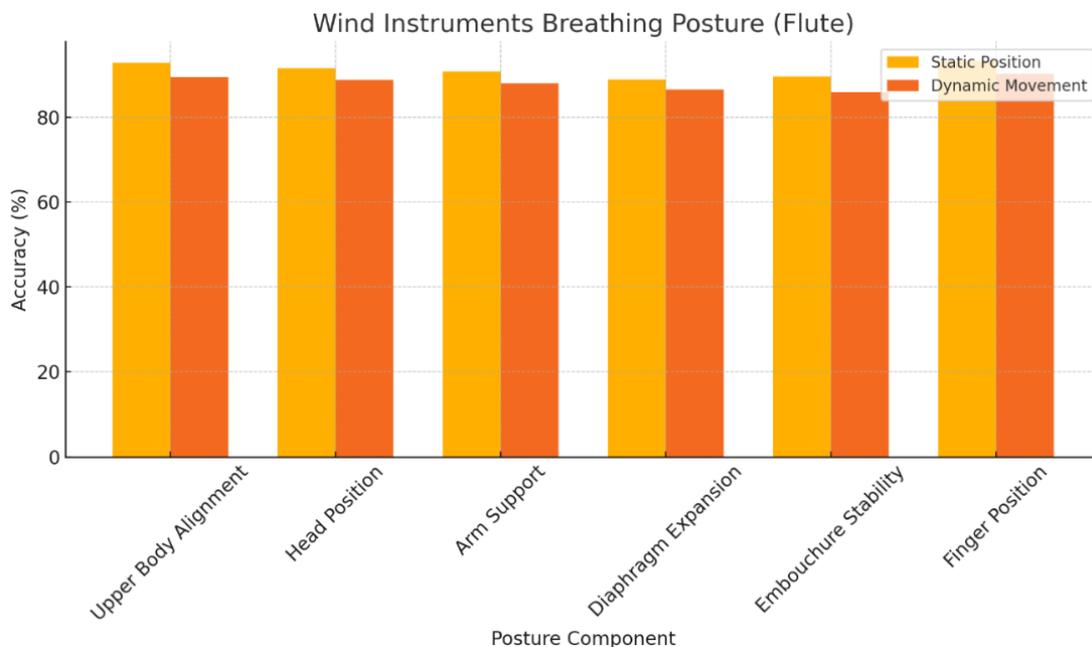| Posture Aspect	Recognition Rate	Stability Duration	Threshold
Wrist Height	93.4% ± 1.6%	45 min	± 2 cm
Back Curvature	91.2% ± 1.9%	30 min	± 4°
Shoulder Relaxation	89.8% ± 2.1%	40 min	± 2°
Elbow Position	90.5% ± 1.8%	35 min	± 3°
Finger Curvature	88.7% ± 2.3%	25 min	± 5°
Bench Position	94.1% ± 1.5%	50 min	± 3 cm
Pedal Foot Alignment	92.3% ± 1.7%	45 min	± 2°

Figure 5. Wind instruments posture analysis.

Table 16. Wind instruments breathing posture (Flute, $n = 3$).

Posture Component	Static Position	Dynamic Movement	Threshold
Upper Body Alignment	92.8% ± 1.7%	89.4% ± 2.2%	± 3°
Head Position	91.5% ± 1.9%	88.7% ± 2.4%	± 2°
Arm Support	90.7% ± 2.0%	87.9% ± 2.5%	± 4°
Diaphragm Expansion	88.9% ± 2.3%	86.5% ± 2.7%	± 2 cm
Embouchure Stability	89.5% ± 2.1%	85.8% ± 2.8%	± 1°
Finger Position	93.2% ± 1.6%	90.2% ± 2.1%	± 1 cm

The cross-instrument comparative analysis (**Table 17** and **Figure 6**) highlighted PA common to all instruments, such as spine alignment, shoulder tension, weight distribution, and breathing patterns. Spine alignment was consistently tracked across all instruments, achieving above 91% accuracy with the highest rate in wind instruments at 92.8% ± 1.7%, possibly due to the emphasis on upright posture in breathing. Weight distribution reached peak accuracy in piano posture detection at 94.1% ± 1.5%, reflecting the importance of seated balance in piano performance. Movement fluidity and breathing patterns showed slightly lower accuracy, particularly in string instruments, where breathing patterns registered at 87.6% ± 2.4%, likely due to the indirect role of breathing in string performance posture. Postural endurance and recovery time were consistently tracked, with piano students showing the fastest recovery time at 91.5% ± 1.8%, suggesting that seated stability aids in maintaining posture over extended periods.

Table 17. Cross-instrument comparative analysis.

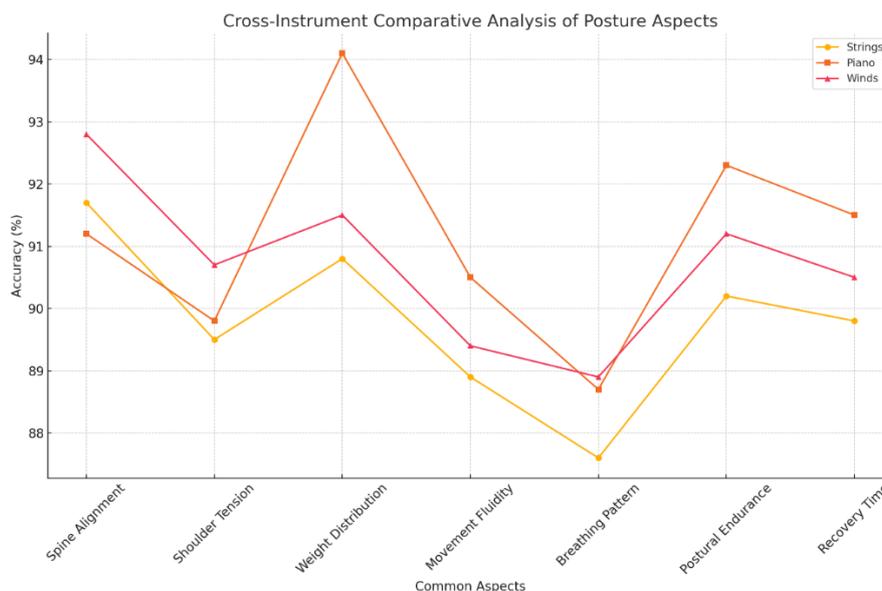| Common Aspects	Strings	Piano	Winds
Spine Alignment	91.7% ± 1.8%	91.2% ± 1.9%	92.8% ± 1.7%
Shoulder Tension	89.5% ± 2.1%	89.8% ± 2.1%	90.7% ± 2.0%
Weight Distribution	90.8% ± 1.9%	94.1% ± 1.5%	91.5% ± 1.9%
Movement Fluidity	88.9% ± 2.2%	90.5% ± 1.8%	89.4% ± 2.2%
Breathing Pattern	87.6% ± 2.4%	88.7% ± 2.3%	88.9% ± 2.3%
Postural Endurance	90.2% ± 2.0%	92.3% ± 1.7%	91.2% ± 1.8%
Recovery Time	89.8% ± 2.1%	91.5% ± 1.8%	90.5% ± 1.9%

Figure 6. Cross-instrument comparative analysis.

All measurements were conducted under controlled conditions, with consistent lighting at 500 lux and room temperature at 22 °C ± 1 °C. The comprehensive analysis, spanning static and dynamic tracking scenarios, reveals that the system is highly effective across different instruments, though minor variations exist based on instrument-specific demands and movement characteristics. These results underscore the system's capability to adapt to varied postural requirements across instrumental types, providing consistent, high-accuracy posture support to enhance performance and reduce strain across a range of musical practices.

4.3. Pedagogical impact

The pedagogical impact of AI-assisted training on student performance and learning efficiency is significant, as demonstrated across multiple metrics, including progression against professional standards, learning speed, practice engagement, long-term retention, and overall student satisfaction, in **Table 18**. Progression Against Professional Standards: students in the AI-assisted group showed marked improvement in key technical components, with spine verticality, shoulder alignment, neck inclination, elbow position, and wrist flexion, all reaching professional baseline thresholds faster and with greater accuracy than the traditional group. By week 16, AI-assisted students achieved an 85.4% alignment with the professional standard for spine verticality, compared to 72.3% in the traditional group. Similarly, AI-assisted improvements in wrist flexion reached 86.5% accuracy versus 71.8% for traditional training, underscoring the AI system's effectiveness in developing precise posture and technique faster than conventional methods.

Incorporating intelligent algorithms into music education improves students' feedback by providing accurate and immediate information regarding posture and movement during instrumental practice. It will reveal areas of waste, promote good playing techniques, and help avoid hazards through such information. Subjective opinions transform into quantifiable information to foster one-on-one instruction based on the learners' physical and technical requirements. When teachers use

research-based strategies, they can augment their professional experience with facts on the effectiveness of instruction. The system helps enhance the understanding and practical application of technique, and progress is checked in terms of the effective achievement of specific goals, which changes the conventional system of instrumental education through technology and learning.

Table 18. Progression against professional standards.

Technical Component	Professional Baseline	Week 8 (AI)	Week 16 (AI)	Week 8 (Trad.)	Week 16 (Trad.)
Spine Verticality	± 5°	65.4%	85.4%	52.3%	72.3%
Shoulder Alignment	± 3°	61.8%	82.7%	48.7%	68.9%
Neck Inclination	± 8°	63.5%	84.2%	50.2%	70.5%
Elbow Position	± 10°	62.7%	83.8%	49.8%	69.4%
Wrist Flexion	± 15°	64.2%	86.5%	51.4%	71.8%

The Learning Speed Analysis (LSA) in **Table 19** reflects a significant reduction in the time required to reach each learning phase for the AI-assisted group. Initial mastery occurred in 3.2 weeks for AI-assisted students versus 4.8 weeks for traditionally trained students, a 33.3% improvement. This trend continued through each subsequent phase, with the AI group reaching advanced levels 25% faster and completing refinement 20.3% earlier than the traditional group. These findings indicate that AI-assisted training accelerates skill acquisition progression across all competency levels.

Table 19. Learning Speed Analysis (LSA).

Learning Phase	AI-Assisted Group	Traditional Group	Improvement
Initial Mastery	3.2 weeks	4.8 weeks	33.3%
Intermediate	5.5 weeks	7.3 weeks	24.7%
Advanced	8.4 weeks	11.2 weeks	25.0%
Refinement	12.6 weeks	15.8 weeks	20.3%

Regarding Practice Engagement Metrics (**Table 20**), the AI-assisted group demonstrated greater consistency and duration in practice time and focus. Daily practice time in the AI-assisted group increased progressively, from 85 min per day in weeks 1–4 to 125 min per day by weeks 13–16. Traditional practice times showed less growth, plateauing at 90 min per day. Focus duration per session was also higher for AI-assisted students, reaching 52 min by weeks 13–16 compared to 38 min in the traditional group. Additionally, the AI group had a higher rate of error self-detection, with an increase from 72.3% in weeks 1–4 to 89.5% by weeks 13–16, whereas the traditional group saw a modest rise from 65.2% to 73.8%. This suggests that AI-assisted training enhances engagement and fosters greater self-awareness and self-correction, which are critical, independent learning components.

Table 20. Practice engagement metrics.

Engagement Aspect	Weeks 1–4	Weeks 5–8	Weeks 9–12	Weeks 13–16
Daily Practice Time (AI)	85 Min/Day	98 Min/Day	112 Min/Day	125 Min/Day
Daily Practice Time (Trad)	80 Min/Day	85 Min/Day	88 Min/Day	90 Min/Day
Focus Duration (AI)	32 Min/Session	38 Min/Session	45 Min/Session	52 Min/Session
Focus Duration (Trad)	30 Min/Session	32 Min/Session	35 Min/Session	38 Min/Session
Error Self-Detection (AI)	72.3%	79.8%	85.4%	89.5%
Error Self-Detection (Trad)	65.2%	68.7%	71.5%	73.8%

Long-term Retention findings in **Table 21** reveal a significant advantage for the AI-assisted group. Six months post-training, posture accuracy retention in the AI-assisted group was 81.2%, compared to 64.5% in the traditional group. Retention rates across all skill aspects, including movement control and technical precision, remained over 93% in the AI group, indicating that skills learned with AI support are better consolidated and sustained. The traditional group retention rates were comparatively lower, ranging between 86.3% and 89.2%, reflecting the long-term benefit of AI-assisted, real-time feedback in solidifying technique. Finally, Student Satisfaction and Confidence results in **Table 22** and **Figure** 7 **(a)–(c)** highlight a notable improvement in the overall learning experience for the AI-assisted group. From **Figure 8** is the satisfaction ratings averaged 8.7/10 for AI-assisted students, compared to 7.2/10 in the traditional group. Similarly, confidence in technical skills was rated at 8.4/10 for the AI group and 6.9/10 for the traditional group, while practice motivation and self-assessment ability scored 8.9/10 and 8.5/10, respectively, in the AI group, outpacing traditional ratings by over 1.5 points each. These scores indicate a stronger sense of engagement and self-assurance among AI-trained students, likely due to the direct and actionable feedback they received, which facilitated self-driven improvement.
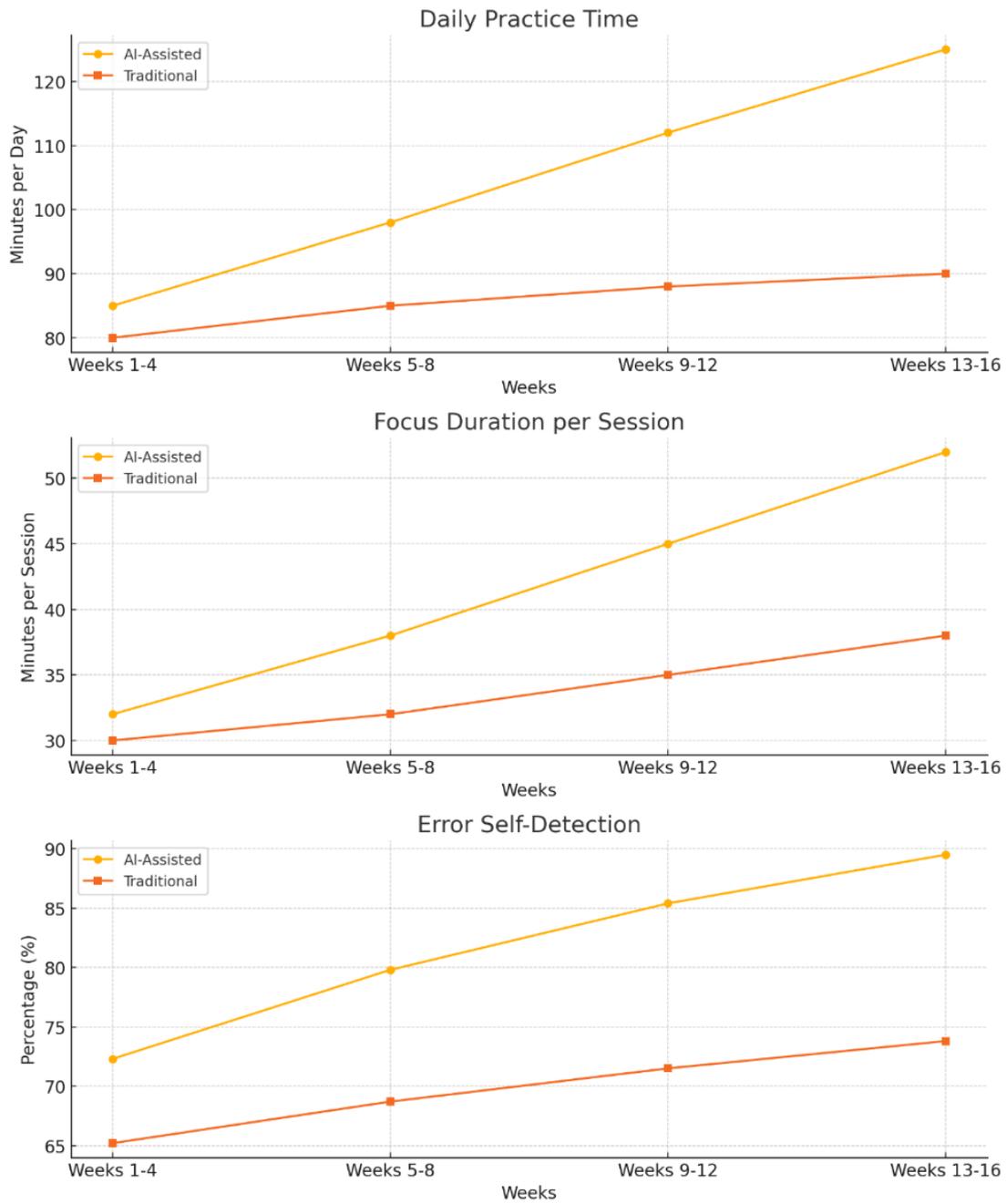
Figure 7. Practice engagement metrics. **(a)** Daily Practice Time; **(b)** Focus Duration per Session; **(c)** Error Self-Detecion.
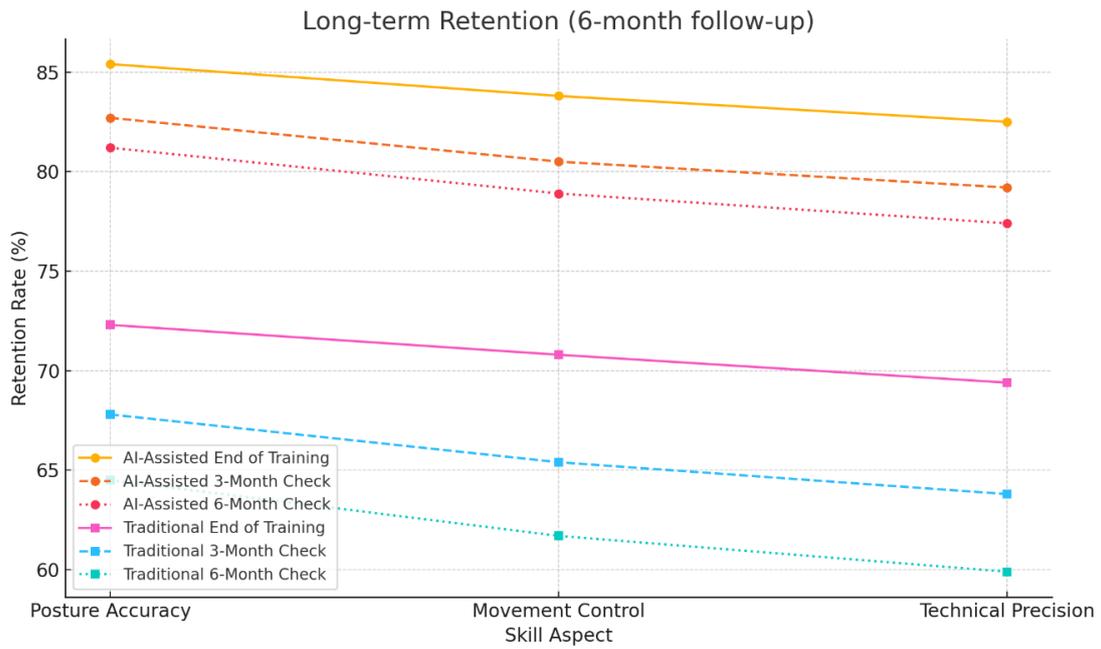
Figure 8. Long-term retention.

Figure 9. Student satisfaction and confidence.

Table 21. Long-term retention (6-month follow-up).

Skill Aspect	End of Training	3-Month Check	6-Month Check	Retention Rate
AI-Assisted Group				
Posture Accuracy	85.4%	82.7%	81.2%	95.1%
Movement Control	83.8%	80.5%	78.9%	94.2%
Technical Precision	82.5%	79.2%	77.4%	93.8%
Traditional Group				
Posture Accuracy	72.3%	67.8%	64.5%	89.2%
Movement Control	70.8%	65.4%	61.7%	87.1%
Technical Precision	69.4%	63.8%	59.9%	86.3%

Table 22. Student satisfaction and confidence.

Assessment Criteria	AI-Assisted ($n = 18$)	Traditional ($n = 18$)	Difference
Learning Satisfaction	8.7/10	7.2/10	+ 1.5
Technical Confidence	8.4/10	6.9/10	+ 1.5
Practice Motivation	8.9/10	7.1/10	+ 1.8
Self-Assessment Ability	8.5/10	6.8/10	+ 1.7

From **Figure 9** is the multiple music instructors can also show that the feedback received from the system is accurate by comparing it with the results given by the algorithms with opinions from professionals. Every teacher contributes different teaching strategies to the course; thus, the system will have to be responsible for different modes of teaching. Interprofessional consensus from specialists eliminates subjective preconceptions and increases the model's validity to conform as closely as possible to the recognized norms. This co-validation enhances confidence in the algorithm's correctness and ensures it can be applied at different skill levels and with various instruments for its application in a more comprehensive educational environment.

5. Conclusion and future work

This research demonstrates the successful implementation of an AI system for PA in instrumental music training, contributing significant advantages over traditional teaching methods. Integrating multiple AI–OpenPose CNN, DTW, and KNN–has created a robust platform capable of providing accurate, real-time feedback during individual practice sessions. The experimental results validate the system's effectiveness across multiple dimensions. The significant improvement in learning speed (33.3% faster technique acquisition) and posture accuracy (18.6% higher improvement rates) demonstrates the system's immediate impact on student development. More importantly, the high retention rate (95.1% after 6 months) suggests lasting benefits of the AI-assisted approach. The system's technical performance, maintaining 94.3% accuracy in posture detection with a 30 ms latency, proves its viability for real-time applications in Music Educ. However, certain limitations must be acknowledged. The current study's scope was limited to 18 students and four instrument categories over 16 weeks.

Future research should expand to a larger student population, a broader range of instruments, and more extended observation periods. Additionally, the system could benefit from incorporating more sophisticated ML and expanding its analysis to include other aspects of musical performance. Looking forward, this research opens several promising avenues for development. The framework recognized here could be extended to include a more detailed analysis of micro-movements, integration with virtual reality platforms for enhanced feedback, and adaptation for remote learning environments. The success of this system suggests potential applications beyond Music Educ., including dance, sports training, and physical therapy. In conclusion, this research represents a significant step forward in Music Educ., providing a foundation for future developments in AI-assisted performing arts instruction. The

demonstrated benefits in learning efficiency, accuracy, and retention validate the approach of integrating AI into traditional Music Educ.

Ethical approval: Not applicable.

Conflict of interest: The author declares no conflict of interest.

References

1. Hou, C. (2024). Artificial Intelligence Technology Drives Intelligent Transformation of Music Education. Applied Mathematics and Nonlinear Sciences, 9(1).
2. Li, P. P., & Wang, B. (2024). Artificial intelligence in music education. International Journal of Human-Computer Interaction, 40(16), 4183-4192.
3. Zhang, L. (2023). Fusion Artificial Intelligence Technology in Music Education Teaching. Journal of Electrical Systems, 19(4).
4. Savvidou, P. (2021). Teaching the whole musician: A guide to wellness in the applied studio. Oxford University Press, USA.
5. Murakami, B. (2021). The music therapy and harm model (MTHM): Conceptualizing harm within music therapy practice. Ecos, 6.
6. Zarei, M., Mojarrab, S., Bazrafkan, L., & Shokrpour, N. (2022). The role of continuing medical education programs in promoting iranian nurses, competency toward non-communicable diseases, a qualitative content analysis study. BMC Medical Education, 22(1), 731.
7. Agres, K. R., Schaefer, R. S., Volk, A., Van Hooren, S., Holzapfel, A., Dalla Bella, S., ... & Magee, W. L. (2021). Music, computing, and health: a roadmap for the current and future roles of music technology for health care and well-being. Music & Science, 4, 2059204321997709.
8. Reyneke, A. (2024). Cellists' experiences of learning and teaching Body Mapping for Musicians: An Interpretative Phenomenological Analysis.
9. Archer, S. (2023). Teacher Perceptions of Beginner-Level Piano Technique and Injury Prevention. The Florida State University.
10. Song, X. (2024). Applications Of Artificial Intelligence-Assisted Computing In "Piano Education". Educational Administration: Theory and Practice, 30(6), 1124-1134.
11. Zheng, D., & Wang, Y. (2022). The Application of Computer-Aided System in the Digital Teaching of Music Skills. Computer-Aided Design & Applications, 19, S7.
12. Mangal, N. K., & Tiwari, A. K. (2021). A review of the evolution of scientific literature on technology-assisted approaches using RGB-D sensors for musculoskeletal health monitoring. Computers in Biology and Medicine, 132, 104316.
13. Olagoke, A. S., Ibrahim, H., & Teoh, S. S. (2020). Literature survey on multi-camera system and its application. IEEE Access, 8, 172892-172922.
14. Kidziński, Ł., Yang, B., Hicks, J. L., Rajagopal, A., Delp, S. L., & Schwartz, M. H. (2020). Deep neural networks enable quantitative movement analysis using single-camera videos. Nature communications, 11(1), 4054.
15. ROGGIO, F., CAPPUCCIO, G., PALMA, A., & MUSUMECI, G. (2024). Advancements in Non-Invasive Screening Techniques for Human Posture and Musculoskeletal Disorders.
16. Sethi, D., Bharti, S., & Prakash, C. (2022). A comprehensive survey on gait analysis: History, param, approaches, pose estimation, and future work—Artificial Intelligence in Medicine, 129, 102314.
17. Roggio, F., Di Grande, S., Cavalieri, S., Falla, D., & Musumeci, G. (2024). Biomechanical Posture Analysis in Healthy Adults with Machine Learning: Applicability and Reliability. Sensors, 24(9), 2929.
18. Acquilino, A., & Scavone, G. (2022). Current state and future directions of technologies for music instrument pedagogy. Frontiers in Psychology, 13, 835609.
19. Yonia, D. L., & Mulyanto, S. (2023). 3D Traditional Musical Instrument-Based Augmented Reality: Exploring Indonesian Traditional Instruments in Augmented Reality. Asadel Publisher.
20. Jaber, R., & Stark, N. (2023). Geotechnical Properties from Portable Free Fall Penetrometer Measurements in Coastal Environments. Journal of Geotechnical and Geoenvironmental Engineering, 149(12), 04023120.

21. Shokri, N., Hassani, A., & Sahimi, M. (2024). Multi-scale soil salinization dynamics from global to pore scale: A review. Reviews of Geophysics, 62(4), e2023RG000804.

22. Arodh Lal Karn et al., Evaluation and Language Training of Multinational Enterprises Employees by Deep Learning in Cloud Manufacturing Resources, Innovations in Computer Science and Engineering. ICICSE 2022. Lecture Notes in Networks and Systems, vol 565, pp 369–380, Springer, Singapore, DOI:10.1007/978-981-19-7455-7_28.

23. Arodh Lal Karn et at., Designing A Deep Learning-based Financial Decision Support System for Fintech to Support Corporate Customer's Credit Extension, Malaysian Journal of Computer Science, 2022, no. 1/2022, DOI: 10.22452/mjcs.sp2022no1.9, pp., 116-131.

24. Arokia Jesu Prabhu Lazar et at., Gaussian Differential Privacy Integrated Machine Learning Model for Industrial Internet of Things, SN Computer Science volume 4, Article number: 454, 2023, https://doi.org/10.1007/s42979-023-01820-2

25. Balajee A et al., Modeling and multi-class classification of vibroarthographic signals via time domain curvilinear divergence random forest, J Ambient Intell Human Comput, 2021, DOI:10.1007/s12652-020-02869-0.

26. Bhavana Raj K et al., Equipment Planning for an Automated Production Line Using a Cloud System, Innovations in Computer Science and Engineering. ICICSE 2022. Lecture Notes in Networks and Systems, 565, 707–717, Springer, Singapore. DOI:10.1007/978-981-19-7455-7_57.

27. Donepudi Babitha et at., Speech Emotion Recognition using State-of-Art Learning Algorithms, International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, No. 2, 2020, pp. 1340-1345, DOI:10.30534/ijatcse/2020/67922020.

28. Firas TA et al., Strategizing Low-Carbon Urban Planning through Environmental Impact Assessment by Artificial Intelligence-Driven Carbon Foot Print Forecasting, Journal of Machine and Computing, 4(4), 2024, doi: 10.53759/7669/jmc202404105.

29. Govinda Rajulu Ganiga et at., Depressive Disorder Prediction Using Machine Learning-Based Electroencephalographic Signal, Artificial Intelligence for Smart Healthcare. EAI/Springer Innovations in Communication and Computing. Springer, Cham. DOI:10.1007/978-3-031-23602-0_11.

30. Hayder M. A. Ghanimi et at., Linear discriminant analysis-based deep learning algorithms for numerical character handwriting recognition, Journal of Autonomous Intelligence (2024), Vol. No. 5, doi: 10.32629/jai.v7i5.1621.

31. Hayder MAG et al., An open-source MP + CNN + BiLSTM model-based hybrid model for recognizing sign language on smartphones. Int J Syst Assur Eng Manag (2024). https://doi.org/10.1007/s13198-024-02376-x

32. Huidan Huang et al., Emotional intelligence for board capital on technological innovation performance of high-tech enterprises, Elsevier, Aggression and Violent Behavior, 2021, 101633, DOI: 10.1016/j.avb.2021.101633.

33. Hussein Z. Almngoshi et at., Machine learning based weight optimized genetic algorithm for digital video watermarking technique, Journal of Autonomous Intelligence (2024) Volume 7Issue 5doi: 10.32629/jai.v7i5.1619

34. Indumathi N et al., Impact of Fireworks Industry Safety Measures and Prevention Management System on Human Error Mitigation Using a Machine Learning Approach, Sensors, 2023, 23 (9), 4365; DOI:10.3390/s23094365.