

Article

# Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation

Jon Echeverria <sup>1,\*</sup> and Olga C. Santos <sup>2</sup> <sup>1</sup> Computer Science School, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain<sup>2</sup> aDeNu Research Group, Artificial Intelligence Department, Computer Science School, Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain; ocsantos@dia.uned.es

\* Correspondence: jecheverr33@alumno.uned.es

**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 extract the features of some predefined movements of *Ippon Kihon kumite* (a one-step conventional assault) and compare classification metrics with four data mining algorithms, obtaining high values with them.

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



**Citation:** Echeverria, J.; Santos, O.C. Toward Modeling Psychomotor Performance in Karate Combats Using Computer Vision Pose Estimation. *Sensors* **2021**, *21*, 8378. <https://doi.org/10.3390/s21248378>

Academic Editors: Georgios Meditskos, Stefanos Vrochidis and Ioannis Yiannis Kompatsiaris

Received: 28 October 2021  
Accepted: 3 December 2021  
Published: 15 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

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, or other sensors. Due to legal and other technical issues [5], HAR sensor-based systems are increasing their number of projects. The scope of the sensors at present is large, with cheap and good quality sensors to be able to develop any system in any sector or specific field. In this way, HAR-type systems can gather data from diverse type of sensors

such as: accelerometers (e.g., see [9,10]), gyroscopes, which are usually combined with an accelerometer (e.g., see [11,12]), GPS (e.g., see [13,14]), pulse-meters (e.g., see [15,16]), magnetometers (e.g., see [17,18]) and thermometers (e.g., see [19]). As analyzed in [20], 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.

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 an interdisciplinary technology with a multitude of applications at a commercial, social, educational and industrial level. 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 [21–29]. 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 [30].

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 [31]. In sports and exercise in general, both sensors and video processing are being applied to improve training efficiency [32–34] to develop sport and physical exercise systems. Regarding inertial sensor data, [35] 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. Techniques based on computer vision for Physical Activity Recognition (PAR) have mainly used [36]: red-green-blue (RGB) images, optical flow, 2D depth maps, and 3D skeletons. They use diverse algorithms, such as Naive Bayes (NB) [37], Decision Trees (DT) [38], Support Vector Machines (SVM) [39], Nearest Neighbor (NN) [40], Hidden Markov Models (HMM) [41], and Convolutional Neural Networks (CNN) [42].

However, currently, hardly any sensor-based or video-based HAR systems address the problem of modeling the movements performed from a psychomotor perspective [43,44], such that they can be compared with the same user along time or with users of different levels of expertise. This 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 [45], which can be applied along the lifecycle of technology-based educational systems [46]. In [45], two fundamental challenges are pointed out: (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 [47]. 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 [48,49] to offer expert advice in real time for both practitioners on the Karate combat strategy to follow. KUMITRON can also be used to train motion anticipation performance in Karate practice, based on improving peripheral vision, using computer vision filters [50].

The martial arts domain is useful to contextualize the research on modeling psychomotor activity since martial arts require developing and working on psychomotor skills to progress in the practice. In particular, there are several challenges for the development of intelligent psychomotor systems for martial arts [51]: (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.

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 [52]. To understand the impact of Karate on the practitioner, there is a fundamental reading by its founder [53]. 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 [54]. This is why karatekas must polish their fighting techniques to launch clear 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 [55–57]. However, in this 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 has to 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 dynamic movements of Karate combats.

In this context, we pose the following research question: “Is it possible to detect and identify the postures within a movement in a Karate *kumite* using computer vision techniques in order to model psychomotor performance?”.

The main objective of this research is to develop an intelligent system capable of identifying the postures performed during a Karate combat so that personalized feedback can be provided, when needed, to improve the performance in the combat, thus, supporting psychomotor learning with technology. The novelty of this 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, 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. It is 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 [58].

The rest of the paper is structured as follows. Related works are in Section 2. In Section 3, we present the methodology and define the dataset to answer the research question. In Section 4, we explain how the current dataset is obtained. After that, in Section 5 we analyze the dataset and present the results. In Section 6, we discuss the results and suggest some ideas for future work. Finally, conclusions are 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. 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 applied to sport as described in [59]. In particular, different computer vision techniques can be applied to detect athletes, estimate pose, detect movements and recognize actions [60]. In our research, we focus on movement detection. Thus, we have reviewed works that delve into existing methods: skeleton based models, contour-based models, etc. As discussed in [61–63], 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 [64–66]. 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 [67,68], while the martial artist performs the techniques by themselves.

Virtual reality (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, [69] uses computer vision techniques and transports the individual to the monitor or screen making use of background subtraction techniques [70]. In turn, [71] also creates the avatar of the person to be introduced into the virtual reality and, thus, be able to practice martial arts in front of virtual enemies that are watched through VR glasses, but in this case, body sensors are used to monitor the movements.

The application of computer vision algorithms in the martial arts domain not only focus on the identification of the pose and the movement, but there are advances in the prediction of the next attack. For this, the trainer 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 [72]. Moreover, computer vision together with sensor data can be used to record audiovisual teaching material for Physics learning from the interaction of two (human) bodies as in Phy+Aik [73], 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.

Karate [74] is a popular martial art, invited at the Tokyo Olympics, and thus, there have been efforts in applying new technologies to its modeling from a computing perspective to improve psychomotor performance. In this sense, [75] 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 the five expert karatekas who performed the technique [76]. Sensors are also used for the analysis of Karate movements (e.g., [77,78]) using Dynamic Time Warping (DTW) and Support Vector Machines (SVM). In this context, it is also relevant to know that datasets of Karate movements (such as [79]) have been created for public use in other investigations 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 [80], 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 [57].

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 [69,81], which has also been explored in KUMITRON [50] 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. In the particular case of Karate, studies are being carried out, although mainly focus on analyzing how practitioners perform the techniques individually. 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 users, where 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. Moreover, 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 performed. Since computer vision seem to produce good results for HPE, in this paper we are going to 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 created with some *kumite* movements aimed to address the research question raised in the introduction.

### 3. Materials and Methodology

In this section, we present some computer vision algorithms that can be used for pose estimation and present the methodology proposed to answer the research question.

#### 3.1. Materials

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 [82], WrnchAI and OpenPose offer similar features, although WrnchAI seems to be faster than OpenPose. In turn, [83] concludes that AlphaPose is above both when used for weight lifting in 2D vision. Other studies such as [84] find OpenPose superior and more robust when applied to real situations outside of the specific datasets such as MPII [85] and COCO [86] datasets.

From our own review, we conclude that OpenPose is the one that has generated more literature works and seems to have the largest community of developers. OpenPose has been used in multiple areas: sports [87], telerehabilitation [88], HAR [89–91], artistic disciplines [92], identification of multi-person groups [93], and VR [94]. Thus, we have selected OpenPose algorithm for this work due to the following reasons: (i) it is open source, (ii) it can be applied in real situations with new video inputs [84], (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.



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

Pose Estimation Algorithms	Description
<b>AlphaPose</b> ( <a href="https://github.com/MVIG-SJTU/AlphaPose">https://github.com/MVIG-SJTU/AlphaPose</a> , accessed on 28 November 2021)	Presented in 2016 [95], 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.
<b>DeepCut</b> ( <a href="https://github.com/eldar/deepcut">https://github.com/eldar/deepcut</a> , accessed on 28 November 2021)	System developed in 2016 [96] 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.
<b>Deep Pose</b> ( <a href="https://github.com/mitmul/deeppose">https://github.com/mitmul/deeppose</a> , accessed on 28 November 2021)	An algorithm presented in 2014 [97] 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 [98], DeepPose obtained better results than the rest of the works, becoming a benchmark of that moment.
<b>DensePose</b> ( <a href="https://github.com/facebookresearch/DensePose">https://github.com/facebookresearch/DensePose</a> , accessed on 28 November 2021)	It is an algorithm developed in 2018 by members of Facebook [99] 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.
<b>High Resolution Net (HRNet)</b> ( <a href="https://github.com/HRNet/HigherHRNet-Human-Pose-Estimation">https://github.com/HRNet/HigherHRNet-Human-Pose-Estimation</a> , accessed on 28 November 2021)	Neural network architecture for the estimation of human pose developed in 2019 by Microsoft [100]. 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 [101]. It is a useful architecture to implement in the postural analysis of televised events since it makes high-resolution estimates of postures.
<b>OpenPose</b> ( <a href="https://github.com/CMU-Perceptual-Computing-Lab/openpose">https://github.com/CMU-Perceptual-Computing-Lab/openpose</a> , accessed on 28 November 2021)	Computer vision algorithm for the estimation of pose in real time of several people in 2D developed in 2017 [102]. 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.
<b>PoseNet</b> ( <a href="https://github.com/tensorflow/tfjs-models/tree/master/posenet">https://github.com/tensorflow/tfjs-models/tree/master/posenet</a> , accessed on 28 November 2021)	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 [103]. At first, it was aimed at lightweight devices such as mobile phones or browsers, although today it has advanced and improved performance.
<b>WrnchAI</b> ( <a href="https://go.hingehealth.com/wrnch">https://go.hingehealth.com/wrnch</a> , 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 video. Due to its commercial use, we could not find any scientific paper describing it.

### About OpenPose

OpenPose [102] 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 $2 \times 21$ 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).

In this way, we can start 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 [48]). Applying 3D will require the use of depth cameras.

As described in [102], 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:

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

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.

$$\sigma : \mathbb{R}^k \rightarrow (0, 1)^k$$

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, \dots, K. \quad (2)$$

All these steps can be represented by the following diagram in Figure 1.

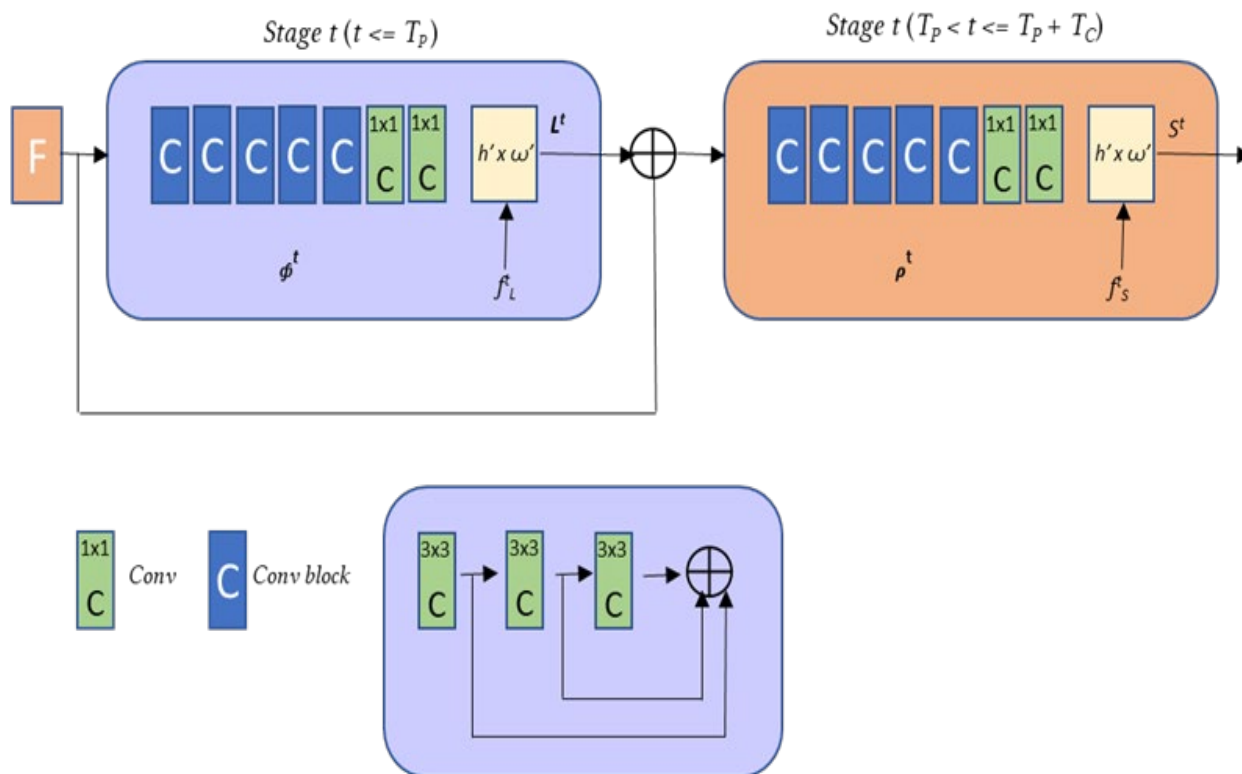


Figure 1. OpenPose's network structure schematic diagram (obtained from [104]).

### 3.2. Methodology

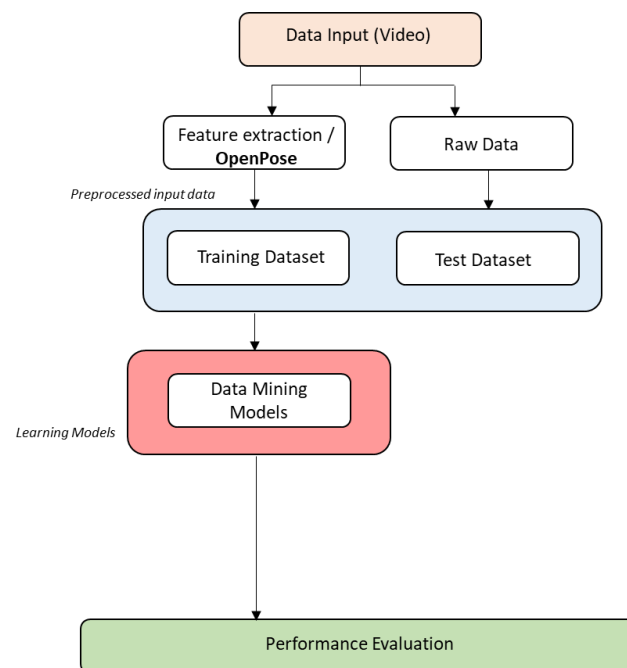
The objective of the research introduced in this paper is to determine if computer vision algorithms (in this case, OpenPose) are useful to identify the movements performed in a Karate *kumite* by both practitioners. To answer the question, a dataset will be prepared following predefined *Kihon Kumite* movements both corresponding to attack and defense as explained in Section 4.2. 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). These two stages are broken down into the following sub-stages, as shown in Figure 1:

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 Figures 3 and 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 [105–107].
4. **Test the movement classifier:** Apply the trained classifier to the non-preprocessed input (raw data) with the movements performed by the karateka.



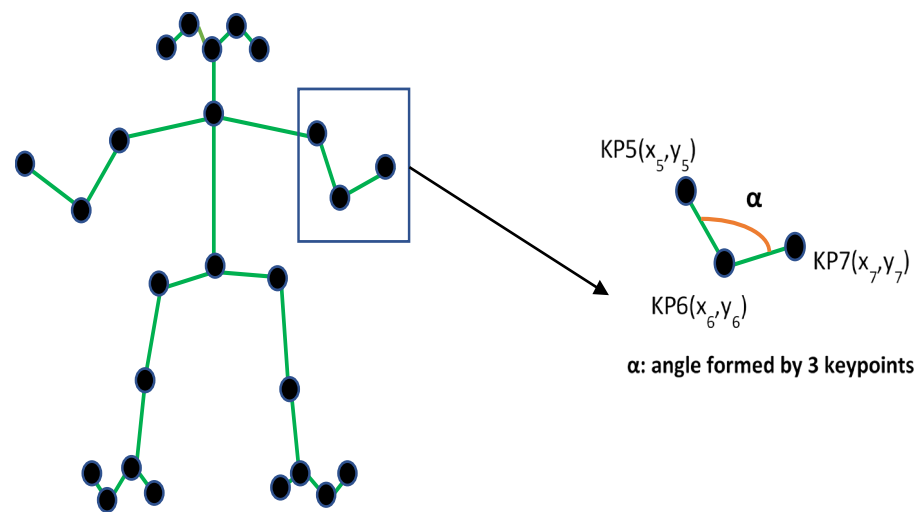
5. **Evaluate the performance of the classifiers:** Compare the results obtained by each algorithm in the classification process with usual machine learning classification metrics, currently those offered by Weka: (i) true positive rate, (ii) false positive rate, (iii) precision (number of true positives that are actually positive compared to the total number of predicted positive values), (iv) recall (number of true positives that the model has classified based on the total number of positive values), (v) F-measure (metric that combines precision and recall), (vi) MMC (Mahalanobis Metric for Clustering: minimizes the distances between similarly labeled inputs while maximizing the distances between differently labeled inputs), ROC area (area under the Receiver Operating Curve: used for classification problems and represents the percentage of true positives against the ratio of false positives), and PRC area (Precision Recall Curve: a plot of precision vs. recall for all potential cut-offs for a test).

This process follows the scheme in Figure 2, which shows the steps proposed for Karate movement recognition using OpenPose algorithm to extract the features from the images and data mining algorithms on these features to classify the movements:



**Figure 2.** Steps for Karate movement recognition using OpenPose.

OpenPose works with 25 keypoints (numbered from 0 to 24, as shown in Section 4.3 in Figure 9) for the extraction of input data, which are used to generate triples, as it can be seen in Figure 3. 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> accessed on 2 December 2021) with the ones for the feet and waist from the Human Foot Keypoint dataset ([https://cmu-perceptual-computing-lab.github.io/foot\\_keypoint\\_dataset/](https://cmu-perceptual-computing-lab.github.io/foot_keypoint_dataset/) accessed on 2 December 2021)).



**Figure 3.** Example of angle feature extraction from the keypoints obtained by combining COCO and Human Foot Keypoint datasets.

The 26 triplets that can be obtained from the 25 keypoints in Figure 3 are grouped and named by their main anatomical part, as shown in Figure 4.

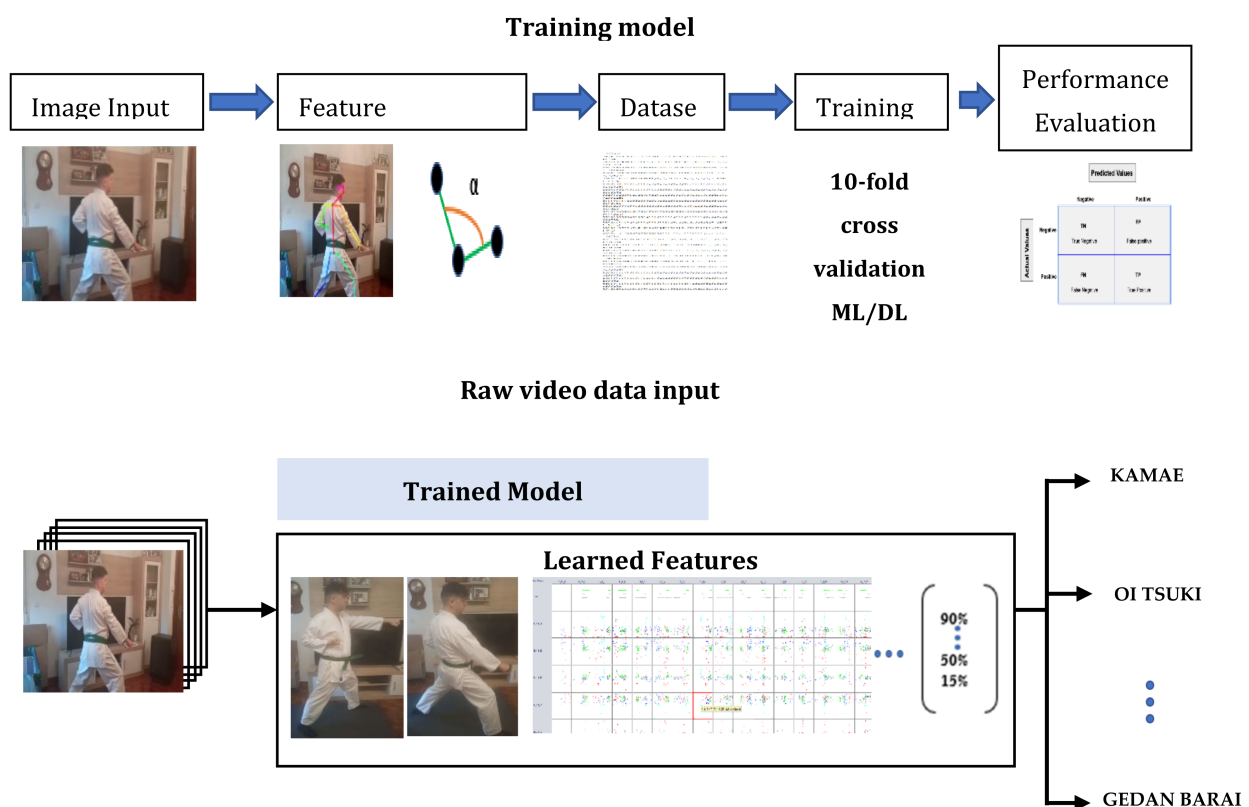
MAIN PART	TRIPLETS OF KEYPOINTS
HEAD	17,15,0
	18,16,0
	15,0,1
	16,0,1
NECK	0,1,2
	0,1,5
SHOULDERS RIGHT ARM	1,2,3
	2,3,4
SHOULDERS LEFT ARM	1,5,6
	5,6,7
SHOULDERS AND TRUNK	2,1,8
	5,1,8
TRUNK AND PELVIS	1,8,9
	1,8,12
LEFT AND RIGHT LEGS	8,9,10
	8,12,13
	9,10,11
	12,13,14
RIGHT FOOT	10,11,24
	10,11,22
	24,11,22
	11,22,23
LEFT FOOT	13,14,21
	13,14,19
	21,14,19
	14,19,20

**Figure 4.** Triplets of keypoints for calculating the angle with OpenPose.

This allows identifying the numbering of the triples with their body position. As an example we have the keypoints numbered 5, 6 and 7 that correspond to the main part “Shoulders Left Arm” as it can be seen in Figure 4. In Figure 3, the angular creation of the commented triplet can be graphically observed. Figure 4 collects all the triples formed by the keypoints in Figure 3, naming them the main body part they represent. The grouping of the keypoints in triples makes it possible to avoid the variability of the points depending on the height of the practitioners, in order to calculate their angle. In this way, a person of 190 cm tall and another of 170 cm tall, will have similar angles for the same Karate posture,

compared to the bigger variation of the keypoints due to the different height of the bodies. This allows the algorithm training to be more efficient and fewer inputs are required to achieve optimal computational performance.

As a result, Figure 5 shows the processing flow proposed to answer the research question 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 used on real *kumite* movements in real time in order to identify the *kumite* posture performed, as shown on the bottom.



**Figure 5.** Training model and the raw data input process flow proposed in this research.

For the selection of the data mining classifiers, a search was made of the types of classifiers applied in general to computer vision. Some works such as [108–112] 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 [113–116]. In particular, a deep learning classification algorithm that is having very good results according to the studies found is the Weka DeepLearning4j algorithm [117–119]. The Weka Deep Learning kit allows to create advanced Neural Network Layers as it is explained in [120].

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 already been used to estimate human pose in other similar experiments [121–123]. Another algorithm that has been applied to this type of classifications is the J48 decision tree [124–126]. 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 [127,128], and the aforementioned DeepLearning4j. Thus, we

have selected two ML and two DL algorithms. Table 3 compares both types of algorithms following [129].

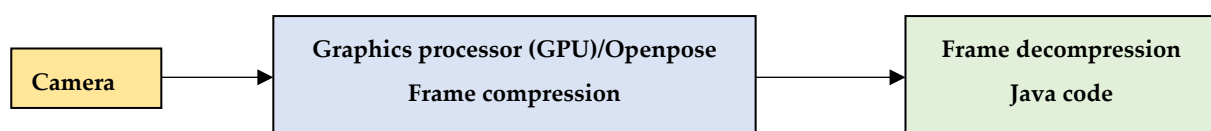
**Table 3.** Machine Learning vs. Deep Learning (obtained from [129]).

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

Due to the type of hardware that was used, which consisted of standard cameras (without depth functionality) available in webcams, mobiles and drones, the research was oriented to the application of the video image in 2D format for its correct processing and postural identification.

### 3.3. Computational Cost

The operation of OpenPose for the application programmed consists of a communication system between OpenPose and the functions designed for the video manipulation, done with the four classes in a Java application listed below. The computational cost associated with the processing is determined through the computational analysis of the following main parts of the code that are represented in Figure 6.



**Figure 6.** Process to calculate the computational cost.

The video input is captured by the OpenPose application libraries that identify the keypoints of each individual that appears in the image. As explained in the official OpenPose paper [102], the runtime of OpenPose consists of two major parts: (1) CNN processing time whose complexity is  $O(1)$ , constant with varying number of people; and (2) multi-person parsing time, whose complexity is  $O(N^2)$ , where  $N$  represents the number of people.

OpenPose communicates with our application by sending information about the identified keypoints and the number of individuals in the image. The Java application, receives the data from OpenPose, dimensioned in “ $N$ ” the problem for the complexity calculation.

There are four main classes:

- *pxStreamReadCamServer*: receives  $M \times N$  pixels image compressed in JPEG:  $(640 * 480) = O(ni)$  to decompress
- *receiveArrayOfDouble*: receives 26 double (angle) =  $26 = K2$
- *stampMatOnJLabel*: display image on a JLabel =  $\sim K1$
- *wrapperClassifier.prediceClase*: deduce position of the angles,  $26 * \text{num layers} = 26 * 16 = K3$

Thus, the cost estimate for these functions is:

- Cost:  $O(ni) + K2 = 26 \Rightarrow 1$  (one) +  $K3 = 26 * 16 \Rightarrow 16 = O(ni) + K$ ;

Therefore, the computational cost of the intelligent movement identification system for an individual and several is:

- Full cost per frame for one person:  $2 * O(ni) + K$ ;

- Full cost per frame for more than one person:  $O(N^2) + 2 * O(ni) + K$ ;

It can be seen how in this case the higher computational cost of the OpenPose algorithm marks the total operational cost of the set of applications, finally reaching a quadratic order.

#### 4. Dataset Construction

##### 4.1. Defining the Dataset Inputs

To answer the research question, first we have to select 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* [130] (very explosive with very fast movements [131]). Multi-step combat (each step is as an attack) is a simple couple exercise that slowly lead 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 with each other but train with themselves. The different types of *kumite* are listed in Table 4, from those with less freedom of movement to free combat.

**Table 4.** 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 ( <i>Kaisho Ippon Kumite</i> ): Unconventional one-step assault. In this type of work one of the karatekas (acting as <i>uke</i> ) performs the attack and the other (as <i>tori</i> ) defends it and counterattacks the <i>uke</i> who defends the counterattack by the <i>tori</i> 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.

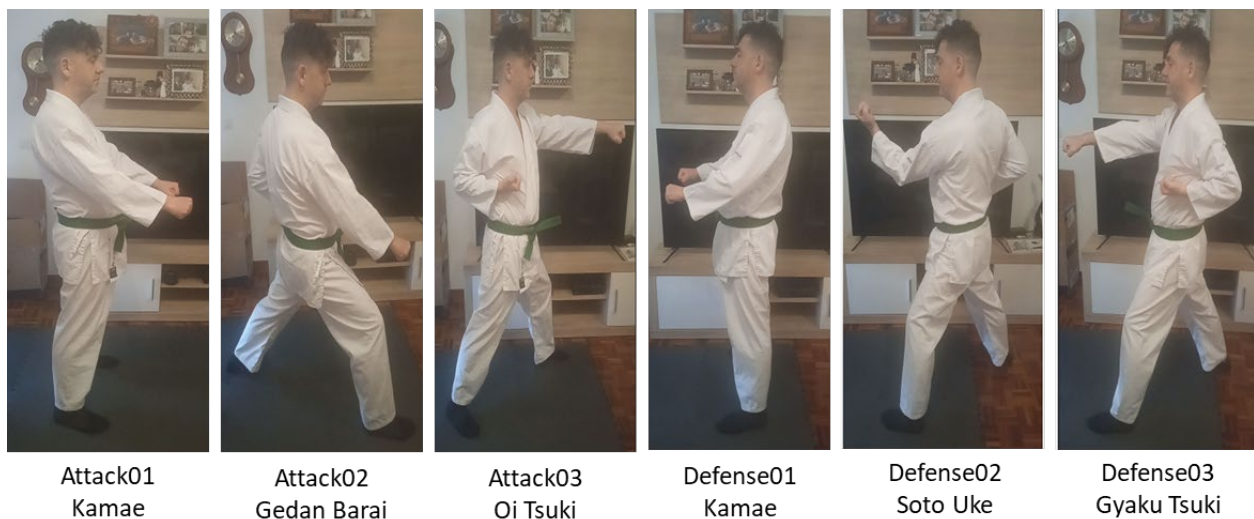
In order to follow a progressive approach in our research, we started with the *Ippon Kihon kumite*, which is a pre-established exercise so that it can facilitate the labeling of movements and their analysis. This is the most basic *kumite* exercise, consisting of the conventional one-step assault. We 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.

##### 4.2. Preparing the Dataset

In this first step of our research, the following attack and defense sequences were defined to create the initial dataset, as shown in Figure 7.

The *Ippon Kihon kumite* 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*.





**Figure 7.** Pictures of the postures in an *Ippon Kihon kumite* selected for the initial dataset.

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). We aim at increasing the difficulty and complexity of the dataset gradually along the research. Thus, for this first dataset, we selected direct upper trunk techniques (not lateral or angular) to facilitate the work of OpenPose classification. In addition, simple techniques with the limbs were selected. 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 5.

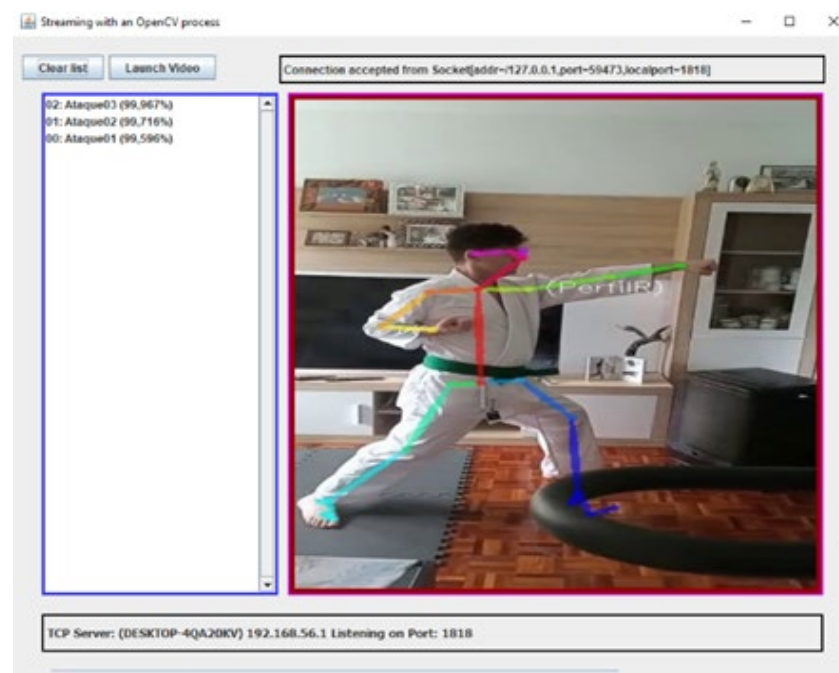
**Table 5.** Postures to be detected in the dataset for the *Ippon Kihon kumite*.

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

#### 4.3. Implemented Application to Obtain the Dataset

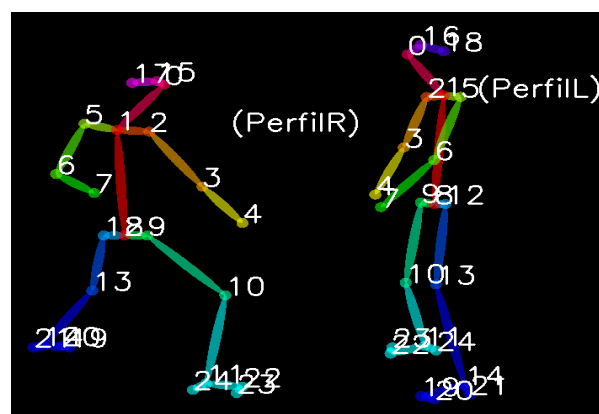
Once the dataset was prepared, a Java application was developed in KUMITRON to collect the data from the practitioners. In particular, in this initial and exploratory collection of data, a green belt participant was video recorded performing the proposed movements, both attack and defense. The video was taken statically for one minute in which the karateka had to be in one for the predefined postures of Table 5 and Figure 7, 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.

Figure 8 shows the interface of the application while recording the *Oi Tsuki* movement performed by the karateka while testing the application. On the left panel, the postures detected are listed chronologically from bottom to top. On the right of the image, the skeleton identified with the OpenPose algorithm is shown on the real image.



**Figure 8.** Performing individually an *Ippon Kihon kumite* movement (*Oi Tsuki* attack) for the recognition tests.

The application can also identify the skeletons of both karatekas when interacting in an *Ippon Kihon kumite*. Figure 9 shows the left practitioner (male, green belt) launching a *Gedan Barai* attack when the one on the right (women, white belt) is waiting in *Kamae*.



**Figure 9.** Applying OpenPose to *Ippon Kihon kumite* training of two participants, where the body keypoints are represented (image obtained from KUMITRON).

## 5. Analysis and Results

The attributes selected for the classification with the data mining algorithms, as explained in Section 3.2, were the 26 triples obtained from the OpenPose keypoints. In this way, it is possible to calculate the angles that are formed by the different joints of the body of each karateka. Figure 10 shows the 26 different attributes considered corresponding to the 26 triplets introduced in Figure 4. They are shown in order as they are sent by the application. The last window shows the classes (postures) that are being identified. Colors are assigned as indicated in Table 6.

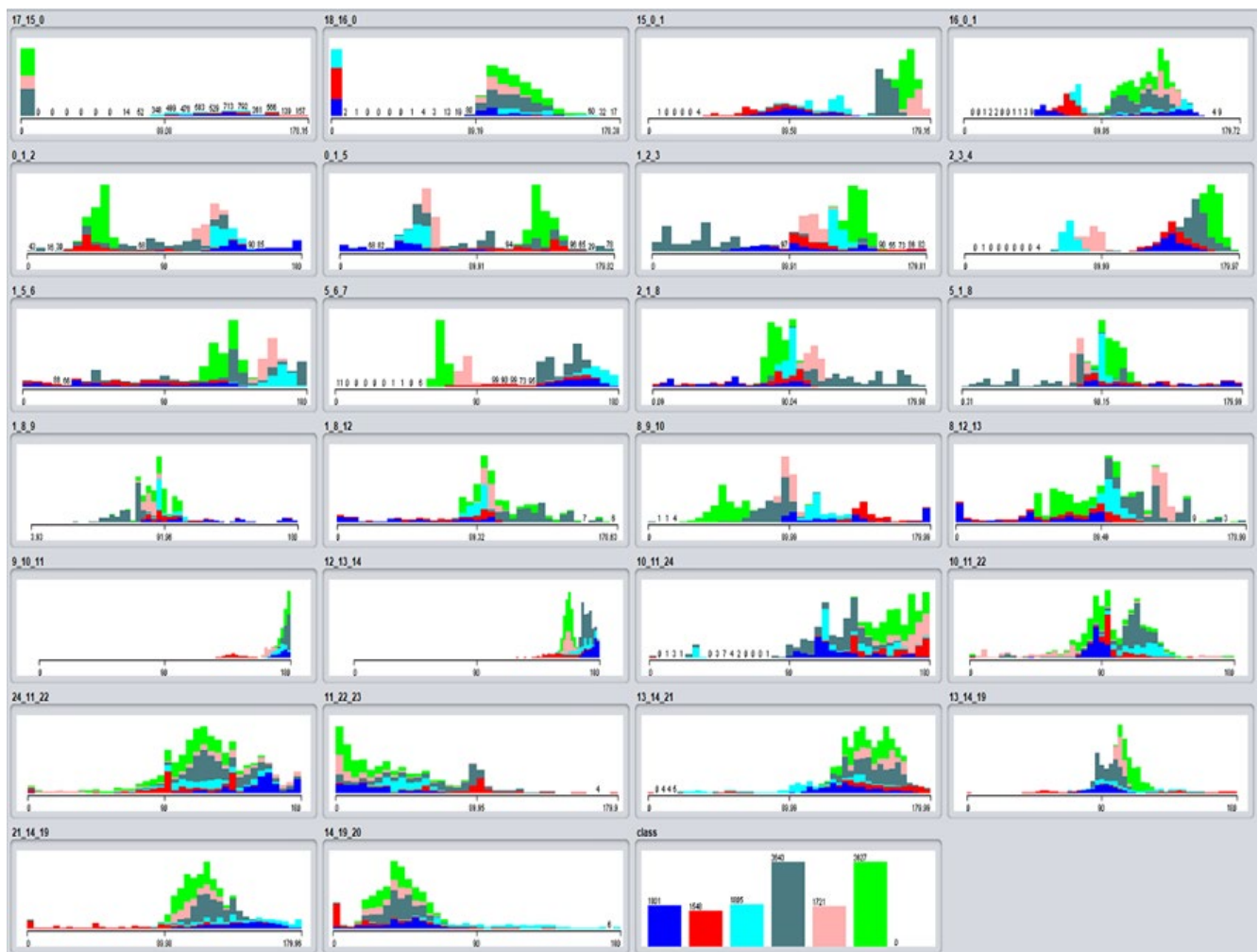


Figure 10. Display of defined attributes selected for identification in defined classes (obtained from Weka).

Table 6. Defined Karate postures with the colors used in Figure 10 and the number of dataset inputs considered.

Class	Color	Number of Dataset Inputs
Attack01: "Kamae"	Blue	1801
Attack02: "Gedan Barai"	Red	1548
Attack03: "Oi Tsuki"	Cyan	1805
Defense01: "Kamae"	Dark green	3643
Defense02: "Soto Uke"	Pink	1721
Defense03: "Gyaku Tsuki"	Light green	3627

For each of the postures (classes) several videos were recorded with KUMITRON to create the initial dataset (as described in Section 4.3) 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 members that remain in the back of the camera.

In Figure 10, it can be observed in a range of 180 degrees (range in which the 26 angles move) the values that each class takes according to the part of the body. This can be interesting in future research to investigate, for example, whether a posture is overloaded and is likely to cause some type of injury because it is generating an angle that what it is convenient for that part of the body.

The statistical results obtained by applying the four algorithms (BayesNet, J48, MLP and DeepLearning4J) using Weka suite to the identification of movements of the *Ippon Kihon kumite* dataset built in this first step of our research are presented in Table 7. The dataset obtained for the expression is publicly accessible (<https://github.com/Kumitron> accessed on 2 December 2021).

**Table 7.** Summary results of the four data mining algorithms using 10-fold cross validation.

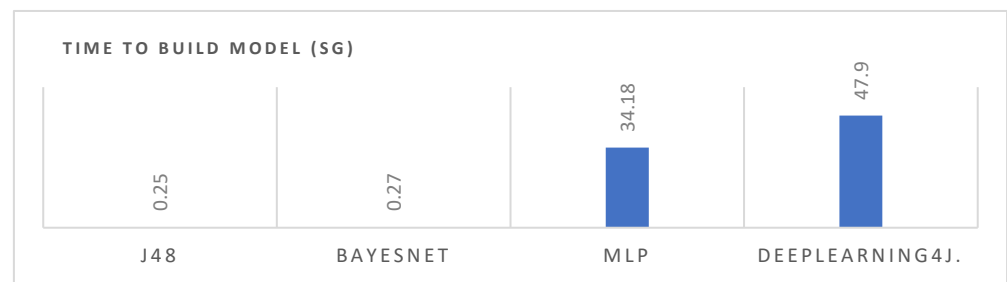
	J48	BayesNet	MLP	DeepLearning4J
Time to build model (seg)	0.5	0.27	34.18	47.9
Correctly Classified Instances	14139	14137	14138	14142
% Correct	99.9576	99.9434	99.9505	99.9788
Incorrectly Classified Instances	6	8	7	3
% Incorrect	0.0424	0.0566	0.0495	0.0212
Kappa statistic	0.9995	0.9993	0.9994	0.9997
Mean absolute error	0.0001	0.0002	0.0006	0.0001
Root mean squared error	0.0119	0.0137	0.0118	0.0074
Relative absolute error	0.0539	0.0702	0.2148	0.0521
Root relative squared error%	3.2416	3.7228	3.2134	2.0249
Total number of Instances	14145	14145	14145	14145

Table 7 shows that the classifying algorithms have obtained a good classification performance without significant differences among them. Nonetheless, the performance of the algorithms will need to be reevaluated when more users and more movements are included in the dataset and it becomes more complex. The detail of the evaluation metrics results by type of algorithm is shown in Table 8.

**Table 8.** Evaluation metrics obtained for the four algorithms.

Algorithm	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
BayesNet	1.000	0.000	0.999	0.999	0.999	0.999	1.000	1.000
J48	1.000	0.000	1.000	1.000	0.999	0.999	1.000	0.999
MLP	1.000	0.000	1.000	1.000	1.000	0.999	1.000	1.000
DeepLearning4J	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000

As expected from Table 7, evaluation metrics are good in all the algorithms. DeepLearning4J seems to perform slightly better. However, another important variable to consider in the performance analysis between algorithms is the time it takes to build the models. Figure 11 shows that, from the comparison of the four algorithms, the differences in processing time between the ML and DL algorithms are high. As expected, neural networks take much longer, and since performance results are similar, they do not seem necessary for the current analysis. However, we will analyze how both the processing time evolves in these algorithms, with a greater number of inputs and varying conditions of speed of movements in future versions of the dataset.



**Figure 11.** Comparative times between the applied algorithms to build the models.

The detailed performance results by type of movement (class) for each algorithm is shown in Tables 9–12 reporting the values of the metrics computed by Weka. These results were already good in the overall analysis, and no significant difference was found between types of movements in any of the algorithms.

The metrics computed for the BayesNet algorithm are reported in Table 9.

**Table 9.** BayesNet algorithm—metrics computed by movement (class).

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	0.999	1.000	0.999	0.999	1.000	1.000	Attack01
	1.000	0.000	0.996	1.000	0.998	0.998	1.000	0.999	Attack02
	0.996	0.000	1.000	0.996	0.998	0.997	1.000	1.000	Attack03
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense01
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense02
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense03
<b>Avg.</b>	1.000	0.000	0.999	0.999	0.999	0.999	1.000	1.000	

The metrics computed for the J48 algorithm are reported in Table 10.

**Table 10.** J48 algorithm—metrics computed by movement (class).

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.999	0.000	0.998	0.999	0.999	0.998	0.999	0.997	Attack01
	0.998	0.000	0.999	0.998	0.998	0.998	0.999	0.997	Attack02
	0.999	0.000	1.000	0.999	1.000	1.000	1.000	1.000	Attack03
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense01
	1.000	0.000	0.999	1.000	1.000	1.000	1.000	0.999	Defense02
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense03
<b>Avg.</b>	1.000	0.000	1.000	1.000	0.999	0.999	1.000	0.999	

The metrics computed for the MLP algorithm are reported in Table 11.



**Table 11.** MLP algorithm—metrics computed by movement (class).

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.999	0.000	0.999	0.999	0.999	0.999	0.999	1.000	Attack01
	0.997	0.000	0.998	0.997	0.998	0.997	1.000	1.000	Attack02
	0.999	0.000	0.999	0.999	0.999	0.999	1.000	1.000	Attack03
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense01
	1.000	0.000	0.999	1.000	1.000	1.000	1.000	0.999	Defense02
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense03
<b>Avg.</b>	1.000	0.000	1.000	1.000	1.000	0.999	1.000	1.000	

The metrics computed for the BayesNet algorithm are reported in Table 12.

**Table 12.** Deep Learning4j—metrics computed by movement (class).

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.999	0.000	0.999	0.999	0.999	0.999	0.999	1.000	Attack01
	0.999	0.000	1.000	0.999	0.999	0.999	1.000	1.000	Attack02
	1.000	0.000	0.999	1.000	0.999	0.999	1.000	1.000	Attack03
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense01
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	0.999	Defense02
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	Defense03
<b>Avg.</b>	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

### Network Hyperparameters

For the application of the classification models, possible methods of optimization of the hyperparameters were investigated [132]. Since the results obtained with the current dataset and default parameters in Weka were good, no optimization was performed, and default values were left. Table 13 shows the scripts used to call the algorithms in Weka.

**Table 13.** Network hyperparameters using Weka.

Network Hyperparameters	
<b>BayesNet</b>	weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 --P 1 -S BAYES -E Weka.classifiers.bayes.net.estimate.SimpleEstimator --A 0.5
<b>J48</b>	weka.classifiers.trees.J48 -C 0.25 -M 2
<b>MLP</b>	weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
<b>DeepLearning4j</b>	weka.dl4j.inference.Dl4jCNNExplorer -custom-model "weka.dl4j.inference.CustomModelSetup -channels 3 -height 224 -width 224 -model-file C:\\Users\\Johni\\wekafiles\\packages" -decoder "weka.dl4j.inference.ModelOutputDecoder -builtIn IMAGENET -classMapFile C:\\Users\\Johni\\wekafiles\\packages" -saliency-map "weka.dl4j.interpretability.WekaScoreCAM -bs 1 -normalize -output C:\\Users\\Johni\\wekafiles\\packages -target-classes -1" -zooModel "weka.dl4j.zoo.Dl4jResNet50 -channelsLast false -pretrained IMAGENET"

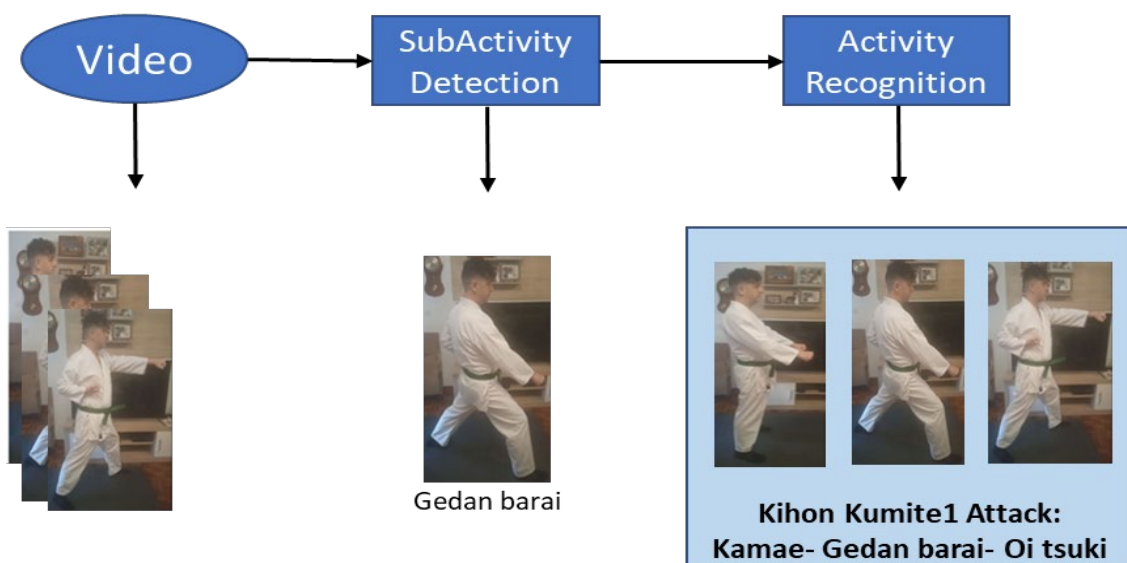
## 6. Discussion

Recalling the classification of human activities introduced in Section 1 [1], this reeseearch focuses on type iv, that is, group activities, which focus on the interaction of actions between two people. 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 up 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.

The current approach 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 [133]. 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 12.



**Figure 12.** Subactivity recognition in a sequence of *Ippon Kihon kumite* techniques.

This can be useful in order 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 has to 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 in order 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 [47,134], 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 keypoints 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 [50].

The progress in the classification of the movements during a *kumite* will be integrated into KUMITRON intelligent psychomotor system [48]. 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 [135]), 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 [136]) through the appropriate sensorial channel (vibrotactile feedback should be explored due to the potential discussed in [137]). 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.

## 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 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., BayesNet, J48, MLP and DeepLearning4j in Weka) have a precision of above 99% for the current (and limited) dataset. With this percentage of success, it is expected that progress can be made with the inclusion of more *Kihon kumite* sequences in a new version of the dataset to analyze whether they can be identified within a *Kihon kumite* fighting exercise. In this way, the research question would be satisfied for the training level of *Kihon kumite*.

**Author Contributions:** Conceptualization, O.C.S. and J.E.; Methodology, O.C.S. and J.E.; 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.

**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<sup>2</sup>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<sup>2</sup>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.

**Data Availability Statement:** Not applicable.

**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.

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

## References

1. Vrigkas, M.; Nikou, C.; Kakadiaris, I.A. A review of human activity recognition methods. *Front. Robot. AI* **2015**, *2*, 1–28. [[CrossRef](#)]
2. Yang, Y.; Saleemi, I.; Shah, M. Discovering motion primitives for unsupervised grouping and one-shot learning of human actions, gestures, and expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1635–1648. [[CrossRef](#)] [[PubMed](#)]
3. Ni, B.; Moulin, P.; Yang, X.; Yan, S. Motion Part Regularization: Improving Action Recognition via Trajectory Group Selection. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2015**, 3698–3706. [[CrossRef](#)]
4. Patron-Perez, A.; Marszalek, M.; Reid, I.; Zisserman, A. Structured learning of human interactions in TV shows. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2441–2453. [[CrossRef](#)]
5. Li, J.H.; Tian, L.; Wang, H.; An, Y.; Wang, K.; Yu, L. Segmentation and Recognition of Basic and Transitional Activities for Continuous Physical Human Activity. *IEEE Access* **2019**, *7*, 42565–42576. [[CrossRef](#)]
6. Martinez, H.P.; Yannakakis, G.N.; Hallam, J. Don't classify ratings of affect; Rank Them! *IEEE Trans. Affect. Comput.* **2014**, *5*, 314–326. [[CrossRef](#)]
7. Lan, T.; Sigal, L.; Mori, G. Social roles in hierarchical models for human activity recognition. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 1354–1361. [[CrossRef](#)]
8. Nweke, H.F.; Teh, Y.W.; Al-garadi, M.A.; Alo, U.R. Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst. Appl.* **2018**, *105*, 233–261. [[CrossRef](#)]
9. Marinho, L.B.; de Souza Junior, A.H.; Filho, P.P.R. A new approach to human activity recognition using machine learning techniques. *Adv. Intell. Syst. Comput.* **2017**, *557*, 529–538. [[CrossRef](#)]
10. Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidiú, R.; Fuks, H. Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. *Lect. Notes Comput. Sci.* **2012**, *7589*, 52–61. [[CrossRef](#)]
11. Masum, A.K.M.; Jannat, S.; Bahadur, E.H.; Alam, M.G.R.; Khan, S.I.; Alam, M.R. Human Activity Recognition Using Smartphone Sensors: A Dense Neural Network Approach. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology, 2019, ICASERT 2019, Dhaka, Bangladesh, 3–5 May 2019; 2019; pp. 1–6. [[CrossRef](#)]

12. Bhuiyan, R.A.; Ahmed, N.; Amiruzzaman, M.; Islam, M.R. A robust feature extraction model for human activity characterization using 3-axis accelerometer and gyroscope data. *Sensors* **2020**, *20*, 6990. [CrossRef]
13. Sekiguchi, R.; Abe, K.; Shogo, S.; Kumano, M.; Asakura, D.; Okabe, R.; Kariya, T.; Kawakatsu, M. Phased Human Activity Recognition based on GPS. In Proceedings of the UbiComp/ISWC 2021—Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, New York, NY, USA, 21–26 September 2021; pp. 396–400. [CrossRef]
14. Zhang, Z.; Poslad, S. Improved use of foot force sensors and mobile phone GPS for mobility activity recognition. *IEEE Sens. J.* **2014**, *14*, 4340–4347. [CrossRef]
15. Ulyanov, S.S.; Tuchin, V.V. Pulse-wave monitoring by means of focused laser beams scattered by skin surface and membranes. *Static Dyn. Light Scatt. Med. Biol.* **1993**, *1884*, 160. [CrossRef]
16. Pan, S.; Mirshekari, M.; Fagert, J.; Ramirez, C.G.; Chung, A.J.; Hu, C.C.; Shen, J.P.; Zhang, P.; Noh, H.Y. Characterizing human activity induced impulse and slip-pulse excitations through structural vibration. *J. Sound Vib.* **2018**, *414*, 61–80. [CrossRef]
17. Zhang, M.; Sawchuk, A.A. A preliminary study of sensing appliance usage for human activity recognition using mobile magnetometer. In Proceedings of the UbiComp '12—Proceedings of the 2012 ACM Conference on Ubiquitous Computing, Pittsburgh, PA, USA, 5–8 September 2012; pp. 745–748. [CrossRef]
18. Altun, K.; Barshan, B. Human activity recognition using inertial/magnetic sensor units. *Lect. Notes Comput. Sci.* **2010**, *6219 LNCS*, 38–51. [CrossRef]
19. Hoang, M.L.; Carratù, M.; Paciello, V.; Pietrosanto, A. 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* **2021**, *21*, 2313. [CrossRef]
20. Santos, O.C. Artificial Intelligence in Psychomotor Learning: Modeling Human Motion from Inertial Sensor Data. *World Sci.* **2019**, *28*, 1940006. [CrossRef]
21. Nandakumar, N.; Manzoor, K.; Agarwal, S.; Pillai, J.J.; Gujar, S.K.; Sair, H.I.; Venkataraman, A. Automated eloquent cortex localization in brain tumor patients using multi-task graph neural networks. *Med. Image Anal.* **2021**, *74*, 102203. [CrossRef]
22. Aggarwal, J.K.; Cai, Q. Human Motion Analysis: A Review. *Comput. Vis. Image Underst.* **1999**, *73*, 428–440. [CrossRef]
23. Aggarwal, J.K.; Ryoo, M.S. Human activity analysis: A review. *ACM Comput. Surv.* **2011**, *43*, 43. [CrossRef]
24. Prati, A.; Shan, C.; Wang, K.I.K. Sensors, vision and networks: From video surveillance to activity recognition and health monitoring. *J. Ambient Intell. Smart Environ.* **2019**, *11*, 5–22. [CrossRef]
25. Roitberg, A.; Somani, N.; Perzylo, A.; Rickert, M.; Knoll, A. Multimodal human activity recognition for industrial manufacturing processes in robotic workcells. In Proceedings of the ICMI 2015—Proceedings of the 2015 ACM International Conference Multimodal Interaction, Seattle, WA, USA, 9–13 November 2015; pp. 259–266. [CrossRef]
26. Piyathilaka, L.; Kodagoda, S. Human Activity Recognition for Domestic Robots. In *Field and Service Robotics*; Springer: Cham, Switzerland, 2015; Volume 105, pp. 395–408. [CrossRef]
27. Osmani, V.; Balasubramaniam, S.; Botvich, D. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *J. Netw. Comput. Appl.* **2008**, *31*, 628–655. [CrossRef]
28. Subasi, A.; Radhwan, M.; Kurdi, R.; Khateeb, K. IoT based mobile healthcare system for human activity recognition. In Proceedings of the 15th Learning & Technology Conference (L & T 2018), Jeddah, Saudi Arabia, 25–26 February 2018; pp. 29–34. [CrossRef]
29. Wang, Y.; Cang, S.; Yu, H. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **2019**, *137*, 167–190. [CrossRef]
30. Rashid, O.; Al-Hamadi, A.; Michaelis, B. A framework for the integration of gesture and posture recognition using HMM and SVM. In Proceedings of the 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems ICIS, Shanghai, China, 20–22 November 2009; pp. 572–577. [CrossRef]
31. Chu, W.C.C.; Shih, C.; Chou, W.Y.; Ahamed, S.I.; Hsiung, P.A. Artificial Intelligence of Things in Sports Science: Weight Training as an Example. *Computer* **2019**, *52*, 52–61. [CrossRef]
32. Zalluhoglu, C.; Ikizler-Cinbis, N. Collective Sports: A multi-task dataset for collective activity recognition. *Image Vis. Comput.* **2020**, *94*, 103870. [CrossRef]
33. Kautz, T.; Groh, B.H.; Eskofier, B.M. Sensor fusion for multi-player activity recognition in game sports. *KDD Work. Large-Scale Sport. Anal.* **2015**, pp. 1–4. Available online: <https://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2015/Kautz15-SFF.pdf> (accessed on 2 December 2021).
34. Sharma, A.; Al-Dala'In, T.; Alsadoon, G.; Alwan, A. Use of wearable technologies for analysis of activity recognition for sports. In Proceedings of the CITISIA 2020—IEEE Conference on Innovative Technologies in Intelligent Systems and Industrial Applications, Sydney, Australia, 25–27 November 2020. Available online: <https://doi.org/10.1109/CITISIA50690.2020.9371779> (accessed on 20 November 2021). [CrossRef]
35. Camomilla, V.; Bergamini, E.; Fantozzi, S.; Vannozzi, G. Trends Supporting the In-Field Use of Wearable Inertial Sensors for Sport Performance Evaluation: A Systematic Review. *Sensors* **2018**, *18*, 873. [CrossRef]
36. Xia, K.; Wang, H.; Xu, M.; Li, Z.; He, S.; Tang, Y. Racquet sports recognition using a hybrid clustering model learned from integrated wearable sensor. *Sensors* **2020**, *20*, 1638. [CrossRef]
37. Wickramasinghe, I. Naive Bayes approach to predict the winner of an ODI cricket game. *J. Sport. Anal.* **2020**, *6*, 75–84. [CrossRef]



38. Jaser, E.; Christmas, W.; Kittler, J. Temporal post-processing of decision tree outputs for sports video categorisation. *Lect. Notes Comput. Sci.* **2004**, *3138*, 495–503. [CrossRef]
39. Sadlier, D.A.; O'Connor, N.E. Event detection in field sports video using audio-visual features and a support vector machine. *IEEE Trans. Circuits Syst. Video Technol.* **2005**, *15*, 1225–1233. [CrossRef]
40. Nurwanto, F.; Ardiyanto, I.; Wibirama, S. Light sport exercise detection based on smartwatch and smartphone using k-Nearest Neighbor and Dynamic Time Warping algorithm. In Proceedings of the 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 5–6 October 2016. [CrossRef]
41. Hoettinger, H.; Mally, F.; Sabo, A. Activity Recognition in Surfing—A Comparative Study between Hidden Markov Model and Support Vector Machine. *Procedia Eng.* **2016**, *147*, 912–917. [CrossRef]
42. Minhas, R.A.; Javed, A.; Irtaza, A.; Mahmood, M.T.; Joo, Y.B. Shot classification of field sports videos using AlexNet Convolutional Neural Network. *Appl. Sci.* **2019**, *9*, 483. [CrossRef]
43. Neagu, L.M.; Rigaud, E.; Travadel, S.; Dascalu, M.; Rughinis, R.V. 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)*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 335–341.
44. Casas-Ortiz, A.; Santos, O.C. Intelligent systems for psychomotor learning. In *Handbook of Artificial Intelligence in Education*; du Boulay, B., Mitrovic, A., Yacef, K., Eds.; Edward Edgar Publishing: Northampton, MA, USA, 2022; In progress.
45. Santos, O.C. Training the Body: The Potential of AIED to Support Personalized Motor Skills Learning. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 730–755. [CrossRef]
46. Santos, O.C.; Boticario, J.G.; Van Rosmalen, P. The Full Life Cycle of Adaptation in aLFanet eLearning Environment. 2004. Available online: [https://tc.computer.org/tclt/wp-content/uploads/sites/5/2016/12/learn\\_tech\\_october2004.pdf](https://tc.computer.org/tclt/wp-content/uploads/sites/5/2016/12/learn_tech_october2004.pdf) (accessed on 27 November 2021).
47. Casas-Ortiz, A.; Santos, O.C. KSAS: A Mobile App with Neural Networks to Guide the Learning of Motor Skills. In Proceedings of the XIX Conference of the Spanish Association for the Artificial Intelligence (CAEPIA 20/21). Competition on Mobile Apps with A.I. Techniques, Malaga, Spain, 22–24 September 2021. Available online: [https://caepia20-21.uma.es/inicio\\_files/caepia20-21-actas.pdf](https://caepia20-21.uma.es/inicio_files/caepia20-21-actas.pdf) (accessed on 27 November 2021).
48. Echeverria, J.; Santos, O.C. KUMITRON: Artificial intelligence system to monitor karate fights that synchronize aerial images with physiological and inertial signals. In Proceedings of the IUI '21 Companion International Conference on Intelligent User Interfaces, College Station, TX, USA, 13–17 April 2021; pp. 37–39. [CrossRef]
49. Echeverria, J.; Santos, O.C. KUMITRON: A Multimodal Psychomotor Intelligent Learning System to Provide Personalized Support when Training Karate Combats. MAIED'21 Workshop. The First International Workshop on Multimodal Artificial Intelligence in Education. 2021. Available online: <http://ceur-ws.org/Vol-2902/paper7.pdf> (accessed on 27 September 2021).
50. Echeverria, J.; Santos, O.C. Punch Anticipation in a Karate Combat with Computer Vision. In Proceedings of the UMAP 21—Adjunct 29th ACM Conference on User Modeling, Adaptation and Personalization, Utrecht, The Netherlands, 12–25 June 2021; pp. 61–67. [CrossRef]
51. Santos, O.C. Psychomotor Learning in Martial Arts: An opportunity for User Modeling, Adaptation and Personalization. In Proceedings of the UMAP 2017—Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, New York, NY, USA, 9–12 July 2017; pp. 89–92. [CrossRef]
52. Santos, O.C.; Corbi, A. 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* **2019**, *7*, 176458–176469. [CrossRef]
53. Funakoshi, G. *My Way of Life*, 1st ed; Kodansha International Ltd.: San Francisco, CA, USA, 1975.
54. World Karate Federation. Karate Competition Rules. 2020. Available online: [https://www.wkf.net/pdf/WKF\\_Competition%20Rules\\_2020\\_EN.pdf](https://www.wkf.net/pdf/WKF_Competition%20Rules_2020_EN.pdf) (accessed on 27 November 2021).
55. Hachaj, T.; Ogiela, M.R. Application of Hidden Markov Models and Gesture Description Language classifiers to Oyama karate techniques recognition. In Proceedings of the 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, IMIS 2015, Santa Catarina, Brazil, 8–10 July 2015; pp. 160–165. [CrossRef]
56. Yıldız, S. Relationship Between Functional Movement Screen and Some Athletic Abilities in Karate Athletes. *J. Educ. Train. Stud.* **2018**, *6*, 66. [CrossRef]
57. Hachaj, T.; Ogiela, M.R.; Koptyra, K. Application of assistive computer vision methods to Oyama karate techniques recognition. *Symmetry* **2015**, *7*, 1670–1698. [CrossRef]
58. Santos, O.C. *Beyond Cognitive and Affective Issues: Designing Smart Learning Environments for Psychomotor Personalized Learning*; Learning, Design, and Technology; Spector, M., Lockee, B., Childress, M., Eds.; Springer: New York, NY, USA, 2016; pp. 1–24. Available online: [https://link.springer.com/referenceworkentry/10.1007%2F978-3-319-17727-4\\_8-1](https://link.springer.com/referenceworkentry/10.1007%2F978-3-319-17727-4_8-1) (accessed on 27 November 2021).
59. Zhang, F.; Zhu, X.; Ye, M. Fast human pose estimation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3512–3521. [CrossRef]
60. Piccardi, M.; Jan, T. Recent advances in computer vision. *Ind. Phys.* **2003**, *9*, 18–21.
61. Chen, X.; Pang, A.; Yang, W.; Ma, Y.; Xu, L.; Yu, J. SportsCap: Monocular 3D Human Motion Capture and Fine-Grained Understanding in Challenging Sports Videos. *Int. J. Comput. Vis.* **2021**, *129*, 2846–2864. [CrossRef]

62. Shingade, A.; Ghotkar, A. Animation of 3D Human Model Using Markerless Motion Capture Applied To Sports. *Int. J. Comput. Graph. Animat.* **2014**, *4*, 27–39. [CrossRef]
63. Bridgeman, L.; Volino, M.; Guillemaut, J.Y.; Hilton, A. Multi-person 3D pose estimation and tracking in sports. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 2487–2496. [CrossRef]
64. Xiaojie, S.; Qilei, L.; Tao, Y.; Weidong, G.; Newman, L. Mocap data editing via movement notations. In Proceedings of the Ninth International Conference on Computer Aided Design and Computer Graphics (CAD-CG'05), Hong Kong, China, 7–10 December 2005; pp. 463–468. [CrossRef]
65. Thành, N.T.; Hùng, L.V.; Công, P.T. An Evaluation of Pose Estimation in Video of Traditional Martial Arts Presentation. *J. Res. Dev. Inf. Commun. Technol.* **2019**, *2019*, 114–126. [CrossRef]
66. Mohd Jelani, N.A.; Zulkifli, A.N.; Ismail, S.; Yusoff, M.F. A Review of Virtual Reality and Motion Capture in Martial Arts Training. *Int. J. Interact. Digit. Media* **2020**, *5*, 22–25. Available online: <http://repo.uum.edu.my/26996/> (accessed on 20 November 2021).
67. Zhang, W.; Liu, Z.; Zhou, L.; Leung, H.; Chan, A.B. Martial Arts, Dancing and Sports dataset: A challenging stereo and multi-view dataset for 3D human pose estimation. *Image Vis. Comput.* **2017**, *61*, 22–39. [CrossRef]
68. Kaharuddin, M.Z.; Khairu Razak, S.B.; Kushairi, M.I.; Abd Rahman, M.S.; An, W.C.; Ngali, Z.; Siswanto, W.A.; Salleh, S.M.; Yusup, E.M. Biomechanics Analysis of Combat Sport (Silat) by Using Motion Capture System. *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *166*, 12028. [CrossRef]
69. Petri, K.; Emmermacher, P.; Danneberg, M.; Masik, S.; Eckardt, F.; Weichelt, S.; Bandow, N.; Witte, K. Training using virtual reality improves response behavior in karate kumite. *Sport. Eng.* **2019**, *22*, 2. [CrossRef]
70. Toyama, K.; Krumm, J.; Brumitt, B.; Meyers, B. Wallflower: Principles and practice of background maintenance. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 255–261. [CrossRef]
71. Takala, T.M.; Hirao, Y.; Morikawa, H.; Kawai, T. Martial Arts Training in Virtual Reality with Full-body Tracking and Physically Simulated Opponents. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; p. 858. [CrossRef]
72. Hämäläinen, P.; Ilmonen, T.; Höysniemi, J.; Lindholm, M.; Nykänen, A. Martial arts in artificial reality. In Proceedings of the CHI 2005 Technology, Safety, Community: Conference Proceedings—Conference on Human Factors in Computing Systems, Safety, Portland, OR, USA, 2–7 April 2005; pp. 781–790. [CrossRef]
73. Corbi, A.; Santos, O.C.; Burgos, D. Intelligent Framework for Learning Physics with Aikido (Martial Art) and Registered Sensors. *Sensors* **2019**, *19*, 3681. [CrossRef]
74. Cowie, M.; Dyson, R. A Short History of Karate. 2016, p. 156. Available online: [www.kenkyoha.com](http://www.kenkyoha.com) (accessed on 1 December 2021). [CrossRef]
75. Hariri, S.; Sadeghi, H. Biomechanical Analysis of Mawashi-Geri in Technique in Karate: Review Article. *Int. J. Sport Stud. Heal.* **2018**, 1–8, in press. [CrossRef]
76. Witte, K.; Emmermacher, P.; Langenbeck, N.; Perl, J. Visualized movement patterns and their analysis to classify similarities—demonstrated by the karate kick Mae-geri. *Kinesiology* **2012**, *44*, 155–165.
77. Hachaj, T.; Piekarczyk, M.; Ogiela, M.R. Human actions analysis: Templates generation, matching and visualization applied to motion capture of highly-skilled karate athletes. *Sensors* **2017**, *17*, 2590. [CrossRef]
78. Labintsev, A.; Khasanshin, I.; Balashov, D.; Bocharov, M.; Bublikov, K. Recognition punches in karate using acceleration sensors and convolution neural networks. *IEEE Access* **2021**, *9*, 138106–138119. [CrossRef]
79. Kolykhalova, K.; Camurri, A.; Volpe, G.; Sanguineti, M.; Puppo, E.; Niewiadomski, R. A multimodal dataset for the analysis of movement qualities in karate martial art. In Proceedings of the 2015 7th International Conference on Intelligent Technologies for Interactive Entertainment, INTETAIN 2015, Torino, Italy, 10–12 June 2015; pp. 74–78. [CrossRef]
80. Goethel, M.F.; Ervilha, U.F.; Moreira, P.V.S.; de Paula Silva, V.; Bendillati, A.R.; Cardozo, A.C.; Gonçalves, M. Coordinative intra-segment indicators of karate performance. *Arch. Budo* **2019**, *15*, 203–211.
81. Petri, K.; Bandow, N.; Masik, S.; Witte, K. Improvement of Early Recognition of Attacks in Karate Kumite Due to Training in Virtual Reality. *J. Sport Area* **2019**, *4*, 294–308. [CrossRef]
82. Gupta, V. Pose Detection Comparison: WrnchAI vs OpenPose. 2019. Available online: <https://learnopencv.com/pose-detection-comparison-wrnchai-vs-openpose/> (accessed on 2 December 2021).
83. Eivindsen, J.E. Human Pose Estimation Assisted Fitness Technique Evaluation System. Master's Thesis, NTNU, Norwegian University of Science and Technology, Trondheim, Norway, 2020. Available online: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2777528?locale-attribute=en> (accessed on 2 December 2021).
84. Carissimi, N.; Rota, P.; Beyan, C.; Murino, V. Filling the gaps: Predicting missing joints of human poses using denoising autoencoders. *Lect. Notes Comput. Sci.* **2019**, *11130 LNCS*, 364–379. [CrossRef]
85. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2D Human Pose Estimation: New Benchmark and State of the Art Analysis. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014. Available online: <http://human-pose.mpi-inf.mpg.de/> (accessed on 2 December 2021).
86. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 740–755. [CrossRef]

87. Park, H.J.; Baek, J.W.; Kim, J.H. Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose. In Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), Hong Kong, China, 20–21 August 2020; pp. 1389–1394. [CrossRef]
88. Rosique, F.; Losilla, F.; Navarro, P.J. Applying Vision-Based Pose Estimation in a Telerehabilitation Application. *Appl. Sci.* **2021**, *11*, 9132. [CrossRef]
89. Chen, W.; Jiang, Z.; Guo, H.; Ni, X. Fall Detection Based on Key Points of Human-Skeleton Using OpenPose. *Symmetry* **2020**, *12*, 744. Available online: <https://www.mdpi.com/2073-8994/12/5/744> (accessed on 24 June 2021). [CrossRef]
90. Yunus, A.P.; Shirai, N.C.; Morita, K.; Wakabayashi, T. Human Motion Prediction by 2D Human Pose Estimation using OpenPose. 2020. Available online: <https://easychair.org/publications/preprint/8P4x> (accessed on 2 December 2021).
91. Lin, C.B.; Dong, Z.; Kuan, W.K.; Huang, Y.F. A framework for fall detection based on OpenPose skeleton and LSTM/GRU models. *Appl. Sci.* **2021**, *11*, 329. [CrossRef]
92. Zhou, Q.; Wang, J.; Wu, P.; Qi, Y. Application Development of Dance Pose Recognition Based on Embedded Artificial Intelligence Equipment. *J. Phys. Conf. Ser.* **2021**, *1757*, 012011. [CrossRef]
93. Xing, J.; Zhang, J.; Xue, C. Multi person pose estimation based on improved openpose model. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *768*, 072071. [CrossRef]
94. Bajireanu, R.; Pereira, J.A.R.; Veiga, R.J.M.; Sardo, J.D.P.; Cardoso, P.J.S.; Lam, R.; Rodrigues, J.M.F. Mobile human shape superimposition: An initial approach using OpenPose. *Int. J. Comput.* **2018**, *3*, 1–8. Available online: <http://www.ias.org/iaras/journals/ijc> (accessed on 22 May 2021).
95. Fang, H.S.; Xie, S.; Tai, Y.W.; Lu, C. RMPE: Regional Multi-Person Pose Estimation. In Proceedings of the the IEEE international Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
96. Pishchulin, L.; Insafutdinov, E.; Tang, S.; Andres, B.; Andriluka, M.; Gehler, P.; Schiele, B. DeepCut: Joint subset partition and labeling for multi person pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4929–4937. [CrossRef]
97. Erhan, D.; Szegedy, C.; Toshev, A.; Anguelov, D. Scalable object detection using deep neural networks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–26 June 2014; pp. 2155–2162. [CrossRef]
98. Toshev, A.; Szegedy, C. DeepPose: Human Pose Estimation via Deep Neural Networks. 2014. Available online: [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2014/html/Toshev\\_DeepPose\\_Human\\_Pose\\_2014\\_CVPR\\_paper.html](https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Toshev_DeepPose_Human_Pose_2014_CVPR_paper.html) (accessed on 12 June 2021).
99. Güler, R.A.; Neverova, N.; Kokkinos, I. DensePose: Dense Human Pose Estimation in the Wild. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7297–7306. [CrossRef]
100. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef]
101. Xiao, G.; Lu, W. Joint COCO and Mapillary Workshop at ICCV 2019: COCO Keypoint Detection Challenge Track. 2019. Available online: [http://cocodataset.org/files/keypoints\\_2019\\_reports/ByteDanceHRNet.pdf](http://cocodataset.org/files/keypoints_2019_reports/ByteDanceHRNet.pdf) (accessed on 25 August 2021).
102. Cao, Z.; Simon, T.; Wei, S.-E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [CrossRef]
103. Kendall, A.; Grimes, M.; Cipolla, R. PoseNet: A convolutional network for real-time 6-dof camera relocalization. In Proceedings of the 2015 International Conference on Computer Vision, ICCV 2015, Santiago, Chile, 7–13 December 2015; pp. 2938–2946. [CrossRef]
104. Zhu, L. Computer Vision-Driven Evaluation System for Assisted Decision-Making in Sports Training Lijin. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 1–7. [CrossRef]
105. Berrar, D. Cross-validation. *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.* **2018**, *1–3*, 542–545. [CrossRef]
106. Rodríguez, J.D.; Pérez, A.; Lozano, J.A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 569–575. [CrossRef]
107. Wong, T.T.; Yeh, P.Y. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1586–1594. [CrossRef]
108. Wang, P.; Cai, R.; Yang, S. News video classification using multimodal classifiers and text-biased combination strategies. *Qinghua Daxue Xuebao J. Tsinghua Univ.* **2005**, *45*, 475–478.
109. Yang, J.; Yan, R.; Hauptmann, A.G. Adapting SVM classifiers to data with shifted distributions. In Proceedings of the Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007), Omaha, NE, USA, 28–31 October 2007; pp. 69–74. [CrossRef]
110. Yin, P.; Criminisi, A.; Winn, J.; Essa, I. Tree-based classifiers for bilayer video segmentation. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8. [CrossRef]
111. Sivic, J.; Everingham, M.; Zisserman, A. Who are you?—Learning person specific classifiers from video. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1145–1152. [CrossRef]
112. Mia, M.R.; Mia, M.J.; Majumder, A.; Supriya, S.; Habib, M.T. Computer vision based local fruit recognition. *Int. J. Eng. Adv. Technol.* **2019**, *9*, 2810–2820. [CrossRef]



113. Ponti, M.A.; Ribeiro, L.S.F.; Nazare, T.S.; Bui, T.; Collomosse, J. Everything You Wanted to Know about Deep Learning for Computer Vision but Were Afraid to Ask. In Proceedings of the 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), Niteroi, Brazil, 17–18 October 2017; pp. 17–41. [\[CrossRef\]](#)
114. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep Learning for Computer Vision: A Brief Review. *Comput. Intell. Neurosci.* **2018**, *2018*, 7068349. [\[CrossRef\]](#)
115. Islam, S.M.S.; Rahman, S.; Rahman, M.M.; Dey, E.K.; Shoyaib, M. Application of deep learning to computer vision: A comprehensive study. In Proceedings of the 5th International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, Bangladesh, 13–14 May 2016; pp. 592–597. [\[CrossRef\]](#)
116. Wu, Q.; Liu, Y.; Li, Q.; Jin, S.; Li, F. The application of deep learning in computer vision. In Proceedings of the 2017 Chinese Automation Congress (CAC), Jinan, China, 20–22 October 2017; pp. 6522–6527. [\[CrossRef\]](#)
117. Van Dao, D.; Jaafari, A.; Bayat, M.; Mafi-Gholami, D.; Qi, C.; Moayed, H.; Van Phong, T.; Ly, H.B.; Le, T.T.; Trinh, P.T.; et al. A spatially explicit deep learning neural network model for the prediction of landslide susceptibility. *Catena* **2020**, *188*, 104451. [\[CrossRef\]](#)
118. Grekow, J. Music emotion recognition using recurrent neural networks and pretrained models. *J. Intell. Inf. Syst.* **2021**, 1–16. [\[CrossRef\]](#)
119. Li, T.; Fong, S.; Siu, S.W.; Yang, X.-S.; Liu, L.-S.; Mohammed, S. White learning methodology: A case study of cancer-related disease factors analysis in real-time PACS environment. *Comput. Methods Programs Biomed.* **2020**, *197*, 105724. [\[CrossRef\]](#)
120. Vanam, M.K.; Amirali Jiwani, B.; Swathi, A.; Madhavi, V. High performance machine learning and data science based implementation using Weka. *Mater. Today Proc.* **2021**. [\[CrossRef\]](#)
121. Ramírez, I.; Cuesta-Infante, A.; Schiavi, E.; Pantrigo, J.J. Bayesian capsule networks for 3D human pose estimation from single 2D images. *Neurocomputing* **2020**, *379*, 64–73. [\[CrossRef\]](#)
122. Wang, Y.-K.; Cheng, K.-Y. A Two-Stage Bayesian Network Method for 3D Human Pose Estimation from Monocular Image Sequences. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 16. [\[CrossRef\]](#)
123. Lehrmann, A.M.; Gehler, P.V.; Nowozin, S. A Non-parametric Bayesian Network Prior of Human Pose. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013. [\[CrossRef\]](#)
124. Wang, Y.; Li, J.; Zhang, Y.; Sinnott, R.O. Identifying lameness in horses through deep learning. In Proceedings of the 36th Annual ACM Symposium on Applied Computing, Gwangju, Korea, 22–26 March 2021; pp. 976–985. [\[CrossRef\]](#)
125. Park, J.H.K.Y.-S. A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems. *J. Korea Soc. Comput. Inf.* **2018**, *23*, 9–16. [\[CrossRef\]](#)
126. Elteren, T.; van Zant, T. Real-Time Human Pose and Gesture Recognition for Autonomous Robots Using a Single Structured Light 3D-Scanner. *Intell. Environ. Workshops* **2012**, 213–220. [\[CrossRef\]](#)
127. Szczuko, P. Deep neural networks for human pose estimation from a very low resolution depth image. *Multimed. Tools Appl.* **2019**, *78*, 29357–29377. [\[CrossRef\]](#)
128. Park, S.; Hwang, J.; Kwak, N. 3D human pose estimation using convolutional neural networks with 2D pose information. *Lect. Notes Comput. Sci.* **2016**, *9915 LNCS*, 156–169. [\[CrossRef\]](#)
129. Rahmad, N.A.; Amir As'ari, M.; Ghazali, N.F.; Shahar, N.; Anis, N.; Sufri, J. A Survey of Video Based Action Recognition in Sports. *Indones. J. Electr. Eng. Comput. Sci.* **2018**, *11*, 987–993. [\[CrossRef\]](#)
130. Kanazawa, H. Karate Fighting Techniques: The Complete Kumite. 2013. Available online: <https://www.amazon.es/Karate-Fighting-Techniques-Complete-Kumite/dp/1568365160> (accessed on 30 July 2021).
131. Chaabène, H.; Hachana, Y.; Franchini, E.; Mkaouer, B.; Chamari, T. Physical and Physiological Profile of Elite Karate Athletes. *Sport. Med.* **2012**, *42*, 829–843.
132. Kotthoff, L.; Thornton, C.; Hoos, H.H.; Hutter, F. Leyton-Brown, K. Auto-WEKA: Automatic Model Selection and Hyperparameter Optimization in WEKA. In *Automated Machine Learning*; Springer: Cham, Switzerland, 2019; pp. 81–95. [\[CrossRef\]](#)
133. Deotale, D.; Verma, M.; Suresh, P. Human Activity Recognition in Untrimmed Video using Deep Learning for Sports Domain. *SSRN Electron. J.* **2021**, 596–607. [\[CrossRef\]](#)
134. Zhao, R.; Wang, K.; Su, H.; Ji, Q. Bayesian graph convolution LSTM for skeleton based action recognition. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 6881–6891. [\[CrossRef\]](#)
135. Santos, O.C.; Boticario, J.G. Practical guidelines for designing and evaluating educationally oriented recommendations. *Comput. Educ.* **2015**, *81*, 354–374. [\[CrossRef\]](#)
136. Santos, O.C.; Urias-Rivas, R.; Rodriguez-Sanchez, M.C.; Boticario, J.G. An Open Sensing and Acting Platform for Context-Aware Affective Support in Ambient Intelligent Educational Settings. *IEEE Sens. J.* **2016**, *16*, 3865–3874. [\[CrossRef\]](#)
137. Santos, O.C. Toward personalized vibrotactile support when learning motor skills. *Algorithms* **2017**, *10*, 15. [\[CrossRef\]](#)