International Journal of Human Movement and Sports Sciences 12(4): 636-645, 2024 http://www.hrpub.org
DOI: 10.13189/saj.2024.120404

Analysis of Human Movement Learning Using Machine Learning Techniques

Ahmed Abatal[1,*], Hasnaa Rezki[2], Ahmed Elhamal[3], Adil Korchi[4]

[1]Faculty of Sciences and Techniques, Hassan Premier University, Morocco
[2]Faculty of Sciences, Hassan II University Ain Chock, Casablanca, Morocco
[3]Faculty of Letters and Humanities, Sidi Mohamed Ben Abdellah University, Morocco
[4]Faculty of Juridical, Economic and Social Sciences, Chouaib Doukkali University, Morocco

Received October 29, 2023; Revised May 17, 2024; Accepted June 16, 2024

Cite This Paper in the Following Citation Styles

(a): *[1] Ahmed Abatal, Hasnaa Rezki, Ahmed Elhamal, Adil Korchi , "Analysis of Human Movement Learning Using Machine Learning Techniques," International Journal of Human Movement and Sports Sciences, Vol. 12, No. 4, pp. 636 - 645, 2024. DOI: 10.13189/saj.2024.120404.*

(b): *Ahmed Abatal, Hasnaa Rezki, Ahmed Elhamal, Adil Korchi (2024). Analysis of Human Movement Learning Using Machine Learning Techniques. International Journal of Human Movement and Sports Sciences, 12(4), 636 - 645. DOI: 10.13189/saj.2024.120404.*

Abstract This research explores the transformative potential of integrating machine learning with human movement learning in education, revolutionizing personalized learning experiences. An advanced system, nestled within a dynamic 3D virtual environment, serves as an analysis tool for human motion and a support mechanism for educators and learners. Central to this investigation is addressing challenges in movement modeling and analysis, particularly in diverse educational contexts like music instrument teaching. Through strategic integration of machine learning algorithms, the system anticipates actions, discerns user situations, and offers refined models, providing valuable feedback for learners and aiding teachers in assessments. Beyond technical proficiency, the ambition is to offer a holistic tool enriching human movement analysis across various learning scenarios, aiming to pave the way for personalized learning to become fundamental in education. In Lute lessons, capturing diverse movements from instructors and students is crucial. Multiple renditions of movements by experts contribute to building robust target gesture models. Filtering steps ensure data integrity, identifying and removing errors swiftly. Challenges in automating the detection of playing styles for intelligent music coaching are addressed by introducing a multimodal dataset covering four Lute techniques, enhancing the development of intelligent software. This dataset encompasses a rich array of human movement data within music instrument teaching, recorded in a 3D virtual environment. Each data point is annotated with contextual information, facilitating nuanced analysis of the relationship between human movement and musical expression. This serves as a foundation for applying machine learning algorithms in personalized music education.

Keywords Movement Detection, Motion, Machine Learning, Clustering, Unsupervised Learning

1. Introduction

Nowadays, motion capture is used more and more in multiple fields, such as video games, animated films, virtual reality, sports, medicine, industry and education. Thanks to advances in electronics, Human-Machine Interfaces (HMI) and data processing, it is reasonable to think that capturing, editing and sharing 3D human movements will soon be democratized. This can have a strong education impact in all areas that involve gestures. In fact, the human movement carries information: it is possible, for example, to extract low-level information related to kinetics and dynamics [1].

The gesture also carries semantic information; in the context of verbal or non-verbal communication [2] [3]. In addition to this, it is also possible to infer high-level information, such as emotions [4] [5], intention and action [6] [7]. Tracing the learner's activities involves the generation of a large amount of movement data.

Processing and analyzing this information manually is not an option. The development of automatic processing techniques (e.g. machine learning) can help alleviate this task [8]. These methods can analyze high dimensional data for classification, feature extraction, regression problem solving, etc. [9]. In learning situations, these algorithms have already been widely used to study the actions and behaviors of learners [10].

However, there hasn't been much research done on automatically analyzing student movement to glean pedagogically valuable information. This can be explained by a number of technical and scientific obstacles [11]. The goal of this work is to investigate the requirements, difficulties, and viewpoints associated with the democratization of this strategy and the use of movement, as well as its learning, whether it is the learning's endpoint or a first step toward the ultimate outcome. After that, we'll propose a system for learner's motion analysis. We'll also talk about machine learning and its applications to learning and learner analysis.

2. Related Works

The following table summarizes the objectives of certain research studies in the analysis of human movement learning using machine learning. It also outlines the strengths and weaknesses of the associated papers.

Table 1. Comparative Analysis of Studies on Human Movement Learning Utilizing Machine Learning

Author's name and year of publication	Aim of the paper	Strength	Weakness
Salonen, Sanna. (2021). [12]	This literature review explores challenges and opportunities in using motion capture for 3D animation, emphasizing the inverse relationship between system simplicity and data quality. It suggests the need to simplify systems without compromising data quality, with machine learning offering potential for understanding human motion nuances. The study highlights the importance of both simple systems for consumer applications and complex systems for high-budget productions requiring reliable, high-quality data.	This study excels in its thorough examination of challenges and opportunities in motion capture for 3D animation. Its strengths lie in covering essential fundamentals, addressing key challenges like accessibility and data quality, and recognizing the dual importance of both simple systems for consumers and complex systems for high-budget productions. The study's insights offer valuable considerations for advancing motion capture technology in the field of 3D animation.	One potential weakness of this study is its limited exploration of strategies to reconcile the inverse relationship between system simplicity and data quality in motion capture. The study could benefit from providing specific examples or case studies to illustrate its findings and enhancing the discussion on machine learning solutions with concrete insights into their current applications.
Pan, Indranil., et al., (2020) [13]	This study aims to explore recent advancements in integrating simulations, machine learning, and statistics, emphasizing data-centric engineering approaches. It reviews key trends, application scenarios, and opportunities in this emerging field, outlining challenges in achieving integration. Additionally, the study discusses translational bottlenecks and addresses the long-term upskilling requirements for the workforce and future graduates in this interdisciplinary landscape.	This study excels in comprehensively exploring recent advancements in integrating simulations, machine learning, and statistics. It stands out for its inclusive view of mechanistic models to data-driven approaches, emphasizing the transformative impact of hybrid, data-centric engineering. The study effectively reviews key trends, application scenarios, and opportunities while realistically addressing challenges and highlighting the importance of upskilling for both the current workforce and future graduates.	A potential weakness of this study is its identification of challenges and bottlenecks in integrating simulations, machine learning, and statistics without offering specific strategies for overcoming these obstacles. Additionally, the discussion on translational aspects and upskilling requirements lacks concrete examples or case studies, which could enhance the study's practical applicability. Addressing these aspects could strengthen the overall study.
Lannan, Nate, et al. (2020) [14]	This study aims to improve depth-based motion capture (D-Mocap) quality by combining deep learning and Tobit Kalman filtering to address challenges such as depth sensing limitations and self-occlusion. Using a trained convolutional autoencoder, it recovers valid human motion from corrupted input, and the Tobit Kalman filter adds kinematic and dynamic constraints. The study explores two structural paradigms for handling various data errors, demonstrating effectiveness in both simulated and real-world human motion data.	This study innovatively enhances depth-based motion capture (D-Mocap) quality for applications like biomedicine and 3D animation. By combining deep learning and Tobit Kalman filtering, it addresses challenges like depth sensing limitations and self-occlusion. The use of a trained convolutional autoencoder enables the recovery of valid human motion from corrupted input. The study's versatility is demonstrated in handling various data errors, showing effectiveness in both simulated and real-world human motion data. Overall, it contributes significantly to motion capture and analysis.	A potential weakness of this study is the limited discussion on potential drawbacks or scenarios where the proposed approach might face challenges. A more explicit exploration of its limitations and a detailed comparison with existing methods could enhance the study's completeness and provide a more balanced understanding of its findings.

Table 1 continued

Orton, Indigo JD. (2020) [15]	This dissertation enhances early detection of psychological distress, particularly depression, through novel meta-features in the bodily modality, such as speed. The method involves extracting pose estimation, detecting gestures, and aggregating meta information for prediction. The study introduces a dataset of 65 videos for feature development, demonstrating the effectiveness of these new features with an 82.70% F1 score in predicting depression.	This study excels in innovatively detecting psychological distress, specifically depression, through novel meta-features in the bodily modality. The systematic methodology, including pose estimation and gesture detection, proves effective, as demonstrated by an 82.70% F1 score in predicting depression within a new dataset of 65 video recordings.	A potential weakness of this study is the limited exploration of drawbacks or challenges associated with the introduced approach. The focus on strengths and results could be balanced with a more comprehensive examination of potential limitations and a direct comparison with existing methods for a clearer understanding of the proposed approach.
Jha, Debesh, et al. (2022) [16]	This paper explores the impact of ubiquitous sensors and IoT in soccer, emphasizing real-time player tracking using machine learning and video processing. It discusses FIFA's approval of tracking systems, enabling performance evaluation during games and practices. The study reviews video analytics, summarizes real-time algorithms, and explores crowdsourcing, tactical performance, and distributed computing in video analytics, proposing future research perspectives.	This paper's strength is in its thorough examination of how ubiquitous sensors and IoT technologies, particularly in soccer, impact real-time player tracking using machine learning and video processing. It discusses FIFA's approval of tracking systems, increased data collection during games and practice, and explores video analytics, crowdsourcing, tactical performance, and distributed computing. The proposed future research perspectives enhance the paper's overall strength.	The paper has a limited discussion of the challenges and practical drawbacks associated with the use of ubiquitous sensors and IoT technologies in soccer. A more detailed exploration of potential limitations and concrete examples or case studies illustrating the practical application of these technologies could provide a more balanced perspective and enhance the paper's overall analysis.
Ng, Kia-Chuan. (2004) [17]	The paper explores ongoing developments like stage augmentation with virtual and augmented realities and investigates gesture analysis correlating musical and physical gestures in interactive multimedia performances.	This paper introduces the Music via Motion (MvM) framework, mapping physical movements to multimedia events, with practical implementations like virtual musical instrument interfaces. The inclusion of a distributed multimedia-mapping server enhances the paper's robustness, and ongoing developments, such as stage augmentation, showcase its forward-looking perspective in interactive multimedia performances.	The limited discussion of challenges or potential drawbacks associated with the Music via Motion (MvM) framework. A more detailed exploration of limitations, practical challenges, or areas where the framework might face constraints could provide a more balanced view. Additionally, specific examples or case studies illustrating instances where the framework may encounter difficulties in real-world applications would enhance the paper's completeness.
Pietro Picerno (2019) [18]	This paper proposes using smartphones and exergame controllers as BYOD solutions for interactive learning in sport and exercise sciences. It focuses on capturing data during physical activities for real-time student feedback and teacher assessment. The conceptual framework integrates these devices into an e-learning platform with Cloud and Fog Computing architecture.	It innovatively suggests using smartphones and exergame controllers as BYOD solutions for interactive learning in sport and exercise sciences. It captures real-time data during physical activities, providing immediate student feedback and enabling teacher assessment. The proposed framework integrates into an e-learning platform with Cloud and Fog Computing architecture, showcasing forward-thinking potential.	The paper lacks a detailed exploration of potential challenges and drawbacks associated with implementing smartphones and exergame controllers as BYOD solutions in sport and exercise sciences interactive learning. A more comprehensive discussion of limitations and practical considerations regarding the proposed e-learning platform with Cloud and Fog Computing architecture would enhance the overall robustness of the paper.

As shown in the related works, Salonen [12], delves into challenges in 3D animation motion capture, emphasizing the delicate balance between system simplicity and data quality. While excelling in fundamental coverage, the study could benefit from a more detailed exploration of strategies to reconcile the inverse relationship between system simplicity and data quality. In 2020, Pan [13] comprehensively explores the integration of simulations, machine learning, and statistics in data-centric engineering, lacking specific strategies to overcome integration challenges and needing more concrete examples for practical applicability. Lannan et al. [14] innovatively enhance depth-based motion capture with deep learning and Tobit Kalman filtering, demonstrating versatility but lacking a more explicit exploration of potential drawbacks and a detailed comparison with existing methods. Orton's study [15] innovatively detects psychological distress through novel meta-features but falls short in a comprehensive exploration of drawbacks or challenges associated with the introduced approach. Jha et al. [16] thoroughly examine the impact of ubiquitous sensors and IoT in soccer, yet the paper could benefit from a more detailed discussion of challenges or practical drawbacks associated with these technologies. Ng Kia-Chuan [17] explores developments in interactive multimedia performances, lacking a detailed discussion of challenges or potential drawbacks associated with the Music via Motion framework. Picerno [18] proposes using smartphones and exergame controllers for interactive learning in sport and exercise sciences, with the paper lacking a detailed exploration of potential challenges and drawbacks associated with this approach.

3. Materials, Methods and System Design

Our methodology follows the proposal of an aid system shown in Figure 1. In this research, we explore the transformative impact of integrating machine learning with human movement learning in education for personalized learning experiences. It aims to introduce a technologically advanced system for analyzing human motion in a 3D virtual environment, addressing challenges in movement modeling, especially in music instrument teaching [19]. Leveraging machine learning algorithms on a dataset, the system seeks to predict actions, detect user situations, and provide valuable feedback for learners and teachers. The overarching goal is to enhance the analysis of human movement in diverse learning scenarios, emphasizing the resolution of technical and scientific challenges. This includes the novel application of machine learning for estimating the excitation point in a classical Lute instrument, considering the external factors influencing sound production and the player's physiology.

In the context of Lute lessons, capturing diverse movements from both the instructor and the student is essential. Recording multiple renditions of the same movement by various experts contributes to building a robust target gesture model [20]. During data capture, potential errors from devices and transmission media necessitate a filtering step to ensure data integrity. If the trace shape is known, regular expressions help correct errors. Swift identification and removal of corrupted data are achieved through outlier detection. The least squares method enables a nuanced comparison of experimental data with a mathematical model. Detecting breaks in movement continuity allows for error correction through extrapolation from adjacent points, potentially prompting modifications or deletions [21]. Concurrently, automating the detection of musical instrument playing styles for intelligent music coaching shows promise but faces early-stage challenges due to limited real-world datasets [22]. To address this, the study introduces a multimodal dataset with some video and audio samples from the web, covering four Lute techniques. Authentically mimicking real-world scenarios, the dataset enhances the development of intelligent software for assessing Lute playstyles, accommodating real-world diversities with different instruments and amplifiers.

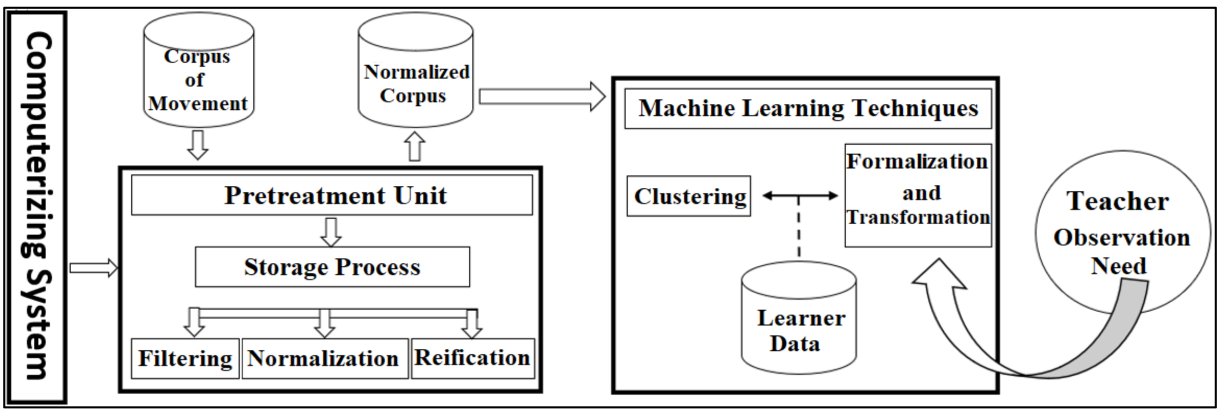

Figure 1. Proposal of an aid system for the analysis of human movements (case of music instrument teaching)

This dataset captures a rich and diverse array of human movement data within the context of music instrument teaching and encompasses a comprehensive range of movements, including finger placements, strumming patterns, and chord transitions, recorded within a 3D virtual environment. The dataset sources input from various skill levels and playing styles to ensure its representativeness and inclusivity. Each data point is annotated with relevant contextual information, such as skill level, tempo, and musical context, providing a holistic view of the movements. The inclusion of such detailed annotations aims to facilitate nuanced analysis and foster a deeper understanding of the intricate relationship between human movement and musical expression [23]. The careful curation of this dataset serves as a crucial foundation for the application of machine learning algorithms, including the Isolation Forest, to uncover hidden patterns and insights in the realm of personalized music education.

Data Description and Statistics

The dataset comprises some distinct categories representing different banjo playing techniques. These techniques encompass legato, hammer-on, alternate picking, sweep picking, tapping, vibrato, pull-off, slide, and bend, each distinguished by its unique style and method of execution. We then, furnish general insights into the dataset, concerning its magnitude and the average time duration for recordings of each of the nine banjo techniques. The collective size of the dataset, housed in the "data" directory, delineates the total size for each file group and the average duration and quantity of recordings per technique class.

In total, we employ three distinct banjos, all of which are outfitted with humbucker pickups. These banjos include Gibson Mastertone RB-250 and RB-800, the Carvin DC-40 Gibson Mastertone RB-250 and RB-800, and the Gold Tone CC-100R and Gold Tone OB-250. This selection of banjos offers a diverse range of sounds for our investigation, enabling us to analyze banjo-playing techniques across various contexts. To replicate the effects of three different amplifiers and further enrich the sound quality, we utilize an amplifier simulation profiler, Kemper Profiler. Specifically, we utilize the amplifier simulations Banjo Preamp Simulator, Tone Shaping Pedals, and Cabinet Simulators to simulate the characteristics of Richwood Banjo Amplifiers, Fender Banjo Amplifiers, and Deering Banjo Amplifiers, respectively.

Machine Learning Experiments

In this section, we outline our training setup for the experiments and describe the two different approaches utilized. The first approach involves developing an SVM classifier, while the second approach employs a Convolutional Neural Network (CNN). For the experiments, we exclusively use the audio and video files from the dataset.

In the Support Vector Machine approach, the audio and video files are segmented into second durations. From each segment, a 125-dimensional feature vector is extracted using the OpenCV (Open Source Computer Vision) and pyAudioAnalysis libraries. This feature vector is computed by averaging and calculating the standard deviation from 1545-dimensional short-term feature vectors. These short-term feature vectors are extracted from each segment at a resolution of 65 ms without any overlap [24]. The resulting 175-dimensional feature vectors serve as input to the SVM model. We opt for the Gaussian activation function as the kernel function, and for each experiment, the parameters for C and gamma are determined through hyperparameter optimization on predefined sets.

Convolutional Neural Networks (CNN)

In this approach, mel-spectrograms are generated from the second video and audio segments using STFT window in terms of capturing temporal information of 65 ms without any overlapping. A total of 145 mel-frequency bins are adopted, resulting in a 3D representation of size 25 × 1145 (timestamps × frequency bins) for each 1-second audio segment. The CNN architecture comprises four convolutional layers followed by a 3-layer linear classifier. Each convolutional layer utilizes 5 × 5 kernels with a stride of 1 × 1 and padding equal to 2, maintaining the spatial dimensions unchanged. The number of channels is doubled in each convolutional layer, starting from 32 and reaching 256 channels. Subsequently, 2D Batch Normalization is applied, followed by a LeakyReLU activation function. Additionally, the spatial dimensions are halved in each convolutional layer using 2D Max Pooling with a kernel size of 2 × 2. The output of the last convolutional layer is a feature map with dimensions 256 × 1 × 8, which, when flattened, yields a resulting 2048-dimensional feature vector [25]. Three linear layers are employed to map the 2048-dimensional feature vector to the final 9-dimensional vector. Each linear layer consists of a linear mapping followed by a LeakyReLU activation function. The output dimensions of these three linear layers are 1048, 256, and 9, respectively.

Figure 1 illustrates a comprehensive computerized aid system within the context of music instrument teaching, facilitating the analysis of human movements throughout the music instruction process and specifically tailored for this purpose. The system utilizes the created dataset to enhance its analytical capabilities.

Starting with the creation and normalization of a movement corpus, involving data filtering, normalization, and reification in a dedicated unit, the core phase employs machine learning components for clustering, formalization, and data transformation. While using various algorithms aligned with pedagogical goals may seem advantageous, it introduces complexities, computational challenges, and risks of overfitting. Therefore, careful consideration of trade-offs and resource implications in system design is crucial. The formalization phase translates identified patterns into structured representations, and data transformation analyzes movements for actionable insights, customizing music education [26]. To construct the movement corpus, the initial step involves choosing a capture method aligned with precision criteria, congestion, and recording time. Ensuring data homogeneity is essential for machine learning algorithms, and various data mining methods can fill gaps in the corpus. Reification of data may be necessary to address teachers' observation needs, transforming low-level information into higher-level characteristics. Segmenting movements into subsets that repeat within the corpus identifies recurring elements, facilitating a standardized corpus after applying these pretreatments [27].

Machine learning algorithms need data to be homogeneous with regard to their dimensions. In this case, there are several data mining methods to fill gaps in the corpus, such as deleting incomplete data, calculating the maximum likelihood, or even estimating missing data points from the reference model, made up of existing points. In order to take into account the observation needs of teachers, reification of the data may prove necessary (for example, moving from low-level information, such as speed or acceleration, to higher-level characteristics such as the emotion of the learner) [28]. Such transformations have been regularly used in the study of gesture, allowing on the one hand, to have more significant data and, on the other hand, to reduce their dimensionality. It is also possible to segment the movement into subsets which are repeated within the corpus. This representation identifies the recurring elements of a gesture, for example, the chords of a Lute piece. The presence (or the absence) of one of these movements can be decisive during learning. Once these pretreatments have been applied, a standardized corpus can be obtained.

An important step is the formalization, the transformation of the teacher's needs into clustering properties or classes for supervised learning. The requirements in terms of observation can be varied. For a Lute teacher, for example, the goal may be to analyze a chord progression that sounds wrong when the student makes it. Therefore, the professor's expertise allows him to direct his observation on the transitions between these agreements [29].

Observing the gesture in real time can be complicated, due to the position of the hands and fingers, thereby blocking the view. From the need (e.g., observe the transitions between chords), we can select the important parts of the data, in order to focus the algorithms on the study of this part of the movement.

Machine learning algorithms, with a preference for unsupervised learning methodologies, are harnessed to fulfill teachers' observational requisites. A pivotal tool in this context includes the Isolation Forest algorithm, K-Means Clustering, Hierarchical Clustering, Self-Organizing Maps, or Principal Component Analysis, renowned for their efficacy in anomaly detection and proving indispensable in scenarios where data lacks labels [30]. These algorithms categorize students into clusters tailored to specific observation needs. In the context of lute teaching, they adeptly pinpoint anomalies in students' playing techniques during machine learning-based instruction. By identifying deviations from established norms, instructors can furnish targeted feedback and propose personalized exercises [31]. The categorization of students into clusters based on playing styles enables the creation of bespoke learning paths, contributing to a dynamic and adaptive lute instruction experience. The algorithm facilitates autonomous discovery of latent patterns, aligning with unsupervised learning principles of clustering and dimensionality reduction. The primary objective is to discern recurrent behavior types based on observation needs, enabling adaptive teaching strategies. The automation of algorithm parameters, underscored for efficiency while exercising caution to prevent a substantial increase in learning time, is grounded in contextual nuances and the unique requirements of teachers. Ultimately, teachers acquire valuable insights to optimize the learning experience, providing specific exercises, such as those emphasizing chord transitions, in response to instances where a learner's movements deviate from established norms.

In summary, the intent of this system is to furnish a supportive tool for the examination of human movement within learning contexts. The discernible challenges, encompassing both technical and scientific aspects, permeate the various stages and modules of the application. Collaborative engagement with the teacher is imperative in the choice and manipulation of algorithms, cluster identification, and result presentation. This collaborative link is pivotal, determining the utilization of data and shaping the outcomes presented to the teacher.

4. Data Analysis and Interpretation

The box plot depicted in Figure 2 describes how the rate of success gets distributed when the participants are grouped based on the three experience levels: Beginner, Advanced, and Intermediate. This is an interesting observation, as not much difference in the success rates for all experience levels exists where the differences in the median are very small. Advanced participants have the least variability of success, represented by the length of the box and whiskers, while Intermediate participants have the most variability. This suggests that while experience did not really differ between the medians of success rate, it did affect the consistency of success.

Figure 3 shows the kernel density estimate over the top of a histogram of the success rate distribution across the entire group of players. This was data plotted with normal distributions, but it was slightly shifted to the left, hence showed that success rates at the lower end of the spectrum were recorded more in subjects. This skew may mean that, in general, while most reach a moderate level of success, a subset may not, and this may be related to underlying issues or factors in the study that contribute to lower performance. Figure 4 shows the number of the participants in each of the experience categories. The three categories are spread almost evenly, hence balance the representation during analysis.

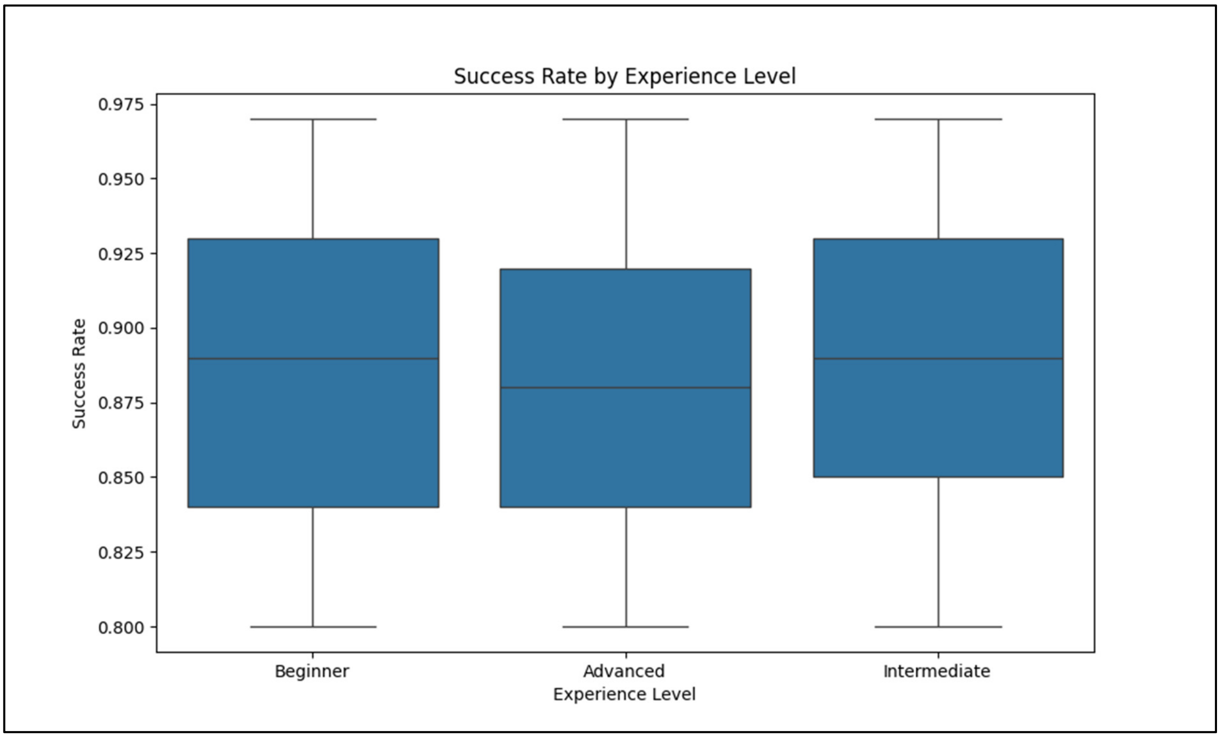

Figure 2. Success Rate by Experience Level

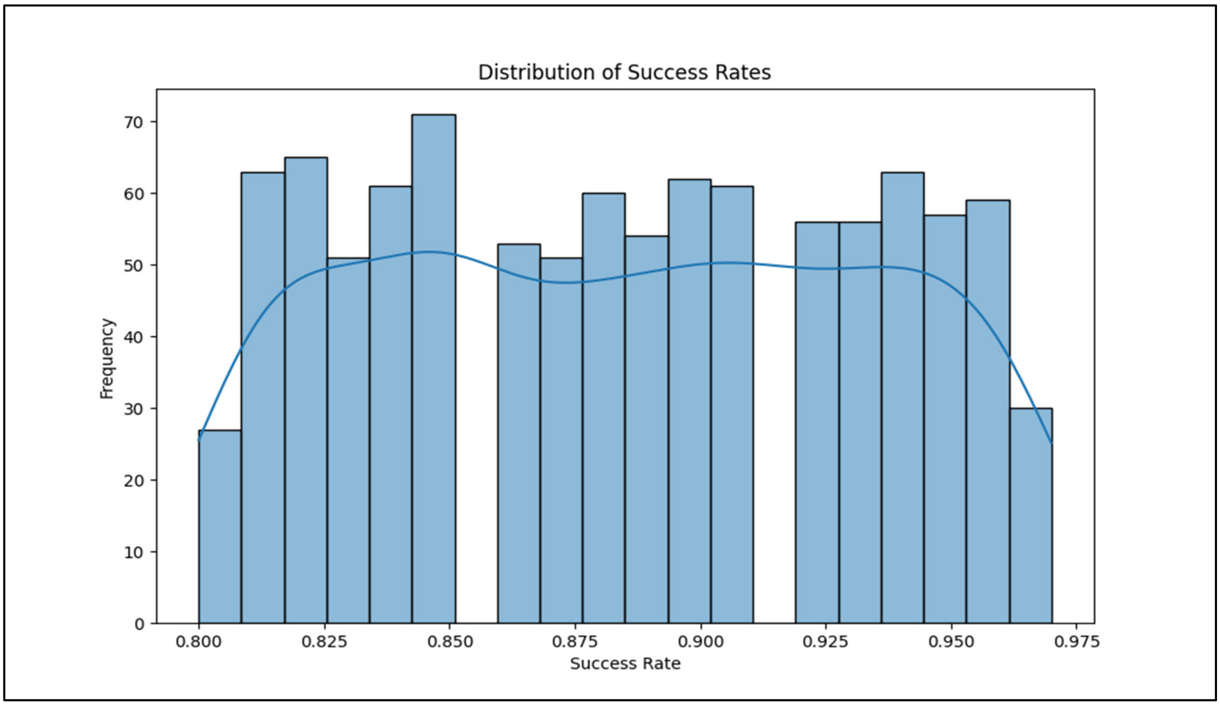

Figure 3. Distribution of Success Rates

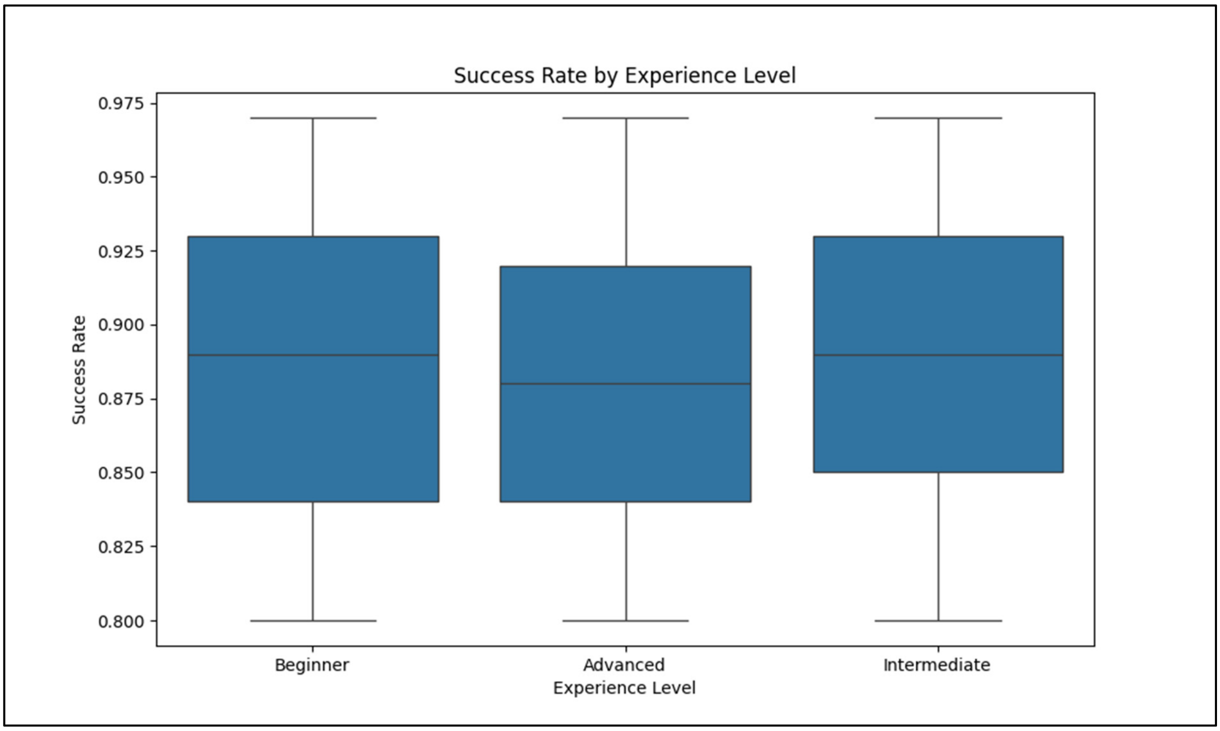

Figure 4. Participant Experience Level Distribution

Linking Graphs to Dataset Insights

Small differences in the number of participants may not matter in analyzing the success rate at an overall level, but this is something that, interestingly, when more minute details are considered, needs to be mentioned with respect to how experience impacts learning outcomes in this instance of movement learning, supported by machine learning techniques. Looking at the dataset and the graphs presented, it is evident that the success rate of the use of machine learning to analyze movements in education has very weak correlation strength with the number of years one might have in experience.

Thus, this rhymes with the argument of the study above, focusing on coming up with a strong system can cater to all users' range of the system with different familiarity levels of the subject of the system. It means that success, when consistent, reaches across experience and points at potential democratization of the system in enabling personalized learning experiences. The distribution of success rates will also bring into focus the need to adjust the system or provide additional assistance for those scholars to reach higher success rates, which his or her peers can easily achieve.

Such kind of results could be taken as a kind of advancement in the future for the system with the design of educators and learners at the heart, ensuring the system serves all users but majorly those finding the curve steeper [32]. Finally, the dataset and the .json files that accompany the publication include visualizations that underpin the transformational potential of machine learning to integrate with human movement learning. Experience level of proficiency has little effect on the median success rate, though it shows great individual spread; and this suggests some potential for personalized system refinements to support learners across experience levels of proficiency. The insight is established for continuous improvement in the educational tools, making sure that they cater to the diversity in the needs of students and educators taking studies in different educational contexts.

5. Discussion

The research explores the transformative integration of machine learning with human movement learning, emphasizing personalized experiences, particularly in practice-oriented teaching in sport sciences. The proposed system, tailored for music instrument teaching, predicts actions, detects user situations, and provides feedback using a curated multimodal dataset. Incorporating devices like Kinect, the study applies machine learning algorithms to enhance the analysis of human movement. The proposed computerized aid system, designed for distance learning, includes a dedicated unit for creating a movement corpus and employs clustering, formalization, and data transformation. Unsupervised learning methodologies categorize students into clusters based on playing styles, enabling adaptive teaching strategies and optimizing the learning experience in diverse scenarios. The study aims to overcome technical challenges, offering a comprehensive and personalized tool for effective education in sport sciences and music.

6. Conclusions

Learning gestures has become a crucial subject of study these days. For the same gesture, the teaching methods can differ from one person to another. In addition, although a subjective assessment is sufficient for certain areas, it is essential to be able to provide an exhaustive assessment of the entire procedure, in order to avoid errors that can be critical. The creation of tools dedicated to automatic movement analysis makes it possible to assist the user in his task of learning, observation and evaluation. Currently, it seems that 3D captured movements are rarely used effectively, in order to generate high-level feedback on learning, both for the learner and the teacher. The cost of acquisition systems, as well as the great dimensionality and heterogeneity of the data produced, are two major obstacles to this development. However, the use of proven machine learning algorithms in the analysis of human actions and behavior could help with this task.

The development of such a system poses technical and scientific challenges, including creating a problem-specific corpus of movements, handling heterogeneous data, managing data dimensionality, formalizing teacher observation needs, and selecting suitable algorithms. The system must be user-friendly for non-experts and undergo rigorous testing with end users (teachers and learners) to validate its usability. While identified challenges exist, the extent to which the system can address them remains uncertain. The experimentation and evaluation phase will determine the system's success in overcoming these challenges. If successful, there is potential for extending the system to other domains.

REFERENCES

[1] Larboulette, C., & Gibet, S., "A review of computable expressive descriptors of human motion," In Proceedings of the 2nd International Workshop on Movement and Computing, pp. 21-28, 2015,https://doi.org/10.1145/2790994.2790998

[2] Doesborgh, S. J., van de Sandt-Koenderman, M. W., Dippel, D. W., van Harskamp, F., Koudstaal, P. J., & Visch-Brink, E. G., "Effects of semantic treatment on verbal communication and linguistic processing in aphasia after stroke: a randomized controlled trial," Stroke, vol. 35, no. 1, pp. 141-146, 2004, https://doi.org/10.1161/01.STR.0000105460.52928.A6

[3] Hou, J., & Aoki, Y., "A real-time interactive non-verbal communication system through semantic feature extraction," In Proceedings. IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, 2002, vol. 2, pp. 425-428. https://doi.org/10.1109/ICME.2002.1035629

[4] Jia, J., Zhou, S., Yin, Y., Wu, B., Chen, W., Meng, F., & Wang, Y., "Inferring emotions from large-scale internet voice data," IEEE Transactions on Multimedia, vol. 21, no. 7, pp. 1853-1866, 2018, https://doi.org/10.1109/TMM.2018.2887016

[5] Tai, C. H., Chung, K. H., Teng, Y. W., Shu, F. M., & Chang, Y. S., "Inference of Mood State Indices by Using a Multimodal High-Level Information Fusion Technique," IEEE Access, vol. 9, pp. 61256-61268, 2021, https://doi.org/10.1109/ACCESS.2021.3073733

[6] Sun, Ruiqi, et al., "Human action recognition using a convolutional neural network based on skeleton heatmaps from two-stage pose estimation," Biomimetic Intelligence and Robotics, vol. 2, no. 3, 2022, https://doi.org/10.1016/j.birob.2022.100062

[7] Fussell, Stephanie G., and Dothang Truong, "Accepting virtual reality for dynamic learning: An extension of the technology acceptance model," Interactive Learning Environments, vol. 31, no. 9, 2020, https://doi.org/10.1080/10494820.2021.2009880

[8] Ghaddar, Bissan, and Joe Naoum-Sawaya, "High dimensional data classification and feature selection using support vector machines," European Journal of Operational Research, vol. 265, no. 3, pp. 993-1004, 2018, https://doi.org/10.1016/j.ejor.2017.08.040

[9] Pappu, V., & Pardalos, P. M., "High-dimensional data classification," Clusters, Orders, and Trees: Methods and Applications, Springer Optimization and Its Applications, vol. 92, pp. 119-150, 2014,https://doi.org/10.1007/978-1-4939-0742-7_8

[10] Rahman, C. M., & Rashid, T. A., "A new evolutionary algorithm: Learner performance based behavior algorithm," Egyptian Informatics Journal, vol. 22, no. 2, pp. 213-223, 2021, https://doi.org/10.1016/j.eij.2020.08.003

[11] Sminchisescu, C., "3d human motion analysis in monocular video: techniques and challenges," Human Motion: Understanding, Modelling, Capture, and Animation, vol. 36, pp. 185-211, 2008, https://doi.org/10.1007/978-1-4020-6693-1_8

[12] Salonen, Sanna, "Motion capture in 3D animation," BS thesis. 2021, https://trepo.tuni.fi/handle/10024/132406

[13] PAN, Indranil, MASON, Lachlan R., et MATAR, Omar K. "Data-centric Engineering: integrating simulation, machine learning and statistics. Challenges and opportunities," Chemical Engineering Science, vol. 249, p. 117271, 2022, https://doi.org/10.1016/j.ces.2021.117271

[14] Lannan, Nate, et al., "Human motion enhancement using nonlinear Kalman filter assisted convolutional autoencoders," 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), Cincinnati, OH, USA, 2020, pp. 1008-1015,https://doi.org/10.1109/BIBE50027.2020.00171

[15] Orton, Indigo JD. "Vision based body gesture meta features for Affective Computing." arXiv preprintarXiv:2003.00809, 2020, https://doi.org/10.48550/arXiv.2003.00809

[16] Jha, Debesh, et al., "Video Analytics in Elite Soccer: A Distributed Computing Perspective." 2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM), Trondheim, Norway, 2022, pp. 221-225, https://doi.org/10.1109/SAM53842.2022.9827827

[17] Ng, Kia-Chuan, "Music via motion: transdomain mapping

of motion and sound for interactive performances," Proceedings of the IEEE, vol. 92, no. 4, pp. 645-655, 2004, https://doi.org/10.1109/JPROC.2004.825885

[18] Picerno, P., Caliandro, P., Iacovelli, C., Simbolotti, C., Crabolu, M., Pani, D., ... & Cereatti, A., "Upper limb joint kinematics using wearable magnetic and inertial measurement units: an anatomical calibration procedure based on bony landmark identification," Scientific reports, vol. 9, no. 1, 14449, 2019, https://doi.org/10.1038/s41598-019-50759-z

[19] Shimada, Soshi, et al., "Neural monocular 3d human motion capture with physical awareness," ACM Transactions on Graphics (ToG), vol. 40, no. 4, pp. 1-15, 2020, https://doi.org/10.1145/3450626.3459825

[20] Ho, Indy Man Kit, Kai Yuen Cheong, and Anthony Weldon. "Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques." Plos one, vol. 16, no. 4, 2021, https://doi.org/10.1371/journal.pone.0249423

[21] Fan, X., Cai, Y., Yang, Y., Xu, T., Li, Y., Zhang, S., & Zhang, F., "Detection of scene-irrelevant head movements via eye-head coordination information," Virtual Reality & Intelligent Hardware, vol. 3, no. 6, pp. 501-514, 2021, https://doi.org/10.1016/j.vrih.2021.08.007

[22] Beyan, C., Shahid, M., & Murino, V., "RealVAD: A real-world dataset and a method for voice activity detection by body motion analysis," IEEE Transactions on Multimedia, vol. 23, pp. 2071-2085, 2020, https://doi.org/10.1109/TMM.2020.3007350

[23] Todd, N. P. M., "The kinematics of musical expression," The Journal of the Acoustical Society of America, vol. 97, no. 3, pp. 1940-1949, 1995, https://doi.org/10.1121/1.412067

[24] Gopalakrishnan, A., Mali, A., Kifer, D., Giles, L., & Ororbia, A. G., "A neural temporal model for human motion prediction," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12116-12125, 2019, https://openaccess.thecvf.com/content_CVPR_2019/html/Gopalakrishnan_A_Neural_Temporal_Model_for_Human_Motion_Prediction_CVPR_2019_paper.html

[25] Miura, T., Kaiga, T., Shibata, T., Tajima, K., & Tamamoto, H., "Low-dimensional feature vector extraction from motion capture data by phase plane analysis," Journal of Information Processing, vol. 25, pp. 884-887, 2017, https://doi.org/10.2197/ipsjjip.25.884

[26] Tak, Ankur. "Big Data Analytics in Healthcare: Transforming Information into Actionable Insights," Journal of Health Statistics Reports. SRC/JHSR-121. pp. 116, 2023, https://doi.org/10.47363/JHSR/2022

[27] Gibet, Sylvie, et al., "Corpus of 3D natural movements and sign language primitives of movement," Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop Bielefeld, Germany, September 17–19, 1997 Proceedings. Springer Berlin Heidelberg, 1998, https://doi.org/10.1007/BFb0052993

[28] Haerens, Leen, et al., "Observing physical education teachers' need-supportive interactions in classroom settings," Journal of Sport and Exercise Psychology, vol. 35, no. 1, pp. 3-17, 2013, https://doi.org/10.1123/jsep.35.1.3

[29] Keller, Robert, et al., "Automating the explanation of jazz chord progressions using idiomatic analysis," Computer Music Journal, vol. 37, no. 4, pp. 54-69, 2013, https://doi.org/10.1162/COMJ_a_00201

[30] Hariri, Sahand, Matias Carrasco Kind, and Robert J. Brunner, "Extended isolation forest," IEEE transactions on knowledge and data engineering, vol. 33, no. 4, pp. 1479-1489, 2019, https://doi.org/10.1109/TKDE.2019.2947676

[31] Qiu, Chen, et al., "Neural transformation learning for deep anomaly detection beyond images," International conference on machine learning. PMLR, 2021, https://proceedings.mlr.press/v139/qiu21a.html

[32] Vitaliy Kashuba, et al., "The formation of human movement and sports skills in processing sports-pedagogical and biomedical data in masters of sports", International Journal of Human Movement and Sports Sciences, vol. 8, no. 5, pp. 249-257, 2020, https://doi.org/10.13189/saj.2020.080513