Article

Improving Teaching and Learning in Higher Education through Machine Learning: Proof of Concept' of AI's Ability to Assess the Use of Key Microskills

Christopher Dann [1], Shirley O'Neill [1], Seyum Getenet [1,*], Subrata Chakraborty [2], Khaled Saleh [3] and Kun Yu [4]

1   School of Education, University of Southern Queensland, Springfield Central, Ipswich 4300, Australia; chris.dann@unisq.edu.au (C.D.); shirley.o'neill@unisq.edu.au (S.O.)
2   School of Science and Technology, University of New England, Sydney 2351, Australia; subrata.chakraborty@une.edu.au
3   School of Information and Physical Sciences, The University of Newcastle, Newcastle 2300, Australia; khaled.saleh@newcastle.edu.au
4   Data Science Institute, University of Technology Sydney, Sydney 2007, Australia; kun.yu@uts.edu.au
*   Correspondence: seyum.getenet@unisq.edu.au

Abstract: Advances in artificial intelligence (AI), including intelligent machines, are opening new possibilities to support teaching and learning in higher education. This research has found a 'proof of concept' in the application of machine learning in the assessment of educators' use of four key microskills, drawn from an internationally established framework. The analysis of teaching videos where these microskills were demonstrated multiple times in front of a green screen or in a space formed the data set. Multiple videos of this nature were recorded to allow for increased analysis and deconstruction of the video components to enable the application of machine learning. The results showed how AI can be used to support the collaborative and reflective practice of educators in a time when online teaching has become the norm. Having achieved a 'proof of concept', this research has laid the groundwork to allow for the whole framework of ten microskills to be applied in this way thus adding a new dimension to its use. Providing such critical information that is not currently available in such a systematic and personalised way to educators in the higher education sector can also support the validity of formative assessment practices.

Keywords: artificial intelligence; higher education; machine learning; microskills; reflective practice

1. Introduction

The fields of teaching and learning and artificial intelligence (AI) are broad fields of knowledge, and the combination of these fields has the potential to fundamentally change the approaches and presentation of higher education programs [1]. Artificial intelligence is rapidly appearing in professional practice worldwide, and higher education and, more broadly, education is no exception. AI advances have been foregrounded as an important part of our educational future through increasing investment globally to address the continuing need for improved outcomes. AI has experienced a compound annual growth rate of nearly 48%, with global trends showing similar growth patterns [2,3].

Although research on AI has been in existence for many years, the move to use it to impact pedagogical approaches in higher education is now possible with the introduction of low-cost, easily transferable videos of teaching practice [4]. It is even more relevant in times such as COVID-19, where higher education teaching and learning has needed to move online. In addition, there is now a considerable amount of commercial investment in the use of AI [5]. In addition, a country's uptake of technology for educational and work purposes has been shown to increase Gross Domestic Product (GDP). Thus, it is timely

to investigate the potential of AI and machine learning's use in higher education, where challenges to the provision of effective online learning have been most recently felt, and how it can support pedagogical strategies and reflective practice. In particular, the research aims to apply AI to illuminate the pedagogical effectiveness of educators' teaching online in higher education. Thus, this research addresses the following question:

Can educators' pedagogical actions be identified and classified to develop an algorithm that provides feedback on their performance in specific key microskills for professional learning?

Microskills are specific, fundamental skills that teachers need and use during the process of teaching a lesson, and they are essential for effective classroom management and student engagement [6]. To answer the research question and ensure that computers can identify and classify teaching actions, we must first determine whether a computer can replicate the human ability to categorise these actions as they occur during teaching. This process may raise ethical concerns, including researcher and research bias [7], especially in the current era of narrow AI, which focuses on specific features such as providing Natural Language Processing (NLP) for chatbots that respond to texts in chat rooms and emulate human conversation [8]. However, the most relevant considerations for this project are, firstly, ensuring the validity of the selected microskills as they must be relevant to a diverse workforce. Secondly, there must be accuracy and reliability in the human analysis of the evidence of macroskills in the selected videos.

It is clear from the literature that if a computer is provided with enough unbiased data to describe clearly distinguishable cues, then it is possible for it to identify actions within a video as required by the present project [9–12]. For instance, data bias may occur when the criteria are unclear to multiple researchers thus preventing clear categorisation of a specific skill. If applied to the provision of data applicable to evaluating educators' performance, which forms part of the current praxis in higher education, the results would have the potential to address the well-recognised need for reform in this area [13]. That is, the identification of categories of successful pedagogical actions of experienced educators within a video can illuminate a contested area and provide a core resource to guide learning and practice. Moreover, it supports achieving consistency in the approach to and understanding of lectures/video presentations in higher education. Such observable pedagogical skills need to be clearly distinguishable in the study of educator performance and are central to this present research. Moreover, given the broad nature of teaching, and the fact that educators comprise a diverse group, the researchers sought to take these into account in selecting the Richmond microskills framework [6], which has been proven to have international applicability as it has identified the specific skills that all teachers need and implement during the process of teaching a lesson, such as establishing expectations and giving instructions.

Thus, the research was underpinned by [6], an internationally established explanatory framework of ten microskills for teaching. This framework provided a standardised structure to create the classification system to enable the process of video analysis. The research objectives were to identify the following:

- the elements that need to be considered to use AI in assessment of teaching video
- if machine learning is able to distinguish between and correctly classify specific microskills
- whether a structured classification system for AI can be generated to support the valid analysis of microskills in higher education teaching
- whether reliable data can be produced from videos of practice to inform an automated feed of information applicable to educator reflective practice.

Thus, in keeping with [2,3]'s prediction that the use of AI in higher education will increase in the upcoming years, this research contributes to the field through its exploration of the machine learning process in relation to much needed innovation in being able to assess the presence of educators' use of the microskills necessary for effective pedagogical practices.

2. Literature Review

2.1. Machine Learning

AI is often referred to in terms of being able to provide solutions to problems deciphered by machines rather than humans. For instance, ref. [1] defined AI as a computing system that is able to engage in humanlike processes, such as learning, adapting, synthesising, and self-correction. They see machine learning as a subset of AI that "includes software able to recognise patterns, make predictions, and apply newly discovered patterns to situations" (p. 2). This subset, which this research focuses upon, recognises the necessity to develop the machine learning element first to establish the AI problem-solving process. This includes the concepts of unsupervised and supervised classification and profiling. The supervision level relates to the involvement of human intervention in the classification method, where humans create supervised classifications for the computer. In contrast, unsupervised refers to the process where the computer generates its own classifications.

As ref. [14] noted, machine learning is a method within AI, the results of which have already been used to enhance the lives of individuals, for example, in supporting people's daily activities through the provision of chatbots. In addition, the AI products initially developed through a machine learning process include applications like Apple's Siri and Amazon's Alexa. Such a personalisation of processes has been achieved through a machine's capacity for synthesising large amounts of data. This involves developing software that can recognise patterns, make predictions, and apply newly discovered patterns to situations that have not been previously considered. New managerial data and student engagement data produced by current software are already being enhanced by the availability of video data and student-contributed information that can inform learning and teaching [15].

2.2. Importance of AI in Higher Education

A comprehensive review of the literature on the use of AI and other advances and the conventional analytics in education reveal a broad spectrum of applications, yet there is limited evidence of AI systems' research being conducted at the scale of educational settings [16]. In particular, the application of AI in higher education teaching practices shows some promising development, particularly in the use of machine learning and video analysis. This underscores the growing recognition of the need for teacher training in AI technologies, which justifies the adoption of a microskills framework for this research.

Currently, the following three main categories of AI applications are prevalent in education: personal tutors (providing individualised teaching and support to students, adapting to their learning pace and style to offer tailored educational experiences), intelligent support for collaborative learning (facilitating group learning activities by enhancing communication, coordination, and cooperation among students), and intelligent virtual reality (virtual reality systems to create immersive learning environments) [17]. These categories align with the learner-facing, teacher-facing, and system-facing perspectives of AI tools [14]. However, despite these advancements, it remains challenging to find substantial evidence of AI's impact on enhancing teaching processes and practices, particularly regarding the humanistic aspects of teaching. Nevertheless, the pursuit of analytics that inform the quality of teaching is both significant and longstanding. Research has varied broadly, from studies involving teacher–student interactions to investigations into the use of microskills by educators [15]. As a result, 'Education Analytics' has emerged as a distinct field, with sub-specialties like educational data mining, which are expected to revolutionise education [18], and this research aims to build upon these developments.

One successful application of data analytics in higher education administration is the prediction of student churn or the students' potential to drop their courses of study. The rise of technology in higher education classrooms, including online learning, has facilitated data collection. However, most advancements have focused on structured data, such as grades and courses, which provide information primarily for educational management rather than offering deep pedagogical insights. Deep knowledge is derived from data

that can demonstrate an educator's use of Richmond's macroskills, verifying the evidence-based pedagogical practices essential for learning. Consequently, administrators often prioritise learning management system outputs consistent with a managerial view over pedagogical practices.

To address this issue, ref. [17] suggests that partnerships between educators and AI developers would help communities of practice to better understand the critical issues and identify opportunities for innovation. Such collaborations could shift the focus towards essential pedagogical practices. However, these partnerships face potential disruption from existing institutional systems that prioritise data collection on teaching tools, like Blackboard and Moodle, over exploring how evidence-based pedagogical practices can be leveraged, as intended in this project.

2.3. Positioning Video for AI Feed on Teaching in Higher Education

The use of video as a teaching tool has long been used in both classroom and online education [19], but the use of video data in analytics is a more recent phenomenon such that the field is in its nascent stages. The shift required to move from video being perceived as an audio–visual medium with a focus on content to a source of data analytics relevant to machine learning and, ultimately, the use of AI means significant reconceptualisation in the field [20]. According to [21], it reflects a move from videoing others for teaching purposes to videoing oneself to critique against a microskill standard in order to build personal capacity. Thus, this opportunity to increase pedagogical feed through video data analytics can lead to new and innovative ways of using the information contained within a video, which is particularly relevant to reforming teaching in higher education [9–12].

Video is a record of a situation and is the capture of 25–30 frames of the activity per second. All videos include images that are digital representations of scenes that capture movement, as well as separate streams of sound and audio data. Basically, the analysis of videos may involve identifying, transforming, and tracking these data streams for purposes relevant to a growing number of disciplines apart from teaching, e.g., medicine and the social sciences [22,23]. Machine learning uses this information to create sets of data that a computer can understand. Video is turned into a set of time stamped images that can be used as analytically prepared data. These data include the identification of low-level features like humans and objects, the detection of relationships between these features, as well as time and, finally, the extraction of variables with time-stamped values. Such data can influence the machine recall of the content in the video so patterns can be recognised for identification in other videos when presented. However, this process is only developing in its application to the higher education field of teaching, although strategies to analyse the quality of teaching in general are not new [24]. The current context of 'video manipulation' is outlined in the next section.

2.4. Performance Enhancing Video

Advancements using human direction and corresponding behaviours have been achieved in fields such as psychology, biomechanics, sports, commercial video games (e.g., Kinect for Xbox 360) and video surveillance and security [25]. However, the kinesthetic fields, such as sports and dance, have led the way in utilising video analysis to improve performance for several years but educators are rarely involved. The high video coverage of sports has enabled the collection of data analytics relevant to the needs of high-stakes worldwide premier league sports, which has prompted clubs to use video analyses to optimise performance and, most recently, the support of machine learning. Video images have also been utilised in biomechanics to study muscular activity to improve human movement [25–29]. However, this study of human movement is a separate area to human motion capture, upon which the present research focuses, as this depends on the ability to capture human motion through the use of behavioural biometrics [27,29]. Thus, when applied to teaching, the angle of the head, the movements of the arms, and the angle of the body leaning into or away from the student all become important factors to consider.

2.4.1. Using Images in Video

There have been various approaches to using images in the video, but these have not been applied to teaching. Ref. [30] employed a Pyramidal Motion Features (PMF) technique with Adaboost. Such adaptive boosting in machine learning can yield high recognition rates for certain databases, although the approach can be computationally expensive. Similarly, ref. [31] used temporal localisation to identify activities, which has the advantage of being computationally efficient and simple, and applicable to real-world scenarios, such as surveillance. While this is not appropriate for analysis of online databases, such as those that contain teaching analysis, ref. [32] technique of the spatiotemporal steerable pyramid (STSP) is more relevant. It offers both the preservation of shape and motion information as well as efficient results on the three activities' recognition datasets. However, it does not perform as well on complex actions and backgrounds, which is critical in the study of teaching because of the complexities of movement and the learning environment.

2.4.2. Teaching Video Availability

With respect to data on teaching videos, there are limited databases of teaching actions [33] and multiple mannerisms have not been created as yet. Nevertheless, the literature demonstrates that there is a range of specialised technological advancements in the recognition of body movement that is worth utilising to explore teaching in higher education. There are also several datasets and databases where initial training and testing could be undertaken, in parallel to data collection. Such datasets also provide pointers on what sort of data must be collected and what form they should take. Examples include the Weizmann Human Action dataset of actions [34], such as bending, running, walking, and skipping, and the UCF dataset of sports-based actions. This existing knowledge supports the feasibility of adaptation to other teaching skills as proposed in this paper.

2.4.3. Extracting Data from Video for AI

In the last two decades, several researchers have developed tools to help track and extract a sufficient quantity of quality learning events that would lay the foundation for video-based analytics for education [35,36]. In laying the foundation for such work, ref. [36] developed techniques for extracting and analysing the textual contents from instructional videos, while [37] developed an efficient video indexing engine called InVideo that makes it amenable for searchability, and both quantitative and qualitative analyses. Along similar lines, ref. [38] proposed and demonstrated the effectiveness of tracking student interactivity using educational video games. Further, ref. [35] developed tools for tracking learning objectives. While the field is nascent, its potential is being well recognised, with researchers investigating how analytics could become part of an integrated instructional solution [39–41]. By being able to build on this work, the present research provides a strategic response to the need to improve teaching in higher education and addresses [3]'s claims that there is "little evidence for the advancement of pedagogical and psychological learning theories related to AI driven technology" (p. 22) and that practitioners in education are rarely the drivers behind research in this field.

2.5. Application of AI to Teaching in Higher Education

The International Journal of Artificial Intelligence in Education (2016) recently presented a special edition that focused on the various applications of machine learning in the education space but with a specific focus on its use in formative assessment processes. The issue presents the case for improving feed processes to support student learning and builds on [17]'s claim that there is about to be a renaissance in assessment practices because of technological advancements. Its exploration of machine learning applied to tutoring systems specifically examines learner differences and the characteristics of individuals that influence the mindful processes of feed that can be offered through machine learning [42]. While demonstrative of the scope of the field, this research sought to use AI to generate knowledge to inform personal teaching development.

According to [3], four broad areas of AI and machine learning development in higher education have emerged. These are (1) profiling and prediction, (2) intelligent tutoring systems, (3) assessment and evaluation, and (4) adaptive systems and personalisation. Assessment and evaluation utilising machine learning focus on tools to help students when they are confused in their work by providing prompts. Such software has broad applications, ranging from supporting trainee pilots to providing automatic feedback for students to help with writing. Other key areas include using AI for course content management, personalising content in MOOCs and supporting academics to provide repetitive learning tasks, e.g., quizzes in online learning environments. These are seen as part of the "student life cycle" [3] and thus clearly supporting a student-centred approach. Importantly, the project proposed in this research steps outside of this to take a teaching process view and focuses on teaching and developments, specifically regarding microskills, to improve teaching in higher education.

AI for Teacher Educators in Higher Education

Of note for this research is [3], the systematic literature on teaching in higher education regarding AI, which includes a focus on assisting teachers with supervision in collaborative student activities and on sharing tutorial tasks. Although this considers both managing the learning process and reducing the teachers' workload, it does not focus on supporting the teaching act. They point out that no research was found with regard to pedagogical concerns and machine learning or AI, or consideration of the teaching perspective. The point was also made that there was a lack of longitudinal research. Thus, this research seeks to address this gap in terms of identifying if machine learning is able to distinguish between and correctly classify specific microskills. Thus, a search for a suitable framework that was (1) internationally recognised, (2) had the capacity to support the production of data analytics for AI, and (3) could accommodate a culturally diverse set of practitioners was conducted. This highlighted the following internationally developed standards, namely Marzano's The Art and Science of Teaching [43] and Danielson's Framework for Teacher Evaluation [44]. However, each of these was deemed too large in scope regarding the teaching skills and/or too broad in the description of the activity, and thus deemed unsuitable for the project's requirements. In direct contrast, it was found that the scope covered by the categories of [6]'s microskills for teaching framework was found to be both internationally relevant and also concise with educator action foci that were clear and appropriate to support the classification processes. Compared with other microskill approaches to teaching, such as the [10] '5 step microskills model' that has mainly been applied in clinical teaching, the specificity of Richmond's framework is more relevant to the effectiveness of educators in higher education apart from the application to AI. This is reinforced by [45]'s research that has found such alternatives to be inadequate in scope. In contrast, this selected microskills for teaching framework being globally accepted is able to provide a cohesive tie between diverse education contexts thus recognising their commonality and centrality to teaching and learning. Of the 252 factors identified, micro-teaching/video review of lessons is positioned thirteenth, with an effect size of 0.88. Moreover, this places microskills above 'teacher clarity', 'scaffolding', and 'deliberate practice' thus highlighting them together with a video review of lessons as a highly critical influence on student learning and thus a workable choice for this research.

3. Materials and Methods

Richmond's [6] ten specific microskills found to be necessary for effective teaching provided the underpinning framework for the research. The use of this framework provided the scope for investigating the key skills that all teachers need to do to maximise their effectiveness and, in turn, improve students' learning [6].

The microskills relate to language overall and are relevant to the management of students. During learning, they are categorised within the following three main areas:

(1) the language of expectation, (2) the language of acknowledgment, and (3) the language of correction, as listed below:

The language of expectation:

1. Establishing expectations.
2. Giving instructions.
3. Waiting and scanning.
4. Cueing with parallel acknowledgment.

The language of acknowledgement:

1. Body language encouraging.
2. Descriptive encouraging.

The language of correction:

1. Selective attending.
2. Redirecting to the learning.
3. Giving a choice.
4. Following through.

Considering the research question as to whether educators' pedagogical actions can be identified and classified for the production of an algorithm that can provide a feed of information for professional learning indicative of their performance of specific key microskills, the first four microskills in the 'language of expectations category' were tested. This included ascertaining the ability to deliver the relevant data analytics for 'proof of concept'. The participants (the researchers and tertiary educators in this preliminary stage) were required to produce a video demonstrating their interpretation of the teacher's actions as represented by the microskills. This initial test of the framework indicated that of the 10 microskills, only the first 4 would be able to be used. Those in the category of the 'language of acknowledgement' (5 to 6) and the 'language of correction' (7 to 10) were found to be non-viable in all the trial's teaching videos as well as in a sample of publicly available videos since the evidence of educators' use of these microskills depends on how students respond compared with the first four that are initiated by the educator. Hence, since microskills 5 to 10 depended on videoing students' responses, which would require substantial expansion of the study and were not necessary to answer the research question, microskills 1 to 4 became the focus of the study. In light of the successful 'proof of concept', the researchers plan to replicate the study to include students' responses.

3.1. Theoretical Underpinning for Implementing AI in Higher Education

In keeping with the research direction of focusing on pedagogical advancement in higher education, this paper presents a pilot study that seeks to explicitly build a machine learning process informed by the underlying philosophical position of critical realism. This position combines the belief that the real world does exist independently of our perceptions with the epistemological perspective of constructivism. From this combined perspective, the authors positioned themselves to think about the issues and insights of the phenomenon central to this research. Constructivism builds on the premise of the social construction of reality [46] and brings with it the advantage of allowing close collaboration between the researcher and the educators. It allows the educators to describe their views of reality and enables the researchers to understand the actions of the educators better. It is through this lens that new understandings can be socially constructed within this case, supported by machine learning analyses of human movement, speech, and intonation.

With a view to build on this preliminary proof of concept research, the project sought to carefully consider and select aspects within the case, which, in this study, included videos of teaching that can illustrate the selected specific classifications. Given the global nature of teaching, the cultural intonation and diversity, and perceived differences in teaching approaches, it was critical to ensure that, initially, the videos were selected according to a set of reproducible tests. Thus, in order to allow the researchers to explore differences between cases with the goal of making comparisons, it was critical that cases were chosen carefully

so that the researchers were cognizant of the similarities and differences between them. Hodge and Sharp [47] described case studies as either intrinsic, instrumental, or collective. This study used the instrumental case as it intended to gain insight and understanding of if and how the selected microskills could be identified and labelled from a set of teaching performances in video. It is based on the proposition that experienced educators make decisions about the effects of teaching based on prior knowledge, and, in turn, these decisions can inform the feedback they give to the performer. As noted, the framework comprises descriptions of microskills that have been found to represent effective teaching practice internationally, so applicable to this underpinning philosophy. The application of this framework is described in the following section.

In testing for the 'proof of concept', the study aimed to employ machine learning to identify each of the selected microskills in a set of videos of teaching. A pilot collection of visual data in the form of video recorded by a member of the research team performing the first four micro teaching skills was compiled. These data were then used to generate an algorithm and test the process of video collection and the production of data analytics. The following section presents the participant information, video selection requirements, and, finally, the four stages that were used to extract data from the selected videos.

3.2. Participants

The research involved three educators in distinct roles. One educator, previously mentioned, was the educator featured in the videos selected for analysis. The other two participants contributed to developing the video analysis process and generating data analytics. All data collection adhered to the ethical approvals from the educators' respective universities. Since selecting videos for machine learning is a crucial aspect of the project, the following section elaborates on the project's methodology, detailing the video selection and classification processes.

3.3. Video Selection Process

The selection of video for the purposes of this research was to test for the 'proof of concept' by conducting initial algorithm training. This was essential to the project in that if successful, it would provide the basis for the extension of the research once the machine learning process had been established and the algorithm developed.

3.3.1. Analysis of Category One Case Study—Provision of Data for Proof of Concept

The initial videos were of a single practitioner and research team member demonstrating each of the four micro skills multiple times in front of a green screen or in a space without students. Multiple videos of this nature were then recorded to allow for increased analysis and deconstruction of the video components. These replicated the educator's work in higher education, which involved teaching online as well as face-to-face. The deconstruction included sounds and movement, but excluded facial expression, as this facet was able to be incorporated through the existing algorithms. Each video had the educator verbally nominate the microskill and then demonstrate the skill and verbally indicate the end of the skill demonstration.

From this research project's perspective, all educators were able to access personal teaching materials and any documentation they felt was necessary to inform their knowledge of each of the microskills. In addition, a basic set of documents informing the definition of each micro skill was produced and accompanied by any supporting material the educators generated during the project.

3.3.2. Ensuring the Project's Sustainability—Ethical Requirements and Continued Implementation

It was vital for the project to be conducted to allow for the natural extension of the research should 'proof of concept' be attained. This was achieved through being conscious of the fact that, in the next phase, the selection of videos would need to be broadened

such that three criteria would need to be met. Firstly, the video would need to be ethically collected and ethically approved for public scrutiny, so it should be made publicly available as well as freely accessible online to enable the replication of the study. Secondly, the video would need to involve educators actually in the action of teaching, where there was subject content, although video content matter was not required to prove that the microskills were being demonstrated through machine learning. The final criterion was that the teaching had to have clear audio and vision of the teacher and be free of text annotations. This would ensure a fair pre-selection process for the video and help assure relevance.

Sitting within the video criteria here was also the need for a set of technical requirements. Each video needed to be viewed on a computer with a 'second by second' indication of time so that educators could associate actions with the seconds within the video. However, an advantage was that the video was not required to be in high definition for machine learning analysis to be applied. Thus, the initial selection of videos needed to undergo data cleaning (sorting) by the researchers according to the presence of the educator and students and active teaching dialogue. In addition, any other teaching descriptions or edited annotations included in the video needed to be removed so that the resultant video would be pure in terms of the teaching in a learning situation, paving the way for the identification of the microskills.

3.4. Video Processing Stages

Challenges that the researchers faced in carrying out the supervised classification for machine learning of the act of teaching included hidden human body parts (including self-occlusion), fast human movements in complex background scenes, light sensitivity and motion ambiguities. For example, in this application, the students in the learning environment needed to be identified and excluded as part of the background scene so that the machine learning process could focus instead on the educator's activities. However, as discussed, several techniques can address these specific issues to obtain the optimal pedagogical and activity dataset, where complex and subtle actions can be distinguished and delineated. This requires the same sensitivity and precision that is evidenced in sports- or dance-based action capture and reinforces the need for supervised classification processes in the initial phase as with the present research. The video extraction stages of the project involved the software's detection of visual movement of which there were four common stages foundational to any machine learning with video [26,27,48]. These stages consist of (i) initialisation, (ii) tracking, (iii) pose estimation, and (iv) interpretation, and are described below to explain the extraction method used to analyse the project's initial video data collected.

3.4.1. Visual Stage One—Initialisation

The first stage involved initialisation by building the initial humanoid model of appearance, shape, and kinematic structure to be set up for subsequent stages. One approach has been based on building models using prior knowledge and manually identified joint locations [49–51]. However, the field has been progressing towards automatic model construction using a range of techniques, and this is often based on 3D depth body shape, sensors for body scans and joint information, using multiple perspectives. In addition, commercial marker-based frameworks and motion capture databases will also be employed in a priori mapping of 'image to pose' space. The models employed primitive shapes, such as cylinders, cones, and ellipsoids [52], and used polygonal mesh surfaces to define kinematic skeletons [53]. Multiple perspectives have also been employed to improve the fidelity of identification and imaging [54,55]. In addition, through utilising databases, the researchers were able to learn about the range of body shapes, statistically [56].

3.4.2. Visual Stage Two—Tracking

The application of machine learning also needs to take into account the figure–ground separation in the video; thus, tracking is a key part of the process. Tracking is a spatio-

temporal step that has two parts, namely segmenting or differentiating the subjects from the background and tracking or detecting the sequences of the segments across consecutive frames. The basic idea of figure ground separation is based on distinguishing the figure-based attributes, such as colour, shade, and intensity differences, as well as fitting geometric representations and kernels. The standard colour differences may be enhanced by intensities [57] or normalised [58]. A kernel-based approach is another technique that offers the advantages of being able to handle cluttered backgrounds and dynamics, evolving scenes (e.g., moving educators) at a higher speed [59,60]. This approach represents each background pixel by a function of bits. Neighbourhood relationships (the relationships between items close together in a video) could also be encoded. Statistical and machine learning techniques (classification and autoregression) are also typically applied to represent a scene, and present-day algorithms permit moving backgrounds. Similarly, segmentation models also make use of the humanoid appearance of humans to distinguish humans from the background, as well as distinguishing appearances based on individual differences. Such appearance-based approaches can also be either context-free or context enriched. A number of these approaches also help with automatic tracking and hence serve twin purposes. In addition, temporal correspondence works on the back of segmentation. Although one challenge with tracking human movement in classroom settings is the occurrence of multiple people and partial occlusion of images, both local and probabilistic approaches for bypassing these issues have been formulated [61,62].

3.4.3. Visual Stage Three—Pose Estimation

This stage addresses the important step of the software's ability to detect human poses. 'Pose estimation' helps synthesise and evaluate the underlying skeletal structure of the pose, often using a high-level human model. The range of pose estimation algorithms can be based on model-free, direct, and indirect methods [27,48]. Algorithms such as 2D estimation of body parts, e.g., head, torso, and extremities, are used to develop pose estimations that learn and map from 2D images to 3D poses [63–69]. Ref. [69], for example, presented a discriminative density propagation algorithm based on a Bayesian Mixture of Expert model (BME). It has the advantage of efficiency in detecting human poses but fails to track if the object is occluded adequately.

3.4.4. Visual Stage Four—Interpretation

The final stage requires the software to identify the people in the video and their behaviour. This recognition can be logically and hierarchically separated into tasks required for interpretation. Each level drills down into details. The first level involves scene interpretation where the whole picture needs to be interpreted without identifying and discriminating constituent parts, such as people and things. The whole human body or body parts also need to be recognisable (holistic recognition, level two). Similarly, at the third level, complex actions can be recognised through the use of grammar with action primitives applied to specific tasks or semantic depiction of a scene [70]. For example, for scene interpretation, refs. [71,72] presented ways to detect irregularities.

3.5. Experimental Study

Temporal Segment Network (TSN) [73] is a versatile and adaptable video-level framework designed for learning action models in videos. As illustrated in Figure 1, it seeks to capture long-range temporal structures through a segment-based sampling and aggregation method. This approach enables the TSN framework to effectively learn action models by utilising the entire video.
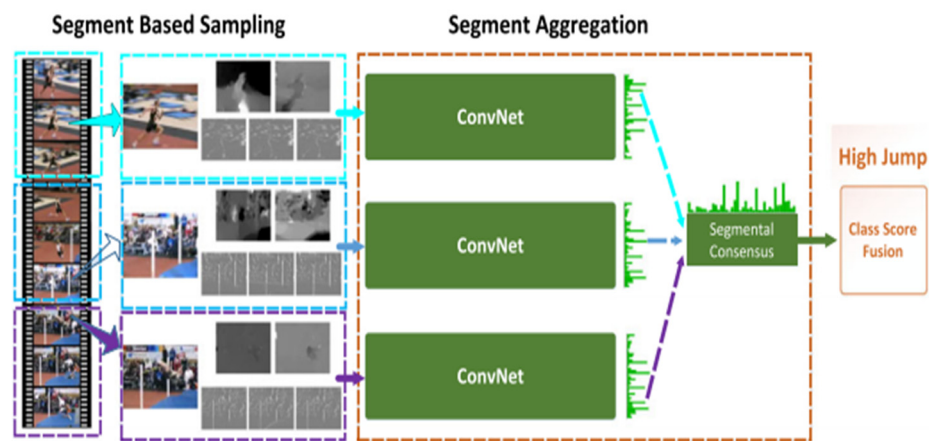
Figure 1. Temporal Segment Network workflow.

Our study analysed 21 videos, each lasting 1 to 20 min. The videos were categorised as follows: Four videos focused on cueing with parallel acknowledgment skills, four videos on establishing expectation skills, eight videos on giving instruction skills, and five videos on waiting and scanning skills.

A TSN was employed to acquire video-level labels. Specifically, we trained a TSN for each micro-skill class. For each class, half of the videos were used for training, and the other half for testing. We compared the performance of two features—colour and optical flow. For training, each model was initialised with pre-trained weights obtained through a cross-modality training technique. The classification results are shown in Table 1. The average accuracy achieved with the colour feature was 25%, whereas the motion feature yielded an accuracy of 62.5%. These results indicate that colour is not a strong feature, as most of the videos were filmed in the same environment by the same person, resulting in minimal colour variation among them.

Table 1. Classifying microskills with TSN.

Class	Classification	
	Color	**Motion**
Establishing expectations	0%	100%
Giving instructions	50%	50%
Waiting and scanning	50%	100%
Cueing with parallel acknowledgment	0%	0%

Based on insights from TSN, we also adopted a hybrid approach to recognise the microskills. As depicted in Figure 2, this approach involves two modules. The first module, the facial feature extractor, captures and acquires face-related representations from each frame in the input sequence. Since a person's face can reveal significant information about their personal behaviours [74], our facial feature extractor module focused on identifying facial features indicative of micro-teaching skills. This module utilised OpenFace 2.0, a state-of-the-art facial behaviour analysis toolkit [75].
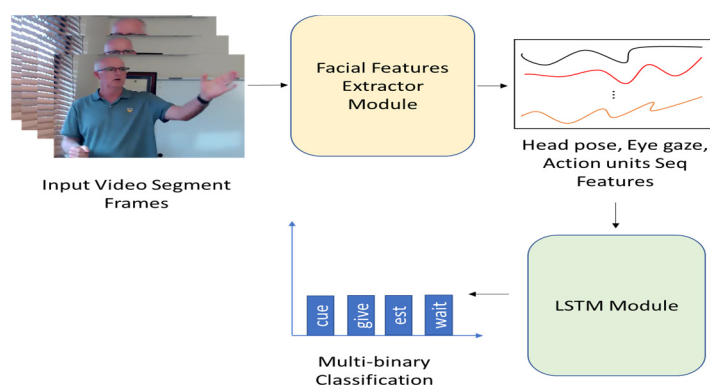
Figure 2. The hybrid approach for video content classification.

The second module of our framework, the temporal modelling module, processes the sequence of facial features and captures the temporal dependencies between them. For this module, we rely on Long Short-Term Memory (LSTM) [73], one of the most successful and widely adopted deep recurrent neural network architectures for modelling sequential data.

To evaluate the performance, each video was split into small chunks of 1 s duration without overlapping. This resulted in approximately 60 videos for training and 70 videos for testing. Given the imbalanced nature of the processed dataset and the correlation between the four classes of essential skills, we trained two separate binary classifiers for each pair of essential skills—(Establishing Expectation, Waiting and Scanning) and (Giving Instructions, Cueing with Parallel Acknowledgement). Table 2 shows the classification results of the hybrid framework. As observed, the two classifiers, combined with the rich facial features, have achieved robust results in terms of precision, recall, and F1 score.

Table 2. Classifying microskills with a hybrid approach.

Class	Precision	Recall	F1-Score
Cueing with parallel acknowledgement	92%	69%	79%
Giving instructions	81%	95%	88%
Cueing with parallel acknowledgment	92%	75%	83%
Waiting and scanning	84%	95%	89%

3.6. Ethical Considerations

There are two key ethical considerations for this research. Firstly, working with the ethical boundaries of the institutions which has been granted for this project (No. ETH19-4205) and, secondly, the ethical implications of AI implementation and development. The project's focus on personal learning and evidence of microskills requires various ethical issues to be addressed. These include the de-identification of individuals and thus the reduction in individual performance to a form of data that can be managed without personal or professional impacts on the educators providing the video. Furthermore, we ensured no video clip or image would be released without consent from the participating educators, and the consent was acquired via the signed consent declaration from them. To maintain the confidentiality, all the data were anonymised before further processing. The impact of the machine learning output also needed to be considered. The researchers for this project were also cognizant of the possible impact on the personal teaching efficacy of those who participate in and undertake feedback by the resultant machine learning process. These considerations have therefore driven the supervised classification and research design. At each stage of the project, the outputs were supervised and correlated to the human classifications to increase the reliability and accuracy of the machine learning process. This indicated an ethical consideration of the implications of such processes as well.

The ethical challenges within the project also included the collection of videos of teaching which may have identifiable images of individuals. In addition, the storage of video requires very large storage spaces and raises issues when the videos need to be transmitted or accessed across the globe. Thus, consideration needs to be given to the management and security of the video, ownership of the storage facility, access to the storage facility, and how the backup systems work. These ethical challenges relate to the dissemination of data and participant privacy and involve issues pertaining to privacy and child protection in data collection. Thus, appropriate ethical clearances were obtained and factored into the methods being applied.

4. Discussion and Conclusions

In answer to the research question, the project results clearly confirm that educators' pedagogical actions in videos, in terms of the four selected microskills, were able to be identified and classified, and an algorithm was produced to provide a feed of information for professional learning and reflective practice. The research identified the elements that need to be considered when applying AI to an assessment of teaching videos and, in turn, proved that machine learning is able to both distinguish between and correctly classify the targeted microskills. Furthermore, the objectives to identify whether a structured classification system for AI could be generated to support the valid analysis of the targeted microskills was also achieved as was confirmation of the reliability of the data produced. While the process demonstrated sound alignment between the machine learning feedback and human feedback, it is acknowledged that following the success of this 'proof of concept', further research is required. Not only is this needed to increase the machine-learning's range of microskills to include those related to student responses (microskills 5–10) but also to further validate its acceptance of educators' culturally and linguistically diverse backgrounds. However, it is also necessary to acknowledge that the study was limited to the higher education context in Australia and to the ecological context of the related learning environment where the videos were initially sourced. It is also limited to the cultural and diverse characteristics of the educators and students who were engaged in the classification process, although a reasonable representation for this initial work has been achieved, cross-institutionally, culturally and linguistically among the research team. These limitations are crucial to the development of further research as future studies must also document the cultural and contextual demographics of people involved to ensure the resultant classification systems can meet the demands of the diverse global cohorts of educators in higher education.

Thus, as the literature has shown, this research, in seeking to position the act of teaching in higher education within the field of AI and machine learning, is very timely. It provides support for the argument that pedagogical affordances are possible through the use of machine learning. These stem from the ability to identify evidence in videos of internationally established microskills required for teaching to be effective. The methods used were developed by a multi-disciplinary team that includes educators and machine learning developers to address the lack of understanding of teaching and learning by AI and machine learning teams that act in isolation from practitioners.

Recognising that the microskills investigated here are applicable to teaching in higher education acknowledges [17]'s view that there is a renaissance in the assessment that recognises a need to focus on the actions of the educator rather than those of the learner currently prevalent in the literature. In conclusion, this research has responded to the limited application of machine learning on the act of teaching in the higher education sector in the AI space and has highlighted the growing importance of the need for formative assessment of teaching and reflective practice in higher education and how it may be achieved. In the light of achieving 'proof of concept', this research has also laid the groundwork for future research that can expand the suite of microskills as indicators of teaching effectiveness and strengthen their applicability to teaching and assessment in higher education, aside from

highlighting the importance of the need for professional development for educators in higher education and modelling them for educators in general.

This study can have implications for teacher education, particularly in professional development and reflective practice. The successful use of machine learning to identify and classify key microskills in teaching videos shows AI's potential to enhance pedagogical practices in higher education. The reliable alignment between AI and human feedback suggests that AI can be a valuable tool for formative assessment. This study also emphasises the need to include diverse microskills and consider cultural and linguistic diversity. By proving the concept, this study lays the groundwork for broader AI applications in teacher education, advocating for its integration into professional development to improve teaching strategies and learning outcomes. Future research should expand these findings across different educational contexts and diverse populations.

5. Recommendations

The results showed how AI could be used to support the collaborative and reflective practice of educators at a time when online teaching has become the norm in response to unpredictable contexts such as COVID-19. This research can lay the groundwork to allow for the whole framework of ten microskills to be applied in this way thus adding a new dimension to its use. Providing such a critical feed of information not currently available in such a systematic and personalised way to educators in the higher education sector can also support the validity of formative assessment practices.

References

1. Popenici, S.A.D.; Kerr, S. Exploring the impact of artificial intelligence on teaching and learning in higher education. *Res. Pract. Technol. Enhanc. Learn.* **2017**, *12*, 22. [CrossRef] [PubMed]
2. Feijóo, C.; Kwon, Y.; Bauer, J.M.; Bohlin, E.; Howell, B.; Jain, R.; Potgieter, P.; Vu, K.; Whalley, J.; Xia, J. Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy. *Telecommun. Policy* **2020**, *44*, 101988. [CrossRef]
3. Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. Systematic Review of Research on Artificial Intelligence Applications in Higher Education–Where are The Educators? *Int. J. Educ. Technol. High. Educ.* **2019**, *16*, 39. [CrossRef]
4. Brame, C.J. Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sci. Educ.* **2016**, *15*, es6. [CrossRef] [PubMed]
5. Arnold, Z.; Rahkovsky, I.; Huang, T. *Tracking AI Investment: Initial Findings from the Private Markets*; Center for Security and Emerging Technology: Washington, DC, USA, 2020.
6. Richmond, C. *Teach More, Manage Less: A Minimalist Approach to Behaviour Management*; Scholastic: New York, NY, USA, 2007.
7. Müller, V.C. Ethics of artificial intelligence and robotics. In *The Stanford Encyclopedia of Philosophy*; Metaphysics Research Lab, Stanford University: Stanford, CA, USA, 2020.
8. Dimitriadou, E.; Lanitis, A. A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learn. Environ.* **2023**, *10*, 12. [CrossRef]

9.   Li, Y.; Li, Y.; Vasconcelos, N. Resound: Towards action recognition without representation bias. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

10.   Neher, J.O.; Gordon, K.C.; Meyer, B.; Stevens, N. A five-step "microskills" model of clinical teaching. *J. Am. Board Fam. Pract.* **1992**, *5*, 419–424. [PubMed]

11.   Noben, I.; Deinum, J.F.; Hofman, W.H.A. Quality of teaching in higher education: Reviewing teaching behaviour through classroom observations. *Int. J. Acad. Dev.* **2022**, *27*, 31–44. [CrossRef]

12.   Rice, R.E. From Athens and Berlin to LA: Faculty scholarship and the changing academy. In *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*; Springer: Dordrecht, The Netherlands, 2007; pp. 11–21.

13.   Patience, A. It's Time to Reform Australia's Higher Education System. *Education* **2020**. Available online: https://johnmenadue.com/higher-education-reform-australia-by-allan-patience/ (accessed on 26 December 2023).

14.   Baker, T.; Smith, L.; Anissa, N. Educ-AI-Tion Rebooted? Exploring the Future of Artificial Intelligence in Schools and Colleges Reports. 2020. Available online: https://www.nesta.org.uk/report/education-rebooted/ (accessed on 12 May 2019).

15.   Dann, C.E.; O'Neill, S. Are you feeding back or is it taking students forward?: Changing the traditional narrative to ensure a dialogic approach in formative assessment. In *Technology-Enhanced Formative Assessment Practices in Higher Education*; IGI Global: Hershey, PA, USA, 2020; pp. 275–296.

16.   Baker, R.S. Stupid tutoring systems, intelligent humans. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 600–614. [CrossRef]

17.   Luckin, R.; Holmes, W. *Intelligence Unleashed: An Argument for AI in Education*; UCL Knowledge Lab: London, UK, 2016.

18.   Tulasi, B. Significance of Big Data and Analytics in Higher Education. *Int. J. Comput. Appl.* **2013**, *68*, 21–23. [CrossRef]

19.   Rideout, V.J.; Foehr, U.G.; Roberts, D.F. *Generation M^2: Media in the Lives of 8-to 18-Year-Olds*; Henry J. Kaiser Family Foundation: San Francisco, CA, USA, 2010.

20.   Choi, H.J.; Johnson, S.D. The effect of context-based video instruction on learning and motivation in online courses. *Am. J. Distance Educ.* **2005**, *19*, 215–227. [CrossRef]

21.   Marcus, A.; Wilder, D.A. A comparison of peer video modeling and self video modeling to teach textual responses in children with autism. *J. Appl. Behav. Anal.* **2009**, *42*, 335–341. [CrossRef] [PubMed]

22.   D'Souza, M.; Munster, C.E.P.V.; Dorn, J.F.; Dorier, A.; Kamm, C.P.; Steinheimer, S.; Dahlke, F.; Uitdehaag, B.M.J.; Kappos, L.; Johnson, M. Autoencoder as a new method for maintaining data privacy while analyzing videos of patients with motor dysfunction: proof-of-concept study. *J. Med. Internet Res.* **2020**, *22*, e16669. [CrossRef] [PubMed]

23.   Nassauer, A.; Legewie, N.M. analyzing 21st century video data on situational dynamics—Issues and challenges in video data analysis. *Soc. Sci.* **2019**, *8*, 100. [CrossRef]

24.   Chua, Y.H.V.; Dauwels, J.; Tan, S.C. Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now? In Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Tempe, AZ, USA, 4–8 March 2019.

25.   Ko, T. A survey on behavior analysis in video surveillance for homeland security applications. In Proceedings of the 2008 37th IEEE Applied Imagery Pattern Recognition Workshop, Washington, DC, USA, 15–17 October 2008; pp. 1–8.

26.   Afsar, P.; Cortez, P.; Santos, H. Automatic visual detection of human behavior: A review from 2000 to 2014. *Expert Syst. Appl.* **2015**, *42*, 6935–6956. [CrossRef]

27.   Moeslund, T.B.; Hilton, A.; Krüger, V. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **2006**, *104*, 90–126. [CrossRef]

28.   Turaga, P.; Chellappa, R.; Subrahmanian, V.S.; Udrea, O. Machine recognition of human activities: A survey. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1473–1488. [CrossRef]

29.   Yampolskiy, R.V.; Govindaraju, V. Behavioural biometrics: A survey and classification. *Int. J. Biom.* **2008**, *1*, 81–113. [CrossRef]

30.   Liu, L.; Shao, L.; Rockett, P. Boosted key-frame selection and correlated pyramidal motion-feature representation for human action recognition. *Pattern Recognit.* **2013**, *46*, 1810–1818. [CrossRef]

31.   Shao, L.; Jones, S.; Li, X. Efficient Search and Localization of Human Actions in Video Databases. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *24*, 504–512. [CrossRef]

32.   Zhen, X.; Shao, L.; Li, X. Action recognition by spatio-temporal oriented energies. *Inf. Sci.* **2014**, *281*, 295–309. [CrossRef]

33.   Sharma, V.; Gupta, M.; Kumar, A.; Mishra, D. EduNet: A New Video Dataset for Understanding Human Activity in the Classroom Environment. *Sensors* **2021**, *21*, 5699. [CrossRef]

34.   Blank, M.; Gorelick, L.; Shechtman, E.; Irani, M.; Basri, R. Actions as space-time shapes. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, China, 17–21 October 2005.

35.   Dyckhoff, A.L.; Zielke, D.; Bültmann, M.; Chatti, M.A.; Schroeder, U. Design and implementation of a learning analytics toolkit for teachers. *J. Educ. Technol. Soc.* **2012**, *15*, 58–76.

36.   Haubold, A.; Kender, J. Analysis and Visualization of Index Words from Audio Transcripts of Instructional Videos. In Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering, Miami, FL, USA, 13–15 December 2004.

37.   Wang, S.; Behrmann, M. Video indexing and automatic transcript creation. In Proceedings of the 2nd International Conference on Education Research, Shanghai, China, 22–24 June 2010. Available online: http://kihd.gmu.edu/assets/docs/kihd/activ/new/Wang_Behrman_press.pdf (accessed on 21 December 2023).

38.   Serrano-Laguna, Á.; Torrente, J.; Moreno-Ger, P.; Fernández-Manjón, B. Tracing a Little for Big Improvements: Application of Learning Analytics and Videogames for Student Assessment. *Procedia Comput. Sci.* **2012**, *15*, 203–209. [CrossRef]

39. Elias, T. Learning analytics: Definitions, processes and potential. *Citado* **2011**, *4*, 27–28.
40. Picciano, A.G. The Evolution of Big Data and Learning Analytics in American Higher Education. *J. Asynchronous Learn. Netw.* **2012**, *16*, 9–20. [CrossRef]
41. Wang, S.P.; Kelly, W. Video-based Big Data Analytics in Cyberlearning. *J. Learn. Anal.* **2017**, *4*, 36–46. [CrossRef]
42. Narciss, S. Conditions and effects of feedback viewed through the lens of the interactive tutoring feedback model. In *Scaling up Assessment for Learning in Higher Education*; Springer: Singapore, 2017; pp. 173–189.
43. Marzano, R.J. *The Art and Science of Teaching: A Comprehensive Framework for Effective Instruction*; ASCD: Alexandria, VA, USA, 2007.
44. Evans, B.R.; Wills, F.; Moretti, M. Editor and Section Editor's Perspective Article: A Look at the Danielson Framework for Teacher Evaluation. *J. Natl. Assoc. Altern. Certif.* **2015**, *10*, 21–26.
45. Ridley, C.R.; Kelly, S.M.; Mollen, D. Microskills training: Evolution, reexamination, and call for reform. *Couns. Psychol.* **2011**, *39*, 800–824. [CrossRef]
46. Berger, P.; Luckmann, T. *The Social Construction of Reality, in Social Theory Re-Wired*; Routledge: Abingdon, UK, 2023; pp. 92–101.
47. Hodge, K.; Sharp, L.-A. Case studies: What are they. In *Routledge Handbook of Qualitative Research in Sport and Exercise*; Routledge: Abingdon, UK, 2016; pp. 62–74.
48. Moeslund, T.B.; Granum, E. A Survey of Computer Vision-Based Human Motion Capture. *Comput. Vis. Image Underst.* **2001**, *81*, 231–268. [CrossRef]
49. Kakadiaris, I.A.; Barrón, C. On the improvement of anthropometry and pose estimation from a single uncalibrated image. *Mach. Vis. Appl.* **2003**, *14*, 229–236. [CrossRef]
50. Parameswaran, V.; Chellappa, R. View independent human body pose estimation from a single perspective image. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004.
51. Parameswaran, V.; Chellappa, R. View Invariance for Human Action Recognition. *Int. J. Comput. Vis.* **2006**, *66*, 83–101. [CrossRef]
52. Plankers, R.; Fua, P. Articulated soft objects for multiview shape and motion capture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1182–1187. [CrossRef]
53. Carranza, J.; Theobalt, C.; Magnor, M.A.; Seidel, H.-P. Free-viewpoint video of human actors. *ACM Trans. Graph.* **2003**, *22*, 569–577. [CrossRef]
54. Starck; Hilton. Model-based multiple view reconstruction of people. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003.
55. Ménier, C.; Boyer, E.; Raffin, B. 3D skeleton-based body pose recovery. In Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06), Chapel Hill, NC, USA, 14–16 June 2006.
56. Magnenat-Thalmann, N.; Seo, H. Data-driven approaches to digital human modeling. In Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, Thessaloniki, Greece, 9 September 2004.
57. Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1337–1342. [CrossRef]
58. McKenna, S.J.; Jabri, S.; Duric, Z.; Wechsler, H. Tracking interacting people. In Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), Grenoble, France, 28–30 March 2000.
59. Heikkila, M.; Pietikainen, M. A texture-based method for modeling the background and detecting moving objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 657–662. [CrossRef] [PubMed]
60. Elgammal, A.; Harwood, D.; Davis, L. Non-parametric model for background subtraction. In *Computer Vision—ECCV 2000, Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland, 26 June–1 July 2000*; Proceedings, Part II 6; Springer: Berlin/Heidelberg, Germany, 2000.
61. Khan, S.M.; Shah, M. Tracking Multiple Occluding People by Localizing on Multiple Scene Planes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 505–519. [CrossRef] [PubMed]
62. Park, S.; Aggarwal, J. Simultaneous tracking of multiple body parts of interacting persons. *Comput. Vis. Image Underst.* **2005**, *102*, 1–21. [CrossRef]
63. Agarwal, A.; Triggs, B. 3D human pose from silhouettes by relevance vector regression. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004.
64. Brand, M. Shadow puppetry. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999.
65. Howe, N.R. Silhouette lookup for automatic pose tracking. In Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, 27 June–2 July 2004.
66. Micilotta, A.S.; Ong, E.J.; Bowden, R. Detection and Tracking of Humans by Probabilistic Body Part Assembly. In Proceedings of the British Machine Vision Conference, Oxford, UK, 5–8 September 2005.
67. Ramanan, D.; Forsyth, D.A.; Zisserman, A. Strike a pose: Tracking people by finding stylized poses. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005.
68. Shakhnarovich; Viola; Darrell. Fast pose estimation with parameter-sensitive hashing. In Proceedings of the Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003.

69. Sminchisescu, C.; Kanaujia, A.; Metaxas, D.N. BM^{3}E: Discriminative Density Propagation for Visual Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2030–2044. [CrossRef]

70. Krüger, V.; Grest, D. Using hidden markov models for recognizing action primitives in complex actions. In *Image Analysis, Proceedings of the 15th Scandinavian Conference, SCIA 2007, Aalborg, Denmark, 10–14 June 2007*; Springer: Berlin/Heidelberg, Germany, 2007.

71. Boiman, O.; Irani, M. Detecting Irregularities in Images and in Video. *Int. J. Comput. Vis.* **2007**, *74*, 17–31. [CrossRef]

72. Chowdhury, A.K.R.; Chellappa, R. A factorization approach for activity recognition. In Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition Workshop, Madison, WI, USA, 16–22 June 2003.

73. Wang, L.; Xiong, Y.; Wang, Z.; Qiao, Y.; Lin, D.; Tang, X.; Van Gool, L. Temporal segment networks for action recognition in videos. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 2740–2755. [CrossRef] [PubMed]

74. Ricciardelli, P.; Bayliss, A.P.; Actis-Grosso, R. Editorial: Reading Faces and Bodies: Behavioral and Neural Processes Underlying the Understanding of, and Interaction with, Others. *Front. Psychol.* **2016**, *7*, 1923. [CrossRef]

75. Tadas, B.; Amir, Z.; Chong, L.Y.; Philippe, L.M. OpenFace 2.0: Facial Behavior Analysis Toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018.