Exploring Pose Estimation with Computer Vision Processing to Model Psychomotor Performance in Karate Combats

Author: Jon Echeverria San Millán

Supervisor: Olga C. Santos

Universidad Nacional de Educación a Distancia (UNED)

Escuela Técnica Superior de Ingeniería Informática

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN INTELIGENCIA ARTIFICIAL

September 2022

Madrid, Spain

Prefacio

Para la defensa del Trabajo Fin de Máster (TFM) en el Máster en Investigación en Inteligencia Artificial de la UNED presento una memoria tipo artículo que consiste en una ampliación del artículo titulado "*Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation*" que he publicado en diciembre de 2021 en la revista Sensors, indexada en JCR SCIE 2021 en Q2. Esta ampliación profundiza la investigación sobre los algoritmos de minería de datos más adecuados para modelar la ejecución de movimientos en contextos complejos donde dos usuarios interaccionan entre si realizando movimientos rápidos, explosivos y dependientes entre ellos, y evalúa una aplicación que clasifica movimientos ya predeterminados de karate recibidos a través de vídeo en tiempo real, incluyendo ejecuciones con y sin errores para analizar el rendimiento psicomotor de los karatekas.

Así, con este trabajo de investigación he conseguido avanzar el estado del arte del HPE (del inglés Human Pose Estimation) en la identificación de movimientos en tiempo real y en modo "multipersona" que involucra a dos o más personas en la ejecución de movimientos, haciendo uso de técnicas de visión artificial y minería de datos. Se avanza en el modelado del movimiento humano monitorizando una actividad deportiva con movimientos rápidos y explosivos para mejorar y avanzar en el aprendizaje de las habilidades motoras que se desarrollan en el arte marcial objeto de estudio. Se mejora la clasificación de las imágenes de vídeo recibidas a través de una aplicación que utiliza el algoritmo visión artificial OpenPose y clasifica las posturas con algoritmos de deep learning. Esta aplicación se ha integrado en el sistema psicomotor inteligente KUMITRON que estoy desarrollando para ofrecer feedback personalizado a los usuarios para mejorar las habilidades motoras implicadas.

A lo largo del desarrollo del TFM he tenido la oportunidad de publicar los avances de mi investigación en varios congresos indexados en el ranking de GII-GRIN-SCIE (GGS), donde también he recibido retroalimentación de otros investigadores que me han ayudado a avanzar. Concretamente:

- **ACM IUI 2021** (26th Annual Conference on Intelligent User Interface). El congreso ACM IUI está indexado en Clase 2 (CORE A). Título de la contribución: "*KUMITRON: Artificial Intelligence System to Monitor Karate Fights that Synchronize Aerial Images with Physiological and Inertial Signals*"[1].
- **ACM UMAP 2021** (29th Conference on User Modeling, Adaptation and Personalization). El congreso ACM UMAP está indexado en Clase 3 (CORE B). Título de la contribución: "*Punch Anticipation in a Karate Combat with Computer Vision*"[2].
- **AIED 2021** (22nd International Conference on Artificial Intelligence in Education). Seleccionado para su presentación como Evento Interactivo[3], que posteriormente se

[1] https://dl.acm.org/doi/10.1145/3397482.3450730
[2] https://dl.acm.org/doi/10.1145/3450614.3461688
[3] https://aied2021.science.uu.nl/interactive-events/

recogió como ejemplo de sistema inteligente en el Showcase de IAIED[4]. El congreso AIED está indexado en Clase 3 (CORE B). Título de la contribución: "*KUMITRON: Learning in Pairs Karate related skills with Artificial Intelligence support*"[5].
- **MAIED 2021** (First International Workshop on Multimodal Artificial Intelligence in Education at AIED 2021). El congreso AIED está indexado en Clase 3 (CORE B). Título de la contribución: "*KUMITRON: A Multimodal Psychomotor Intelligent Learning System to Provide Personalized Support when Training Karate Combats*" [6].

Además, he participado (junto con otro alumno del máster que actualmente está haciendo el doctorado) en la elaboración de un capítulo titulado "*Intelligent Systems for Psychomotor Learning: A Systematic Review and two Cases of Study*" para el "Handbook of Artificial Intelligence in Education" que se va a publicar en breve en la editorial Edward Elgar Publishing y que está editado por tres investigadores de gran relevancia del campo, los profesores Benedict du Boulay, Tanja Mitrovic y Kalina Yacef. En este capítulo se presenta una revisión sistemática del estado del arte del campo de los sistemas inteligentes de aprendizaje psicomotor, se plantea un marco para guiar su desarrollo y se contextualiza en dicho marco mi propuesta de investigación, consistente en el sistema psicomotor inteligente KUMITRON. La editorial donde se va a publicar el libro ocupa la posición 28 de 259 en el ranking de editoriales SPI.

También he participado activamente en varios eventos de divulgación científica mostrando los avances del trabajo de investigación que iba realizando, tanto en la Semana de la Ciencia y la Innovación de Madrid 2020[7], como en dos ediciones de la Feria AULA (2021[8] y 2022[9]), donde mi trabajo de investigación en el Máster ha servido para mostrar las posibilidades de investigar al estudiar en la UNED a estudiantes de bachillerato para ayudarles a decidir sus estudios universitarios. Por su parte, en la Feria de Madrid es Ciencia de 2022[10] mostré a los participantes el funcionamiento del sistema KUMITRON y pudieron probarlo por sí mismos[11]. Incluso me han entrevistado en el programa de radio Radio 5.0[12]. También se han recogido reseñas de mi trabajo de investigación en el máster en medios de comunicación como Nova Ciencia[13] y la propia UNED[14].

---

[4] https://iaied.org/showcase
[5] https://aied2021.science.uu.nl/interactive-events/
[6] http://ceur-ws.org/Vol-2902/paper7.pdf
[7] http://www.madrimasd.org/semanaciencia2020/actividad/tecnologias-inteligentes-en-las-artes-marciales-sensores-algoritmos-y-feedback
[8] https://canal.uned.es/video/607ede0cb60923170814dcac
[9] http://portal.uned.es/portal/page?_pageid=93,71537978&_dad=portal&_schema=PORTAL
[10] https://extension.uned.es/actividad/idactividad/26445
[11] https://twitter.com/ETSIIUNED/status/1500026008426188800
[12] https://www.rtve.es/alacarta/audios/50/50-sensei-virtual-telespara-ver-juegos-olimpicos/5894461/
[13] https://novaciencia.es/inteligencia-artificial-aplicada-al-karate/
[14] http://portal.uned.es/portal/page?_pageid=93,71398470&_dad=portal&_schema=PORTAL

La infraestructura de recogida de datos está en proceso de solicitud de patente. Actualmente, está publicada en la Oficina Española de Patentes y Marcas[15]. También está publicado como WO2022008768 A1 (13.01.2022), y Número de Solicitud: OEPM PCT/ES2021/070433 (14.06.2021).

Además, se ha recibido financiación del programa Ekintzaile 2021 para ser parte de un programa de emprendimiento recibiendo ayudas a fondo perdido de hasta 30.000 Euros para la creación de startups con una base tecnológica e innovadora. El programa Ekintzaile, gestionado por el Grupo SPRI y con un presupuesto de 3,1 millones de euros. También se está recibiendo asesoramiento del COIE de la UNED orientado a la creación de una empresa de base tecnológica centrada en el sistema inteligente psicomotor KUMITRON, dentro del VIII edición del Programa de Creación de Empresas[16].

Antes de terminar, me gustaría destacar que la investigación realizada hasta la fecha ha tenido ya impacto, puesto que además de quedar finalista en el Premio Toribio Echevarría 2021[17], algunos de mis trabajos han sido ya citados por otros investigadores, destacando las siguientes citas recibidas en publicaciones científicas:

- *"Keep Me in the Loop: Real-Time Feedback with Multimodal Data"* (https://doi.org/10.1007/s40593-021-00281-z). Revista International Journal of Artificial Intelligence in Education. Publicado online en noviembre 2021.
- *"Unified End-to-End YOLOv5-HR-TCM Framework for Automatic 2D/3D Human Pose Estimation for Real-Time Applications"*[18]. Revista JCR SCIE Sensors. Publicado online en julio 2022.
- *"A Survey on Deep Learning Architectures in Human Activities Recognition Application in Sports Science, Healthcare, and Security"*[19]. Proceedings of the ICR'22 International Conference on Innovations in Computing Research (ICR 2022). Publicado online en agosto de 2022.

Por último, me gustaría manifestar que todos estas contribuciones no habrían sido posibles sin el apoyo, dedicación y orientaciones de la profesora Olga Santos, mi tutora en el máster y directora del TFM, quién me ha brindado la posibilidad de combinar mis inquietudes científicas con una de mis aficiones. Esta investigación está enmarcada en el escenario "SC4 habilidades motoras" liderado por la Dra. Santos en el proyecto de investigación INT^{2}AFF (INTelligent INTra-subject development approach to improve actions in AFFect-aware adaptive educational systems) financiado por el Ministerio de Ministerio de Ciencia, Innovación y Universidades (PGC2018-102279-B-100).

En Elgoibar, a 16 de septiembre de 2022
Jon Echeverría San Millán

---

15 https://invenes.oepm.es/InvenesWeb/detalle?referencia=PCT/ES2021/070433
16 https://www2.uned.es/bici/Curso2021-2022/220221/18-anexoII.pdf
17 https://www.toribioechevarria.com/finalistas-2021/
18 https://www.mdpi.com/1424-8220/22/14/5419
19 https://link.springer.com/chapter/10.1007/978-3-031-14054-9_13

Article

Exploring Pose Estimation with Computer Vision Processing to Model Psychomotor Performance in Karate Combats[1]

Jon Echeverria [1,*] and Olga C. Santos [2]

[1] Artificial Intelligence Department, Computer Science School, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain; jecheverr33@alumno.uned.es; jecheverria@dia.uned.es

[2] aDeNu Research Group, Artificial Intelligence Department, Computer Science School, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain; ocsantos@dia.uned.es

* Correspondence author

Abstract: Technological advances enable the design of systems that interact more closely with humans in a multitude of previously unsuspected fields. Martial arts are not outside the application of these techniques. From the point of view of the modeling of human movement in relation to the learning of complex motor skills, martial arts are of interest because they are articulated around a system of movements that are predefined, or at least, bounded, and governed by the laws of Physics. Their execution must be learned after continuous practice over time. Literature suggests that artificial intelligence algorithms, such as those used for computer vision, can model the movements performed. Thus, they can be compared with a good execution as well as analyze their temporal evolution during learning. We are exploring the application of this approach to model psychomotor performance in karate combats (called *kumites*), which are characterized by the explosiveness of their movements. In addition, modeling psychomotor performance in a *kumite* requires the modeling of the joint interaction of two participants, while most current research efforts in human movement computing focus on the modeling of movements performed individually. Thus, in this work, we explore how to apply a pose estimation algorithm to identify attack and defense movements performed by both karatekas in an *ippon kihon kumite* (a karate combat characterized by one-step conventional assault) and how to model their psychomotor performance. For this, we compare, using an error threshold, the differences in the angles between the execution in the model (recorded in the dataset) and the current execution. These comparisons can decrease the error threshold along the evolution of the karatekas, thus allowing to measure the psychomotor learning progress. In addition, postural identification of both karatekas during real *kumites* have also been made to confirm the viability of our proposal.

Keywords: human activity recognition (HAR); computer vision; deep learning; human pose estimation (HPE); OpenPose; martial arts; karate

1. Introduction

The range of possible human movements is enormous. Even the simplest gesture is the product of the activation and interaction of numerous and different motor units, which in turn can respond in very different ways, according to the stimuli that trigger its action. It is understood that except for the so-called genetic gestures, such as suckling action in

[1] This document, elaborated for the Master Thesis of the Research Master on Artificial Intelligence at UNED, extends significantly the work already published in Sensors (JCR SCIE indexed journal, Q2 in 2021) with the advances carried out during the first year of the Master. The details of that paper are: Echeverria J., Santos O.C. "Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation". Sensors. 2021; 21(24):8378. https://doi.org/10.3390/s21248378.

babies or other innate reflexes, all others are learned movements. In this context, new technologies could be used to **support the learning of these motor skills when their execution is complex and has to be practiced over and over until mastery**.

Human activity recognition (HAR) techniques have proliferated and focused on recognizing, identifying and classifying inputs through sensory signals, images or video, and are used to determine the type of activity that the person being analyzed is performing [1]. Following [1], human activities can be classified according to: (i) *gestures* (primitive movements of the body parts of a person that may correspond to a particular action of this person [2]); (ii) *atomic actions* (movements of a person describing a certain motion that may be part of more complex activities [3]); (iii) *human-to-object* or *human-to-human interactions* (human activities that involve two or more persons or objects [4]); (iv) *group actions* (activities performed by a group of persons [5]); (v) *behaviors* (physical actions that are associated with the emotions, personality, and psychological state of the individual [6]); and (vi) *events* (high-level activities that describe social actions between individuals and indicate the intention or the social role of a person [7].

HAR-type techniques are usually divided into two main groups [5], [8]: (i) HAR models based on image or video, and (ii) those that are based on signals collected from accelerometer, gyroscope, and/or other sensors. Both approaches have advantages and weaknesses, although they can be compatible and offer a hybrid solution to human modeling. There are papers such as [9] where an exhaustive analysis is made of the techniques used by both approaches, the existing public datasets, ongoing research and technological challenges. In **Figure 1** both approaches for the recognition of human activity are shown with an indication of the pre-processing and feature engineering methods approaches most used in each kind of HAR system.

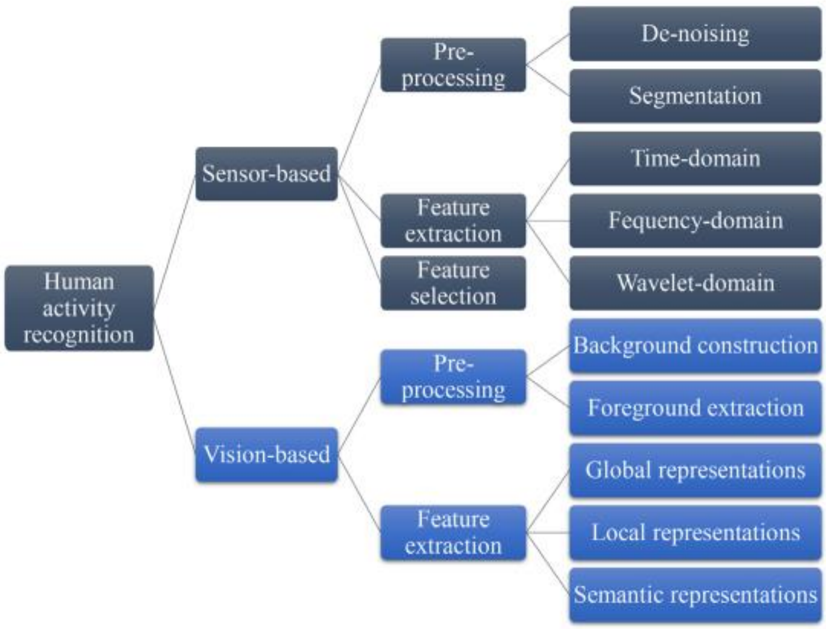


Figure 1: Standard pre-processing and feature engineering methods for sensor-based HAR and vision-based HAR (obtained from [9])

Due to legal and other technical issues [5], HAR sensor-based systems are increasing their usage. The scope of the sensors at present is large, with cheap and good quality sensors able to develop any system in any sector or specific field. Compilation papers ([10], [11], [12]) analyze the works carried out in this field up to now. In this way, HAR-type systems can gather data from diverse type of sensors such as: accelerometers (e.g., see [13], [14]), gyroscopes, which are usually combined with an accelerometer (e.g., see [15], [16]), GPS (e.g., see [17], [18]), pulse-meters (e.g., see [19], [20]), magnetometers (e.g., see [21],

[22]) and thermometers (e.g., see [23]). As analyzed in [24], the field is still emerging, and inertial-based sensors such as accelerometers and gyroscopes can be used to (i) recognize specific motion learning units and (ii) assess learning performance in a motion unit. However, they have some disadvantages compared to video-based approaches, such as personal satisfaction (uncomfortable to wear), need for multiple devices for team activities, imperfect sign, and more [25], [26].

In turn, the models developed based on images and video can provide more meaningful information about the movement, as they can record the movements performed by the skeleton of the body. It is applicable to many aspects of the recognition and modeling of human activity, such as medical, rehabilitation, sports, surveillance cameras, dancing, human–machine interfaces, art and entertainment, and robotics [27], [28], [29], [30], [31], [32], [33], [34], [35]. In particular, gesture and posture recognition and analysis is essential for various applications such as rehabilitation, sign language, recognition of driving fatigue, device control and others [36]. The creation of international challenges such as the ImageNet Large Scale Visual Recognition [37] in the field of computer vision analysis with deep learning techniques has contributed to the development of solutions using new convolutional algorithms: LeNet [38], AlexNet [39], VGG [40], ResNet [41], GoogleNet [42], etc. In this way, some of these networks are used by pose estimation algorithms. In particular, VGG is the backbone of Openpose, an algorithm that will be discussed in more detail later, and that will serve to help us in the identification and classification of movements and postures of this work.

The world of physical exercise and sport is susceptible to the application of these techniques. What is called the Artificial Intelligence of Things (AIoT) has already emerged, applicable to different sports [43]. In sports and exercise in general, both sensors and video processing are being applied to improve training efficiency ([44], [45], [46]) to develop sport and physical exercise systems. Regarding inertial sensor data, [47] provides a systematic review of the field, showing that sensors are being applied in the vast majority of sports, both worn by athletes and in sport tools. However, sensor based approaches in sport have some limitations, such as comfort, since many athletes refuse to wear sensors. In turn, several techniques based on computer vision for Physical Activity Recognition (PAR) have been used [48]: red-green-blue (RGB) images, optical flow, 2D depth maps, and 3D skeletons. They use diverse algorithms, such as Naive Bayes [49], Decision Trees [50], Support Vector Machines [51], Nearest Neighbor [52], Hidden Markov Models [53], and Convolutional Neural Networks [54]. A lot of money is generated around sports, and this means that computer vision technology is developing new tools to market in this field [55]. One of the most outstanding challenges of applying computer vision to sport today [56] is dealing with the high speed of the movements, both of athletes and accessories (ball, sport-practitioner...).

However and despite existing challenges to detect the movements performed in complex and high-speed HAR scenarios such as those in the sport domain, **currently any sensor-based or video-based HAR systems hardly address the problem of modeling the movements performed by the users from a psychomotor perspective** [57]**,** [58]**.** Modeling user psychomotor performance *requires comparing the current execution of the user with the same user along time (within-subject or intra-subject approach) or with other users (between-subject or inter-subject approach).* This modeling of the user performance would allow to provide personalized guidance to the user to improve the execution of the movements, as proposed in the sensing-modeling-designing-delivering (SMDD) framework described in [59], which can be applied along the lifecycle of technology-based educational systems [60].

In this sense, the following challenges when developing intelligent psychomotor learning systems are pointed out in [59]: (1) modeling psychomotor interaction and (2) providing adequate personalized psychomotor support. One system that follows the SMDD psychomotor framework is KSAS, which uses the inertial sensors from a smartphone to identify wrong movements in a sequence of predefined arm movements in a blocking set of American Kenpo Karate [61]. In turn, and also following the SMDD psychomotor framework, we are developing an intelligent infrastructure called KUMITRON that simultaneously gathers both sensor and video data from two karate practitioners [62], [63] to offer expert advice in real time for both practitioners on the karate combat strategy to follow. By applying sensors and computer vision we get the advantages of both approaches for human movement computing. KUMITRON can also be used to train motion anticipation performance in karate practice, based on improving peripheral vision, using computer vision filters [64].

The martial arts domain is useful to contextualize the research on modeling the user performance level on the psychomotor activity since martial arts require developing and working on psychomotor skills to progress in the practice. In fact, in martial arts the performance level of the practitioners is defined using colored belts [65]. From the point of view of the modeling of human movement in relation to the learning of high-speed and complex motor skills, martial arts are of interest because they are articulated around a system of movements that are predefined, or at least, bounded, and governed by the laws of Physics [66].

Nonetheless, there are several challenges for the development of intelligent psychomotor systems for martial arts [67]: (1) improve movement modeling and, specifically, movement modeling in combat; (2) improve interaction design to make the virtual learning environment more realistic (where available); (3) design motion modeling algorithms that are sensitive to relevant motion characteristics but insensitive to sensor inaccuracies, especially when using low cost wearables; and (4) create virtual reality environments where realistic force feedback can be provided.

Karate is one of the most known martial arts [68]. To understand the impact of karate on the practitioner, there is a fundamental reading by its founder [69]. The physical work and coordination of movements can be practiced individually by performing forms of movements called "*katas*", which are developed to train from the most basic to the most advanced movements. In turn, the karate combat (called "*kumite*") is developed by earning points on attack techniques that are applied to the opponent in combat. The different scores that a karateka (i.e., karate practitioner) can obtain are: *Ippon* (3 points), *Wazari* (2 points) and *Yuko* (1 point). The criterion for obtaining a certain point in *kumite* is conditioned by several factors, among others, that the stroke is technically well executed [70] (thus, properly taking advantage of the physics in karate [71]). Therefore karatekas must polish their fighting techniques to launch clear and high-speed winning shots that are rewarded with a good score. The training of the combat technique can be performed through the *katas (*individual movements that simulate fighting against an imaginary opponent), or through *Kihon kumite* exercises, where one practitioner has to apply the techniques in front of a partner, with the pressure that this entails.

There are some works that develop and apply movement modeling to the study of the technique performed by karatekas, but only individually [72], [73], [74]. However, in our work, **we are exploring if it is also feasible to develop and apply computer vision techniques to other types of exercises in which the karateka must apply the techniques with the pressure of an opponent**, such as in *Kihon kumite* exercises. Thus, in the current work, we focus on how to model the movements performed in a *kumite* by processing video recordings obtained with KUMITRON when performing the complex and high-speed dynamic movements of karate combats to be able to model the user performance, and thus, to compare the current execution with previous ones of the same user or with executions from other users. To approach the processing, we divide the movement into the different stances or postures (or poses, as commonly used in the technical jargon) the user takes when performing the *kumite*.

In this context, we address the following **research question**: "*Is it possible to detect and identify the postures (poses) within a karate combat among two karate practitioners using computer vision techniques in order to model the psychomotor performance of each practitioner, that is, to* *compare the current execution with previous ones of the same user or with executions from other users?".*

To resolve this issue, our long-term approach is oriented to combine both the perspective of sensors and computer vision, as proposed in [62]. However, in the current research reported in this work we focus only on the progress we have made so far regarding the computer vision algorithms. Thus, the main objective of the current work is to explore the algorithm support required to develop an intelligent system capable of identifying the postures performed during a karate combat and compare them with the way they were performed in previous *kumite* executions of the same or different users. In this way, personalized feedback can be provided, when needed, to improve the performance in the combat, thus, supporting psychomotor learning with technology. **The novelty of our current research lies in applying computer vision to an explosive activity where an individual interacts with another while performing rapid and strong movements that change quickly in reaction to the opponent's actions.** According to the review of the field in human movement computing that is reported next, computer vision has been applied to identify postures performed individually in sports in general and karate in particular, but we have not found postural identification in the interaction of two or more individuals performing the same activity together.

Thus, and in addition to contributing to the development of an innovative intelligent psychomotor system based on data-driven knowledge, our research seeks to offer a series of advantages in the learning of martial arts, such as studying the movements of practitioners in front of opponents of different heights, in real time and in a fluid way, which can have an impact in the modeling of the user performance in the *kumite*. It is also intended to use the advances of this study to improve the technique of karate practitioners applying explosive movements, which is essential for the assimilation of the technique and hence, achieve improvement in the performance of the movements, according to psychomotor theories [75].

The rest of the paper is structured as follows. Related works are in Section 2. In Section 3 we select the computer vision algorithm used and present the methodology to identify the user poses within the movements executed as well as the psychomotor modeling approach. In Section 4, we justify the need to develop our own dataset and describe the dataset obtained to carry out the research. After that, in Section 5 we analyze the results obtained in this research. In Section 6, we discuss the results, the limitations and suggest some ideas for future work. Finally, conclusions are gathered in Section 7.

2. Related Works

The introduction of new technologies and computational approaches is providing more opportunities for human movement recognition techniques that are applied to sports. The extraction of activity data, together with their analysis with data mining algorithms is making the training efficiency higher. Barla [76] describes seven applications of human pose estimation: i) AI-powered personal trainers, ii) robotics, iii) motion capture, iv) augmented reality, v) athlete pose detection, vi) motion tracking for gaming, and vii) infant motion analysis. In this context, we align with the fifth point, where pose detection can help players to improve their technique and achieve better results. Apart from that, pose detection can be used to analyze and learn about the strength and weaknesses of the opponent, which is key for professional athletes and their trainers.

Through the acronym HPE (Human Pose Estimation), studies have been found in which different technologies are applied that seek to model the movement of people when performing physical activities. Human modeling technologies have been applied for the analysis of human movements in sport as described in [77]. In particular, different computer vision techniques can be applied to detect athletes, estimate pose, detect movements and recognize actions [78]. Currently, mainstream sports have begun to use HPE techniques in practice and competition [79], [80], [56].

In our research, we focus on posture detection within a movement to support the modeling of the user performance by comparing different executions of the same postures of the movement by different users or by the same user along time. Thus, we have reviewed works that delve into existing methods: skeleton based models, contour-based models, etc. As discussed in [81], [82], [83], 3D computer vision methodologies can be used for the estimation of the pose of an athlete in 3D, where the coordinates of the three axes (x, y, z) are necessary, and this requires the use of depth-type cameras, capable of estimating the depth of the image and video received.

Martial arts, similar to sport in general, is a discipline where human modeling techniques and HPE are being applied [84], [85], [86]. Motion capture (mocap) approaches are sometimes used and can add sensors to computer vision for motion modeling, combining the interpretation of the video image with the interpretation of the signals (accelerometer, gyroscope, magnetometer, GPS). In any case, the pose estimation work is usually performed individually [87], [88], while the martial artist performs the techniques by themselves.

The most recent HPE methods are classified into bottom-up and top-down approaches, depending on how they operate. Bottom-up approaches [89], [90], [91], [92] first detect individual body joints and then group them into people. On the other hand, top-down approaches [93], [94], [95], [96] first detect people bounding boxes and then predict their joint locations within each region. We will work the bottom up approach, which is OpenPose's way of extracting and allocating keypoints.

Other approaches to develop martial art systems relay on virtual reality (VR). In particular, VR techniques have been used to project the avatar or the image of martial artists to a virtual environment programmed in a video game for individual practice. For instance, [97] combines VR and computer vision in karate training concluding that VR training is useful to improve response behavior in young karate athletes. Following this line of research, [98] has analyzed the improvement of "Soto Uke", a defensive karate technique, using VR, concluding that VR is a suitable tool to acquire sport-specific techniques, especially for beginners. In this sense, although the improvement of motor activity using VR has not yet been substantially confirmed, improvements have been found in training in areas such as: i) motivation of practitioners to use new technologies, ii) better training programs (the virtual opponent never gets tired), iii) variety of exercises for the different practitioners of a sports class that frees the instructor to control many people at the same time. Thus, and in addition to karate [99], VR is being explored in different sports such as juggling [100], throwing darts [101] and baseball batting [102]. Moreover, hybrid systems for martial arts training [103] are being designed that combine VR and corporal lords.

The application of computer vision algorithms in the martial arts domain not only focuses on the identification of the pose and the movement of the users, but there are also advances in the prediction of the next attack, which can also depend on the user fighting style. For this, the user can be monitored using residual RGB and CNN network frames to which LSTM neural networks are applied that predict the next attack movement in 2D [104]. Moreover, computer vision together with sensor data can also be used to record audiovisual teaching material for Physics learning from the interaction of two (human) bodies as in Phy+Aik [105], where Aikido techniques practiced in pairs are monitored and used to show Physics concepts of circular motion when applying a defensive technique to the attack received.

The classification of postures by computer vision is not only of interest to help novices and experts when training, but this technology is also being used in sport competitions, for instance, to support referees in Kung Fu championships ([106], [107]). Karate was included in the Tokyo Olympics, and thus, there have been efforts in applying new technologies to the modeling of its movements from a computing perspective to improve the practitioners' psychomotor performance. In this sense, [108] has reviewed the technologies used in twelve articles to analyze the *"mawasi geri"* (side kick) technique, finding that several kinds of inputs, such as 3D video image, inertial sensors (accelerometers, gyroscopes, magnetometers) and EMG sensors can be used to study the speed, position, movements of body parts, working muscles, etc. There are also studies on the *"mae geri"* movement (forward kick) using the Vicon optical system (with twelve MX-13 cameras) to create pattern plots and perform statistical comparison among expert karatekas who performed the technique [109]. Sensors are also used for the analysis of karate movements (e.g., [110], [111], [112]) using Dynamic Time Warping and Support Vector Machines. In this context, it is also relevant to know that datasets of karate movements (such as [113]) have been created for public use as in the above works, but they only record individual movements.

Kinematics has also been used to analyze intra-segment coordination due to the importance of speed and precision of blows in karate as in [114], where a Vicon camera system consisting of seven cameras (T10 model) is used. These cameras capture the markers that the practitioners wear, and which are divided into sub-elite and elite groups. In this way, a comparison of the technical skill among both groups is made. Using a gesture description language classifier and comparing it with Markov models, different karate techniques individually performed are analyzed to determine the precision of the model [74].

In addition to modeling the techniques performed individually, some works have focused on the attributes needed to improve the performance in martial arts practice. For instance, VR glasses and video have been used to improve peripheral vision and anticipation [97], [115], which has also been explored in KUMITRON [64] with computer vision filters.

The conclusion that we reached after the literature review carried out is that new technologies are being introduced in many sectors, including sports, and hence, martial arts are not an exception. The computation approaches that can be applied vary. In some cases, the signals obtained from sensors are processed, in others computer vision algorithms are used, and sometimes both are combined (as in the intelligent psychomotor system called KUMITRON that we are developing). In the case of karate, studies are being carried out, although mainly focus on analyzing how practitioners perform the techniques individually, and some have developed public datasets for this purpose. Thus, **an opportunity has been identified to study if existing computational approaches can be applied to model the movements during the joint practice of several practitioners performing quick and explosive movements.** In this context, in addition to the individual physical challenge of making the movement in the correct way, other variables such as orientation, fatigue, adaptation to the anatomy of another person, or the affective state can be of relevance to model the user psychomotor performance when the motor interaction is performed in pairs and the execution of one of the practitioners depends on the movement performed by the other. Nevertheless, **none of the works reviewed uses the modeling of the motion to evaluate the psychomotor performance such that appropriate feedback can be provided when needed to improve the technique executed**. Since computer vision seem to produce good results for HPE, in this paper we explore existing computer vision algorithms that can be used to estimate the pose of karatekas in a combat, select one and use it on a dataset that we had to create with some *kumite* movements aimed to address the research question raised in the introduction, which focuses on detecting and identifying the karateka poses within a *kumite* movement in order to compare the current execution with previous ones of the same user or with executions from other users.

3. Approach

The modeling of the psychomotor performance is key to support the learning of motor skills [116] and is one of the steps defined in the SMDD framework [59]. Following the SMDD framework, in the current work, the sensing of the movement is done with computer vision techniques. The modeling normally involves a comparison of the movements of an apprentice against the movement of an expert, but it can compare different executions of the same user or with other users. In any case, a recent systematic review [58] highlights the need for more research to design appropriate psychomotor feedback, and the importance of students being able to get advice to improve their motor skills, both with immediate delivery or getting feedback after the training has been completed.

In order to address the research question of the current work with the focus on how to ***model the psychomotor performance of each karate practitioner in the kumite comparing the current execution with previous ones of the same user or with executions from other users for each of the postures identified***, we have studied different technical aspects of karate. According to the literature [117], [118] importance is given to the correct posture and settlement of the lower body.

In this section, we present some computer vision algorithms that can be used for HPE and present the methodological approach proposed to answer the research question.

3.1. Computer Vision Algorithms for Human Pose Estimation

HPE is a fundamental computer vision problem aiming to extract the posture of human bodies from input images by estimating the locations and connections of body segments [119], [120], [121], [122], [123]. They can follow both pre-processing approaches mentioned in **Figure 1**. On the one hand, the background construction works recording videos of the activity from a fixed camera. On the other hand, for foreground extraction-based segmentation, human activities are recorded by a pan-tilt-zoom camera or camera mounted on moving objects, such as moving robots, cars, and unmanned aerial vehicles. Depending on the activity, one or the other should be applied. For example, for cycling there is no other option than to work with mobile cameras, although their efficiency is lower for tracking than fixed ones. **Figure 2** shows a classification of HPE based on the 2D or 3D dimensions considered in the video image, as well as if there is one several people on the image.

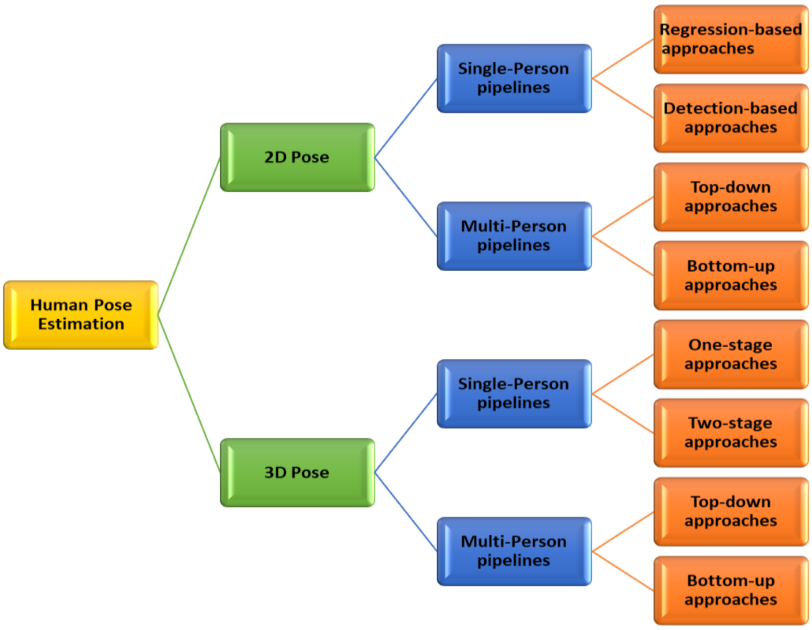


Figure 2: Taxonomy of human pose estimation HPE (obtained from [124])

For this work we have selected the simplest technique, which is to recognize the movement made by a person in 2D, and then increase the complexity by applying the multiperson in 2D. To work with 3D, we need to use specific hardware (depth-type cameras), so if the first approach is positive, further investigation and acquisition of hardware will follow.

After the analysis of the state of the art, we have identified several computer vision algorithms that produce good results for HPE, and which are listed in **Table 1**. According to the literature, the top three algorithms from **Table 1** offering best results are WrnchAI, OpenPose and AlphaPose. Several comparisons with varied conclusions have been made among them. According to LearnOpenCV [125], WrnchAI and OpenPose offer similar features, although WrnchAI seems to be faster than OpenPose. In turn, [126] concludes that AlphaPose is above both when used for weight lifting in 2D vision. Other studies such as [127] find OpenPose superior and more robust when applied to real situations outside of the specific datasets such as MPII [128] and COCO [129] datasets.

Table 1: Description of existing computer vision algorithms for pose estimation.

Pose Estimation Algorithms	Description
AlphaPose (https://github.com/MVIG-SJTU/AlphaPose, accessed on 28 November 2021)	Presented in 2016 [130], it is an algorithm that allows estimating the pose of one or more individuals. It is the first open source system that has reached the following records: 80+ mAP (82.1 mAP) on MPII dataset and 70+ mAP (72.3 mAP) on COCO dataset. This means that the algorithm is more precise in detecting keypoints in comparison with others. AlphaPose is free to use and distribute as long as it is not used for commercial purposes.
DeepCut (https://github.com/eldar/deepcut, accessed on 28 November 2021)	System developed in 2016 [131] presented as a multi-person computer vision system, with deeper, stronger and faster features compared to the state of the art at that time. It works bottom-up for image treatment. The way of working is to detect the people who are in an image to later predict the joint locations. It can be applied to both images and video of sports such as baseball, athletics or soccer.
Deep Pose (https://github.com/mitmul/deeppose, accessed on 28 November 2021)	An algorithm presented in 2014 [132] that estimates the human pose using Deep Neural Networks (DNN). To do this, a regression based on DNN is performed to estimate the joints. In challenges of precision in the classification of images [133], DeepPose obtained better results than the rest of the works, becoming a benchmark of that moment.
DensePose (https://github.com/facebookresearch/DensePose, accessed on 28 November 2021)	It is an algorithm developed in 2018 by members of Facebook [134] that maps the pixels of the human body in 2D to turn it into a 3D surface that covers the human body. It serves one or more individuals. It is being used to determine the surface of the human body for different purposes such as trying on virtually an article of clothing on the avatar created for oneself.
High Resolution Net (HRNet) (https://github.com/HRNet/HigherHRNet-Human-Pose-Estimation, accessed on 28 November 2021)	Neural network architecture for the estimation of human pose developed in 2019 by Microsoft [135]. It is also used for semantic segmentation and object detection. Despite being a relatively new model, it is becoming a benchmark in the field of computer vision algorithms. It has been the winner in several computer vision tournaments, for example in ICCV2019 [136]. It is a useful architecture to implement in the postural analysis of televised events since it makes high-resolution estimates of postures.
OpenPose (https://github.com/CMU-Perceptual-Computing-Lab/openpose, accessed on 28 November 2021)	Computer vision algorithm for the estimation of pose in real time of several people in 2D developed in 2017 [137]. It has undergone functionalities extensions, and currently allows to be used in 3D, hand point detection, face detection, and work with Unity. The OpenPose API allows obtaining the image from various devices: recorded video, streaming video, webcam, etc. Other hardware is also supported, such as CUDA GPUs, OpenCL GPUs, and CPU-only devices.
PoseNet (https://github.com/tensorflow/tfjs-models/tree/master/posenet,	It is a pose estimator for a single person or several people, offering 17 keypoints with which to model the human body. It was developed in 2015 [138]. At first, it was aimed at lightweight devices

Pose Estimation Algorithms	Description
accessed on 28 November 2021)	such as mobile phones or browsers, although today it has advanced and improved performance.
WrnchAI (https://go.hinge-health.com/wrnch, accessed on 28 November 2021)	WrnchAI is a human deposit estimation algorithm developed by a company based in Canada in 2014 and released only under license. It can be used for one or several individuals making use of the low latency engine, being a system compatible with all types of videos. Due to its commercial use, we could not find any scientific paper describing it.
Yolo-Pose (https://github.com/TexasInstruments/edgeai-yolov5)	Developed by a team of engineers [139] at Texas Instruments Inc. YOLO is a region network that enables real-time object detection. It is fast and precise and can process images at 30 FPS. Furthermore it is possible to tradeoff between speed and accuracy simply by changing the size of the model without needing to retrain it. The main advantages of YOLO compared to other classifiers is that it looks at the entire image only once at the time of the test, so the predictions are informed by the global context of the image, and it also makes predictions using only a single neural network.

From our own review, we conclude that OpenPose is the HPE algorithm that has generated more literature works and seems to have the largest community of developers. OpenPose has been used in multiple areas: sports [140], telerehabilitation [141], HAR [142], [143], [144], artistic disciplines [145], identification of multi-person groups [146], and VR [147]. Thus, **we have selected OpenPose algorithm for the HAR processing of this research** due to the following reasons: (i) it is open source, (ii) it can be applied in real situations with new video inputs [127], (iii) there is a large number of projects available with code and examples, (iv) it is widely reported in scientific papers, (v) there is a strong developers community, and (vi) the API gives users the flexibility of selecting source images from camera fields, webcams, and others.

3.1.1 About OpenPose

OpenPose [137] is a computer vision algorithm proposed by the Cognitive Computing Laboratory at Carnegie Mellon University for the real-time estimation of the shape of the bodies, faces and hands of various people. OpenPose provides 2D and 3D multi-person hotspot detection, as well as a calibration toolbox for estimating specific region parameters. OpenPose accepts many types of input, which can be images, videos, webcams, etc. Similarly, its output is also varied, which can be PNG, JPG, AVI or JSON, XML and YML. The input and output parameters can also be adjusted for different needs. OpenPose provides a C ++ API and works on both CPU and GPU (including versions compatible with AMD graphics cards). The main characteristics are summarized in **Table 2**.

Table 2: Technical characteristics of the OpenPose algorithm (obtained from https://github.com/CMU-Perceptual-Computing-Lab/openpose, accessed on 28 November 2021).

	OpenPose Features
Main functionality (with a plain camera)	Detection of keypoints of several people in real time 2D. Body/foot keypoint estimate of 15, 18 or 25 keypoints, including 6 foot keypoints. Execution time invariable with respect to the number of people detected. Handheld keypoint estimate of 2 × 21 keypoints. The execution time depends on the number of people detected. Estimation of keypoints of faces of 70 keypoints. The execution time depends on the number of people detected.
Real-time single-person 3D keypoint detection	3D triangulation of multiple unique views. Synchronization of Flir cameras managed. Compatible with Flir/Point Gray cameras.
Calibration Toolbox	Estimation of the distortion, intrinsic and extrinsic parameters of the camera.
Input	Image, Video, Webcam, Flir/Point Gray, IP Camera, and support for adding your own custom input source (e.g., depth camera).
Output	Basic image + keypoint display/save (PNG, JPG, AVI...), keypoint save (JSON, XML, YML...), keypoints as array class and support to add your own code custom output (e.g., some fancy user interface).

As mentioned above, we have started by exploring the 2D solutions that OpenPose offers so that it can be used with different plain cameras such as the one in a webcam, a mobile phone or even the camera of a drone (which is used in KUMITRON system [62]). Applying 3D will require the use of depth cameras.

As described in [137], OpenPose algorithm works as follows:

1. Deep learning bases the estimation of pose on variations of Convolutional Neural Networks (CNN). These architectures have a strong mathematical basis on which these models are built:

$$(f * g) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau = \int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau \quad (1)$$

2. Apply ReLu (REctified Linear Unit). The rectifier function is applied to increase the non-linearity in the CNN.
3. Group. It is based in spatial invariance, a concept in which the location of an object in an image does not affect the ability of the neural network to detect its specific characteristics. Thus, the clustering allows CNN to detect features in multiple images regardless of the lighting difference in the pictures and the different angles of the images.
4. Flattening. Once the grouped featured map is obtained, the next step is to flatten it. The flattening involves transforming the entire grouped feature map matrix into a single column which is then fed to the neural network for processing.
5. Full connection. After flattening, the flattened feature map is passed through a network neuronal. This step is made up of the input layer, the fully connected layer, and the output layer. The output layer is where the predicted classes are provided. The final values produced by the neural network do not usually add up to one. However, it is important that these values are reduced to numbers between zero and one, which represent the probability of each class. This is the role of the SoftMax function.

$$\sigma\colon \mathbb{R}^k \to (0,1)^k$$

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{1\Leftarrow 1}^{k} e^{z_k}} \; for\; j = 1,\ldots..,K. \tag{2}$$

3.2. Methodological approach

The objective of the research introduced in this paper is to determine if computer vision algorithms (in this case, OpenPose) are useful to identify the movements performed in a Karate *kumite* by both practitioners so that their performance can be modeled and then used to provide some personalized support if needed. To answer the question, first we need to identify each of the postures performed by each of the karate practitioners along the movements in the combat. This corresponds to the modeling of atomic actions [3] according to [1]. Then, we need to identify the differences in the execution between several executions of the movements by the same or different users. Thus, a dataset is needed. As explained in Section 4.2, a dataset has been prepared following predefined *Kihon Kumite* movements both corresponding to attack and defense postures.

To process the dataset, the process defined in the scheme in **Figure 3** has been followed, which shows the steps proposed for karate movement recognition using OpenPose algorithm to extract the features from the images and used data mining algorithms on these features to classify the movements.

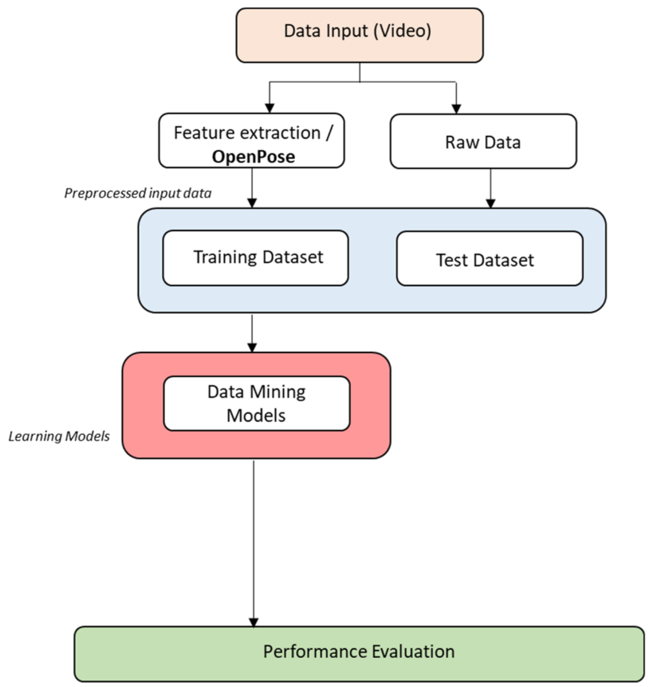


Figure 3: Steps for Karate movement recognition using OpenPose

In Section 3.3 we outline the experiments prepared to evaluate the research question.

3.2.1. Processing steps

In order to classify the movements performed during the *kumite*, the processing of the dataset consists of two clearly differentiated parts: (i) the training of the classification algorithms of the system using the features extracted with the OpenPose algorithm, and (ii) the application of the system to the input of non-preprocessed movements (raw data). Following the classification represented previously in **Figure 1**, the proposed approach applies semantic representation for the feature extraction based on imitating the human perception of an activity. These two stages identified in **Figure 3** are broken down into the following sub-stages:

1. **Acquisition of data input:** Record the movements to create the dataset to be used in the experiment, and later to test it.
2. **Feature extraction:** Applying the OpenPose algorithm to the dataset to group anatomical positions of the body (called keypoints) into triplets to calculate the angle, which allow generating a pre-processed input data file for algorithm training (see **Figure 4**). OpenPose allows to have the 2D position of each point (x, y). Thus, by having the coordinates of three consecutive points, the angle formed by those three with respect to the central one is calculated. An example is provided next.
3. **Train a movement classifier:** With the pre-processed data from point 2, train data mining algorithms to classify and identify the movements. Several data mining algorithms can be used for the classification in the current work generating the corresponding learning models (see below). For the evaluation of the classification performance, 10-fold cross validation is proposed, following [148], [149], [150].
4. **Test the movement classifier:** Apply the trained classifier to the non-preprocessed input (raw data) with the movements performed by the karateka .
5. **Evaluate the performance of the classifiers:** Compare the results obtained by each algorithm in the classification process with usual data mining metrics ([151], [152], [153]).

OpenPose works with 25 keypoints (numbered from 0 to 24) for the extraction of input data, which are used to generate triples, as it can be seen in **Figure 4**. OpenPose keypoints are grouped into triplets of three consecutive points, which are the input attributes (angles) for classifying the output classes. To define the keypoints, OpenPose expanded the 18 keypoints of COCO dataset (https://cocodataset.org/#home) with the ones for the feet and waist from the Human Foot Keypoint dataset (https://cmu-perceptual-computing-lab.github.io/foot_keypoint_dataset/).

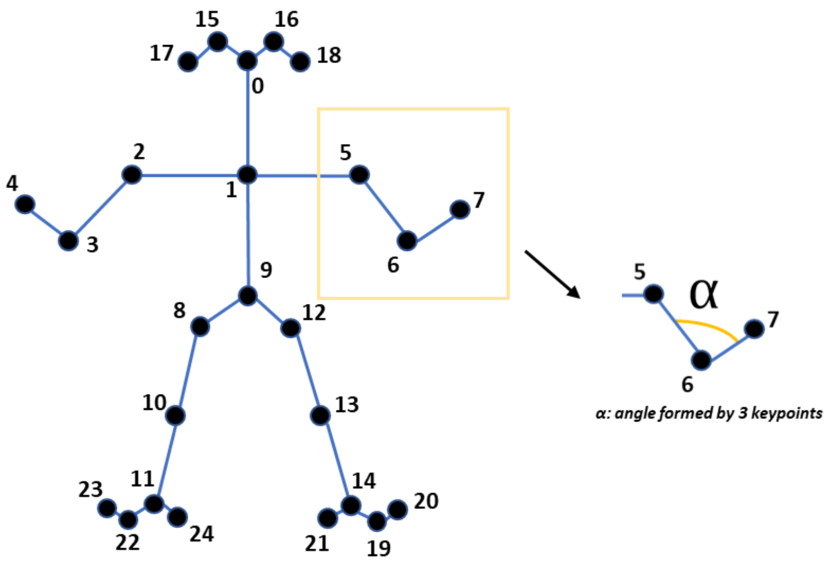


Figure 4: Example of angle feature extraction from the keypoints obtained by combining COCO and Human Foot Keypoint datasets

For the creation of a dataset, the relevant points are grouped three by three in the formation of angles, and each grouping is an attribute of the dataset. As an example, in **Figure 4** the keypoints numbered 5, 6 and 7 that correspond to the main part "Shoulders

Arm Left" are identified separately. The grouping of the keypoints in triples makes it possible to avoid the variability of the points depending on the height of the practitioners, by calculating their angle. In this way, a person 190 cm tall and another 170 cm tall will have similar angles for the same karate posture, compared to the greater variation of the keypoints due to the different height of the bodies. This allows the algorithm to be trained more efficiently and requires fewer inputs to achieve optimal computational performance.

As a result, **Figure 5** shows the processing flow proposed regarding the identification of *kumite* postures. On the top part of the figure, the data models are trained to identify all the movements selected for the study. This is done by generating a dataset with the angles of each karate technique, which is trained according to different data mining algorithms. Subsequently, these trained algorithms are to be used on real *kumite* movements obtained in real time to identify the *kumite* posture performed, as shown on the bottom.

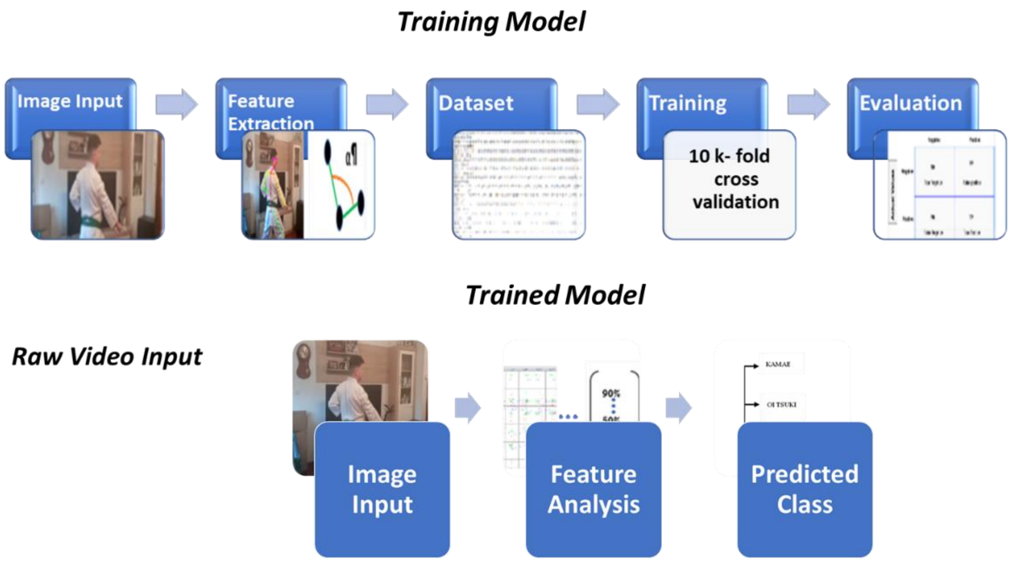


Figure 5: Training model (top) and the raw data input process flow (bottom) proposed in this research.

3.2.2. Data mining algorithms

To select the data mining algorithms the types of classifiers applied in general to computer vision have been studied. Some works such as [154], [155], [156], [157], [158] use machine learning (ML) algorithms, mainly decision trees and Bayesian networks. However, deep learning (DL) techniques are more and more used for the identification and qualification of video images as in [159], [160], [161], [162]. In particular, a deep learning classification algorithm that is having very good results according to the studies found is the Weka DeepLearning4j algorithm ([163], [164], [165]). The Weka Deep Learning kit allows to create advanced Neural Network Layers as it is explained in [166].

In addition to a general analysis of ML and DL algorithms, specific application to sports, human modeling and video image processing were also sought. The BayesNet algorithm has also been used to estimate human pose in other similar experiments [167], [168], [169]. Another algorithm that has been applied to this type of classifications is the J48 decision tree [170], [171], [172]. These two ML algorithms were selected for the classification of the movements, as well as two neural network algorithms included in the WEKA application: the MultiLayerPerceptron (MLP) algorithm, which has also been used in different works as well as similar algorithms [173], [174], and the aforementioned DeepLearning4J.

Note that for the implementation of the ML and DL classification models, a first approach was first made using the WEKA suite. This is a tool that provides ease of work for testing quickly and efficiently. After preliminary tests with WEKA, the Keras and Scikit learn Python libraries were implemented to move forward and implement more efficient and personalized models. Thus, once the preliminary tests with WEKA were carried out, the algorithms were implemented in Python. Because Python is currently a very popular language with many libraries and resources, it combines advanced DM libraries such as NumPy and ScyPi, allowing Python to be one of the best performing languages for data mining projects. In particular, for this work, PyCharm Community 2020.3.5 has also been used as an IDE, to which is added the Data Analysis part using Google Colab, where the algorithms are trained with the usual Python libraries for data such as: NumPy, Pandas, SciPy, Matplotlib, TensorFlow and others.

Once the programming libraries to be used have been decided, we advance in the implementation and selection of models. The algorithms to implement depend on the type of problem to be solved. In this case, it is a matter of identifying the karate postures performed, which are unique, in such a way that an entry will not be, for example, an "orange" and a "pear"; or it is a “pear”, or it is an “orange” In addition, the number of postures to identify is predefined, so a binary model solution cannot be applied. As a similar problem, the classification of *Iris* flower types has been found, a typical problem in ML. For the resolution of this multiclass problem, there is literature that provides guidance on the type of appropriate algorithm to use in the literature ([175], [176], [177]). After studying the Python Scikit-learn[2] libraries, the algorithms that could help us in this task were chosen: i) K Nearest Neighbors (KNN) ii) Random Forest (RF) iii) Support Vector Machine (SVM) iv) Naïve Bayes(NB) v) Logistic Regression(LR) vi) Decision Tree(DT).

DL algorithms can be described as particular instances of a fairly simple recipe, applied to solve a particular machine learning problem. The big difference between DL and the rest of the learning techniques is the capacity for abstraction. Its ability to form hierarchies of concepts using multiple levels of representation. Abstraction is the way humans deal with the complexity of the world around us. DL techniques aim to automate this abstraction process, creating new abstractions on top of existing ones. The most common topology of neural networks is made up of multiple layers, each of which receives as input the output of the previous layer. Deep Learning techniques start from the basic idea that if we are able to successfully learn multiple levels of representation, we will be able to generalize correctly. In this case we were talking about a multiclass classification problem in which we assigned to each image (obtaining the keypoints) a single label from a set of possible ones.

One of the networks selected for the job has been the multilayer perceptron (MLP), that is a feedforward artificial neural network model that maps input data sets to a set of appropriate outputs. An MLP consists of multiple layers and each layer is fully connected to the following one. The nodes of the layers are neurons with nonlinear activation functions, except for the nodes of the input layer. Between the input and the output layer there may be one or more nonlinear hidden layers. In addition, a neural network, "NN1" has been designed intuitively to try to improve the results and have the flexibility to modify and make improvements depending on the needs.

Thus, although statistical comparisons can be applied to ML models, the comparison between ML and DL models already seems more difficult since there are many parameters to consider when deciding which model is more suitable for our project. The **Table 3** shows different characteristics between ML and DL.

[2] https://scikit-learn.org/stable/modules/multiclass.html#multiclass-classification

Table 3: Machine Learning vs. Deep Learning (obtained from [178])

Characteristics	Machine Learning	Deep Learning
Data Requirement	Small/Medium	Large
Accuracy	High accuracy	Medium accuracy
Preprocessing phase	Needed	Not needed
Training time	Short time	Takes longer time
Interpretability	From easy (Tree, logistic) to difficult (SVM)	From difficult to impossible
Hardware requirement	Trains on CPU	Requires GPU

3.3. Studies to answer the research question

As explained in **Figure 5**, the first step consists of creating a dataset, and selecting the classification model based on the comparative performance results between them. Once this is done, video tests are performed on the application to verify, on the one hand, that it classifies the positions well, does not confuse them, and does not attribute a dataset label to other positions. That is, a posture outside the dataset is not confused with one from the dataset. In order to answer the research question posed in this work, we have proposed the following studies.

- **S1. Results of classifying the correct postures.** Videos of the postures of the dataset will be shown to the application, without intermediate postures (movements between scent and scent) and only those defined in the dataset are classified. The objective is to verify that the application is able to correctly identify the postures in video as with the tests of the models and the dataset. A 1-minute video has been prepared, with 10 seconds of each posture, to record the correction of the classification.
- **S2. Classification of scents that are not in the dataset.** Some modifications are going to be created from the postures of the datasets, to check if similar postures with few modifications with respect to another of the data, is not confused by the application. That is, a posture outside the dataset is identified in any way, avoiding false predictions. This test is important, since during the *ippon kihon kumite* intermediate postures are performed that can confuse the application and give a false posture.
- **S3. Posture classification of two karatekas.** In this study, the videos of two karatekas during a real time kumite interaction. The postures made are those of the designed dataset, and to avoid noise, the images are passed without intermediate postures. It is verified that the received predictions coincide with the postures of the two karatekas.

4. Dataset Construction

As introduced in Section 2, there are available datasets of karate movements. However, after reviewing all the ones we have come across, none of them includes the movements we required for our research as they only record individual movements. We need recorded karate techniques from both practitioners together to develop a pattern of *kumite* sequences. Thus, we had to create our own one[3].

[3] https://raw.githubusercontent.com/Kumitron/OpenPoseDataset/main/JohnyXX_LyR_juntos_2.csv

4.1. Available Online Open Datasets

There are currently multiple repositories of public datasets on the Internet. In fact Google has created a specific search engine for public datasets (https://datasetsearch.research.google.com/). Other well-known datasets repositories are: Kaggle[4] repository, UCI Machine Learning Repository[5] or Microsoft Datasets[6].

Regarding human movement modeling, there are also several datasets such as COCO[7] and MPII[8], where there are images and videos recorded of people performing different types of activities. Regarding martial arts in general and karate in particular, many of the scientific papers create their own dataset for their projects [179], [180]. In **Table 4**, the repositories of public datasets of martial arts that we have found are compiled. Despite not having found datasets that fit our needs (i.e., recordings of *kumite* interactions), we found interesting repositories that may help us in future work or extensions of our research.

Table 4: Martial arts open datasets

N.	Dataset	Description	Karate movements
1	Martial Arts, Dancing and Sports (MADS) [87]	Two martial art masters, two dancers and an athlete performed these actions while being recorded with either multiple cameras or a stereo depth camera. The depth data (stereo image) was captured at 10 fps or 20 fps. The resolution of the images is 1024 × 768. The MADS dataset contains 5 action categories (Tai-chi, Karate, Jazz dance, Hip-hop dance, and Sports), totalling about 53,000 frames.	Katas and special movements: 3000 frames in total. 1. "Second Basic Form" ("Fukyugata Ni") 2. "Third Basic Form" ("Fukyugata Sandan") 3. "Third Peace Form" ("Pinan Sandan") 4. "Fifth Peace Form" ("Pinan Godan") 5. "Straddle Stance" ("Naihanchi Nidan") 6. "Horse Riding Stance" ("Naihanchi Nidan Sandan")
2	NTU RGB+D 120 [181]	NTU RGB+D 120 is a large-scale dataset for RGB+D human action recognition, which is collected from 106 distinct subjects and contains more than 114 thousand video samples and 8 million frames. This dataset contains 120 different action classes including daily, mutual, and health-related activities.	HAR videos using striking (general) techniques: 1. A51: kicking 2. A50: punch/slap 3. A102: side kick
3	Infomus Karate dataset [182]	Karate dataset. Contains full-body movements of athletes performing katas. Data includes 3D positions of 25 body joints at 250 Hz. In total 28 trials were obtained.	Karate Shotokan katas. 7 athletes, performing 2 katas two times: Bassai Dai and Heian Yondan. Athletes perform katas with their own speed and rhythm.
4	Motion Database (http://gdl.org.pl/) [183]	Karate dataset. 3D selected katas and techniques. Recording was made with Shadow 2.0 MoCap system.	3 karate styles movements are recorded: Shotokan, Shorin-Ryu,

4 https://www.kaggle.com/

5 https://archive.ics.uci.edu/ml/index.php

6 https://msropendata.com

7 https://cocodataset.org/#home

8 http://human-pose.mpi-inf.mpg.de/

N.	Dataset	Description	Karate movements
			Kyokushinkai. Katas and specific techniques (kicks, punches, defenses, etc,).
5	MS-KARD [184]	Karate dataset. MS-KARD consisting of multi-stream data for 23 karate action with 2,814,930 frames and 5,623,734 sensor data samples for karate action recognition. To the best of our knowledge, it is the first of its kind where data has been recorded with 2 orthogonal RGB cameras and 3 wearable inertial sensors.	Goju Ryu karate. 23 karate techniques composed of 6 kicking techniques (Yoko Geri, Tobi Geri, Ushiro mawashi Geri, Yoko Tobi Geri, Mae Geri (Kokomi), Hiza Geri), 6 basic stances (Heiko Dachi, Heisoku Dachi, Musubi Dachi, Kiba Dachi, Zenkutsu Dachi, Kosa Dachi), 7 hand techniques (Jodan Zuki, Heiko Zuki, Oi Zuki, Shuto Uchi, Teisho Uchi, Ura-ken Uchi, Mawashi Empi) and 4 blocking techniques (Gedan Barai uke, Mawashi uke, Soto uke, Age uke).

4.2 Defining the Dataset Inputs

To answer the research question, first we have selected the postures to be recorded in the video images to generate the dataset with the interaction of the karatekas. There are different forms of *kumite* in karate, from multi-step combat (to practice) to free *kumite* [185] (very explosive with very fast movements [186]). Multi-step combat (each step is as an attack) is a simple couple exercise that slowly leads the karateka to an increasingly free action, according to predefined rules. It does not necessarily have to be equated with competition: it is more of a pair exercise in which the participant together with a partner develops a better understanding of the psychomotor technique. Participants do not compete but train together. The different types of *kumite* are listed in **Table 5**, from those with less freedom of movement to free combat.

Table 5: Types of kumite.

KIHON-KUMITE (multi-step combat)	*IPPON KIHON KUMITE*: One-step conventional assault.
	SAMBON KIHON KUMITE: Three-step conventional assault.
	GOHON KIHON KUMITE: Five-step conventional assault.
KUMITE	*JYU IPPON KUMITE*: Free and flexible assault one step away. It can have different work types: (i) announcing height and type of attack, (ii) announcing height, (iii) announcing type of attack, and (iv) unannounced.
	URA IPPON KUMITE (*Kaisho Ippon Kumite*): Unconventional one-step assault. In this type of work one of the karatekas (acting as *uke*) performs the attack and the other (as *tori*) defends it and counterattacks the *uke* who defends the counterattack by the *tori* and ends up counterattacking. There are three working types: (i) announcing the attack and with the pre-established counterattack, (ii) announcing the attack and with the free counterattack, and (iii) unannounced.
	JIYU KUMITE: Free and flexible combat.
	SHIAI KUMITE: Regulated combat for competition.

In order to follow a progressive approach in our research, we started with the *ippon kihon kumite*, which is a pre-established exercise so that it can facilitate the labeling of movements and their analysis. This is the most basic kumite exercise, consisting of the conventional one-step assault. We will use it to compare the results obtained from the application of the data mining algorithms on the feature extracted from the videos recorded and processed using the OpenPose algorithm. The following attack and defense sequences were defined to create the initial dataset, as shown in **Figure 6**.

Figure 6: Pictures of the postures in an *ippon kihon kumite* selected for the initial dataset.

The *ippon kihon* k*umite* sequence proposed for this study would actually be (i) *Gedan Barai,* (ii) *Oi Tsuki,* (iii) *Soto Uke, and* (iv) *Gyaku Tsuki*. However, to calibrate the algorithm and enrich the dataset, we also took two postures *Kamae* (which is the starting posture), both for attack and defense, although they are not properly part of the *ippon Kihon kumite*.

Since monitoring the simultaneous interaction of two karatekas in movement is complex, we have made this first dataset as simple as possible to familiarize ourselves with the algorithm, learn and understand the strengths and weaknesses of OpenPose and explore its potential in the human movement computing scenario addressed in this research (performing martial arts combat techniques between two practitioners aimed to explore if it is possible to model their psychomotor performance in terms of the mistakes made when performing each of the attack and defense postures in the *kumite*). We aim at increasing the difficulty and complexity of the dataset gradually along the research. Thus, for this first dataset, we selected direct techniques (not lateral or angular) with the limbs to facilitate the work of OpenPose classification. Thus, when working in 2D, lateral and circular blows that could acquire angles that could be difficult to calculate in their trajectory and execution, are avoided. The sequence of movements in pairs of attack and defense is shown in **Table 6**.

Table 6: Postures to be detected in the dataset for the *ippon kihon kumite*.

	Attack		Defense
1	Kamae	1	Kamae
2	Gedan Barai	2	Soto Uke
3	Oi Tsuki	3	Gyaku Tsuki

4.3. Preparing the Dataset for posture classification

The created dataset consists in the angles of the triplets obtained by applying the OpenPose algorithm to the *ippon kihon kumite* postures defined in Table 6. The data was extracted from the video recorded with postures of the two karate practitioners. Each posture is recorded from two perspectives, from a left profile and a right profile. Two videos are made from each perspective for each posture to avoid collisions of the body members.

Thus, for each of the postures (classes) several videos were recorded with KUMITRON psychomotor system (see next subsection for more details) to create the initial dataset (as described below) and obtain the video inputs of the postures for the OpenPose algorithm indicated in Table 6. The postures were recorded in different ways. On the one hand, one minute long videos were made where the camera moved from one side to the other while the karateka was completely still maintaining the posture. To generate postural variability, other videos were created where the karateka performed the selected postures for the *ippon kihon kumite*, stopping at the one that was to be recorded. In this way, it is assumed that the recorded videos can offer more difference between keypoints than just making static recordings.

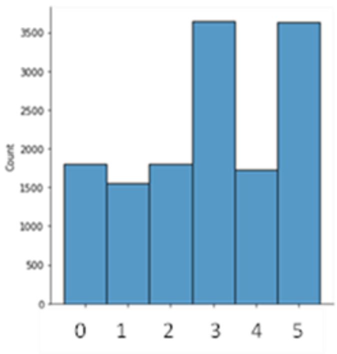
From the recorded videos, the inputs generated come out at a rate of 25 frames per second. Thus, from one second of video, 25 inputs are generated for every second. A feature in OpenPose to clean the inputs with values very different from the average ones was used to deal with the body parts that remain in the back of the camera.

The dataset consists of 26 features and the multiple class of 6 different values. The features of the dataset can be seen in the **Table 7**, where the triplets of corresponding keypoints are indicated.

Table 7: The dataset features associated with the keypoints they represent

Main Part	Features	Keypoints
HEAD	A00	17,15,0
	A01	18,16,0
	A02	15,0,1
	A03	16,0,1
NECK	A04	0,1,2
	A05	0,1,5
SHOULDERS RIGHT ARM	A06	1,2,3
	A07	2,3,4
SHOULDERS LEFT ARM	A08	1,5,6
	A09	5,6,7
SHOULDERS AND TRUNK	A10	2,1,8
	A11	5,1,8
TRUNK AND PELVIS	A12	1,8,9
	A13	1,8,12
LEFT AND RIGHT LEGS	A14	8,9,10
	A15	8,12,13
	A16	9,10,11
	A17	12,13,14
RIGHT FOOT	A18	10,11,24
	A19	10,11,22
	A20	24,11,22
	A21	11,22,23
LEFT FOOT	A22	13,14,21
	A23	13,14,19
	A24	21,14,19
	A25	14,19,20

The same video time was recorded for all postures, although finally, not all classes have the same number of data record, which is due to several factors such as cleaning outliers and some low-quality recording. In the following **Figure 7** it is possible to see the number of data records that each posture or type class of the dataset has.



Nº	Class type
0	Kamae (Attack)
1	Gedan Barai
2	Oi Tsuki
3	Kamae (Defense)
4	Soto Uke
5	Gyaku Tsuki

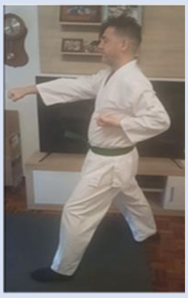
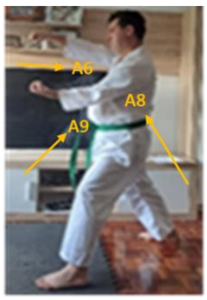
Figure 7: Number of data records that each posture or type class is contained in the dataset

4.4 Preparing the Dataset for psychomotor performance modelling

To explore how to address the modeling of the psychomotor performance and respond to study S2 (section 3.3), new postures have been generated to the first version of the dataset described in [187] with modifications of angles (triplets of keypoints) over those of the original dataset. Thus, three postures shown in **Table 8** were generated enhanced with a greater degree of modification in each.

Changes were done by modifying OpenPose angles. Thus, defensive posture 3 (Gyaku Tsuki) has been selected for the study of the modeling of the psychomotor performance and its attributes have been modified from less to more to check the degree of sensitivity the processing approach shows with respect to similar postures. These angular modifications in posture are shown in **Table 8**.

Table 8: New stances with modified angles compared to Defense03

	Modifications of features of dataset scents		
Base Scent	 Defense03 Gyaku Tsuki		
Modified scent	 M1	 M2	 M3
Modified features **Feature number and key-points triplet**	A6(1,2,3)	A6(1,2,3) A7(2,3,4)	A6(1,2,3) A8(1,5,6) A9(5,6,7)

Progressively modifying the angles is intended to allow measuring the sensitivity of the application with respect to this parameter, which allows to adjust the hyperparameters of the application and of the prediction algorithms (classification) in such a way that we get as close as possible to a 100% ranking in classified positions.

This study allows us to have a measurement of psychomotor performance, since it can observe the comparison of the movement performed by the karateka with respect to the one previously stored in the dataset.

4.5. Implemented Application to Obtain the Dataset

An application was developed to connect the video image with the OpenPose algorithm and with the ML/DL libraries for the processing. A first prototype of KUMITRON was made in Java, which used the WEKA library. Thus, this application was capable of extracting the body keypoints from the image received, computing the angles of the defined triples and training the classification algorithms. In addition, when receiving new inputs, it displays on the panel the posture that it identifies as previously trained. In a second version, the application was coded using Python and applying the Keras and Scikit-learn ML/DL libraries.

In this initial and exploratory collection of data to build the dataset with the predefined postures, a green belt participant was video recorded performing the proposed kumite movements, both attack and defense. The video was taken statically for one minute in which the karateka had to be in one for the predefined postures the *kihon kumite* and the camera was moved to capture different possible angles of the shot in 2D. In addition, videos were also taken dynamically in which the karateka went from one posture to another and waited 30 seconds in the last one.

The computer vision part is incorporated into KUMITRON, an application that combines sensors and computer vision[62]. The **Figure 8** shows how *kumite* activity is monitored by combining artificial vision and receiving information from the signals of sensors in the panels on the left.

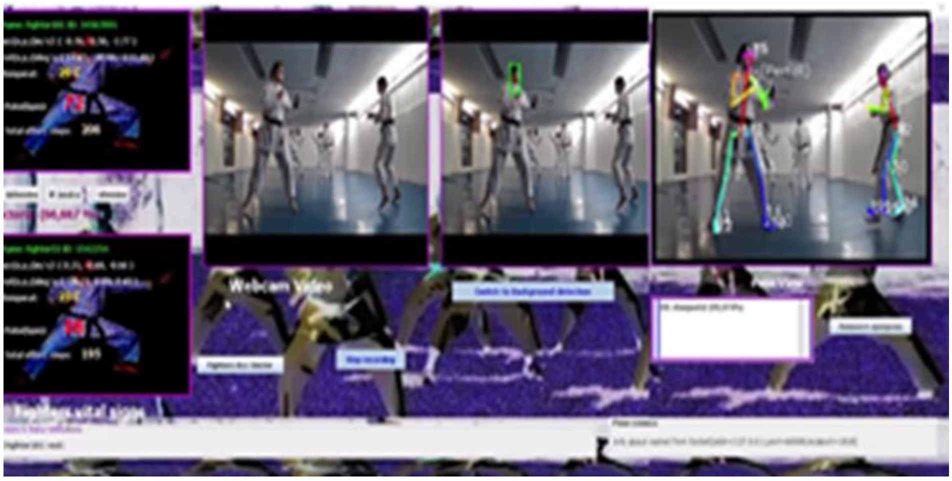


Figure 8: KUMITRON application recording a session of *kumite* in Dojo

5. Analysis and Results

Here we present the results obtained both in the classification of the postures and for the experiments proposed to study the modeling of the psychomotor performance.

5.1 Results training and evaluating selected data mining models

Regarding the ML experiments, the following algorithms are applied as mentioned in Section 3: Random Forest (RF), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Logistic Regression (LR) and Decision Tree (DT).

The best results in training are obtained by RF (Accuracy: 99.8485%) and DT (Accuracy: 99.5455%). The results are deepened by applying statistical tests: 5x2cv paired t test and the McNemar. The comparison made with both the t-Student test and McNemar test do not allow the null hypothesis to be ruled out, meaning that the results obtained with the different classifiers (DT and RF) do not perform significantly differently.

Regarding the neural networks analysis, the results of the standard network of Keras library for MLP has been 98.50% in Accuracy, which is a very good result, although slightly lower than the ML algorithms. On the other hand, the manually created neural network (NN1) has obtained an Accuracy of 98.59%, somewhat higher than the standard MLP. The hyperparameters selected are the Adaptive Moment Estimation and the “Sparse categorical crossentropy” for the loss function, suitable for speeding up the execution of training and saving memory.

The characteristics and results of NN1 are presented in **Table 9**. For the first layer, a "flatten" has been chosen, which receives inputs from 26 angles. The second is a hidden “Dense” type, which is followed by a batch normalization layer. The last output layer is responsible for returning the Type of position thanks to the “Softmax” type activation function, which re-turns an array of 6 possibilities that together add up to 1. Thus, each of the nodes contains an approximation of the probability that the angles with which you are working belong to one of the 6 attack/defense classes.

Table 9: NN1 arquitecture and accuracy results

Model	Layer (type)	Trainable params	Total params	Train acc.	Valid. acc.	Test acc.
NN1	Flatten + Dense + BatchNormalization +Drop + 2Dense	2.881	2.981	98,61	98,59	98,68

5.1.1. Global Results

To compare the results of the best models (i.e., DT and RF for ML and MLP and the NN for DL), the classification report are presented in the following Tables (**Table 10**, **Table 11**, **Table 12**, **Table 13**), which shows the performance and evaluation of the data classification metrics by the different models. The metrics displayed are: i) Precision: the ratio of true positives to the sum of true and false positives, ii) Recall: the ratio of true positives to the sum of true positives and false negatives, iii) F1 Score: the weighted harmonic mean of precision and recall, and iv) Support: Support is the number of actual occurrences of the class in the dataset.

The first column in the tables identifies each of the six postures being classified, that is: 0 is Kamae (Attack01), 1 is Gedan Barai (Attack02), 2 is Oi Tsuki (Attack03), 3 is Kamae (Defense01), 4 is Soto Uke (Defense02), 5 is Gyaku Tsuki (Defense03).

Table 10: RF Classification Report

Classification Report				
	Precision	**Recall**	**F1-score**	**Support**
0: Kamae (Attack01)	1,00	1,00	1,00	529
1: Gedan Barai (Attack02)	0,99	0,99	0,99	439
2: Oi Tsuki (Attack03)	1,00	0,99	1,00	530
3: Kamae (Defense01)	1,00	1,00	1,00	1114
4: Soto Uke (Defense02)	1,00	1,00	1,00	548
5: Gyaku Tsuki (Defense03)	1,00	1,00	1,00	1084
Accuracy			1,00	4244
Macro avg	1,00	1,00	1,00	4244
Weighted avg	1,00	1,00	1,00	4244

Table 11: DT Classification Report

Classification Report				
	Precision	Recall	F1-score	Support
0: Kamae (Attack01)	0,99	0,99	0,99	529
1: Gedan Barai (Attack02)	1,00	0,99	1,00	439
2: Oi Tsuki (Attack03)	1,00	1,00	1,00	530
3: Kamae (Defense01)	1,00	0,99	0,99	1114
4: Soto Uke (Defense02)	0,99	1,00	0,99	548
5: Gyaku Tsuki (Defense03)	1,00	1,00	1,00	1084
Accuracy			1,00	4244
Macro avg	0,99	0,99	0,99	4244
Weighted avg	1,00	1,00	1,00	4244

Table 12: MLP Classification Report

Classification Report				
	Precision	Recall	F1-score	Support
0: Kamae (Attack01)	0,97	0,98	0,97	529
1: Gedan Barai (Attack02)	0,96	0,95	0,96	439
2: Oi Tsuki (Attack03)	0,98	0,98	0,98	530
3: Kamae (Defense01)	0,99	0,99	0,99	1114
4: Soto Uke (Defense02)	0,99	0,99	0,99	548
5: Gyaku Tsuki (Defense03)	0,99	0,99	0,99	1084
Accuracy			0,99	4244
Macro avg	0,98	0,98	0,98	4244
Weighted avg	0,99	0,99	0,99	4244

Table 13: NN1 Classification Report

Classification Report				
	Precision	Recall	F1-score	Support
0: Kamae (Attack01)	0,98	0,96	0,97	529
1: Gedan Barai (Attack02)	0,95	0,96	0,95	439
2: Oi Tsuki (Attack03)	0,99	0,98	0,98	530
3: Kamae (Defense01)	0,99	1,00	0,99	1114
4: Soto Uke (Defense02)	0,99	1,00	1,00	548
5: Gyaku Tsuki (Defense03)	1,00	1,00	1,00	1084
Accuracy			0,99	4244
Macro avg	0,98	0,98	0,98	4244
Weighted avg	0,99	0,99	0,99	4244

As it can be seen in the results, the model with the best performance has been the RF, since it obtains 100% in all grouped results, including the Accuracy metric. It is a hard-to-beat performance. The rest of the models have also performed well as it can also be seen in the tables, all above 0.98 in the grouped results.

Since all the models perform very well in the classification of the postures, we have decided to select the NN1 model as the classifier for the current work, since despite offering a slightly lower performance than ML algorithms, it has more possibilities of making parameter modifications for its improvement. In addition, we think that if the experimentation is positive, the dataset will increase significantly, which will make it necessary to implement a DL algorithm, and we think that it is better to have one already designed from the initial phase and avoid having to take steps back later.

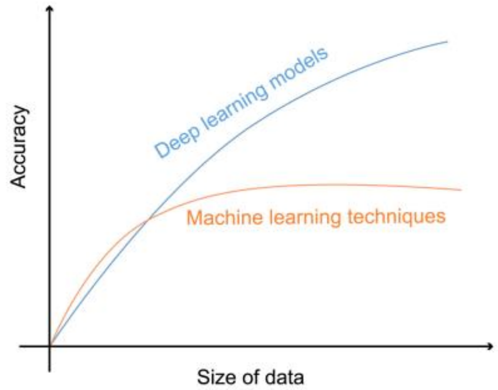


Figure 9: The relation between the accuracy of ML approaches compared to the accuracy of DL models with respect to data size (obtained from [188])

Figure 9 clearly shows the divergent performance between the ML and DL models depending on the size of the dataset, as described in[188]. Therefore, having obtained such a small difference in training and testing, the DL approach has been chosen to save future work.

5.2 Results of the experimentation in in the Psychomotor Modeling Studies

Once the algorithm has been selected, we can carry out the experimental tests proposed in 3.3. The NN1 model is used as the classifier.

5.2.1 Study S1.

For this first study, KUMITRON application receives the images of the 6 postures defined in the dataset. Each posture lasts 10 seconds, and it is verified that the algorithm correctly identifies the posture during the entire 10 seconds, without offering any erroneous scent.

Intermediate posture images have not been passed to avoid introducing inputs other than the classes defined by the dataset. In this way, it is intended to isolate the test to measure the reliability of the application with respect to the dataset and see if the result is close to that obtained only by the models. Results are shown in **Table 14**.

Table 14: S1 Results of identifying video postures of *ippon kihon kumite* postures designed

Nº	Tori(Attack)/ Uke(Defense)	Scent	Identification
0	Tori (Attack01)	Kamae	OK
1	Tori (Attack02)	Gedan Barai	OK
2	Tori (Attack03)	Oi Tsuki	OK
3	Uke (Defense01)	Kamae	OK
4	Uke (Defense02)	Soto Uke	OK
5	Uke (Defense03)	Gyaku Tsuki	OK

The postures have been correctly classified by the application, as expected after the good results obtained by the classifiers. During the video time there has been no identification error, being 100% successful.

5.2.2. Study S2

The S1 study has been a success, but it remains to be seen what happens when the application receives postures performed by the karatekas different from those in the dataset.

For the verification of the images with respect to that of the dataset, a comparison is made between the keypoints and angles where a differential percentage has been defined at each angle from which it is determined that the posture is different. This method to verify the postures, has made the verification efficient, and the results have been positive in the 3 cases. The 3 positions analyzed have been identified as incorrect. The angular variability has been determined at 15% with respect to the original, since many times in Karate, especially beginners, they do not do the posture correctly. The percentage of angular variability is different depending on the level of the practitioner, so that it is useful for calculating psychomotor performance. For example, if you are a novice, the threshold angle can be higher, and as you level up, the application lowers it.

M1, M2, M3 have been recognized as incorrect postures, as one or more angles exceed the defined 15% threshold. The results are shown in **Table 15:** Comparison between the Defense03 posture and others with angle modifications.

Table 15: Comparison between the Defense03 posture and others with angle modifications

Comparative results of the modified angles with respect to the original	
M1	A6(1,2,3): >> 15% angle
	A6(1,2,3): >> 15% angle
M2	A6(1,2,3): >> 15% angle
	A7(2,3,4): >> 15% angle
	A6(1,2,3): >> 15% angle
M3	A8(1,5,6): >> 15% angle
	A9(5,6,7): >> 15% angle

5.2.3. Study S3

It was verified that the application was able to identify the postures in a multi-person way, in this case in a couple. Working in pairs presents the handicap of collisions, since many times the body of an individual can cover the parts of the body of the other person.

In this case, the defined *ippon kihon kumite* left distance between the karateka, and there were hardly any collisions. So the predictions made for the postures analyzed were correct.

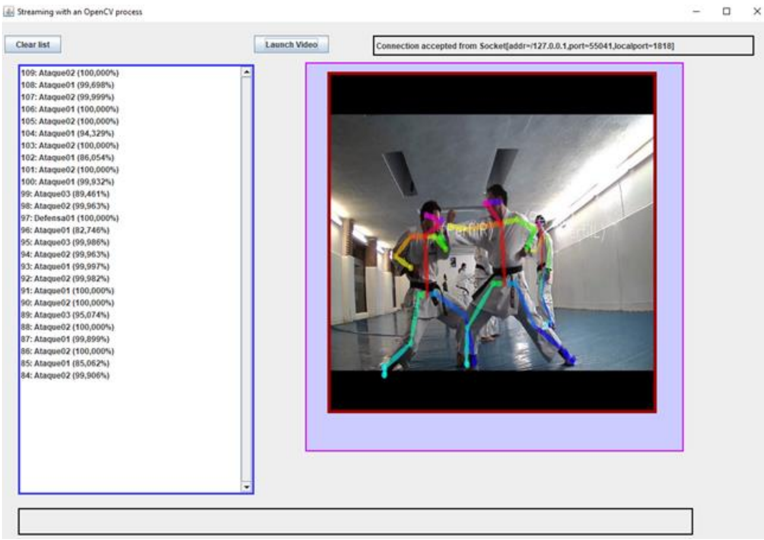


Figure 10: Application recognizing postures with recorded video of two karatekas applying the defined *ippon kihon kumite* postures

In particular, **Figure 10** shows an image of the application annealing video images of a recording of *kumite* training session held at the Zaldupe sports center (Ondarroa[24]) on 11/29/2021. It is observed how postures performed by the karatekas of the Dojo who had been informed of the movements to be performed are identified. On the other hand, we should also comment that we discovered that Openpose is a very demanding algorithm at a computational level, and we observed a slight "lag" in the video processing. This real world experienced confirmed the viability of our approach to be used on the wild.

6. Discussion

Recalling the classification of human activities introduced in Section 1 [1], the problem we are addressing is at level (iii) human-to-human interactions (human activities that involve two or more persons and that to address it, we reduce it to i) when classifying the movement, and ii) when identifying different postures.

That is the line of research that makes this work different from those found in the state of the art. The application of OpenPose algorithm to *kumite* recorded videos aims to facilitate improving the practitioners' technique against different opponents when applying the movements freely and in real time. This is important for the assimilation of the techniques in all kinds of scenarios, including sport practice and personal defense situations. The modeling of movements for psychomotor performance can provide a useful tool to learn Karate that opens new lines of research.

To start with, it can allow studying and improving the combat strategy, having exact measurements of the distance necessary to know when a fighter can be hit by the opponent's blows, as well as the distance necessary to reach the opponent with a blow. This allows training in the gym in a scientific and precise way for combat preparation and distance taking on the mat. In addition, studies can be carried out on how some factors such as fatigue can create variability in the movements developed within a combat in each one of its phases. This is important because this allows for deciding on the choice of certain fighting techniques at the beginning or end of the *kumite* to win. Modeling and capturing tagged movements can also be used to produce datasets and to apply them in other areas such as cinema or video games, which are economically attractive sectors.

It has been verified that the calibration of the models has been correct and that they have achieved a very high performance in the classification test.

[24] https://www.ondarroa.eus/eu-ES/Zerbitzuak/Barnekoa/Kirolak/Orriak/PolideportivoZaldupe.aspx

The studies carried out (S1, S2, S3) have been positive, which encourages us to continue the line of research and endorses the initial modeling hypothesis for karatekas developing *kumite* in pairs.

Even so, to advance in adding more postures and faster movements, we believe that there are improvements that can be applied. For example many of the keypoints offered by the dataset are unnecessary for the modeling of the activity, and that they are left over from the outset and generate noise in the classification. For example, the keypoints of the head hardly provide any information since the work is done in 2D with a profile video shot, and the karateka always look straight ahead. The soles of the feet also, when working in 2D, it is difficult to obtain the angles of such small members, in addition to the fact that the soles of the feet normally in the postures always look forward. In this way we think that the processing demand detected in S3 will be reduced when performing it in a real environment.

To advance in extensions, cases of success in the implementation of this type of techniques are being studied through papers ([188], [189], [190]) to implement the necessary improvements to improve the classification performance. Papers ([191], [192], [193]) related to metrics applied to the evaluation of computer vision algorithms are also being studied, studying how to obtain metrics such as: percentage of Correct Parts (PCP), percentage of Correct Keypoints (PCK), the Average Precision (AP).

Papers have been found successful cases applying this technology to explosive sports such as hurdling [190], badminton [194] and soccer ([195], [196]).

Regarding the dataset, we think that a series of modifications are able to be made that would improve the performance. We think it is important that the postures of the dataset are performed by an art professional, with an advanced belt in the system. In addition to increasing the size of the dataset, different techniques can be applied such as DataAugmentation techniques [197] to augment the inputs and create a large dataset. Also the weighing of the attributes by weights and a feature selection study can give a leap in quality in the results.

The processing performance is also an aspect that is going to be investigated to avoid the "little lag" that the application currently offers. To solve this point, the OpenPose documentation[26] has been consulted, and it has been verified that there are many queries on this aspect, and a section of parameters to be tweaked to improve performance has been enabled. We are currently testing to see if we are able to improve with these changes.

In addition, updates and improvements to the algorithms are coming out every time, in such a way that the performance features they offer are improved. In this case, a light version of OpenPose [198] has been released, which improves processing speed.

According to its author, he has modified the VGG convolutional network for another of the MobileNet family in feature extraction. With this modification and others, the accuracy versus network complexity ratio was increased in more than 6.5 times due to the use of dilated MobileNet v1 feature extractor with depthwise separable convolutions and design of lightweight refinement stage with residual connections. The solutions run at:

- 28 fps on mini PC Intel® NUC, which consumes little power and has 45 watt CPU TDP.
- 26 fps on a usual CPU without the need of a graphic card.

The accuracy of the optimized version nearly matches the baseline: Average Precision (AP) drop is less than 1%. In addition, we will also work on improving the efficiency of the programming made for the application and try to improve the software architecture used to improve performance.

There are more options such as FastPose, it is an open source library that can perform a 2D/3D pose estimation in real time. In this work [199], authors mentioned that FastPose is 46% smaller and 47% faster (forward time) than OpenPose.

[26] https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_06_maximizing_openpose_speed.html

Another research direction we aim to explore to progress in our research is to label not only single postures (techniques) within the *ippon kihon kumite* (as done here) but the complete sequences of the movements. There is some work that can guide the technical implementation of this approach, where classes are first identified and then become subclasses of a superclass [200]. In the work reported in this paper, classes have been defined as the specific techniques to be identified. In the next step, the idea is that these techniques are considered as subclasses, being part of a superior class or superclass. In this way, when the system identifies a concatenation of specific movements, it should be able to classify the attack/defense sequence (default *ippon kihon kumite*) that is being carried out as shown in **Figure 11**.

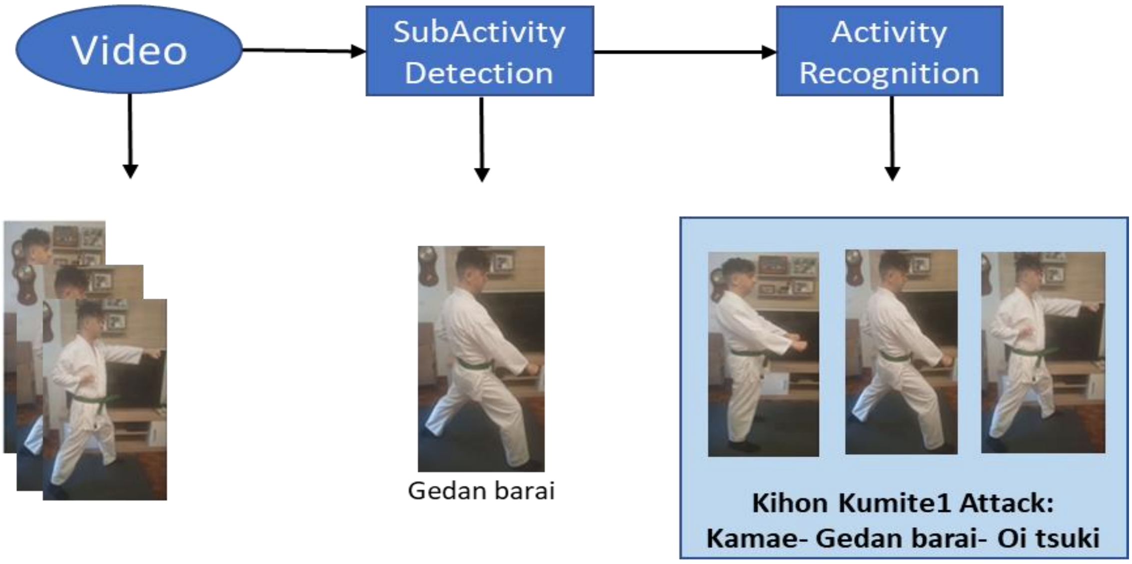


Figure 11: Sub activity recognition in a sequence of *ippon kihon kumite* techniques.

This can be useful to add personalized support to the *kumite* training based on the analysis of the psychomotor performance both comparing the current movements with a good execution and analyzing the temporal evolution of the execution of the movements. The system will be able to recognize if the karateka is developing the movements of a certain section correctly, not only one technique, but the entire series. Thus, training is expected to help practitioners assimilate the concatenations of movements in attack and defense. Furthermore, having super classes identified by the system makes it possible to know in advance which next subclass will be carried out, in such a way that the following movements that will be carried out by a karateka can be monitored through the system, provided that they follow the pattern of some predefined concatenated movements. This is of special interest to be able to work the anticipation training for the defender.

For this, a methodology must be followed for creating super classes that bring together the different concatenated movements. In principle, it seems to be technically possible, as it has been introduced above, but more work is necessary to be able to answer the question with absolute security, since to recognize a superclass, the application must recognize all subclasses without any errors or identification of any wrong postures in the sequence. This requires high identification and classification precision. Moreover, the system needs to have in memory the different movements of each defined superclass, which means that it knows which movement should be the next to perform. In this way, training can be enriched with attack anticipation work in a dynamic and natural way.

For future work, and in addition to exploring other classification algorithms, such as the LSTM algorithm that have reported good results in other works ([61][201]), it would be interesting to add more combinations of movements to create a wider base of movements and to compare the results between different *ippon kihon kumite* movements. Thus, the importance of the different keypoints from OpenPose will be studied (including the analysis of the movements that are more difficult to identify) as well as if it is possible to eliminate some to make the application lighter. For the *ippon kihon kumite* sequence of postures that has been used in the current work, the keypointsp of the legs does not seem to be especially important since the labeled movements are few, and the monitored postures are well identified only with the upper part of the body.

Next, the idea is to extend the current research to faster movements that are performed in other types of *kumite*, following the order of difficulty of these, from *kihon sambon kumite* to competition *kumite*, or *kumite* with free movements. This will allow the calibration of the technical characteristics of the algorithm used (currently OpenPose) to check what type of image speed is capable of working with while still obtaining satisfactory results. This would also be an important step forward in adding elements to anticipation and peripheral vision training. Such attributes have already been started to be explored in our research using OpenCV algorithms [64].

The progress in the classification of the movements during a *kumite* will be integrated into KUMITRON intelligent psychomotor system [62]. KUMITRON collects both video and sensor data from karatekas' practice in a combat, models the movement information, and after designing the different types of feedback that can be required (e.g., with TORMES methodology [202]), delivers the appropriate one for each karateka in each situation (e.g., taking into account the karatekas' affective state during the practice, which can be obtained with the physiological sensors available in KUMITRON following a similar approach as in [203]) through the appropriate sensorial channel (vibrotactile feedback should be explored due to the potential discussed in [204]). In this way, it is expected that computer vision support in KUMITRON can help karatekas learn how to perform the techniques in a *kumite* with the explosiveness required to win the point, making rapid and strong movements that quickly react in real time and in a fluid way to the opponent's technique, adapting also the movement to the opponent's anatomy. The psychomotor performance of the karatekas is to be evaluated both comparing the current movements with a good execution and analyzing the temporal evolution of the execution of the movements.

6.1 Limitations

The limitations found for this work have been mainly technical. On the one hand, the fact of not having a depth camera has conditioned the work to start the approach by treating the image in 2D. To obtain a useful 2D video image for the dataset, video was recorded from both sides of the practitioners, thus collecting body angles from 2 perspectives. On the other hand, OpenPose is a very heavy algorithm that needs a lot of computing power and graphics card, which requires a computer with a powerful graphics card capable of obtaining keypoints and analyzing images efficiently.

The dataset has also been a limitation, since even though there are multiple free datasets on the web, we have not found karate datasets that had images of the objective work in this TFM. Keep in mind that we had the handicap that the dataset we wanted to create was not one of individual movements, but of a *kihon kumite* that involves the movements of two karatekas simultaneously. In this way, we have been forced to create our own dataset. Moreover, the prepared dataset has not been made with expert movement, since it has been obtained from recorded video of 2 low-ranking karateka (green belt and white belt). This means that the created pattern may not be correct and may have some variability with respect to the optimum, but it serves the purposes of testing the processing approach.

7. Conclusions

The main objective of this work was to carry out a first step in our research to assess if computer vision algorithms allow identifying the postures performed by karatekas within the explosive movements that are developed during a *kumite*. The selection of *kumite* was not accidental. It was chosen because it challenges image processing in human movement computing, and there is little scientific literature on modeling the psychomotor performance in activities that involve the joint participation of several individuals, as in a karate combat. In addition, the movements performed by a karateka in front of an opponent may vary with respect to performing them alone through *katas* due to factors such as fear, concentration or adaptation to the opponent's physique (e.g., height).

The results obtained from the training of the classification algorithms with the features extracted from the recorded videos of different *kihon kumite* postures and their application to non labelled images have been 100% satisfactory. It has been observed that the four algorithms used to classify the features extracted with OpenPose algorithm for the detection of movements (i.e., DecisionTree, RandomForest, MLP and DL model) have a precision of above 98% for the current (and limited) dataset.

The studies carried out have also been positive, so the research hypothesis is demonstrated, and it is concluded that human modeling through artificial vision is possible.

We are aware that as we increase the demands of the system by adding new classes to the dataset and increasing the speed of movements, it will be necessary to make technical refinements that allow this increase in computational demand to be processed. Thus, similar solutions are being investigated both in karate and in sports with explosive movements to investigate the type of technical solutions they offer. In particular, the improvement of the OpenPose algorithm regarding processing speed is being studied, replacing it with lighter algorithms that offer a similar performance or even in some papers it is announced that it is better.

Author Contributions: Conceptualization, O.C.S. and J.E.; Methodology, J.E. and O.C.S.; Software, J.E.; Validation, J.E. and O.C.S.; Formal Analysis, J.E.; Investigation, J.E.; Resources, J.E. and O.C.S.; Data Curation, J.E.; Writing—Original Draft Preparation, J.E.; Writing—Review and Editing, O.C.S. and J.E.; Visualization, J.E.; Project Administration, O.C.S. and J.E.; Funding Acquisition, O.C.S and J.E. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: The authors would like to thank the Karate practitioners who have given feedback along the design and development of KUMITRON, especially Victor Sierra and Marcos Martín at IDavinci, who have been following this research from the beginning.

Funding: This research is partially supported by the Spanish Ministry of Science, Innovation and Universities, the Spanish Agency of Research and the European Regional Development Fund (ERDF) through the project INT^{2}AFF, grant number PGC2018-102279-B-I00. KUMITRON received on September 2021 funding from the Department of Economic Sponsorship, Tourism and Rural Areas of the Regional Council of Gipuzkoa half shared with the European Regional Development Fund (ERDF) with the motto "A way to make Europe".

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and the INT^{2}AFF project, which was approved by the Ethics Committee of the Spanish National University of Distance Education – UNED (date of approval: 18th July 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, "A review of human activity recognition methods," *Front. Robot. AI*, vol. 2, no. NOV, pp. 1–28, 2015, doi: 10.3389/frobt.2015.00028.

[2] Y. Yang, I. Saleemi, and M. Shah, "Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1635–1648, 2013, doi: 10.1109/TPAMI.2012.253.

[3] B. Ni, P. Moulin, X. Yang, and S. Yan, "Motion Part Regularization: Improving Action Recognition via Trajectory Group Selection."

[4] A. Patron-Perez, M. Marszalek, I. Reid, and A. Zisserman, "Structured learning of human interactions in TV shows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2441–2453, 2012, doi: 10.1109/TPAMI.2012.24.

[5] J. H. Li, L. Tian, H. Wang, Y. An, K. Wang, and L. Yu, "Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity," *IEEE Access*, vol. 7, pp. 42565–42576, 2019, doi: 10.1109/ACCESS.2019.2905575.

[6] H. P. Martinez, G. N. Yannakakis, and J. Hallam, "Don't classify ratings of affect; Rank Them!," *IEEE Trans. Affect. Comput.*, vol. 5, no. 3, pp. 314–326, Jul. 2014, doi: 10.1109/TAFFC.2014.2352268.

[7] T. Lan, L. Sigal, and G. Mori, "Social roles in hierarchical models for human activity recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1354–1361, 2012, doi: 10.1109/CVPR.2012.6247821.

[8] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, 2018, doi: 10.1016/j.eswa.2018.03.056.

[9] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit.*, vol. 108, p. 107561, Dec. 2020, doi: 10.1016/J.PATCOG.2020.107561.

[10] S. Zhang *et al.*, "Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances," *Sensors* , vol. 22, no. 4. 2022, doi: 10.3390/s22041476.

[11] E. De-La-Hoz-Franco, P. Ariza-Colpas, J. M. Quero, and M. Espinilla, "Sensor-Based Datasets for Human Activity Recognition – A Systematic Review of Literature," *IEEE Access*, vol. 6, pp. 59192–59210, 2018, doi: 10.1109/ACCESS.2018.2873502.

[12] F. Fereidoonian, F. Firouzi, and B. Farahani, "Human Activity Recognition: From Sensors to Applications," in *2020 International Conference on Omni-layer Intelligent Systems (COINS)*, 2020, pp. 1–8, doi: 10.1109/COINS49042.2020.9191417.

[13] L. B. Marinho, A. H. de Souza Junior, and P. P. R. Filho, "A new approach to human activity recognition using machine learning techniques," *Adv. Intell. Syst. Comput.*, vol. 557, pp. 529–538, 2017, doi: 10.1007/978-3-319-53480-0_52.

[14] W. Ugulino, D. Cardador, K. Vega, E. Velloso, R. Milidiú, and H. Fuks, "Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7589, no. January, 2012, doi: 10.1007/978-3-642-34459-6.

[15] A. K. M. Masum, S. Jannat, E. H. Bahadur, M. G. R. Alam, S. I. Khan, and M. R. Alam, "Human Activity Recognition Using Smartphone Sensors: A Dense Neural Network Approach," Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology, Dhaka, Bangladesh, 3–5 May, 2019. doi: 10.1109/ICASERT.2019.8934657.

[16] R. A. Bhuiyan, N. Ahmed, M. Amiruzzaman, and M. R. Islam, "A robust feature extraction model for human activity characterization using 3-axis accelerometer and gyroscope data," *Sensors (Switzerland)*, vol. 20, no. 23, pp. 1–17, 2020, doi: 10.3390/s20236990.

[17] R. Sekiguchi *et al.*, "Phased Human Activity Recognition based on GPS," *UbiComp/ISWC 2021 - Adjun. Proc. 2021 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Proc. 2021 ACM Int. Symp. Wearable Comput.*, pp. 396–400, Sep. 2021, doi: 10.1145/3460418.3479382.

[18] Z. Zhang and S. Poslad, "Improved use of foot force sensors and mobile phone GPS for mobility activity recognition," *IEEE Sens. J.*, vol. 14, no. 12, pp. 4340–4347, Dec. 2014, doi: 10.1109/JSEN.2014.2331463.

[19] S. S. Ulyanov and V. V. Tuchin, "Pulse-wave monitoring by means of focused laser beams scattered by skin surface and membranes," *Static Dyn. Light Scatt. Med. Biol.*, vol. 1884, p. 160, Jul. 1993, doi: 10.1117/12.148363.

[20] S. Pan *et al.*, "Characterizing human activity induced impulse and slip-pulse excitations through structural vibration," *J. Sound Vib.*, vol. 414, pp. 61–80, Feb. 2018, doi: 10.1016/J.JSV.2017.10.034.

[21] M. Zhang and A. A. Sawchuk, "A preliminary study of sensing appliance usage for human activity recognition using mobile magnetometer," *UbiComp'12 - Proc. 2012 ACM Conf. Ubiquitous Comput.*, pp. 745–748, 2012, doi: 10.1145/2370216.2370380.

[22] K. Altun and B. Barshan, "Human activity recognition using inertial/magnetic sensor units," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif.*

Intell. Lect. Notes Bioinformatics), vol. 6219 LNCS, pp. 38–51, 2010, doi: 10.1007/978-3-642-14715-9_5.

[23] M. L. Hoang, M. Carratù, V. Paciello, and A. Pietrosanto, "Body temperature—indoor condition monitor and activity recognition by mems accelerometer based on IoT-alert system for people in quarantine due to COVID-19," *Sensors*, vol. 21, no. 7, 2021, doi: 10.3390/s21072313.

[24] O. C. Santos, "Artificial Intelligence in Psychomotor Learning: Modeling Human Motion from Inertial Sensor Data," *World Sci.*, vol. 28, no. 4, Jul. 2019, doi: 10.1142/S0218213019400062.

[25] F. Serpush, M. B. Menhaj, B. Masoumi, and B. Karasfi, "Wearable Sensor-Based Human Activity Recognition in the Smart Healthcare System," *Comput. Intell. Neurosci.*, vol. 2022, p. 1391906, 2022, doi: 10.1155/2022/1391906.

[26] R. Kolkar and V. Geetha, "Issues and Challenges in Various Sensor-Based Modalities in Human Activity Recognition System," in *Applications of Advanced Computing in Systems*, 2021, pp. 171–179.

[27] N. Nandakumar *et al.*, "Automated eloquent cortex localization in brain tumor patients using multi-task graph neural networks," *Med. Image Anal.*, vol. 74, p. 102203, Dec. 2021, doi: 10.1016/J.MEDIA.2021.102203.

[28] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review," *Comput. Vis. Image Underst.*, vol. 73, no. 3, pp. 428–440, Mar. 1999, doi: 10.1006/CVIU.1998.0744.

[29] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv*, vol. 43, no. 16, p. 43, 2011, doi: 10.1145/1922649.1922653.

[30] A. Prati, C. Shan, and K. I. K. Wang, "Sensors, vision and networks: From video surveillance to activity recognition and health monitoring," *J. Ambient Intell. Smart Environ.*, vol. 11, no. 1, pp. 5–22, 2019, doi: 10.3233/AIS-180510.

[31] A. Roitberg, N. Somani, A. Perzylo, M. Rickert, and A. Knoll, "Multimodal human activity recognition for industrial manufacturing processes in robotic workcells," *ICMI 2015 - Proc. 2015 ACM Int. Conf. Multimodal Interact.*, pp. 259–266, Nov. 2015, doi: 10.1145/2818346.2820738.

[32] L. Mejias, P. Corke, and J. Roberts, "Field and service robotics: Results of the 9th international conference," in *Springer Tracts in Advanced Robotics*, 2015, vol. 105, no. December, doi: 10.1007/978-3-319-07488-7.

[33] V. Osmani, S. Balasubramaniam, and D. Botvich, "Human activity recognition in pervasive health-care: Supporting efficient remote collaboration," *J. Netw. Comput. Appl.*, vol. 31, no. 4, pp. 628–655, Nov. 2008, doi: 10.1016/J.JNCA.2007.11.002.

[34] A. Subasi, M. Radhwan, R. Kurdi, and K. Khateeb, "IoT based mobile healthcare system for human activity recognition," *2018 15th Learn. Technol. Conf. L T 2018*, pp. 29–34, May 2018, doi: 10.1109/LT.2018.8368507.

[35] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Syst. Appl.*, vol. 137, pp. 167–190, Dec. 2019, doi: 10.1016/J.ESWA.2019.04.057.

[36] O. Rashid, A. Al-Hamadi, and B. Michaelis, "A framework for the integration of gesture and posture recognition using HMM and SVM," *Proc. - 2009 IEEE Int. Conf. Intell. Comput. Intell. Syst. ICIS 2009*, vol. 4, pp. 572–577, 2009, doi: 10.1109/ICICISYS.2009.5357615.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.

[38] Y. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989, doi: 10.1162/neco.1989.1.4.541.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 2012, Accessed: Jun. 23, 2022. [Online]. Available: http://code.google.com/p/cuda-convnet/.

[40] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Sep. 2014, doi: 10.48550/arxiv.1409.1556.

[41] C. Szegedy *et al.*, "Going Deeper with Convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June-2015, pp. 1–9, Sep. 2014, doi: 10.48550/arxiv.1409.4842.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 770–778, Dec. 2015, doi: 10.48550/arxiv.1512.03385.

[43] W. C. C. Chu, C. Shih, W. Y. Chou, S. I. Ahamed, and P. A. Hsiung, "Artificial Intelligence of Things in Sports Science: Weight Training as an Example," *Computer (Long. Beach. Calif).*, vol. 52, no. 11, pp. 52–61, 2019, doi: 10.1109/MC.2019.2933772.

[44] C. Zalluhoglu and N. Ikizler-Cinbis, "Collective Sports: A multi-task dataset for collective activity recognition," *Image Vis. Comput.*, vol. 94, p. 103870, Feb. 2020, doi: 10.1016/J.IMAVIS.2020.103870.

[45] T. Kautz, B. H. Groh, and B. M. Eskofier, "Sensor fusion for multi-player activity recognition in game sports," *KDD Work. Large-Scale Sport. Anal.*, pp. 1–4, 2015.

[46] A. Sharma, T. Al-Dala'In, G. Alsadoon, and A. Alwan, "Use of wearable technologies for analysis of activity recognition for sports," *CITISIA 2020 - IEEE Conf. Innov. Technol. Intell. Syst. Ind. Appl. Proc.*, Nov. 2020, doi: 10.1109/CITISIA50690.2020.9371779.

[47] V. Camomilla, E. Bergamini, S. Fantozzi, and G. Vannozzi, "Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review," *Sensors*, vol. 18, no. 3, p. 873, Mar. 2018, doi: 10.3390/s18030873.

[48] K. Xia, H. Wang, M. Xu, Z. Li, S. He, and Y. Tang, "Racquet sports recognition using a hybrid clustering model learned from integrated wearable sensor," *Sensors (Switzerland)*, vol. 20, no. 6, 2020, doi: 10.3390/s20061638.

[49] I. Wickramasinghe, "Naive Bayes approach to predict the winner of an ODI cricket game," *J. Sport. Anal.*, vol. 6, no. 2, pp. 75–84, 2020, doi: 10.3233/jsa-200436.

[50] E. Jaser, W. Christmas, and J. Kittler, "Temporal post-processing of decision tree outputs for sports video categorisation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3138, pp. 495–503, 2004, doi: 10.1007/978-3-540-27868-9_53.

[51] D. A. Sadlier and N. E. O'Connor, "Event detection in field sports video using audio-visual features and a support vector machine," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 10, pp. 1225–1233, Oct. 2005, doi: 10.1109/TCSVT.2005.854237.

[52] F. Nurwanto, I. Ardiyanto, and S. Wibirama, "Light sport exercise detection based on smartwatch and smartphone using k-Nearest Neighbor and Dynamic Time Warping algorithm," *Proc. 2016 8th Int. Conf. Inf. Technol. Electr. Eng. Empower. Technol. Better Futur. ICITEE 2016*, Feb. 2017, doi: 10.1109/ICITEED.2016.7863299.

[53] H. Hoettinger, F. Mally, and A. Sabo, "Activity Recognition in Surfing - A Comparative Study between Hidden Markov Model and Support Vector Machine," *Procedia Eng.*, vol. 147, pp. 912–917, Jan. 2016, doi: 10.1016/J.PROENG.2016.06.279.

[54] R. A. Minhas, A. Javed, A. Irtaza, M. T. Mahmood, and Y. B. Joo, "Shot classification of field sports videos using AlexNet Convolutional Neural Network," *Appl. Sci.*, vol. 9, no. 3, 2019, doi: 10.3390/app9030483.

[55] G. Thomas, R. Gade, T. B. Moeslund, P. Carr, and A. Hilton, "Computer vision for sports: Current applications and research topics," *Comput. Vis. Image Underst.*, vol. 159, pp. 3–18, Jun. 2017, doi: 10.1016/J.CVIU.2017.04.011.

[56] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions," *Appl. Sci. 2022, Vol. 12, Page 4429*, vol. 12, no. 9, p. 4429, Apr. 2022, doi: 10.3390/APP12094429.

[57] L. M. Neagu, E. Rigaud, S. Travadel, M. Dascalu, and R. V. Rughinis, "Intelligent tutoring systems for psychomotor training – a systematic literature review," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Jun. 2020, vol. 12149 LNCS, pp. 335–341, doi: 10.1007/978-3-030-49663-0_40.

[58] A. Casas-Ortiz, O. C. Santos, and J. Echeverria, *Intelligent systems for psychomotor learning*. Edward Edgar Publishing. In progress., 2022.

[59] O. C. Santos, "Training the Body: The Potential of AIED to Support Personalized Motor Skills Learning," *Int. J. Artif. Intell. Educ.*, vol. 26, no. 2, pp. 730–755, Jun. 2016, doi: 10.1007/s40593-016-0103-2.

[60] O. C. Santos, J. G. Boticario, and P. Van Rosmalen, "The full life cycle of adaptation in aLFanet e-learning environment," in *IEEE Learning Technology Newsletter, 4*, 2004, pp. 59–61, Accessed: Nov. 27, 2021. [Online]. Available: https://research.ou.nl/en/publications/the-full-life-cycle-of-adaptation-in-alfanet-elearning-environmen.

[61] A. Casas-Ortiz and O. C. Santos, "KSAS: A Mobile App with Neural Networks to Guide the Learning of Motor Skills," in *XIX Conference of the Spanish Association for Artificial Intelligence (CAEPIA)*, 2021, pp. 997–1000.

[62] J. Echeverria and O. C. Santos, "KUMITRON: Artificial intelligence system to monitor karate fights that synchronize aerial images with physiological and inertial signals," *IUI '21 Companion Proc. Int. Conf. Intell. User Interfaces*, pp. 37–39, Apr. 2021, doi: 10.1145/3397482.3450730.

[63] J. Echeverria and O. C. Santos, "KUMITRON: A Multimodal Psychomotor Intelligent Learning System to Provide Personalized Support when Training Karate Combats," in *MAIEd'21 Workshop. The First International Workshop on Multimodal Artificial Intelligence in Education*, 2021, p. 12, Accessed: Sep.

27, 2021. [Online]. Available: http://ceur-ws.org/Vol-2902/paper7.pdf.

[64] J. Echeverria and O. C. Santos, "Punch Anticipation in a Karate Combat with Computer Vision," *UMAP 2021 - Adjun. Publ. 29th ACM Conf. User Model. Adapt. Pers.*, pp. 61–67, Jun. 2021, doi: 10.1145/3450614.3461688.

[65] J. C. Cox, "Traditional Asian Martial Arts Training: A Review," *Quest*, vol. 45, no. 3, pp. 366–388, Aug. 1993, doi: 10.1080/00336297.1993.10484094.

[66] O. C. Santos and A. Corbi, "Can Aikido Help with the Comprehension of Physics? A First Step towards the Design of Intelligent Psychomotor Systems for STEAM Kinesthetic Learning Scenarios," *IEEE Access*, vol. 7, pp. 176458–176469, 2019, doi: 10.1109/ACCESS.2019.2957947.

[67] O. C. Santos, "Psychomotor learning in martial arts: An opportunity for user modeling, adaptation and personalization," in *UMAP 2017 - Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, Jul. 2017, pp. 89–92, doi: 10.1145/3099023.3099107.

[68] M. Cowie and R. Dyson, "A Short History of Karate," *A Short Hist. Am. Revolut. War*, 2020, doi: 10.5040/9780755603817.0010.

[69] G. Funakoshi, *MY WAY OF LIFE*, 1st pbk. e. San Francisco: Kodansha International Ltd., 1975.

[70] World Karate Federation, "Karate Competition Rules," 2019.

[71] S. R. Wilk, R. E. McNair, and M. S. Feld, "The physics of karate," *Am. J. Phys.*, vol. 51, no. 9, p. 783, Jul. 1998, doi: 10.1119/1.13498.

[72] T. Hachaj and M. R. Ogiela, "Application of Hidden Markov Models and Gesture Description Language classifiers to Oyama karate techniques recognition," *Proc. - 2015 9th Int. Conf. Innov. Mob. Internet Serv. Ubiquitous Comput. IMIS 2015*, pp. 160–165, 2015, doi: 10.1109/IMIS.2015.26.

[73] S. Yıldız, "Relationship Between Functional Movement Screen and Some Athletic Abilities in Karate Athletes," *J. Educ. Train. Stud.*, vol. 6, no. 8, p. 66, 2018, doi: 10.11114/jets.v6i8.3352.

[74] T. Hachaj, M. R. Ogiela, and K. Koptyra, "Application of assistive computer vision methods to Oyama karate techniques recognition," *Symmetry (Basel).*, vol. 7, no. 4, pp. 1670–1698, 2015, doi: 10.3390/sym7041670.

[75] O. C. Santos, *Beyond Cognitive and Affective Issues: Designing Smart Learning Environments for Psychomotor Personalized Learning*. Learning, Design, and Technology. An International Compendium of Theory, Research, Practice, and Policy, 2016.

[76] N. Barla, "Human Pose Estimation: Deep Learning Approach [2022 Guide]," 2022. https://www.v7labs.com/blog/human-pose-estimation-guide (accessed Aug. 09, 2022).

[77] F. Zhang, X. Zhu, and M. Ye, "Fast human pose estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 3512–3521, 2019, doi: 10.1109/CVPR.2019.00363.

[78] M. Piccardi and T. Jan, "Recent advances in computer vision," *Ind. Phys.*, vol. 9, no. 1, pp. 18–21, 2003.

[79] L. ZATOLOKINA, "Human Pose Estimation Technology 2022 Guide," 2022. https://mobidev.biz/blog/human-pose-estimation-technology-guide (accessed Aug. 09, 2022).

[80] K. Host and M. Ivašić-Kos, "An overview of Human Action Recognition in sports based on Computer Vision," *Heliyon*, vol. 8, no. 6, p. e09633, Jun. 2022, doi: 10.1016/J.HELIYON.2022.E09633.

[81] X. Chen, A. Pang, W. Yang, Y. Ma, L. Xu, and J. Yu, "SportsCap: Monocular 3D Human Motion Capture and Fine-Grained Understanding in Challenging Sports Videos," *Int. J. Comput. Vis.*, vol. 129, no. 10, pp. 2846–2864, 2021, doi: 10.1007/s11263-021-01486-4.

[82] A. Shingade and A. Ghotkar, "Animation of 3D Human Model Using Markerless Motion Capture Applied To Sports," *Int. J. Comput. Graph. Animat.*, vol. 4, no. 1, pp. 27–39, 2014, doi: 10.5121/ijcga.2014.4103.

[83] L. Bridgeman, M. Volino, J. Y. Guillemaut, and A. Hilton, "Multi-person 3D pose estimation and tracking in sports," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 2487–2496, 2019, doi: 10.1109/CVPRW.2019.00304.

[84] S. Xiaojie, L. Qilei, Y. Tao, G. Weidong, and L. Newman, "Mocap data editing via movement notations," *Proc. - Ninth Int. Conf. Comput. Aided Des. Comput. Graph. CAD/CG 2005*, vol. 2005, pp. 463–468, 2005, doi: 10.1109/CAD-CG.2005.58.

[85] N. T. Thành, L. V. Hùng, and P. T. Công, "An Evaluation of Pose Estimation in Video of Traditional Martial Arts Presentation," *J. Res. Dev. Inf. Commun. Technol.*, vol. 2019, no. 2, pp. 114–126, 2019, doi: 10.32913/mic-ict-research.v2019.n2.864.

[86] N. A. Mohd Jelani, A. N. Zulkifli, S. Ismail, and M. F. Yusoff, "A Review of virtual reality and motion capture in martial arts training," *Int. J. Interact. Digit. Media*, vol. 5, no. 1, pp. 22–25, 2020, [Online]. Available: http://repo.uum.edu.my/26996/.

[87] W. Zhang, Z. Liu, L. Zhou, H. Leung, and A. B. Chan, "Martial Arts, Dancing and Sports dataset: A challenging stereo and multi-view dataset for 3D

human pose estimation," *Image Vis. Comput.*, vol. 61, pp. 22–39, 2017, doi: 10.1016/j.imavis.2017.02.002.

[88] M. Z. Kaharuddin *et al.*, "Biomechanics Analysis of Combat Sport (Silat) by Using Motion Capture System," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 166, no. 1, 2017, doi: 10.1088/1757-899X/165/1/012028.

[89] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1302–1310, Nov. 2016, doi: 10.48550/arxiv.1611.08050.

[90] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9910 LNCS, pp. 34–50, May 2016, doi: 10.48550/arxiv.1605.03170.

[91] A. Newell, Z. Huang, and J. Deng, "Associative Embedding: End-to-End Learning for Joint Detection and Grouping," in *Advances in Neural Information Processing Systems*, 2017, vol. 30, [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/8edd72158ccd2a879f79cb2538568fdc-Paper.pdf.

[92] L. Pishchulin *et al.*, "DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 4929–4937, Nov. 2015, doi: 10.48550/arxiv.1511.06645.

[93] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, and J. Sun, "Cascaded Pyramid Network for Multi-Person Pose Estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7103–7112, Nov. 2017, doi: 10.48550/arxiv.1711.07319.

[94] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988, doi: 10.1109/ICCV.2017.322.

[95] G. Papandreou *et al.*, "Towards Accurate Multi-person Pose Estimation in the Wild," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 3711–3719, Jan. 2017, doi: 10.48550/arxiv.1701.01779.

[96] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep High-Resolution Representation Learning for Human Pose Estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 5686–5696, Feb. 2019, doi: 10.48550/arxiv.1902.09212.

[97] K. Petri *et al.*, "Training using virtual reality improves response behavior in karate kumite," *Sport. Eng.*, vol. 22, pp. 1–12, 2019, doi: 10.1007/s12283-019-0299-0.

[98] S. Pastel, K. P. C. H. Chen, A. Milena, W. Cáceres, and M. S. C. Nübel, "Training in virtual reality enables learning of a complex sports movement," *Virtual Real.*, no. 0123456789, 2022, doi: 10.1007/s10055-022-00679-7.

[99] K. Petri, S. Masik, M. Danneberg, P. Emmermacher, and K. Witte, "Possibilities to Use a Virtual Opponent for Enhancements of Reactions and Perception of Young Karate Athletes," *Int. J. Comput. Sci. Sport*, vol. 18, no. 2, pp. 20–33, 2019, doi: doi:10.2478/ijcss-2019-0011.

[100] R. Lammfromm and D. Gopher, "Transfer of Skill from a Virtual Reality Trainer to Real Juggling," *BIO Web Conf.*, vol. 1, p. 00054, 2011, doi: 10.1051/BIOCONF/20110100054.

[101] J. Tirp, C. Steingröver, N. Wattie, J. Baker, and J. Schorer, "Virtual realities as optimal learning environments in sport—A transfer study of virtual and real dart throwing.," *Psychol. Test Assess. Model.*, vol. 57, no. 1, pp. 57–69, 2015.

[102] R. Gray, "Transfer of Training from Virtual to Real Baseball Batting," *Front. Psychol.*, vol. 8, 2017, doi: 10.3389/fpsyg.2017.02183.

[103] T. M. Takala, Y. Hirao, H. Morikawa, and T. Kawai, "Martial Arts Training in Virtual Reality with Full-body Tracking and Physically Simulated Opponents," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, May 2020, pp. 858–858, doi: 10.1109/vrw50115.2020.00282.

[104] P. Hämäläinen, T. Ilmonen, J. Höysniemi, M. Lindholm, and A. Nykänen, "Martial arts in artificial reality," *CHI 2005 Technol. Safety, Community Conf. Proc. - Conf. Hum. Factors Comput. Syst.*, no. 1, pp. 781–790, 2005, doi: 10.1145/1054972.1055081.

[105] A. Corbi, O. C. Santos, and D. Burgos, "Intelligent Framework for Learning Physics with Aikido (Martial Art) and Registered Sensors," *Sensors*, vol. 19, no. 17, p. 3681, Aug. 2019, doi: 10.3390/s19173681.

[106] J. Ji and F. Liang, "Influence of Embedded Microprocessor Wireless Communication and Computer Vision in Wushu Competition Referees' Decision Support," *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/2121573.

[107] J. Lu and M. Xie, "A Wushu Referee's Decision Support System Using Error Recognition Theory," *Mob. Inf. Syst.*, vol. 2022, pp. 1–8, Jun. 2022, doi: 10.1155/2022/7717254.

[108] S. Hariri and H. Sadeghi, "Biomechanical Analysis of Mawashi-Geri in Technique in Karate: Review Article," *Int. J. Sport Stud. Heal.*, vol. In Press, no. In Press, pp. 1–8, 2018, doi: 10.5812/intjssh.84349.

[109] K. Witte, P. Emmermacher, N. Langenbeck, and J. Perl, "Visualized movement patterns and their analysis to classify similarities-demonstrated by the karate kick Mae-geri," *Kinesiology*, vol. 44, no. 2, pp. 155–165, 2012.

[110] T. Hachaj, M. Piekarczyk, and M. Ogiela, "Human Actions Analysis: Templates Generation, Matching and Visualization Applied to Motion Capture of Highly-Skilled Karate Athletes," *Sensors*, vol. 17, no. 11, p. 2590, Nov. 2017, doi: 10.3390/s17112590.

[111] A. Labintsev, I. Khasanshin, D. Balashov, M. Bocharov, and K. Bublikov, "Recognition punches in karate using acceleration sensors and convolution neural networks," *IEEE Access*, vol. 9, pp. 1–1, 2021, doi: 10.1109/access.2021.3118038.

[112] T. Hachaj, M. R. Ogiela, M. Piekarczyk, and K. Koptyra, "Advanced human motion analysis and visualization: Comparison of mawashi-geri kick of two elite karate athletes," *2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc.*, vol. 2018-Janua, pp. 1–7, 2018, doi: 10.1109/SSCI.2017.8285269.

[113] K. Kolykhalova, A. Camurri, G. Volpe, M. Sanguineti, E. Puppo, and R. Niewiadomski, "A multimodal dataset for the analysis of movement qualities in karate martial art," *Proc. 2015 7th Int. Conf. Intell. Technol. Interact. Entertain. INTETAIN 2015*, vol. 17, pp. 74–78, 2015, doi: 10.4108/icst.intetain.2015.260039.

[114] M. F. Goethel *et al.*, "Coordinative intra-segment indicators of karate performance," *Arch. Budo*, vol. 15, no. February, pp. 203–211, 2019.

[115] K. Petri, N. Bandow, S. Masik, and K. Witte, "Improvement of Early Recognition of Attacks in Karate Kumite Due to Training in Virtual Reality," *J. Sport Area*, vol. 4, no. 2, pp. 294–308, Nov. 2019, doi: 10.25299/sportarea.2019.vol4(2).3370.

[116] O. C. Santos and M. H. Eddy, "Modeling psychomotor activity: Current approaches and open issues," in *UMAP 2017 - Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, Jul. 2017, pp. 305–310, doi: 10.1145/3099023.3099083.

[117] N. Koropanovski *et al.*, "Anthropometric and physical performance profiles of elite karate kumite and kata competitors," *J. Hum. Kinet.*, vol. 30, no. 1, pp. 107–114, Dec. 2011, doi: 10.2478/V10078-011-0078-X.

[118] C. E. Ogarrio Perkins, "Karate Do formativo," *Karate Do Form.*, 2021, doi: 10.47807/UNISON.12.

[119] X. Xu, Y. Gao, K. Yan, X. Lin, and Q. Zou, "Location-Free Human Pose Estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 13137–13146.

[120] M. Ben Gamra and M. A. Akhloufi, "A review of deep learning techniques for 2D and 3D human pose estimation," *Image Vis. Comput.*, vol. 114, Oct. 2021, doi: 10.1016/J.IMAVIS.2021.104282.

[121] J. Hruthika, P. K. Chaitanya, and G. shiva Chaithanya, "Deep Learning Based Human Pose Estimation Using Opencv," *Int. J. Innov. Eng. Res. Technol.*, vol. 7, no. 12, pp. 246–253, 2020.

[122] M. Li and J. Zhao, "Human Sports Action and Ideological and PoliticalEvaluation by Lightweight Deep Learning Model," *Comput. Intell. Neurosci.*, vol. 2022, p. 5794914, 2022, doi: 10.1155/2022/5794914.

[123] R. Changotra *et al.*, "A Comparative Study of Pose Estimation," in *2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)*, 2022, vol. 2, pp. 210–216, doi: 10.1109/ICIPTM54933.2022.9754071.

[124] L. K. Topham, W. Khan, ljmuacuk Dhiya Al-jumeily, and ljmuacuk Abir Hussain, "Human Body Pose Estimation for Gait Identification: A Comprehensive Survey of Datasets and Models," *ACM Comput. Surv.*, Nov. 2021, doi: 10.1145/3533384.

[125] V. Gupta, "Pose Detection comparison : wrnchAI vs OpenPose | LearnOpenCV," *LearnOpenCV*, 2019, Accessed: Dec. 08, 2021. [Online]. Available: https://learnopencv.com/pose-detection-comparison-wrnchai-vs-openpose/.

[126] J. E. Eivindsen, "Human Pose Estimation Assisted Fitness Technique Evaluation System," Norweigian Universityof Science and technology, 2020.

[127] N. Carissimi, P. Rota, C. Beyan, and V. Murino, "Filling the gaps: Predicting missing joints of human poses using denoising autoencoders," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11130 LNCS, no. January, pp. 364–379, 2019, doi: 10.1007/978-3-030-11012-3_29.

[128] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," 2014, Accessed: Dec. 08, 2021. [Online]. Available: http://human-pose.mpi-inf.mpg.de/.

[129] T. Y. Lin *et al.*, "Microsoft COCO: Common Objects in Context," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes*

Bioinformatics), vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014, doi: 10.1007/978-3-319-10602-1_48.

[130] H.-S. Fang, S. Xie, Y.-W. Tai, C. Lu, S. Jiao Tong University, and T. YouTu, "RMPE: Regional Multi-Person Pose Estimation," Accessed: Nov. 27, 2021. [Online]. Available: https://cvsjtu.wordpress.com/rmpe-regional-multi-person-pose-.

[131] L. Pishchulin *et al.*, "DeepCut: Joint subset partition and labeling for multi person pose estimation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 4929–4937, 2016, doi: 10.1109/CVPR.2016.533.

[132] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2155–2162, 2014, doi: 10.1109/CVPR.2014.276.

[133] A. Toshev and C. Szegedy, "DeepPose_Human_Pose_2014_CVPR_paper.pdf," *2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1653–1660, 2014, [Online]. Available: https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Toshev_DeepPose_Human_Pose_2014_CVPR_paper.html.

[134] R. A. Güler, N. Neverova, and I. Kokkinos, "DensePose: Dense Human Pose Estimation in the Wild," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7297–7306, 2018, doi: 10.1109/CVPR.2018.00762.

[135] J. Wang *et al.*, "Deep High-Resolution Representation Learning for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, 2021, doi: 10.1109/TPAMI.2020.2983686.

[136] W. Xiao, Gong, Lu, "Joint COCO and Mapillary Workshop at ICCV 2019: COCO Keypoint Detection Challenge Track," 2019. [Online]. Available: http://cocodataset.org/files/keypoints_2019_reports/ByteDanceHRNet.pdf.

[137] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," *IEEE Journals Mag. | IEEE Xplore*, Art. no. Vol. 43, 2021, Accessed: Jun. 06, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8765346.

[138] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-dof camera relocalization," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2015 Inter, pp. 2938–2946, 2015, doi: 10.1109/ICCV.2015.336.

[139] D. Maji, S. Nagori, M. Mathew, and D. Poddar, "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022, pp. 2637–2646.

[140] H. J. Park, J. W. Baek, and J. H. Kim, "Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose," *IEEE Int. Conf. Autom. Sci. Eng.*, vol. 2020-Augus, no. August 2020, pp. 1389–1394, 2020, doi: 10.1109/CASE48305.2020.9216833.

[141] F. Rosique, F. Losilla, and P. J. Navarro, "Applying Vision-Based Pose Estimation in a Telerehabilitation Application," *Appl. Sci.*, 2021, doi: https://doi.org/10.3390/app11199132.

[142] W. Chen, Z. Jiang, H. Guo, and X. Ni, "Fall Detection Based on Key Points of Human-Skeleton Using OpenPose," *Symmetry (Basel).*, vol. 12, no. 5, p. 744, 2020, [Online]. Available: https://www.mdpi.com/2073-8994/12/5/744.

[143] A. P. Yunus, N. C. Shirai, K. Morita, and T. Wakabayashi, "Human Motion Prediction by 2D Human Pose Estimation using OpenPose," *EasyChair Prepr.*, 2020.

[144] C. B. Lin, Z. Dong, W. K. Kuan, and Y. F. Huang, "A framework for fall detection based on openpose skeleton and lstm/gru models," *Appl. Sci.*, vol. 11, no. 1, pp. 1–20, 2021, doi: 10.3390/app11010329.

[145] Q. Zhou, J. Wang, P. Wu, and Y. Qi, "Application Development of Dance Pose Recognition Based on Embedded Artificial Intelligence Equipment," *J. Phys. Conf. Ser.*, vol. 1757, no. 1, 2021, doi: 10.1088/1742-6596/1757/1/012011.

[146] J. Xing, J. Zhang, and C. Xue, "Multi person pose estimation based on improved openpose model," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 768, no. 7, 2020, doi: 10.1088/1757-899X/768/7/072071.

[147] R. Bajireanu *et al.*, "Mobile human shape superimposition: an initial approach using OpenPose," *Int. J. Comput.*, vol. 3, pp. 1–8, 2018, [Online]. Available: http://www.iaras.org/iaras/journals/ijc.

[148] D. Berrar, "Cross-Validation Call for Papers for Machine Learning journal: Machine Learning for Soccer View project Cross-validation," doi: 10.1016/B978-0-12-809633-8.20349-X.

[149] J. D. Rodríguez, A. Pérez, and J. A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 569–575, 2010, doi: 10.1109/TPAMI.2009.187.

[150] T. T. Wong and P. Y. Yeh, "Reliable Accuracy Estimates from k-Fold Cross Validation," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 8, pp. 1586–1594, 2020,

doi: 10.1109/TKDE.2019.2912815.

[151] N. H. Nguyen, D. T. A. Nguyen, B. Ma, and J. Hu, “The application of machine learning and deep learning in sport: predicting NBA players’ performance and popularity,” *J. Inf. Telecommun.*, vol. 6, no. 2, pp. 217–235, Apr. 2022, doi: 10.1080/24751839.2021.1977066.

[152] J. Almulla and T. Alam, “Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League,” *IEEE Access*, vol. 8, pp. 213695–213705, 2020, doi: 10.1109/ACCESS.2020.3038601.

[153] N. Seliya, T. M. Khoshgoftaar, and J. Van Hulse, “A Study on the Relationships of Classifier Performance Metrics,” in *2009 21st IEEE International Conference on Tools with Artificial Intelligence*, 2009, pp. 59–66, doi: 10.1109/ICTAI.2009.25.

[154] P. Wang, R. Cai, and S. Yang, “News video classification using multimodal classifiers and text-biased combination strategies,” *Qinghua Daxue Xuebao/Journal Tsinghua Univ.*, vol. 45, no. 4, pp. 475–478, 2005.

[155] J. Yang, R. Yan, and A. G. Hauptmann, “Adapting SVM classifiers to data with shifted distributions,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 69–74, 2007, doi: 10.1109/ICDMW.2007.37.

[156] P. Yin, A. Criminisi, J. Winn, and I. Essa, “Tree-based classifiers for bilayer video segmentation,” 2007, doi: 10.1109/CVPR.2007.383008.

[157] J. Sivic, M. Everingham, and A. Zisserman, “Who are you? - Learning person specific classifiers from video,” *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, vol. 2009 IEEE, pp. 1145–1152, 2009, doi: 10.1109/CVPRW.2009.5206513.

[158] M. R. Mia, M. J. Mia, A. Majumder, S. Supriya, and M. T. Habib, “Computer vision based local fruit recognition,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 1, pp. 2810–2820, 2019, doi: 10.35940/ijeat.A9789.109119.

[159] M. A. Ponti, L. S. F. Ribeiro, T. S. Nazare, T. Bui, and J. Collomosse, “Everything You Wanted to Know about Deep Learning for Computer Vision but Were Afraid to Ask,” *Proc. - 2017 30th SIBGRAPI Conf. Graph. Patterns Images Tutorials SIBGRAPI-T 2017*, vol. 2018-Janua, pp. 17–41, 2017, doi: 10.1109/SIBGRAPI-T.2017.12.

[160] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: A Brief Review,” *Comput. Intell. Neurosci.*, vol. 2018, 2018, doi: 10.1155/2018/7068349.

[161] S. M. S. Islam, S. Rahman, M. M. Rahman, E. K. Dey, and M. Shoyaib, “Application of deep learning to computer vision: A comprehensive study,” *2016 5th Int. Conf. Informatics, Electron. Vision, ICIEV 2016*, pp. 592–597, 2016, doi: 10.1109/ICIEV.2016.7760071.

[162] Q. Wu, Y. Liu, Q. Li, S. Jin, and F. Li, “The application of deep learning in computer vision,” *Proc. - 2017 Chinese Autom. Congr. CAC 2017*, vol. 2017-Janua, pp. 6522–6527, 2017, doi: 10.1109/CAC.2017.8243952.

[163] D. Van Dao *et al.*, “A spatially explicit deep learning neural network model for the prediction of landslide susceptibility,” *Catena*, vol. 188, no. December 2019, p. 104451, 2020, doi: 10.1016/j.catena.2019.104451.

[164] J. Grekow, “Music emotion recognition using recurrent neural networks and pretrained models,” *J. Intell. Inf. Syst.*, 2021, doi: 10.1007/s10844-021-00658-5.

[165] T. Li, S. Fong, S. W. I. Siu, X. she Yang, L. S. Liu, and S. Mohammed, “White learning methodology: A case study of cancer-related disease factors analysis in real-time PACS environment,” *Comput. Methods Programs Biomed.*, vol. 197, p. 105724, 2020, doi: 10.1016/j.cmpb.2020.105724.

[166] M. K. Vanam, B. Amirali Jiwani, A. Swathi, and V. Madhavi, “High performance machine learning and data science based implementation using Weka,” *Mater. Today Proc.*, no. xxxx, 2021, doi: 10.1016/j.matpr.2021.01.470.

[167] I. Ramírez, A. Cuesta-Infante, E. Schiavi, and J. J. Pantrigo, “Bayesian capsule networks for 3D human pose estimation from single 2D images,” *Neurocomputing*, vol. 379, pp. 64–73, Feb. 2020, doi: 10.1016/J.NEUCOM.2019.09.101.

[168] Y.-K. Wang and K.-Y. Cheng, “A Two-Stage Bayesian Network Method for 3D Human Pose Estimation from Monocular Image Sequences,” *EURASIP J. Adv. Signal Process.*, vol. 2010, p. 16, 2010, doi: 10.1155/2010/761460.

[169] A. M. Lehrmann, P. V Gehler, and S. Nowozin, “A Non-parametric Bayesian Network Prior of Human Pose,” 2013, doi: 10.1109/ICCV.2013.162.

[170] Y. Wang, J. Li, Y. Zhang, and R. O. Sinnott, “Identifying lameness in horses through deep learning,” *Proc. ACM Symp. Appl. Comput.*, pp. 976–985, Mar. 2021, doi: 10.1145/3412841.3441973.

[171] J. H. K. Y.-S. Park, “A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems,” *J. Korea Soc. Comput. Inf.*, vol. 23, no. 8, pp. 9–16, 2018, doi: 10.9708/JKSCI.2018.23.08.009.

[172] T. van Elteren and T. van der Zant, "Real-Time Human Pose and Gesture Recognition for Autonomous Robots Using a Single Structured Light 3D-Scanner," *Work. Proc. 8th …*, pp. 213–220, 2012, doi: 10.3233/978-1-61499-080-2-213.

[173] P. Szczuko, "Deep neural networks for human pose estimation from a very low resolution depth image," *Multimed. Tools Appl. 2019 7820*, vol. 78, no. 20, pp. 29357–29377, Mar. 2019, doi: 10.1007/S11042-019-7433-7.

[174] S. Park, J. Hwang, and N. Kwak, "3D human pose estimation using convolutional neural networks with 2D pose information," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9915 LNCS, pp. 156–169, 2016, doi: 10.1007/978-3-319-49409-8_15.

[175] M. De Marsico, A. Petrosino, and S. Ricciardi, "Iris recognition through machine learning techniques: A survey," *Pattern Recognit. Lett.*, vol. 82, pp. 106–115, Oct. 2016, doi: 10.1016/J.PATREC.2016.02.001.

[176] B. T. Chicho, A. M. Abdulazeez, D. Q. Zeebaree, and D. A. Zebari, "Machine Learning Classifiers Based Classification For IRIS Recognition," *Qubahan Acad. J.*, vol. 1, no. 2, pp. 106–118, May 2021, doi: 10.48161/QAJ.V1N2A48.

[177] K. Reitz and T. Schlusser, *The Hitchhiker's Guide To Python*. O'Reilly Media, Inc., 2015.

[178] N. A. Rahmad, M. Amir As'ari, N. F. Ghazali, N. Shahar, N. Anis, and J. Sufri, "A Survey of Video Based Action Recognition in Sports," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 11, no. 3, pp. 987–993, 2018, doi: 10.11591/ijeecs.v11.i3.pp987-993.

[179] A. Aju, C. Mathew, and O. S. G. Prakasi, "PoseNet based Model for Estimation of Karate Poses," *J. Innov. Image Process.*, vol. 4, no. 1, pp. 16–25, May 2022, doi: 10.36548/JIIP.2022.1.002.

[180] B. Gnana Priya and M. Arulselvi, "Action Classification for Karate Dataset Using Deep Learning," *INTELINC*, vol. 18, 2018, Accessed: Sep. 12, 2022. [Online]. Available: www.ijrat.org.

[181] P. Barbosa, P. Cunha, V. H. Carvalho, and F. O. Soares, "Classification of Taekwondo Techniques using Deep Learning Methods: First Insights," 2021.

[182] R. Niewiadomski, M. Mancini, S. Piana, P. Alborno, G. Volpe, and A. Camurri, "Low-Intrusive recognition of expressive movement qualities," *ICMI 2017 - Proc. 19th ACM Int. Conf. Multimodal Interact.*, vol. 2017-January, pp. 230–237, Nov. 2017, doi: 10.1145/3136755.3136757.

[183] T. Hachaj and M. R. Ogiela, "Classification of Karate Kicks with Hidden Markov Models Classifier and Angle-Based Features," *Proc. - 2018 11th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2018*, Feb. 2019, doi: 10.1109/CISP-BMEI.2018.8633251.

[184] S. Yadav *et al.*, "MS-KARD: A Benchmark for Multimodal Karate Action Recognition," Oct. 2022.

[185] Hirokazu Kanazawa, *Karate Fighting Techniques: The Complete Kumite*. 2013.

[186] H. Chaahène, Y. Hachana, E. Franchini, B. Mkaouer, and T. Chamarf, "Physical and Physiological Profile of Elite Karate Athletes," *Sport. Med*, 2012.

[187] J. Echeverria and O. C. Santos, "Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation," *Sensors 2021, Vol. 21, Page 8378*, vol. 21, no. 24, p. 8378, Dec. 2021, doi: 10.3390/S21248378.

[188] F. Es-Sabery, A. Hair, J. Qadir, B. Sainz-De-Abajo, B. García-Zapirain, and I. D. L. Torre-Díez, "Sentence-Level Classification Using Parallel Fuzzy Deep Learning Classifier," *IEEE Access*, vol. 9, pp. 17943–17985, 2021, doi: 10.1109/ACCESS.2021.3053917.

[189] V. Anand Thoutam *et al.*, "Yoga Pose Estimation and Feedback Generation Using Deep Learning," *Comput. Intell. Neurosci.*, vol. 2022, p. 4311350, 2022, doi: 10.1155/2022/4311350.

[190] P. Jafarzadeh, P. Virjonen, P. Nevalainen, F. Farahnakian, and J. Heikkonen, "Pose Estimation of Hurdles Athletes using OpenPose," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2021, pp. 1–6, doi: 10.1109/ICECCME52200.2021.9591066.

[191] T. Duy, N. Milan, and K. Kresovi´c, "A survey of top-down approaches for human pose estimation," Feb. 2022, doi: 10.48550/arxiv.2202.02656.

[192] H. Ren, W. Wang, K. Zhang, D. Wei, Y. Gao, and Y. Sun, "Fast and Lightweight Human Pose Estimation," *IEEE Access*, vol. 9, pp. 49576–49589, 2021, doi: 10.1109/ACCESS.2021.3069102.

[193] W. Kim, J. Sung, D. Saakes, C. Huang, and S. Xiong, "Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose)," *Int. J. Ind. Ergon.*, vol. 84, p. 103164, 2021, doi: https://doi.org/10.1016/j.ergon.2021.103164.

[194] Hui Xiang, "Lightweight Open Pose Based Body Posture Estimation for Badminton Players," *For. Chem. Rev.*, no. SE-Articles, pp. 339–350, Apr. 2022, [Online]. Available: http://forestchemicalsreview.com/index.php/JFCR/article/view/923.

[195] R. G. Reis, D. P. Trachtinguerts, A. K. Sato, R. Y. Takimoto, F. S. G. Tsuzuki, and M. de S. G. Tsuzuki, "Soccer Player Pose Recognition in Games," *Lect. Notes Data Eng. Commun. Technol.*, vol. 146, pp. 565–574, 2023, doi: 10.1007/978-3-031-13588-0_49.

[196] R. Li and B. Bhanu, “Fine-Grained Visual Dribbling Style Analysis for Soccer Videos With Augmented Dribble Energy Image,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2019, pp. 2439–2447, doi: 10.1109/CVPRW.2019.00299.

[197] M. Bowditch and D. van der Haar, “Context-based Data Augmentation for Improved Ballet Pose Recognition,” in *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, 2022, pp. 1–6, doi: 10.1109/icABCD54961.2022.9856109.

[198] D. Osokin, “Real-time 2D Multi-Person Pose Estimation on CPU: Lightweight OpenPose,” *ICPRAM 2019 - Proc. 8th Int. Conf. Pattern Recognit. Appl. Methods*, pp. 744–748, Nov. 2018, doi: 10.48550/arxiv.1811.12004.

[199] F. Zhang, X. Zhu, and M. Ye, “Fast Human Pose Estimation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 3512–3521, Nov. 2018, doi: 10.48550/arxiv.1811.05419.

[200] D. Deotale, M. Verma, and S. P, “Human Activity Recognition in Untrimmed Video using Deep Learning for Sports Domain,” *SSRN Electron. J.*, no. Icicnis, pp. 596–607, 2021, doi: 10.2139/ssrn.3769815.

[201] R. Zhao, K. Wang, H. Su, and Q. Ji, “Bayesian graph convolution LSTM for skeleton based action recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 6881–6891, 2019, doi: 10.1109/ICCV.2019.00698.

[202] O. C. Santos and J. G. Boticario, “Practical guidelines for designing and evaluating educationally oriented recommendations,” *Comput. Educ.*, vol. 81, pp. 354–374, Feb. 2015, doi: 10.1016/j.compedu.2014.10.008.

[203] O. C. Santos, R. Uria-Rivas, M. C. Rodriguez-Sanchez, and J. G. Boticario, “An Open Sensing and Acting Platform for Context-Aware Affective Support in Ambient Intelligent Educational Settings,” *IEEE Sens. J.*, vol. 16, no. 10, pp. 3865–3874, 2016, doi: 10.1109/JSEN.2016.2533266.

[204] O. C. Santos, “Toward personalized vibrotactile support when learning motor skills,” *Algorithms*, vol. 10, no. 1, pp. 1–11, 2017, doi: 10.3390/a10010015.

Apéndice. Actividades de Investigación y Transferencia realizadas por Jon Echeverria San Millán durante el desarrollo del Máster en Investigación en Inteligencia Artificial de la UNED (cursos 2020-2021 y 2021-2022).

A continuación, se incluye un índice con el material incluido en un fichero comprimido que se adjunta junto con esta memoria del TFM y que estructurado por los siguientes apartados:

\Citas_Recibidas
- paper_iui_2021
- paper_sensors_2021

\Congresos
- ACM IUI_2021
- ACM UMAP_2021
- AIED_IE_2021
- AIED_MAIED_2021

\Divulgación
- SemanaCiencia_2020
- DivulgaUNED_2021
- NovaCiencia_2021
- Radio5_2021
- AULA_2021
- AULA_2022
- FeriaMadridCiencia_2022

\Libros
- HandbookAIED_2022

\Patente
- ES2890715A1

\Premio
- Toribio_2021

\Proyectos
- Ekintzaile
- INT2AFF

\Revistas
- Sensors_2021

\Showcase
- IAEID_2021