Classification of Taekwondo Techniques using Deep Learning Methods: First Insights

Paulo Barbosa[1], Pedro Cunha[1,2][a], Vítor Carvalho[1,2][b] and Filomena Soares[2][c]
[1]2Ai - School of Technology, IPCA, Barcelos, Portugal
[2]Algoritmi Research Centre, University of Minho, Guimarães, Portugal

Keywords: Deep Learning, Human Action Recognition, Neural Networks, Computer Vision, Taekwondo.

Abstract: Research in motion analysis area has enabled the development of affordable and easy to access technological solutions. The study presented aims to identify and quantify the movements performed by a taekwondo athlete during training sessions using deep learning techniques applied to the data collected in real time. For this purpose, several approaches and methodologies were tested along with a dataset previously developed in order to define which one presents the best results. Considering the specificities of the movements, usually fast and mostly with a high incidence on the legs, it was concluded that the best results were obtained with convolution layers models, such as, Convolutional Neural Networks (CNN) plus Long Short-Term Memory (LSTM) and Convolutional Long Short-Term Memory (ConvLSTM) deep learning models, with more than 90% in terms of accuracy validation.

1 INTRODUCTION

The taekwondo martial art emerged in Korea and it was presented as a Korean martial art in Barcelona's Olympic Games (1992) becoming an Olympic sport in the Seoul Olympic Games ("História do Taekwondo | Lutas e Artes Marciais," n.d.).

The evaluation of the performance of the athletes is a difficult task for coaches in any sport. Considering this, over time and with the development of technology, several systems have been developed to assist coaches in evaluating athletes' performance. This development has emphasis on sports with high social impact and financial capacity. In Taekwondo, from the research undertaken, despite the technological development there has been no relevant development of tools to aid the evaluation of the performance of the athletes in training environment.

The use of technology in sport to enhance athlete's performance has already been explored in several modalities. Some systems are being developed that allow accessing to information of the athletes movements, such as velocity, acceleration, applied force, displacement, among other characteristics (Arastey, 2020; Cunha, Carvalho, & Soares, 2019; Nadig & Kumar, 2015).

In Taekwondo, essentially, the athlete's performance evaluation during the training is currently made manually by the coach. He/she usually uses on time visual evaluation or videos of the athletes' training sessions which are time consuming tasks that hinders the quick feedback from the coach to change and adapt the training process (Pinto et al., 2018).

The study presented in this paper is part of a project that aims to develop a real-time prototype to assist the performance of Taekwondo athletes during training sessions. The tool consists in a system composed by a framework that works along with a 3D camera that allows to collect athlete movements' data, providing the velocity, acceleration and applied force of the athlete's hands and feet. The wearable devices holders will be placed at the athlete ankles and wrists using velcro, for a less intrusive fit.

The main outputs of this framework are:

- Statistical analysis, where it will be made the identification and quantification of the movements performed by the athlete during the

[a] https://orcid.org/0000-0001-6170-1626
[b] https://orcid.org/0000-0003-4658-5844
[c] https://orcid.org/0000-0002-4438-6713

Barbosa, P., Cunha, P., Carvalho, V. and Soares, F.
Classification of Taekwondo Techniques using Deep Learning Methods: First Insights.
DOI: 10.5220/0010412402010208
In *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021) - Volume 1: BIODEVICES*, pages 201-208
ISBN: 978-989-758-490-9

athlete's training sessions. This will allow analysing the evolution of the athlete's performance through the data collected during training sessions;

- Biomechanics and motion analysis, to calculate acceleration, velocity and the applied force of the movement (Cunha, 2018).

Thus, this study intends to contribute with a new method of identifying and quantifying the movements performed by the taekwondo athlete during training sessions using deep learning methodologies applied to the data collected from the taekwondo athletes' movements in real time.

This paper is organized into five chapters. The second chapter presents the state of the art; in the third chapter, the methodologies used will be described; in the fourth chapter the preliminary results will be presented; and, in the fifth chapter the final comments are presented. The study presented in this paper is described by the flowchart represented in figure 1.

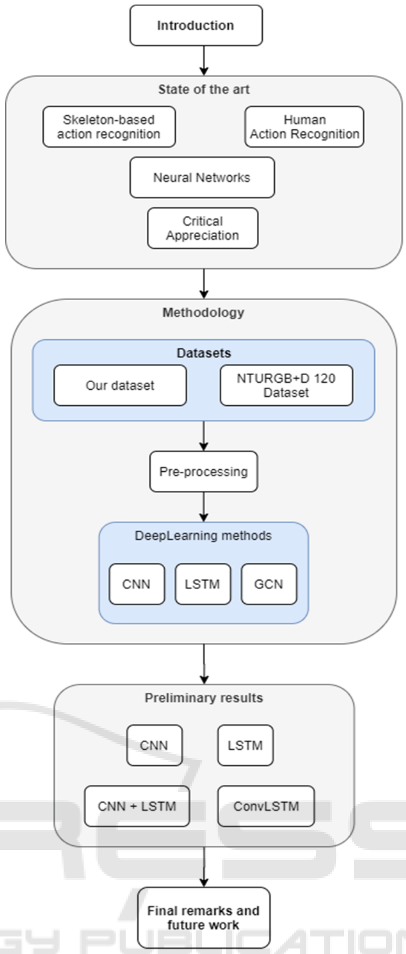

Figure 1: Study Flowchart.

2 STATE OF THE ART

The evaluation of the performance of athletes in sport has been raising the interest of the scientific community through the development and adaptation of new technologies that facilitate this process. According to the literature (Pinto et al., 2018; Zhuang, 2019), there are already systems that analyse the performance of athletes in sports. However, there are few available technological tools for assisting Taekwondo trainings.

In this chapter different approaches used to recognize movements performed by individuals published in different studies will be presented. For that an analysis was made to the references available in the area of the monitorization on movements of the human body, more specifically in the recognition of those same movements, as presented in (P. Wang, Li, Ogunbona, Wan, & Escalera, 2018; Kong & Fu, 2018).

2.1 Human Action Recognition

The Human Action Recognition (HAR) is a process of identifying, analysing and interpreting which kind of actions a person is taking. The studies based on this method allow to better understand how it is possible to categorize movements/actions of the human body.

This has been an area of study that has contributed with a great development in computer vision.

As presented in (H. B. Zhang et al., 2019), in most of the studies carried out, the acquisition of movement data is done through optical sensors, namely depth cameras that facilitates the possibility of extracting human poses that will be then transformed into skeleton data.

The possibility of recognizing human activities through the use of technologies can assist in solving many current problems, such as the identification of violent acts in video surveillance systems, automatic screening of video content by resource extraction, human interaction with the robot, assistance in automatic vehicle task driving, among other challenges (Kong & Fu, 2018).

2.2 Skeleton-based Action Recognition

As a human action recognition specific study area there is the skeleton-based action recognition that uses the data obtained of the human joints localization in a three-dimensional environment to perform motion recognition. Depending on the systems, 20 joints are usually considered to define the human body, as shown in figure 2.

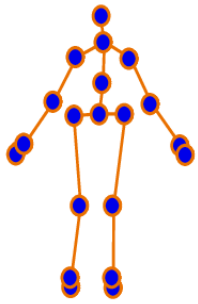

Figure 2: Skeleton of 20 joints(Kacem, Daoudi, Amor, Berretti, & Alvarez-Paiva, 2020).

In order to perform the skeleton-based action recognition, besides the use of the joints data, other information is used to help accomplishing the task. In some cases, motion sensors are used with the purpose of obtaining data, other than the three-dimensional position of the joints, to complement the data collection or to prevent lack of data (eg. caused by occlusions) (Jiang & Yin, 2015) (Y. Zhang, Zhang, Zhang, Bao, & Song, 2018). Other studies used depth sensors to acquire information in addition to the raw information of the Red Green Blue (RGB) sensor and the Infrared (IR) sensor.

In Liu et al. study, different joints may receive different attention from the Long Short-Term Memory (LSTM) neural network. The Attention Mechanism allows the network to be more focused on a specific joint, hands and legs, in this case study (Liu, Wang, Hu, Duan, & Kot, 2017).

2.3 Neural Networks

In the area of pattern categorization most of the studies approach is based on the use of deep learning methods, namely, neural networks. This occurs due to the impressive performance demonstrated on tasks as image classification and object detection (Ren, Liu, Ding, & Liu, 2020).

As well in the area of recognition of human action, recent research tends to present the use of deep learning techniques, specifically neural networks (H. B. Zhang et al., 2019)(Ren et al., 2020). The methods of implementation and the typology vary depending on the objective to be achieved.

Lei Wang and Du Q. Huynh (L. Wang, Huynh, & Koniusz, 2020) show a good comparison between the different deep learning methods applied to action recognition challenges. Of all the presented studies, the most relevant methods were neural networks utilization with different types and architectures such as Convolutional Neural Networks (CNN) networks, LSTM (Ruj, Ryhu, Wlph, & Wudglwlrqdo, 2019)(Zhu et al., 2015) and the most recent Graph ConvNets (GCN) (L. Shi, Zhang, Cheng, & Lu, 2019). Besides those single type architectures there are also other works with hybrid solutions combining different network types (Zhao, Wang, Su, & Ji, 2019) (Sanchez-caballero, Fuentes-jimenez, & Losada-guti, 2020).

As it is possible to undertake from previous studies in the recognition of movements of the human body, there are different approaches depending on the objectives and application areas.

3 METHODOLOGY

As presented in the state-of-the-art chapter, different approaches are used to achieve the identification and interpretation of human motion, each with satisfactory results allowing to find a solution to the research question raised.

Regarding that, in order to find the methodology capable of satisfying the objectives of this study, the different methodologies and approaches used by other studies were tested. To accomplish this task the system presented in figure 3 was designed.

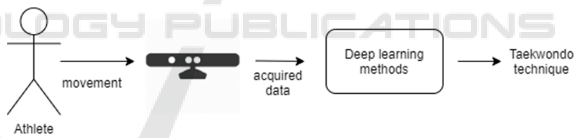

Figure 3: HAR system diagram.

The data acquisition is performed via Orbbec Astra 3D camera and the data processing via deep learning methods to identify and quantify the athlete techniques.

The proposed methodology is discussed in this chapter. First, a brief description of the datasets in use and then the approaches proposed to solve the problem of categorize the type of movement that is been executed by the athlete are presented.

3.1 Taekwondo Movements Dataset

The dataset was developed specifically with data on movements performed by taekwondo athletes. To achieve this task it was used a system previously developed as part of the presented main project (Cunha, Carvalho, & Soares, 2018) which allow to collect the data of the movements according to the positions of the athletes' joints in a three-dimensional environment, more properly the Cartesian coordinates of the several joints. These data acquisition was performed with the support of Braga Sporting Club Taekwondo team and Minho University Taekwondo team athletes.

The obtained dataset consists of eight classes, where each class represents a different technique/movement performed by the athlete.

In figure 4 it is possible to view the raw data along a sequence of 80 samples for the Left Ankle joint movement, in x, y, z coordinates, respectively.

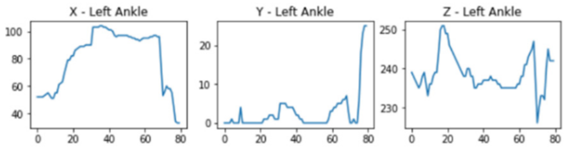

Figure 4: Raw data from Left Ankle Joint.

The main propose of this dataset is to gather information about the taekwondo athletes movements aiming to use on training of deep learning classification methods.

3.2 NTU RGB+D 120 Dataset

Along with the created dataset it was used the NTU RGB+D 120 Dataset, for being a dataset with a higher number of classes (120) of different human actions in everyday life.

The total number of samples in this dataset is 114,480. Each sample is described in 4 different data types, RGB image, depth map, 3D map of the human skeleton and also the IR image (Shahroudy, n.d.). In this study only the information from the 3D skeleton map was considered for training the deep learning models.

3.3 Pre-processing

When using deep learning techniques, the way data it is delivered is vital to achieve better results in recognizing movement patterns. In this way, pre-processing data plays an important role in the entire classification task.

When it comes to sequenced data, the role of pre-processing gains is even more relevant because there are important parameters to define such as the size of the time window as well as the overlap of data in each time window. In this study it was decided to use an 80 points windows size. 80 samples dataset is sufficient to collect all data of two seconds of the movement. As we are dealing with a martial art movement, all movements are performed at a high speed which means that for most athletes, two seconds will be enough time to achieve from the start point to the end of the technique.

3.4 Deep Learning methods

As presented in the state of the art, in order to be able to identify movements it is necessary to use deep learning methodologies. Thus, in this study, some of these methodologies were tested in order to verify which would be the most appropriate according to the available dataset. These methodologies will be presented below, from the convolutional networks to the graphical networks and also the recurrent networks.

3.4.1 Convolutional Neural Networks (CNN)

Since AlexNet won the ImageNet competition in 2012, the use of CNN networks in several deep learning applications has proved to be a success (Ismail Fawaz, Forestier, Weber, Idoumghar, & Muller, 2019). Initially in image classification and object recognition, these good results led these techniques to be also applied to sequential data types. Especially when it comes to video type data or even in 1D raw sequential data (Khan, Sohail, Zahoora, & Qureshi, n.d.). The main idea of convolution layers is to apply convolution throw the image and from that convolution extract features that will be unique or different in each movement class.

Along the with convolutional layers architecture, it is also possible to find studies where image encoding techniques are used in order to be able to transform sequential data into 2D matrix (Yang, Yang, Chen, Lo, & Member, 2019) (Debayle, Hatami, & Gavet, 2018) (Images, 2020) (Dobhal, Shitole, Thomas, & Navada, 2015). The purpose of applying this technique is because CNN are showing good results in pattern recognition in images.

Beyond the image encoding technique as 2D data, Shahroudy [20] has introduced the possibility of motion recognition such as sitting, standing and lying down using an 1D CNN. In this case the CNN layer is applied to a 1D sequential data.

3.4.2 Long Short-term Memory (LSTM)

LSTM is in the category of recurrent neural networks (RNN) developed for data problems referenced to time, such as audio files, GPS path, text recognition, etc. All of these challenges imply that the information of the current moment also takes into account the previous and the later moment(Aditi Mittal, 2019).

For this purpose, recurrent cells were created. Initially with problems such as gradient vanishing, and also the difficulty of contemplating long-term information (old), only considering short-term information (recent), a problem that was solved with the introduction of LSTM cells by Hochreiter & Schmidhuber in 1997 (Olah, 2015). These, as the name implies (Long Short Term) were created with a structure (figure 5), which allows to define the importance of the information transmitted from t-1 thus being able to understand if this same information should be considered in the later cell and if it should be transmitted to t + 1.

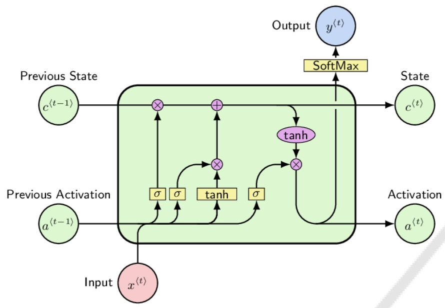

Figure 5: LSTM cell(“LSTM cell Image,” n.d.).

Referring to studies about human activity recognition, Ruj et al. in (Ruj et al., 2019) shown that with this network model it is possible to obtain good results for the classification of human activities such as jogging, sitting, standing, etc. In this work, the authors used the WIDSM dataset with data collected through a simple sensor that allows to identify human movement in x, y and z coordinates.

In Liu et al. the study performed shows good results in classifying activity on data types based in skeleton data [10]. The main focus of the authors was to create the model “Global Context-Aware Attention LSTM”. This model has the particularity of allowing to attribute greater relevance to certain joints of the skeleton. This contributes with a useful functionality to the model because not all joints bring useful information, many of them can even introduce noise in the model. With this proposed strategy, the authors managed to achieve better results in relation to the simpler LSTM model.

3.4.3 Graph Convolution Networks (GCN)

In the real world there are data easily represented by graphs, such as a molecular structure, social networks, in this way, the Euclidean representation of data has been called into question.

The GCN (Graph Convolution Networks) just as CNN allows to apply convolutional layers in order to extract characteristics from the data. These data when represented in graphs are described as nodes and edges, so instead of applying a convolution in a 2D matrix, it is applied in graphs.

The graph is represented by three matrices, Adjacent A, Degree D and Feature X. Thus, the problem data needs to be initially modelled for a graph representation so that later they can be submitted to the neural network model (figure 6).

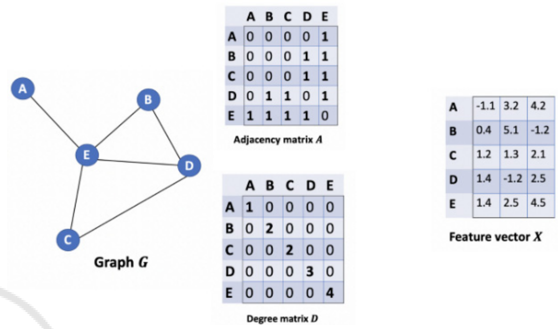

Figure 6: Graph Struct, Adjacency matrix A, Degree matrix D and Feature vector X (“Graph struct,” n.d.).

According to this methodology, Shi et al. (L. Shi et al., 2019) presented one of the first neural network model solutions with graphs for the problem of recognition of human activities. In their study, the authors used the NTU RGB + D 120 Dataset as the dataset, the same one used in this work (detailed in section 3.2). The results of the study allowed them to realize that with the used dataset, and in comparison to other methods, like LSTM, the results obtained were better (+20.8% on cross-view and +21% on cross-subject validation accuracy).

3.4.4 LSTM Hybrid Model

This model has an architecture that implies the use of a first layer of a Convolution Neural Network (CNN). This allows to extract spatial features from the data that will be used as input for the combination with the LSTM layers, where the temporal features will be extracted and allowing the classification of the sequence.

In the same study context, another relevant model should also be mentioned, namely, the ConvLSTM architecture, that consists on a LSTM with embedded Convolution, which allows a good spatial temporal correlation of features (Sanchez-caballero et al., 2020)

4 PRELIMINARY RESULTS

According with the recent studies presented in human action recognition and considering the several approaches used for this study the main methodologies were tested and applied to this study objective.

Thus, the results presented in this chapter were obtained through the development of data processing algorithms and the training of neural network algorithms in Python programming language. The entire algorithm was developed using frameworks such as keras, pandas, matplotlib, numpy, among others. The application of the different approaches on the data of this study and the results obtained will be presented below.

4.1 CNN Approach

This approach required more pre-processing than the others because encoding imaging techniques were applied. The sequence data was transformed into images so that these images could later be submitted to the CNN network (figure 7). There are several techniques to enable this transformation (Z. Wang & Oates, 2015). The technique used in our approach was the Gramian Angular Field (GAF) transformation, basically a polar representation of the time series data that will then be transformed into a 2D image.

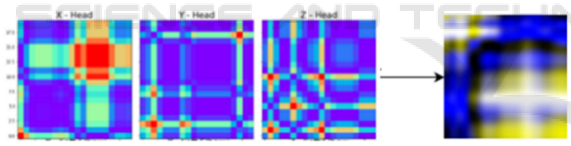

Figure 7: Matricial representation from joint head 1D raw data.

These images will then be the input to our classification algorithm. This algorithm consists of a set of convolutional layers; the images of the different joints will cross these same convolutional layers allowing to extract features that will later be concatenated in order to group the information extracted from the different joints. The set of convolutional layers used is known as VGG16.

This model was proposed by K. Simonyan and A. Zisserman (Simonyan & Zisserman, 2015) with very good results (top-1 and top-5 validation error, 23.7% and 6,8%) in the ImageNet dataset. Thus, it was decided to apply this same model as shown in figure 8.

With this technique it was possible to achieve a classification validation accuracy of 80% for two movement categories, jump up and throw.

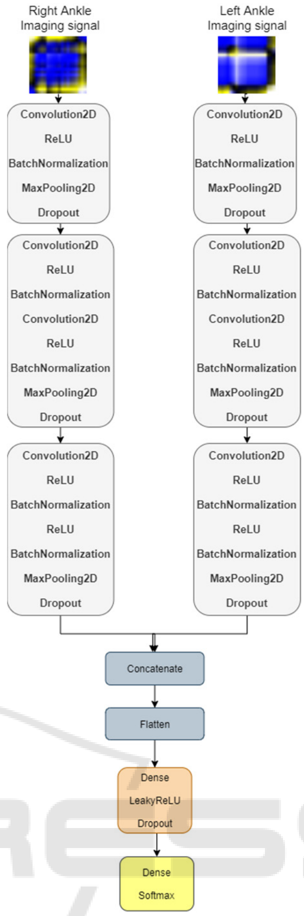

Figure 8: Whole system diagram of the classification algorithm using CNN VGG16.

4.2 LSTM Approach

Based on the good results obtained in other studies in data sequences, it was decided to start by implementing a simple LSTM. Each data sequence was composed of 80 samples with information from 25 joints, resulting in an entry with dimension [80, 25, 3]. In order to simplify the dimensionality of the data, a resize was done where a dimension was removed, it was then [80, 25x3], resulting in a data matrix with dimensions [80, 75]. With this simple LSTM model approach, it was possible to achieve a validation accuracy of 88%.

4.3 CNN+LSTM Approach

Based in the CNN + LSTM Hydro model presented in (X. Shi, Chen, & Wang, 2015) it was possible to achieve a validation accuracy of 93%, allowing to obtain a better result compared to the simple LSTM model.

It was intended to start applying a first convolutional layer of 1D over the data sequence, as shown in figure 9. The result of this convolution will thus be the input of the LSTM layers.

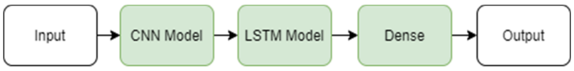

Figure 9: CNN+LSTM Network Architecture.

4.4 ConvLSTM Approach

The ConvLSTM model has a very similar architecture relatively to the model previously presented (CNN+LSTM). This model is composed of a layer embedded in the LSTM cell itself.

As expected, due to similarity with CNN+LSTM, applying the same data the results obtained were very close. Thus, with this model architecture, it was achieved a 92% validation accuracy as result.

4.5 Training Computation

For the study of the methods presented above, they were processed through Anaconda Spyder IDE running on CPU Intel I5 8th Gen with 16GB of RAM.

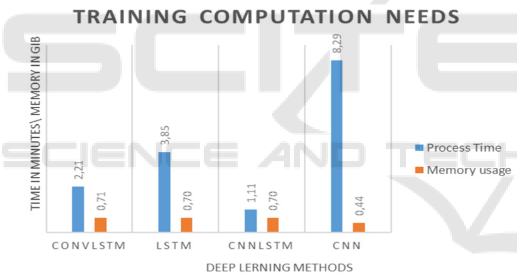

Figure 10: Comparing chart of computation needs during methods training.

Regarding computation needs and as described by figure 10, CNN method it is by far the method that requires the most training time, in addition to its complex pre-processing due to the task of image encoding. The best performance accuracy method, CNN+LSTM it also the one that requires less training time and less required memory.

5 FINAL REMARKS

The study presented in this paper is part of a project that aims to evaluate the performance of taekwondo athletes in real time. Moreover, it aims to test the different approaches used in previous studies in order to identify the adequate methodology using deep learning in the identification of the taekwondo athletes' movements.

The methodologies used by the different approaches presented in the area of HAR applied to our dataset, made it possible to realize that for the purpose of identifying the movements of taekwondo athletes. The best result was obtained with convolution layers models. This fact may be related to the spatial features of the data used in this study. Both methods, CNN+LSTM and ConvLSTM, managed to get results above 90% on accuracy validation. On the other hand, the results obtained by the LSTM method, where the spatial features are not considered, were inferior.

As future work, it is intended to continue the study on deep learning methods to Human Action Recognition, such as CGN methods. Along with that, data collection on the movements of Taekwondo will continue to be carried out, with the aim of increasing the dataset. Finally, we will include the algorithms developed with the study carried out in the framework developed in the global project, which will allow the identification and accounting of athletes' movements in real time.

ACKNOWLEDGEMENTS

This work has been supported by COMPETE: POCI-01-0145-FEDER-007043 and by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. Pedro Cunha thanks FCT for the PhD scholarship SFRH/BD/121994/2016.

REFERENCES

Aditi Mittal. (2019). Understanding RNN and LSTM. Retrieved October 31, 2020, from https://towardsdatascience.com/understanding-rnn-and-lstm-f7cdf6dfc14e

Arastey, G. M. (2020). Computer Vision In Sport. Retrieved from https://www.sportperformanceanalysis.com/article/computer-vision-in-sport

Cunha, P. (2018). Development of a Real-Time Evaluation System For Top Taekwondo Athletes, (c), 140–145.

Cunha, P., Carvalho, V., & Soares, F. (2019). Real-time data movements acquisition of taekwondo athletes: First insights. Lecture Notes in Electrical Engineering, 505, 251–258.

Debayle, J., Hatami, N., & Gavet, Y. (2018). Classification of time-series images using deep convolutional neural networks, (October), 23. https://doi.org/10.1117/12.2309486

Dobhal, T., Shitole, V., Thomas, G., & Navada, G. (2015). Human Activity Recognition using Binary Motion Image and Deep Learning. Procedia Computer Science, 58, 178–185.

Graph struct. (n.d.). Retrieved from https://www.topbots.com/graph-convolutional-networks/

História do Taekwondo | Lutas e Artes Marciais. (n.d.). Retrieved November 11, 2019, from https://lutasartesmarciais.com/artigos/historia-taekwondo

Images, T. C. (2020). Sensor Classification Using Convolutional Neural, (1).

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. Data Mining and Knowledge Discovery, 33(4), 917–963.

Jiang, W., & Yin, Z. (2015). Human activity recognition using wearable sensors by deep convolutional neural networks. MM 2015 - Proceedings of the 2015 ACM Multimedia Conference, 1307–1310.

Kacem, A., Daoudi, M., Amor, B. Ben, Berretti, S., & Alvarez-Paiva, J. C. (2020). A Novel Geometric Framework on Gram Matrix Trajectories for Human Behavior Understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(1), 1–14.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (n.d.). A Survey of the Recent Architectures of Deep Convolutional Neural Networks 1 Introduction, 1–70.

Kong, Y., & Fu, Y. (2018). Human Action Recognition and Prediction: A Survey. ArXiv, 13(9).

Liu, J., Wang, G., Hu, P., Duan, L. Y., & Kot, A. C. (2017). Global context-aware attention LSTM networks for 3D action recognition. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua, 3671–3680.

LSTM cell Image. (n.d.). Retrieved from https://www.researchgate.net/profile/Juan_Victores/publication/334360853/figure/fig1/AS:778955447599106@1562728859405/The-LSTM-cell-internals.png

Nadig, M. S., & Kumar, S. N. (2015). Measurement of Velocity and Acceleration of Human Movement for Analysis of Body Dynamics. International Journal of Advanced Research in Computer Science & Technology (IJARCST 2015), 3(3), 2013–2016.

Olah, C. (2015). Understanding LSTM Networks. Retrieved October 31, 2020, from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Pinto, T., Faria, E., Cunha, P., Soares, F., Carvalho, V., & Carvalho, H. (2018). Recording of occurrences through image processing in Taekwondo training: First insights. Lecture Notes in Computational Vision and Biomechanics, 27, 427–436.

Ren, B., Liu, M., Ding, R., & Liu, H. (2020). A survey on 3d skeleton-based action recognition using learning method. ArXiv, 1–8.

Ruj, D. P. L., Ryhu, L., Wlph, O., & Wudglwlrqdo, Q. R. Z. (2019). Human Activity Recognition Using LSTM-RNN Deep Neural Network Architecture.

Sanchez-caballero, A., Fuentes-jimenez, D., & Losada-guti, C. (2020). Exploiting the ConvLSTM : Human Action Recognition using Raw Depth Video-Based Recurrent Neural Networks, 1–29.

Shahroudy, A. (n.d.). NTU RGB + D : A Large Scale Dataset for 3D Human Activity Analysis, 1010–1019.

Shi, L., Zhang, Y., Cheng, J., & Lu, H. (2019). Two-stream adaptive graph convolutional networks for skeleton-based action recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2019-June, 12018–12027.

Shi, X., Chen, Z., & Wang, H. (2015). Convolutional LSTM Network : A Machine Learning Approach for Precipitation Nowcasting, 1–11.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks For Large-Scale Image Recognition, 1–14.

Wang, L., Huynh, D. Q., & Koniusz, P. (2020). A Comparative Review of Recent Kinect-Based Action Recognition Algorithms. IEEE Transactions on Image Processing, 29, 15–28.

Wang, P., Li, W., Ogunbona, P., Wan, J., & Escalera, S. (2018). RGB-D-based human motion recognition with deep learning: A survey. Computer Vision and Image Understanding, 171(April), 118–139.

Wang, Z., & Oates, T. (2015). Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. AAAI Workshop - Technical Report, WS-15-14(January), 40–46.

Yang, C., Yang, C., Chen, Z., Lo, N., & Member, S. (2019). Multivariate Time Series Data Transformation for Convolutional Neural Network, 188–192.

Zhang, H. B., Zhang, Y. X., Zhong, B., Lei, Q., Yang, L., Du, J. X., & Chen, D. S. (2019). A comprehensive survey of vision-based human action recognition methods. Sensors (Switzerland), 19(5), 1–20. https://doi.org/10.3390/s19051005

Zhang, Y., Zhang, Y., Zhang, Z., Bao, J., & Song, Y. (2018). Human activity recognition based on time series analysis using U-Net. Retrieved from http://arxiv.org/abs/1809.08113

Zhao, R., Wang, K., Su, H., & Ji, Q. (2019). Bayesian graph convolution LSTM for skeleton based action recognition. Proceedings of the IEEE International Conference on Computer Vision, 6881–6891.

Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., & Xie, X. (2015). Co-occurrence Feature Learning for Skeleton based Action Recognition using Regularized Deep LSTM Networks, (i).

Zhuang, Z., & Xue, Y. (2019). Sport-Related Human Activity Detection and Recognition Using a Smartwatch. Sensors, 19(22), 5001.