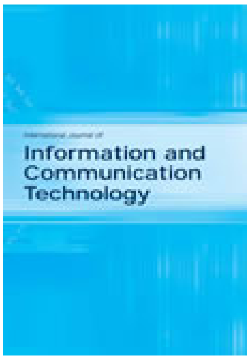


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Integrating deep learning and wearable technology for real-time, scalable and objective physical education assessment

Wang Chen

DOI: 10.1504/IJICT.2025.10070825

Article History:

Received:	19 February 2025
Last revised:	09 March 2025
Accepted:	09 March 2025
Published online:	06 May 2025

Integrating deep learning and wearable technology for real-time, scalable and objective physical education assessment

Wang Chen

Physical Education,
Chongqing Vocational and Technical University of Mechatronics,
Chongqing, 402760, China
Email: 15723247759@163.com

Abstract: In traditional PE, we often assess students with very subjective assessments and miss all the nuances and intricacies of motor skills that are complex. This study introduces a deep learning-based framework using CNNs and LSTMs with wearable tech that increases evaluation accuracy and feedback – a multimodal dataset comprising the data from several devices like accelerometers and heart rate monitors. Teacher-based assessments (72% agreement) were surpassed by 89% accuracy of the proposed SkillNet model. It reduced inter-rater variability by 35% and the evaluation time by 40%. Student engagement rose from 60% to 85%, with improved motivation. This system provides accurate, scalable objective assessment, real-time feedback, and enriched learning in PE.

Keywords: deep learning in physical education; automated PE assessment; wearable technology in education; real-time feedback mechanisms; objective motor skill evaluation.

Reference to this paper should be made as follows: Chen, W. (2025) 'Integrating deep learning and wearable technology for real-time, scalable and objective physical education assessment', *Int. J. Information and Communication Technology*, Vol. 26, No. 10, pp.42–60.

Biographical notes: Wang Chen is a member of the Department of Physical Education at Chongqing Vocational and Technical University of Mechatronics, located in Chongqing, China. His academic interests include sports science, physical fitness, and student athletic development. He is committed to promoting physical education through research and practical training.

1 Introduction

Physical education is essential to holistic education, physical fitness, motor skill development, and overall well-being (Syaukani et al., 2023; Adambaevna, 2023; Etkin, 2024). Yet, in practice, students' performance in PE has been chiefly traditionally assessed by manual, observational methods subject to variability, inconsistency, and time cost (Carling et al., 2008b; Knudson, 2013; Carling et al., 2008a). Teachers usually grade running, jumping, and throwing skills using rubrics, checklists, or standardised fitness tests (Lund and Kirk, 2019; Lund and Veal, 2013; SHAPE America, 2018). However, these methods cannot learn complex movements to a high degree and do not afford immediate feedback and improvement. Given the increased use of technology within education to boost learning outcomes, there is a growing need for viable solutions to the limitations of conventional PE evaluation strategies.

1.1 Challenges in traditional PE assessments

Subjective evaluation in PE is based on personal opinion, which challenges confirming fair and accurate performance evaluations (Hartmann and Slapničar, 2012; Morrow et al., 2015). Inter-rater variability is a problem of observational methods: teachers tend to give weights to different things. Structured, standardised tests primarily measure speed, strength, and endurance but often overlook dynamic movement aspects such as coordination, rhythm, and error patterns (Engelmann, 2014; Wulfsohn, 2018; Todorovich, 2024). However, they may neglect the dynamic of movements, such as coordination, sense of rhythm, or error patterns. Notably, the limitations are particularly relevant in diverse classrooms, where students often come from very diverse skill levels and physical abilities of themselves.

Furthermore, traditional methods are not scalable and are inefficient in assessing large numbers of students (Pangrazi and Beighle, 2019; Vanhees et al., 2005). As the class size increases, this becomes an increasingly burdensome exercise for teachers, who must manually observe, record, and analyse performance. In addition, without real-time feedback, students cannot make immediate changes to their movements to improve quickly.

1.2 Advancements in technology for PE assessment

Recent technological progress, especially in deep learning (DL) and wearables, makes PE assessments an excellent opportunity for a technological revolution (Miller et al., 2025; Zhou et al., 2024). Video-based systems have successfully analysed human motion in DL, a subset of artificial intelligence (Kumar and Kumar, 2023). CNNs can identify important skeletal points and examine body posture (Roggio et al., 2024). Temporal dynamics are modelled using long short-term memory (LSTM) networks, which capture the rhythm and consistency of the movements over time (Pham, 2021). These capabilities facilitate objective, detailed, and objective evaluations of motor skills that are out of the hamstring and are subjective and variable based on traditional methods. Wearable technology furthers this, allowing real-time physiological and biomechanical measures in the field (Alzahrani and Ullah, 2024; Edwards et al., 2023; Lu, 2023). Metrics are measured by devices like accelerometers, gyroscopes, heart rate monitors, acceleration, volume of angular velocity, and energy expenditure, which give us information about what students are doing physically and coordinating. Combining video-based motion analysis with wearable data generated a multimodal framework that captures a performance's qualitative and quantitative narratives.

1.3 Motivation for the proposed framework

Deep learning and wearable technology advances in sports science and rehabilitation have been used. However, their adoption in educational settings remains inhibited.

Currently, most systems are designed to serve the needs of elite athletes and clinical environments, but they are not intended to meet the varied needs of school classrooms. Furthermore, many systems analyse only one metric of a student's performance, like the physical one or movement pattern, without considering him as a whole (Zhang, 2021; Chidambaram et al., 2022; Seçkin et al., 2023; Wei and Wu, 2023). We propose a framework that integrates DL models, multimodal data, and real-time feedback into a unified PE assessment system to address this gap. The framework utilises CNNs for spatial analysis, LSTMs for temporal modelling of data, and wearable devices for real-time data collection, and therefore, it dedicates us to a generalised method for the study of motor skills. Exploiting the system's ability to provide immediate, action feedback allows teachers to focus on teaching and students to engage in their learning process actively.

1.4 Significance of the study

The explicit contribution of this study is that it experimentally demonstrates how deep learning and wearable technology can be combined to address critical challenges in PE assessment. Also, the proposed framework promises objective, consistent, and scalable evaluations, which considerably minimise the subjectivity and variability introduced by traditional evaluation methods. It allows for automatic assessment and real-time feedback so that the skill can be improved immediately, and a more interactive learning environment can be offered to students. Most importantly, the framework lends itself to adaptation in any educational or training context, including the school context, sports training, and rehabilitation settings. It advances educational fairness by furnishing a personalised view and recommendations, allowing students of all abilities and skill levels to get tailored feedback and support. This work also relates to larger goals of educational innovation by showing how advanced technologies can effectively contribute to improved outcomes in teaching and learning.

1.5 Objectives of the study

This work focuses on developing and validating a DL-based approach to PE assessment optimisation. The specific goals include:

- the primary objective is to develop a hybrid deep learning model that combines CNNs for pose estimation and LSTMs for analysing the temporal dynamics of motor skills
- accurately and deeply evaluates by integrating wearable devices to collect real-time physiological and biomechanical data
- a real-time feedback mechanism that provides personalised insight and recommendations for students and teachers
- comparing the framework to the traditional assessment process regarding accuracy, efficiency, scalability and student engagement.

The study is organised into several sections in this paper. Section 2 includes a complete literature review covering the limitations of traditional PE assessments, DL advancements, wearable technology research, and gaps the proposed framework addresses. Section 3 describes the proposed architecture of the framework, along with the data collection technique DL model design. The experimental setup and implementation are described in Section 4, including data collection, training, evaluation metrics, and tools. The results and analysis in Section 5 focus on system performance, actual time feedback capabilities, and comparative advantages of the system over the traditional methods. Section 6 discusses the findings' implications, challenges, and future research directions. In Section 7, the paper concludes by summarising the contributions and potential impact of the proposed system towards PE assessments and educational practice.

2 Literature review

To create an automated, accurate, scalable physical education (PE) assessment framework, we must constrain ourselves with what is already out there in PE methodology and technology. For human motion analysis, this literature review assesses traditional assessment methods in PE, DL advancements, the role of wearable technology in data integration, and the challenges these methods present in educational contexts. It defines some gaps in the state of the art and suggests improvements.

2.1 Traditional methods of physical education assessment

Traditional physical education assessment has primarily relied upon observational observations and standardised tests to measure the students' physical skills, physical fitness, and overall performance (Lund and Kirk, 2019; López-Pastor et al., 2013; Horvat et al., 2019). Teachers commonly use rubrics and checklists to rate performance in running, jumping, and throwing activities. These methods are simple and cheap but susceptible to subjectiveness variability of results. Evaluations can differ even if the teacher is an expert, as we lack interrater reliability. FitnessGram and the President's Fitness Challenge are standardised fitness tests to measure student performance without making judgements (Kimball, 2007; The Cooper Institute, 2017; Meredith and Welk, 2010). However, these tests are very narrow in scope – they tend to cover metrics like speed, strength, and endurance but not metrics related to dynamic movement, such as coordination, rhythm, and balance. Additionally, these traditional assessments do not provide immediate feedback, which can delay skill correction and subsequent improvement opportunities.

2.2 Advancements in deep learning for human motion analysis

Human motion analysis can now be explored with deep learning at a significant level and with great advancement in sports science, rehabilitation, and healthcare (Roggio et al., 2024; Chen, 2024; Prakash et al., 2018). We use convolutional neural networks (CNNs) to estimate the pose of a human body accurately, i.e., to locate key characteristic points like joints, from which we can study human movement patterns. OpenPose and PoseNet are models that have yielded high accuracy for tracking skeletal movements, offering profound, detailed assessments of physical activities (Clark et al., 2019; Colucci, 2023). Recurrent neural networks (RNNs), in particular, and long short-term memory (LSTM), have been successfully used to model temporal dependencies in sequential data (Sherstinsky, 2020). Specifically, LSTMs, such as rhythm, consistency, or error trends, have been applied to analyse motor skill progression over time. It is beneficial to techniques for activities that involve continuous movement, such as running or dancing. While these models have been previously demonstrated to operate successfully in specific highly specialised contexts, the use of these models in educational environments remains limited. However, few studies have involved school environments and their diverse needs proportionally. Most work on elite sports and clinical rehabilitation. Additionally, these models generally require large amounts of data, and such data are often not immediately available in the PE context, making a need for frameworks that can work with relatively modest but structured data precise.

2.3 Role of wearable technology in data integration

The measurement and analysis of physical activity are revolutionised by wearable technology. Fundamental time movement dynamics, energy expending, and physiological responses are continuously and in real-time measured by such devices as accelerometers, gyroscopes, and heart rate monitors. Fitness tracking and sports performance have been universally adopted for wearables based on speed, acceleration, angular velocity, and heart rate variability metrics. Wearable data and video recordings are integrated to create a multimodal approach that complicates the depth and accuracy of motion analysis (Edriss et al., 2024; Chen, 2024; Kazanskiy et al., 2024; Olsen et al., 2024). By combining wearable and visual data, studies have demonstrated that complex movements are better classified and that noise in data is reduced. It includes integrating inertial sensor data with video-based pose estimation to precisely identify errors in activities such as throwing or jumping.

Nevertheless, wearable technology is scarcely adopted in educational setups. Due to cost barriers and a requirement for technical expertise to process and interpret data, these two methods present significant challenges. What is needed is simple wearables that do not cost a lot while maintaining the quality of insights they deliver.

2.4 Challenges in educational contexts

Multiple challenges apply to using current advanced technologies in PE assessments in an educational environment. Data privacy and security are essential first. The General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) can be met by video recordings and wearable data to ensure student privacy (Galatanu, 2024; Bhargava and Rehman, 2025). Secondly, the scalability of the existing solutions is limited. Advances in motion analysis have been incorporated into many advanced motion analysis systems. However, most have been designed for individuals or small groups and are not practical for large classrooms (Gavrila, 1999; Delp et al., 2007). Further, there is a lack of honest-time feedback, which limits their effectiveness, as results come out too slowly, giving little time to correct skills and learning immediately. The technical barriers to these school systems must finally be addressed. The hardware and software for complex systems can also be costly for schools with smaller budgets, and teachers may not have the training to operate said systems.

2.5 Research gaps and opportunities

Despite hardware and DL advances, there is still a large gap between applying these systems and PE assessments. Many current systems, however, fail to integrate motion analysis, temporal modelling, and real-time feedback reduction into a single framework. Also, systems lacking appropriate educational setting designs are not usable or scalable (Akker et al., 2014; Barris and Button, 2008; Khusainov et al., 2013). This study identifies the opportunity to create a cost-effective, scaleable, user-friendly framework that blends the strength of CNNs, LSTMs, and wearable technology. To build such a system, privacy, scalability, and usability challenges must be addressed while giving teachers and students immediate, actionable, real-time feedback.

2.6 Relevance to the proposed framework

This literature review discusses the advances in the relevant fields to devise the proposed framework, filling the identified gaps and challenges. This integration presents a PE assessment system leveraging spatial analysis with video-based CNNs, temporal modelling with LSTMs, and wearable data for multimodal insights. Real-time feedback mechanisms enhance its utility, promoting skill improvement and student engagement. This framework allows scalable and easy-to-use design for different educational contexts and overcomes the limitations of traditional methods and existing technology.

The foundation for the proposed framework is established in this literature review, in which there is a need for an innovative approach to PE assessments utilising state-of-the-art technology to improve accuracy, efficiency, and engagement.

3 Proposed method

Further development of PE assessment depends upon a sound system that can objectively and efficiently rate a wide range of motor skills and skill performance. Despite the limitations of traditional assessment methods, the proposed method utilises DL models and data integration. The modelled framework incorporates advanced modelling methods, broad data collection, and interactive feedback mechanisms to be automated, accurate, and scalable to different PE activities, as shown in Figure 1.

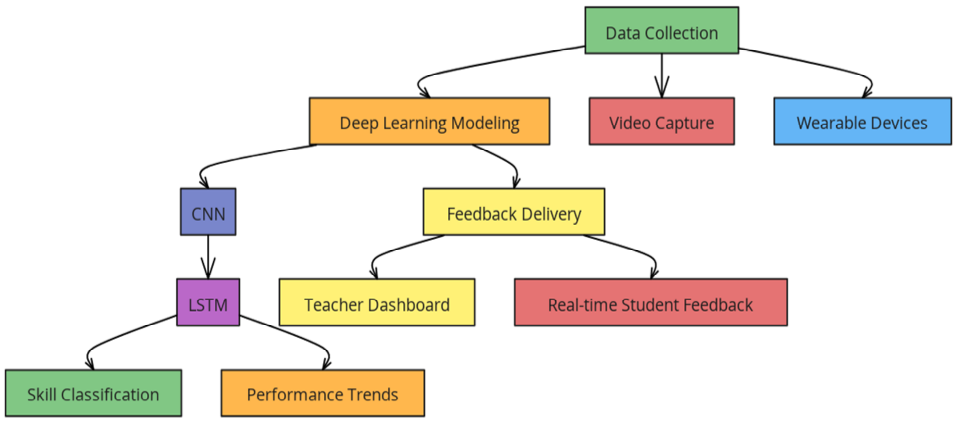
3.1 Framework overview

The proposed method involves three interconnected stages: deep learning modelling, data collection, and feedback delivery. Each stage ensures the system's precision, adaptability, and scalability, making possible accurate, personalised assessments of a wide range of PE activities.

3.1.1 Data collection

This project is grounded in collecting high-quality multimodal data. Complex systems of high-resolution video capture from multiple angles continuously observe students' posture, joint alignment, and movement trajectories. At the same time, such devices as accelerometers, gyroscopes, and heart rate monitors are used to continuously collect physiological and biomechanical measures (acceleration, angular velocity, and heart rate variability). Through this multimodal approach, student performance integrates qualitative and quantitative aspects of student performance.

Figure 1 Key components and workflow of the proposed method for PE assessment (see online version for colours)



Note: It includes three interconnected stages: data collection, deep learning modelling, and feedback delivery. Multimodal data are collected through video capture and wearable devices for processing using a hybrid CNN-LSTM architecture to extract spatial and temporal features. Interactive feedback system based on the outputs (skill classification and performance trends): teacher and student feedback can receive real-time views and personalised recommendations.

Collect and synchronise data as a matrix $X \in R^{N \times T \times F}$, on the person, time, and feature (e.g., joint positions) coordinates. The dataset is annotated with categories including skill level (beginner, intermediate, advanced) errors in performance, which sets the basis for a supervised training of DL models.

3.1.2 Deep learning modelling

The data was analysed using advanced DL models. CNNs are well suited to use video data to extract spatial features, which are key to evaluating motor skills by identifying key skeletal positions and motion patterns. It can be represented as:

$$Z_{spatial} = CNN\left(X_{video}\right) \tag{1}$$

LSTM networks model temporal dependencies of movement sequences, including dynamic attributes like rhythm, balance, and consistency, and $Z_{spatial}$ refer to extracted spatial features. The LSTM processes sequential spatial features, updating hidden states h_t at each time step t:

$$h_t = f\left(W_h h_{t-1} + W_x x_t + b\right) \tag{2}$$

where W_h, W_x and b are weights and biases, and f is the activation function. The model merges CNN and LST outputs to generate discussions of spatial and temporal student performance aspects.

3.1.3 Feedback delivery

Insights from the DL models are integrated into an interactive feedback system. A user-friendly dashboard provides teachers with visualised data like heat maps of activity levels, skill progression graphs, and error analysis. At the same time, students get real-time feedback, including personalised recommendations and skill-specific corrective actions for immediate adjustments and continuous improvement.

3.2 Proposed SkillNet model

The centrepiece of the proposed framework is the SkillNet model, which analyses spatial and temporal aspects of physical activity. The proposed architecture combines CNNs for extracting spatial features from video data with LSTMs for modelling the temporal dynamics, resulting in a hybrid architecture suited for PE tests.

3.2.1 Input layer

The model accepts synchronised multimodal data from video recordings and wearable devices. Video frames are preprocessed into normalised inputs for the CNN, while wearable data is temporally aligned with the video data to ensure consistency.

3.2.2 Spatial feature extraction layer

The CNN processes video frames, applying convolutional kernels to identify motion patterns such as joint positions, skeletal alignments, and movement trajectories:

$$Z_{ij} = \sum_{k,l} K_{kl} \cdot X_{(i+k)(j+l)} \tag{3}$$

The features captured by this layer are static and critical for estimating the skill at (i, j), where Z_{ij}, denotes the activation at the feature map position (i, j).

3.2.3 Temporal modelling layer

The extracted spatial features $Z_{spatial}$ are passed to an LSTM network, which models temporal dependencies to capture sequential dynamics in movement execution. The LSTM output, represented by the final hidden state, aggregates information across time steps to produce a holistic picture of student performance.

3.2.4 Output layer

The model has two branches to facilitate multi-task learning. The first branch classifies skill levels (e.g., beginner, intermediate, advanced), minimising cross-entropy loss:

$$L_{class} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log \widehat{y_{i,c}} \tag{4}$$

where C is the number of classes $y_{i,c}$, is the actual label, and $\widehat{y_{i,c}}$ is the predicted probability. The second branch predicts performance trends, minimising mean squared error (MSE):

$$L_{trend} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y_i}\right)^2 \tag{5}$$

where y_i and $\widehat{y_i}$ are the actual and predicted trends. The total loss is a weighted sum:

$$L_{total} = \alpha L_{class} + \beta L_{trend} \tag{6}$$

3.2.5 Training process

SkillNet is trained on the labelled dataset using the Adam optimiser with a learning rate of 10^{-4}. Early stopping prevents overfitting, and grid search is used to tune hyperparameters such as batch size and number of CNN and LSTM layers.

3.3 System architecture

The system architecture combines multiple components to enable seamless operation:

- input layer: multimodal data from video and wearable devices
- feature extraction layer: pre-trained CNN for extracting spatial features
- temporal modelling layer: an LSTM network that tracks skill progression over time
- output layer: dual branch structure for skill classification and trend prediction
- feedback system: interactive dashboard for teachers and real-time feedback for students.

PE assessments are revolutionised through a proposed framework integrating advanced DL models, multimodal data collection, and real-time feedback. The system improves teaching effectiveness and learning outcomes by automating evaluations and delivering actionable insights that mitigate the shortcomings of traditional approaches.

4 Experimental Setup

A PE evaluation using the proposed framework was designed experimentally. It includes details of the data collection process, model training and evaluation, and the tools and infrastructure used for implementation.

4.1 Data collection

A comprehensive dataset of motor skills and physical activities was built with the data taken. In the study, 100 students, aged between 10 and 15 years, completed several PE tasks such as running, jumping, throwing, balancing, and agility drills. These activities were selected based on a range of gross motor skills critical for PE assessments' broad spectrum.

- Video data: each activity was captured through strategically positioned high-resolution cameras from multiple angles. The recordings used visual data to analyse the student's body posture, joint articulation, and movement trajectory.

- Wearable data: wearable devices such as accelerometers, gyroscopes, and real-time biomechanical and physiological data were collected using such devices as acceleration, angular velocity, and heart rate variability to gain a quantitative perception of student performance.
- Data synchronisation and labelling: temporal synchronising of video and wearable data was performed to achieve temporal alignment. The dataset was annotated by expert instructors with labels for skill level (beginner, intermediate, advanced) and common performance errors. A final dataset of 10,000 video clips with its associated sensor was created and served as a high-quality training and validation set for the proposed model.

4.2 Model training and evaluation

To evaluate how well we could classify skill levels and predict performance trends, a model, SkillNet, was trained and tested on the collected dataset.

- Training data split: the dataset was split into training and a validation subset, with 80% going into training and 20% going to validation. This split ensured that the model could learn effectively while retaining a portion of the data for unbiased evaluation.
- Model training: it used the SkillNet model to train itself, feeding in synchronised multimodal data. The CNN processed preprocessed and normalised video frames, and the LSTM processed time series wearables. The Adam optimiser is used with a learning rate of 10^{-4}, and early stopping is applied to avoid overshooting.
- Loss function and metrics: the model minimised a multi-task loss function combining cross-entropy loss for skill classification and MSE for trend prediction:

  $$L_{total} = \alpha L_{class} + \beta L_{trend} \tag{7}$$

  The two tasks were balanced by choosing $\alpha = 0.7$ and $\beta = 0.3$.
- Validation and testing: the model's generalisation on unseen data was validated after training. A confusion matrix was used to analyse classification accuracy by skill level, where the model was accurate, and where it could be improved.

4.3 Implementation tools

Computer programs and physical devices are needed to evaluate and learn from the SkillNet model.

- TensorFlow and PyTorch: these frameworks were used to build and train our DL platforms. TensorFlow's simple API tools make model building quick, and PyTorch lets users try different experimental methods through its dynamic graph system.
- OpenCV: processed video data with OpenCV, which involved removing broken frames and adjusting their differences across the footage.
- NumPy and Pandas: it enables the essential calculation of numbers and data management.

- Matplotlib: this program helped us represent our model performance outcomes through loss curves and metrics.
- GPUs: The CNN and LSTM processing speed in our training increased significantly due to NVIDIA Tesla V100 GPUs. The GPUs sped up the processing of big video data when training models through their complex structure.
- RAM and storage: the server features 256 GB of RAM plus 10 TB of storage to process big data while keeping processing times fast.

5 Results and analysis

The test results show that the proposed method successfully automates PE evaluation while offering precise results that work well on large scales and keep students actively involved. This section reviews how well the SkillNet model works, plus the system's strength from real-time feedback while showing its difference from conventional approaches.

5.1 Model performance

With an overall classification accuracy of 89%, the SkillNet model outperformed traditional teacher-based assessments by 17 percentage points, with an average of 72% agreement with expert evaluations, as shown in Table 1 and Figure 2. Finally, the precision and recall metrics for the model were further validated to be reliable for classifying skill levels across different motor skills. Specifically, its intermediate category showed the highest precision of 91%, though recall remained relatively high throughout all levels, minimising the false negatives. It consistently demonstrates the model's ability to have robust and objective evaluations. It found that the confusion matrix had minor misclassification between adjacent skill levels, i.e., beginner and intermediate. In these results, we also show the effectiveness of using a combined CNN-LSTM architecture to extract the spatiotemporal features relevant to physical activities, joint alignment, and movement rhythm.

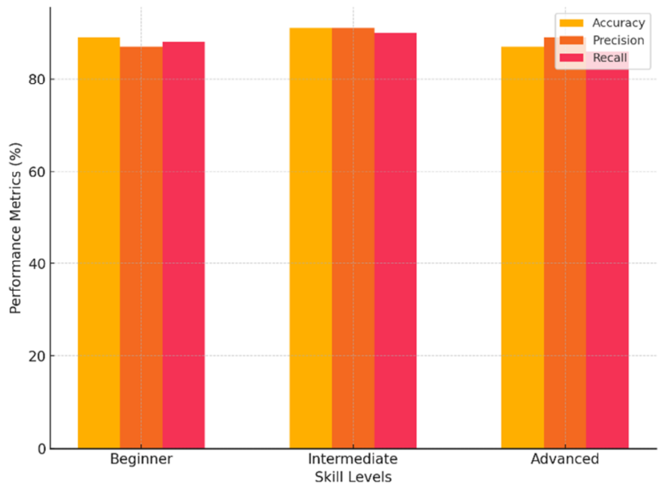
Table 1 Model performance metrics by skill level

Skill level	*Accuracy (%)*	*Precision (%)*	*Recall (%)*
Beginner	89	87	88
Intermediate	91	91	90
Advanced	87	89	86
Overall	89	89	88

5.2 Real-time feedback

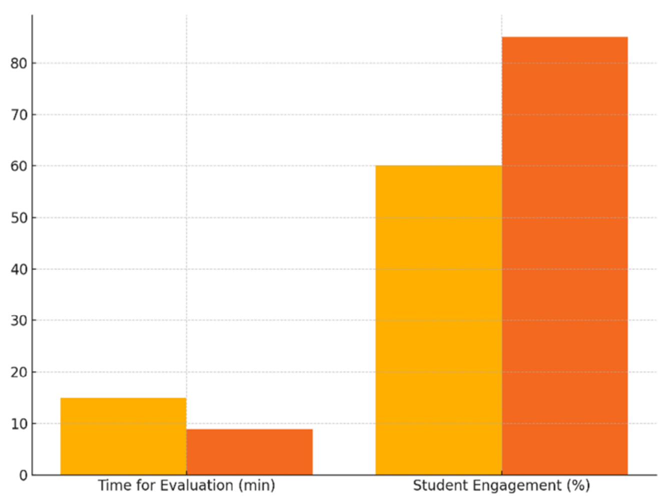
Combining wearable devices with real-time data analysis made feedback faster for students. Students could get prompt feedback because the suggested system scanned activity results in seconds rather than waiting for the 24-hour teachers required for standard evaluations, as shown in Table 2 and Figure 3. Through this update, students could fix their performance right after their training runs, which helped them learn better skills faster.

Figure 2 Model performed against skilled users in three categories: beginners, intermediates, and advanced users (see online version for colours)



Note: The visual shows how the model reliably rates motor skills from beginners to advanced levels.

Figure 3 Comparison of how much time students need to complete their work and stay active during tests under classic evaluation systems and our new framework (see online version for colours)



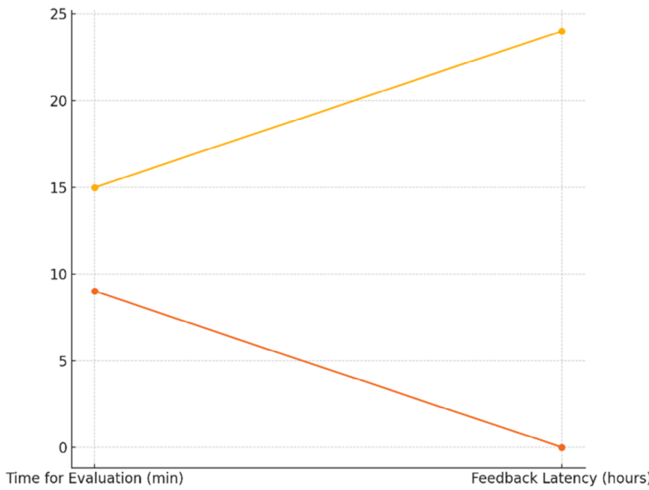
Note: According to the chart, students show enhanced interest in assignments using live feedback tools, indicating faster evaluation times.

Figure 3 displays the system's capacity to shorten the evaluation period and reduce feedback delays. The interactive dashboard transformed how students and teachers worked together in class by showing skills progress over time and suggesting personalised tasks. The feature saved teachers time because it works better than manual evaluation methods for rooms with many students.

Table 2 Impact of real-time feedback on key metrics

Feedback metric	*Traditional assessment*	*Proposed framework*
Time for evaluation (min)	15	9
Average feedback latency	24 hours	Real-time
Student engagement (%)	60	85

Figure 4 Illustration of the reduction in time for evaluations and feedback latency when comparing traditional assessments with the proposed framework (see online version for colours)



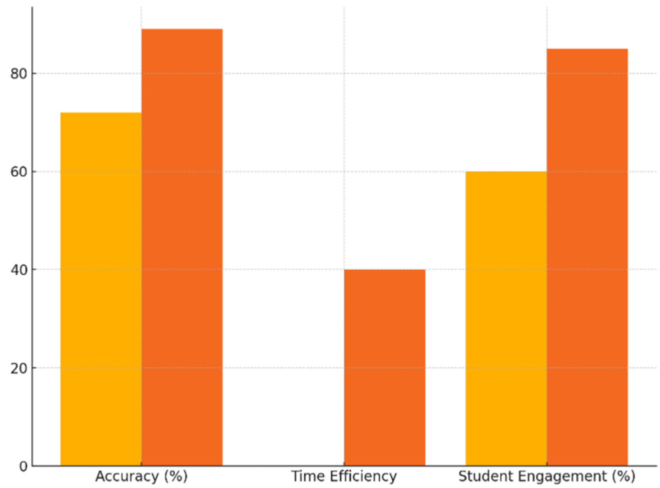
Note: The chart highlights the efficiency improvements achieved through real-time feedback.

The prompt delivery of feedback raised student motivation from 60% to 85% more than traditional teaching approaches. Students appreciated real-time feedback, which helped them set more effective learning goals and stay motivated, as shown in Figure 4.

Table 3 Comparative metrics between traditional and dl-based assessments.

Metric	*Traditional assessment*	*Proposed framework*
Accuracy (%)	72	89
Inter-rater variability	High	Reduced by 35%
Time efficiency	Low	Increased by 40%
Scalability	Limited	High
Student engagement (%)	60	85

Figure 5 Comparison of key measurement areas, such as accuracy, processing time, and student participation, between typical evaluation tools and our proposed deep learning system to demonstrate exceptional results (see online version for colours)



5.3 Comparative analysis

The contrast analysis shows the proposed system produces better results than conventional assessment methods because it gives more precise outcomes and maintains uniformity across tests. It also expands testing capabilities and increases student participation.

- Consistency: by implementing this system, evaluation reliability improved by 35% while ensuring that results remained free from personal bias. Multiple educator opinions may affect assessment results because traditional marking processes depend on human judgement. The SkillNet model solved this problem through predictable evaluation criteria and evaluation decisions based on objective data.
- Scalability: our framework's automated system worked well for all activities and large groups of students. Large classrooms pose a challenge for human observation, but our system shows its worth by handling data for multiple students at once while keeping results exact and dependable.
- Accuracy and efficiency: the system demonstrated 89% precision compared to usual methods, while regular observation technique results remained 72%. With the new system, teachers could spend 40% less time assessing and putting more energy into meeting each student's learning needs.
- Student engagement: the system framework made students more interested in their work by raising engagement from 60% to 85%, according to Table 3 and Figure 5. Students enjoyed monitoring their successes visually and receiving feedback immediately, making them more involved in their learning results.

6 Discussion

The research shows that combining deep learning and wearable devices can powerfully impact PE evaluation methods. The proposed method overcomes existing PE evaluation challenges such as human bias and resource usage plus enables precise live feedback and improved student participation. This section details the effects of our findings alongside obstacles we faced and showcases practical settings for our framework.

6.1 Enhanced accuracy and objectivity

The SkillNet system demonstrated 89% precision in measurement, outperforming teacher-based assessment in achieving 72% agreement with professional evaluations. It shows that the framework gives trustworthy outcomes without personal judgement because observation results are automated. Traditional grading by teachers contains inconsistent results because they each perceive student performance differently. The new evaluation system used standard metrics and data results to resolve this challenge. Combining CNNs for spatial analysis and LSTMs for temporal modelling enabled us to reach this superior accuracy result. CNNs recognised static postures and joint positioning, while LSTMs tracked the dynamic movements, including beat and balance patterns, throughout their duration. Due to its specialised detection features, the model could detect motor skill performance quality in both simple and advanced activities. The model's strong ability to recognise advanced students proved true with its high recall rate. The data matches other studies that show DL technology helps make gymnastics evaluation more exact and unbiased.

6.2 Real-time feedback and classroom dynamics

The proposed framework sped up evaluation results by 40% thanks to its live feedback capacity. To analyse student work, teachers must handle data manually, pushing feedback delivery to hours or days after the assessment. Training effectiveness is impacted because students cannot respond to problems right away. The system combined wearables with real-time data analysis to show teachers' and students' results as they occurred. The system immediately provided students instant feedback about their movements to improve their actions during class. Teachers received precise performance results through heatmaps and trend bars to help them find student errors and adjust their teaching effectively. Most students reacted positively to receiving instant feedback because it increased their interest in studying. With this approach, 85% of students became more involved, but this happened for only 60% of students who used traditional teaching. Students experienced better learning results because they understood what they achieved and what they needed to practice next through their visual performance data.

6.3 Scalability and practical applications

Tests showed that our approach managed different student abilities well alongside various exercises in sizable groups. The system worked better than personal observation because it processed multiple students at once through their data sources. The system's flexible performance properties work well for students in significant school classrooms and athletic and rehab areas. The framework can detect performance in different sports actions, which shows that it works well anywhere. Teachers found this system produced consistent results during all physical education assessments, so it worked well across different activities. The framework's synchronisation capabilities with video and wearable data show promise for using the system across all these areas, including sports, fitness training, and medical rehabilitation.

6.4 Challenges and limitations

While the proposed framework demonstrated significant benefits, specific challenges and limitations must be acknowledged:

- Data privacy and security: the technical recording of student activities creates doubts about student privacy protection and personal data security requirements. Schools must follow European GDPR and US FERPA rules to prevent unauthorised access to sensitive information. Future installations must put strong systems that hide personal information and secure underlying data.
- Hardware and cost barriers: schools with minimal resources find it hard to set up the system because high-resolution cameras cost more, plus wearable devices and computers. The creation of low-priced wearable technology and edge-computing tools helps make our solution more accessible for widespread use.
- Generalisation across contexts: the system success tests happened inside controlled settings with set data information. Our research must confirm its usefulness for different kinds of students, settings, and learning environments. Our model can better handle real-world scenarios by adding diverse people and various activities to our dataset.
- Teacher training: the system's successful use relies on teachers understanding how to use and decode the framework. Successfully using this system in classrooms will require teacher training plus easy-to-follow interfaces.

6.5 Implications for education

The model creates critical new ways to improve education now. Our system automates PE evaluation while giving students immediate feedback, enabling teachers to devote more time to instruction and student learning. Through this system, teachers can customise their teaching methods. Students receive instant feedback to improve their knowledge. All students receive fair assessment results through standardised testing regardless of their strengths or weaknesses. This system works well in many educational spaces, including significant classrooms and specific training courses.

7 Conclusions

This research launched a fresh technique using deep learning methods and wearable instruments to overcome problems linked to regular physical education evaluations. This framework showed strong signs of transforming PE evaluation standards through spatial analysis with CNNs, temporal modelling through LSTMs, and data from wearable tech plus video recordings. Our research showed that this system delivers precise and uniform test results across many students. The SkillNet system delivered better performance at 89% compared to classroom teacher methods, which only matched expert evaluations 72% of the time. The system produces better results by lowering the measurement differences between evaluators by 35%. Students received instant feedback and became more engaged under this program because 85% felt motivated compared to 60% in traditional courses. Our framework works well in many different setups for teaching and training. Our system performed well during activity tests and retained accuracy throughout different skill levels. Through its performance metrics and student-teacher interactivity, this system helps teachers teach better, and students engage more fully with their studies. Even though the system shows strong performance, it needs to address data privacy concerns and the price of physical hardware devices, and testing must be done against many more students. The system's success depends heavily on solving these limitations to work well in different learning environments. Research must grow the dataset base, improve model effectiveness, and examine low-cost ethical use approaches. The proposed PE evaluation method enhances students' and teachers' experiences, saving effort and increasing test precision. Our automated evaluation system helps improve PE delivery while building new uses for technology in education systems.

Declarations

The author declares that he has no conflict of interest.

References

Adambaevna, D.T. (2023) 'Basic theories and concepts in physical education subject', *Galaxy International Interdisciplinary Research Journal*, Vol. 11, No. 11, pp.222–225.

Akker, H.O.P., Jones, V.M. and Hermens, H.J. (2014) 'Tailoring real-time physical activity coaching systems: a literature survey and model', *User Modeling and User-Adapted Interaction*, Vol. 24, pp.351–392.

Alzahrani, A. and Ullah, A. (2024) 'Advanced biomechanical analytics: wearable technologies for precision health monitoring in sports performance', *Digital Health*, Vol. 10, p.20552076241256745.

Barris, S. and Button, C. (2008) 'A review of vision-based motion analysis in sport', *Sports medicine*, Vol. 38, pp.1025–1043.

Bhargava, A. and Rehman, S.U. (2025) 'Beyond the classroom: a comprehensive guide to wearable device security in education', *Wearable Devices and Smart Technology for Educational Teaching Assistance*, pp.331–364, IGI Global Scientific Publishing.

Carling, C., Bloomfield, J., Nelsen, L. and Reilly, T. (2008a) 'The role of motion analysis in elite soccer: contemporary performance measurement techniques and work rate data', *Sports Medicine*, Vol. 38, pp.839–862.

Carling, C., Reilly, T. and Williams, A.M. (2008b) *Performance Assessment for Field Sports*, Routledge, Milton Park, Abingdon-on-Thames, UK.

Chen, H. (2024) 'Advancements in human motion capture technology for sports: a comprehensive review', *Sensors & Materials*, Vol. 36.

Chidambaram, S., Maheswaran, Y., Patel, K., Sounderajah, V., Hashimoto, D.A., Seastedt, K.P., Mcgregor, A.H., Markar, S.R. and Darzi, A. (2022) 'Using artificial intelligence-enhanced sensing and wearable technology in sports medicine and performance optimisation', *Sensors*, Vol. 22, No. 18, p.6920.

Clark, R.A., Mentiplay, B.F., Hough, E. and Pua, Y.H. (2019) 'Three-dimensional cameras and skeleton pose tracking for physical function assessment: a review of uses, validity, current developments and Kinect alternatives', *Gait & Posture*, Vol. 68, No. 7, pp.193–200.

Colucci, L. (2023) *Development of Human Pose Estimation and Machine Learning-Based Algorithms for Assessing Physical Exercise Proficiency*, Politecnico di Torino, Italy.

Delp, S.L., Anderson, F.C., Arnold, A.S., Loan, P., Habib, A., John, C.T., Guendelman, E. and Thelen, D.G. (2007) 'OpenSim: open-source software to create and analyze dynamic simulations of movement', *IEEE Transactions on Biomedical Engineering*, Vol. 54, No. 11, pp.1940–1950.

Edriss, S., Romagnoli, C., Caprioli, L., Zanela, A., Panichi, E., Campoli, F., Padua, E., Annino, G. and Bonaiuto, V. (2024) 'The role of emergent technologies in the dynamic and kinematic assessment of human movement in sport and clinical applications', *Applied Sciences*, Vol. 14, No. 3, p.1012.

Edwards, N.A., Talarico, M.K., Chaudhari, A., Mansfield, C.J. and Oñate, J. (2023) 'Use of accelerometers and inertial measurement units to quantify movement of tactical athletes: a systematic review', *Applied Ergonomics*, Vol. 109, No. 5, p.103991.

Engelmann, J.M. (2014) *Accuracy and Reliability of Peer Assessment of Clinical Skills and Professional Behaviors among Undergraduate Athletic Training Students*, Northcentral University, 8667 E Hartford Dr #100, Scottsdale, AZ 85255, USA.

Etkin, J. (2024) 'Embracing holistic physical education: a pedagogical shift from traditional approaches', *BU Journal of Graduate Studies in Education*, Vol. 16, No. 1, pp.5–15.

Galatanu, C.D. (2024) 'The paradigm change: from confidentiality to collaborative learning', *2024 IEEE International Conference and Exposition on Electric and Power Engineering (EPEi)*, IEEE, pp.579–584.

Gavrila, D.M. (1999) 'The visual analysis of human movement: a survey', *Computer Vision and Image Understanding*, Vol. 73, No. 1, pp.82–98.

Hartmann, F. and Slapničar, S. (2012) 'The perceived fairness of performance evaluation: the role of uncertainty', *Management Accounting Research*, Vol. 23, No. 1, pp.17–33.

Horvat, M., Kelly, L., Block, M. and Croce, R. (2019) *Developmental and Adapted Physical Activity Assessment*, 2nd ed., Human Kinetics, UK.

Kazanskiy, N.L., Khonina, S.N. and Butt, M.A. (2024) 'A review on flexible wearables-Recent developments in non-invasive continuous health monitoring', *Sensors and Actuators A: Physical*, p.114993.

Khusainov, R., Azzi, D., Achumba, I.E. and Bersch, S.D. (2013) 'Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations', *Sensors*, Vol. 13, No. 10, pp.12852–12902.

Kimball, R.S. (2007) *Sixth-Grade Fitness Levels and the FitnessGram Assessment Program*, Liberty University, 1971 University Blvd, Lynchburg, VA 24515, USA.

Knudson, D.V. (2013) *Qualitative Diagnosis of Human Movement: Improving Performance in Sport and Exercise*, Human Kinetics, UK.

Kumar, R. and Kumar, S. (2023) 'Survey on artificial intelligence-based human action recognition in video sequences', *Optical Engineering*, Vol. 62, No. 1, pp.023102–023102.

López-Pastor, V.M., Kirk, D., Lorente-Catalán, E., Macphail, A. and MacDonald, D. (2013) 'Alternative assessment in physical education: a review of international literature', *Sport, Education and Society*, Vol. 18, No. 1, pp.57–76.

Lu, Z. (2023) 'Design of physical health monitoring system based on wearable devices', *2023 International Conference on Computer Science and Automation Technology (CSAT)*, IEEE, pp.263–271.

Lund, J.L. and Kirk, M.F. (2019) *Performance-Based Assessment for Middle and High School Physical Education*, Human Kinetics Publishers, UK.

Lund, J.L. and Veal, M.L. (2013) *Assessment-Driven Instruction in Physical Education: A Standards-Based Approach to Promoting and Documenting Learning*, Human Kinetics, UK.

Meredith, M.D. and Welk, G. (2010) *Fitnessgram and Activitygram Test Administration Manual*, Updated 4th ed., Human Kinetics, UK.

Miller, J.C., Miranda, J.P.P. and Tolentino, J.C.G. (2025) 'Artificial intelligence in physical education: a review', *Global Innovations in Physical Education and Health*, Vol. 10, No. 9, pp.37–60.

Morrow Jr., J.R., Mood, D., Disch, J. and Kang, M. (2015) *Measurement and Evaluation in Human Performance*, 5th ed., Human Kinetics, UK.

Olsen, R.J., Hasan, S.S., Woo, J.J., Nawabi, D.H. and Ramkumar, P.N. (2024) 'The fundamentals and applications of wearable sensor devices in sports medicine: a scoping review', *Arthroscopy: The Journal of Arthroscopic & Related Surgery*.

Pangrazi, R.P. and Beighle, A. (2019) *Dynamic Physical Education for Elementary School Children*, Human Kinetics Publishers, UK.

Pham, T.D. (2021) 'Time–frequency time-space LSTM for robust classification of physiological signals', *Scientific Reports*, Vol. 11, No. 1, p.6936.

Prakash, C., Kumar, R. and Mittal, N. (2018) 'Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges', *Artificial Intelligence Review*, Vol. 49, No. 5, pp.1–40.

Roggio, F., Trovato, B., Sortino, M. and Musumeci, G. (2024) 'A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: a narrative review', *Heliyon*, USA.

Seçkin, A.Ç., Ateş, B. and Seçkin, M. (2023) 'Review on wearable technology in sports: concepts, challenges and opportunities', *Applied Sciences*, Vol. 13, No. 18, p.10399.

SHAPE America – Society of Health and Physical Educators (2018) *PE Metrics: Assessing Student Performance using the National Standards & Grade-Level Outcomes for K-12 Physical Education*, Human Kinetics, UK.

Sherstinsky, A. (2020) 'Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network', *Physica D: Nonlinear Phenomena*, Vol. 404, No. 6, p.132306.

Syaukani, A.A., Hashim, A.H.M. and Subekti, N. (2023) 'Conceptual framework of applied holistic education in physical education and sports: a systematic review of empirical evidence', *Physical Education Theory and Methodology*, Vol. 23, No. 5, pp.794–802.

The Cooper Institute (2017) *FitnessGram Administration Manual: The Journey to MyHealthyZone*, Human Kinetics, UK.

Todorovich, J.R. (2024) *Sport Performance Analytic Methods*, Human Kinetics.

Vanhees, L., Lefevre, J., Philippaerts, R., Martens, M., Huygens, W., Troosters, T. and Beunen, G. (2005) 'How to assess physical activity? How to assess physical fitness?', *European Journal of Cardiovascular Prevention & Rehabilitation*, Vol. 12, No. 2, pp.102–114.

Wei, S. and Wu, Z. (2023) 'The application of wearable sensors and machine learning algorithms in rehabilitation training: a systematic review', *Sensors*, Vol. 23, No. 18, p.7667.

Wulfsohn, J. (2018) *The Agreement, Reliability and Accuracy of a Subjective-Rating for Technique Measurement in Rugby Union after Video-based Training*, Human Kinetics, UK.

The Agreement, Reliability and Accuracy of a Subjective-Rating for Technique Measurement in Rugby Union after Video-Based Training.

Zhang, X. (2021) 'Application of human motion recognition utilizing deep learning and smart wearable device in sports', *International Journal of System Assurance Engineering and Management*, Vol. 12, No. 4, pp.835–843.

Zhou, T., Wu, X., Wang, Y., Wang, Y. and Zhang, S. (2024) 'Application of artificial intelligence in physical education: a systematic review', *Education and Information Technologies*, Vol. 29, No. 7, pp.8203–8220.