

Application of Kinect Technology and Artificial Neural Networks in the Control of Rehabilitation Therapies in People with Knee Injuries

Bisset Gonzales Loayza¹, Alberto Calla Bendita², Mario Huaypuna Cjuno³, Jose Sulla-Torres⁴
Escuela de Ingeniería de Sistemas, Universidad Nacional de San Agustín de Arequipa
Arequipa, Perú

Abstract—In the field of physiotherapy, the recognition of the poses of the human body is obtaining more research so that the patient has an accelerated recovery rate in his rehabilitation. Nowadays, it is not so challenging to have devices like Microsoft Kinect that allow us to interact with the user for the recognition of poses and body gestures. The objective of this work to capture the data of the joints of a person's body through a set of angles using the Kinect device, then artificial neural networks with the Back-Propagation algorithm were used for machine learning, and their precision was determined. The results found on the performance of the neural network show that 99.70% accuracy was achieved in the classification of the patients' postures, which can be used as an alternative in the rehabilitation therapies of patients with knee injuries.

Keywords—Machine learning; artificial neural network; kinect; physiotherapy; rehabilitation

I. INTRODUCTION

The most common injuries in a person's early stage are bruises and fractures, not to mention that they can present pain from different anatomical structures, such as osteochondritis of the knee (Osgood Schlatter disease) [1]. Knee injuries are common in children, and in people who engage in some form of sports activity or overuse their knee, most studies report traumatic knee injuries [2].

According to the National Institute of Statistics and Informatics (INEI), in Peru, 5.2% of the population (1 million 619 thousand people) have some type of disability, "88.6% of the population with a disability did not receive treatment or therapy for rehabilitation and only 11.4% if they received any treatment or therapy. Among those who received treatment or rehabilitation therapy we can mention physical rehabilitation therapies (46.1%), psychological treatment (18.9%), psychiatric treatment (11.3%), speech therapy (11.0 %), emotional support (3.8%), occupational therapy (3.6%), another type (5.4%)".

The rehabilitation process is commonly carried out in a rehabilitation center, in which two main problems arise: a) in the rehabilitation environment, the patient is motivated by the therapist's guidance, while at home he is completely unmotivated, and b) some patients may perform incorrect exercises by compensating for limited movement with other stereotypes. Knee injuries, which are the most common disabling injuries in athletic and physically active individuals,

can be treated with the use of knee braces and bandaging techniques that are widely used to reduce and / or prevent the severity and incidence of injuries of knee as well as the use of kinesiotaping to improve muscle strength and jump performance in the knee [3]. This is one of the main problems that patients can get bored from repeated rehabilitation activities or cannot properly perform the exercises. This can negatively influence patients to enable them to undergo rehabilitation.

On the other hand, there are several studies where the Kinect tool is used to apply its benefit to different forms of rehabilitation [4] such as the proposal to improve specific rehabilitation such as flexion, abduction and extension of the limbs through digital applications [5], in conjunction with artificial neural networks (ANN) as new approaches for the medical evaluation of injured patients [6].

In this context, the proposal is to carry out a prototype system for monitoring/training physical physiotherapy of patients with knee injuries. This system is based on artificial neural networks, where the input data corresponds to the detection of the joints that belong to the knee; for this, the Microsoft Kinect v1.0 tool is used, which provides exact measurements of this data.

The rest of this document is organized as follows: Section II, review of related works on movement recognition with Kinect; Section III, a summary of methods, data set, neural network model; Section IV presents experimental results on the recognition rate of human movements and the effectiveness in monitoring the patient, and the discussion regarding other works. Finally, we conclude the results in Section V.

II. RELATED WORK

The recognition of human movements has been studied for several years [7], and the techniques for machines to learn to recognize them have been evolving. It is where the concept of Machine Learning appears, which, through its techniques, have helped improve the performance and precision of machines to recognize specific human activities [8] [9].

In the study by Da Gama et al. [10], a review of interactive systems is made to help patients in various therapies. Among these technologically advanced systems is Microsoft's Kinect, which has helped lead the way in how user interaction

C. Description of the Physiotherapeutic Monitoring System

Our proposal follows a three-layer architecture. As seen in Fig. 2, the left part of the system interacts with the physiotherapist through a web application that can be accessed from a PC or smartphone; this provides maintainability and scalability; while on the other hand, the patient interacts with an application based on Windows Presentation Foundation (WPF) of Microsoft.NET to create Windows applications of secure development. This application is connected to a Microsoft Kinect v1.0 sensor. Communication between the application server and web client is done through HTTPS, which is recommended for security, since the application treats sensitive private data of people, such as the health of a patient.

Table I shows the data dictionary used in the system.

Both the physiotherapist and the patient must validate their user to enter the system; in the case of the patient, this is very important since their account also maintains information (stored in a database) about their progress with rehabilitation exercises. The physical therapist reviews such progress. The WPF application is the one that implements the Back-Propagation algorithm in order to recognize the patient's movements and detect if the patient performs the proposed exercise correctly or incorrectly.

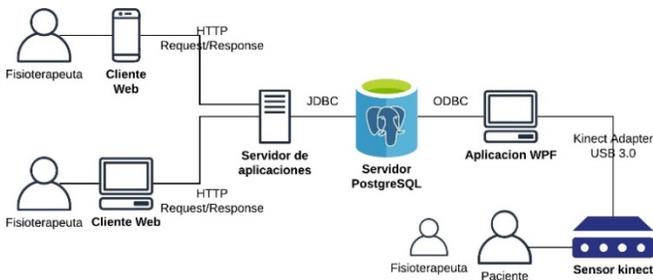


Fig. 2. System's Architecture Proposed.

TABLE I. DESCRIPTION OF DATA USED IN ARCHITECTURE

Entity	Attribute	Type	Length
Patient	Name	varchar	30
Patient	Mother's last name	varchar	30
Patient	Last name	varchar	30
Patient	DNI	Integer	8
Patient	Age	Integer	2
Physiotherapist	Name	varchar	30
Physiotherapist	Mother's last name	varchar	30
Physiotherapist	Last name	varchar	30
Physiotherapist	DNI	Integer	8
Therapy	Code	Integer	4
Therapy	Names	varchar	30
Therapy	Date	Date	-
Therapy	Hour	Datetime	-
Therapy	Punctuation	Float	5, 3
Position	Code	Integer	4
Position	Name	varchar	30
Position	Value	Float	

D. Artificial Neural Networks

The method used to classify human postures involves using an artificial neural network with dense layers. Entries for ANN are expressed by feature vectors associated with human poses captured by Kinect.

Likewise, the output of ANN data is expressed by values 0 and 1, which corresponds to an incorrect and correct posture, respectively. The output is calculated by an activation function $f(x)$. ANN training was performed using the Back-Propagation algorithm, as shown in Eq. 1.

$$\text{sigmoid function: } f(x) = \frac{1}{1+e^{-ax}} \quad (1)$$

Where: a is a parameter of the model.

On the other hand, for the development part of the neural network, it has been decided to use AForge.NET [19]. It has a simple API for the construction of neural networks with different algorithms to train, such as BackPropagation, DeltaRule, ElasticNetwork, among others. AForge.NET is used as a tool to develop the motion detection module, and it is comprised of a set of libraries that are useful for neural networks. Therefore, C# was used for development, and it is easy to integrate into AForge.NET, the tool selected for the motion detection module [20].

Fig. 3 shows the main components of the recognition system.

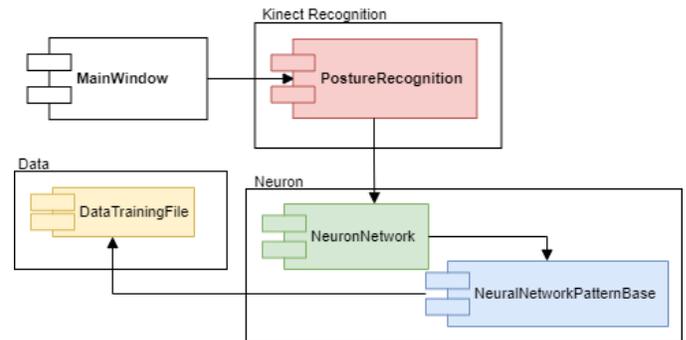


Fig. 3. Main Components of the Recognition System for this Work.

- **MainWindows:** Represents the WPF application that instantiates components of the Kinect SDK. Displays a frame that draws the skeleton in interconnected green segments. Also, here the posture recording processes are executed, managed by concurrent flags and threads. Kinect v1 provides 30 FPS, and since each frame represents a SkeletonFrame (containing the required Skeleton object), it can achieve 30 Skeleton objects per second.
- **PostureRecognition:** Instances a NeuronNetwork object, in addition to having the main training and predict methods, to be used directly by a user.
- **NeuronNetwork:** Uses the ActivationNetwork class, which is the neural network itself, and creates an instance of BackPropagationLearning (both AForge.NET classes) that executes the neural network training.

- **NeuralNetworkPatternBase:** Provides how data is processed before entering the neural network. Our approach is based on the use of angles as in [21], that is, we take the Skeleton objects stored somewhere, then we access the Joins properties of each object, and with these data, we calculate these angles, which serve as inputs to the neural network.

E. Data Collection and Exercises

The data set that we use for the training and evaluation of the model was obtained from the postures performed by the patients with the supervision of a specialist. These data were captured through a Microsoft Kinect device, where the time of each session lasts between 7 and 15 seconds, where the correct posture is first recorded, and then the patient performs the postures to determine if they are doing it correctly.

In Fig. 4, the points recognized by the Kinect device are shown, which correspond to the joints of the human body that consider 20 points of the skeleton. The skeletal tracking system of the Kinect provides for each posture a set of twenty joints expressed by coordinates (x, y, z), where these coordinates are evaluated as vectors for every three joints that give the angles that serve for the predictions of the postures through the supervised neural network. We focus only on the knee joint; it also takes some arm movements for the balance of the patient, to describe the specific physiotherapy postures [22] [23] [24].



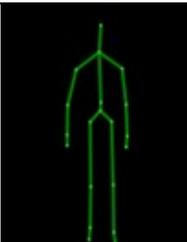
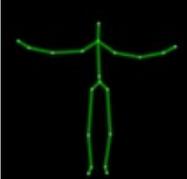
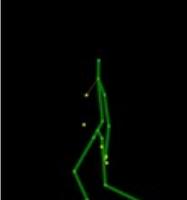
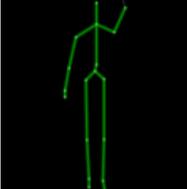
Fig. 4. Capture by Kinect of the Joint of the Second Posture of the Patient.

The coordinates x_i, y_i, z_i ($0 < i < 19$) depend on the size and position of the patient in the scene. The approach used in this work is that with the angles formed between three points through geometry, the angles of the joints are obtained that will give us a better perception of the results, Those are the entries for the network after doing the training is saved in a file in .td format, inside a folder called Training-3, then the trained network is saved in a file for later use.

Table II shows the four postures that are considered for knee therapy. The postures that the child must simulate for his recovery and that the program will tell him if I do the correct posture. The first posture is Standing since the knees must be very straight, posture 2 is Frog here the knee bends an angle of

more than 90 degrees, posture 3 is Lateral Frog here you can see from another view the angles that the knees form if they are uniform and posture 4 in Right knee raised were to raise the bent right knee with the support of the arm lift for the balance of the child. Then the patient will stand in front of the Kinect, and do they pose so that the system displays the name of the pose only if it was done correctly, as shown in Fig. 5.

TABLE II. BODY POSTURE

Body posture	Name
	Standing
	Frog
	Lateral Frog
	Right knee raised

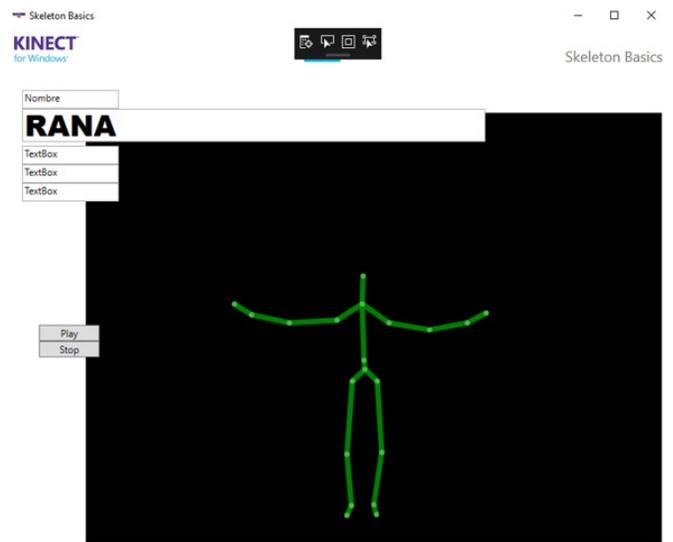


Fig. 5. Recognition of Correct Patient Posture using Kinect.

F. Pre-Processing of Skeleton Data

To improve and optimize the training process is necessary to have the normalized data, we focus on having a representation of the human body independent of the position of the sensor, the distance of the body and the sensor and the structure of the patient's body.

Since the skeleton is represented in three-dimensional space, the first feature extracted is the position. As each joint has three values of three coordinates (x, y, z), and a skeleton consists of 20 joints. However, we only use 14 joints; therefore, our data will have a dimension of 32. This can be seen in Fig. 5, where each join (A, B, C ..., T) that is relevant, is labeled for the capture of the different postures except for the joins L, P, CT, and R.

On the other hand, in the pre-processing, it is necessary that, from the 14 joints, the ten angles are obtained for the characteristic vector. These angles are generated from the union of 2 adjacent vectors, for example, the vectors (\vec{HS}, \vec{Htt}) , (\vec{FQ}, \vec{FW}) they represent angles that are obtained from the GHS and EFQ joins respectively. Therefore, the feature vector consists of 10 angles, as shown in Eq. 2:

$$f = (\theta_1, \theta_2, \dots, \theta_{10}) \in \mathbb{R}^{10} \tag{2}$$

Fig. 6 shows the Joint captured by the Kinect.

G. Model Training

For training, a person executes the postures defined for rehabilitation [9]. This allows 90 samples to be recorded for each pose. In the end, it has been possible to collect (21 poses x 90 samples) 1890 samples distributed among 11 different positions, of which four were used, which are shown in Table II, for more details of how the data of a sample are, before processing it is shown in Fig. 7.

These samples are divided into two random sets: m_{train} for training and m_{test} to calculate the precision of the model.

The method used to classify human postures involves using a dense, layered artificial neural network. The choice of the ANN was made after making a comparison of the precision between different techniques in recognition of postures. Table III shows the comparison of the precision of techniques in recognition of postures.

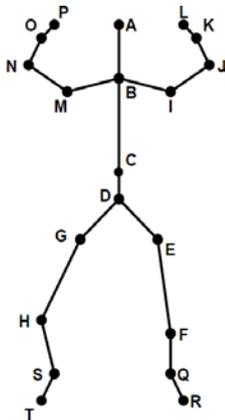


Fig. 6. Joint Captured by the Kinect.

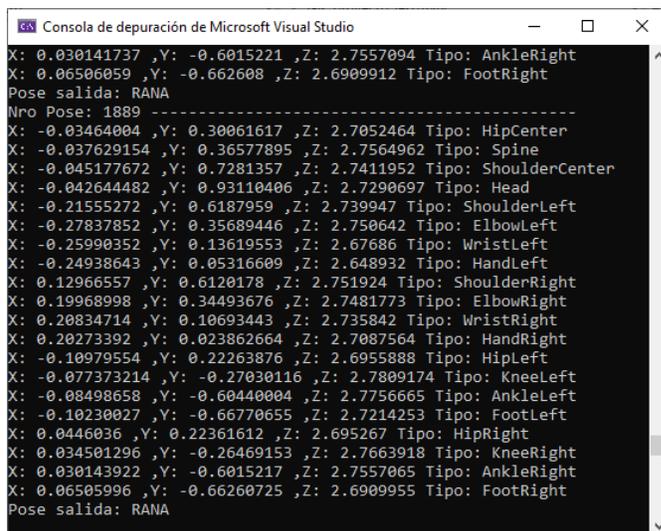


Fig. 7. Sample of the Coordinates of a Frog Posture.

TABLE III. COMPARISON OF THE PRECISION OF TECHNIQUES IN RECOGNITION OF POSTURES

Technique	Accuracy
BLSTM	70.72%
Random Forest (RF)	85.17%
SVM	83.05%
ANN	92.00%

Table III shows that for the BLSTM (bidirectional short-term memory neural network) [25] where the 3D skeleton is detected for each user in the study group, the precision achieved is 70.72%. On the other hand, RF obtained an average general precision of 85.17%, surpassing SVM, which obtained 83.05%, also mentioning that SVM has performance advantages in training time, and RF is more suitable for classifying activities in real-time. However, the use of ANN in the creation of a platform for the recognition of human poses and gestures in order to help patients suffering from stroke disease, an accuracy of 92.00% was obtained [8]. Therefore, we see that ANN's offer excellent precision when dealing with body postures.

Entries for ANN are expressed by feature vectors associated with human poses captured by Kinect. The ANN also has a configuration adapted to the number of input values and the number of output classes. Here are more details:

- the neural network is trained together with the Back-Propagation algorithm.
- the α parameter of the neural network is configured with an initial value of 1.0.
- the network receives inputs consisting of 10 real numbers (angles from 0 to 180 degrees),
- the network has a hidden layer of 20 neurons.
- the network output is made up of 11 neurons that correspond to the total number of positions defined to train the model, and

- the number of iterations for the training, when dealing with angles between 0 to 180 degrees, up to 100,000 iterations had to be used to achieve reasonable precision rates.

Fig. 8 shows the Neural network diagram.

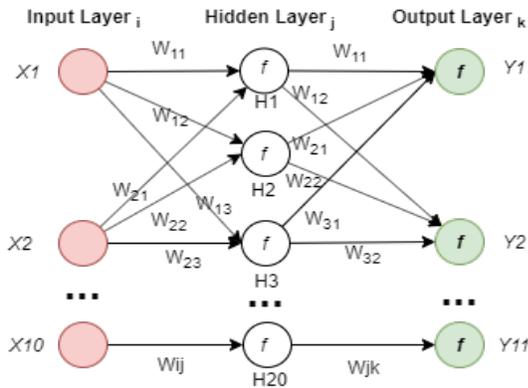


Fig. 8. Neural Network Diagram.

In the testing stage, some postures made by a patient are recorded, each sample is processed, and ten angles that form a vector of characteristics are extracted, and then this vector is entered into the neural network model and the predicted class and its associated precision.

IV. RESULTS AND DISCUSSION

The tests were available, with hundreds of samples distributed among the four positions considered. The results are displayed on the console, following the form: precision and classified posture. Fig. 8 shows the results that were obtained from the tests carried out with the stored objects.

The first seven samples were extracted from each test posture. The prediction results are shown in Table III. For reasons of space, each posture has been identified by a letter:

- A - Right knee raised
- B - Stopped
- C - Frog (sideways)
- D - Frog (from the front)

Fig. 9 shows the sample of tests performed on saved objects.

Table IV shows that the samples reach a significantly high precision (the highest of 99.70% and the lowest of 98.22%, both averages), but a detail to consider is that the most challenging positions to execute present a slight variation between the training data and are therefore less likely to achieve the best accuracy.

On the other hand, in the neural network, we used a Sigmoid activation function, which gave good results for a 2-layer model; also the initial parameter α of 1.0 was not so decisive, this is derived from the results seen by Choubik and Mahmoudi [9] when using a network with a number of different layers (from 1 to 4), where the parameter α did not make the difference but, using another activation function.

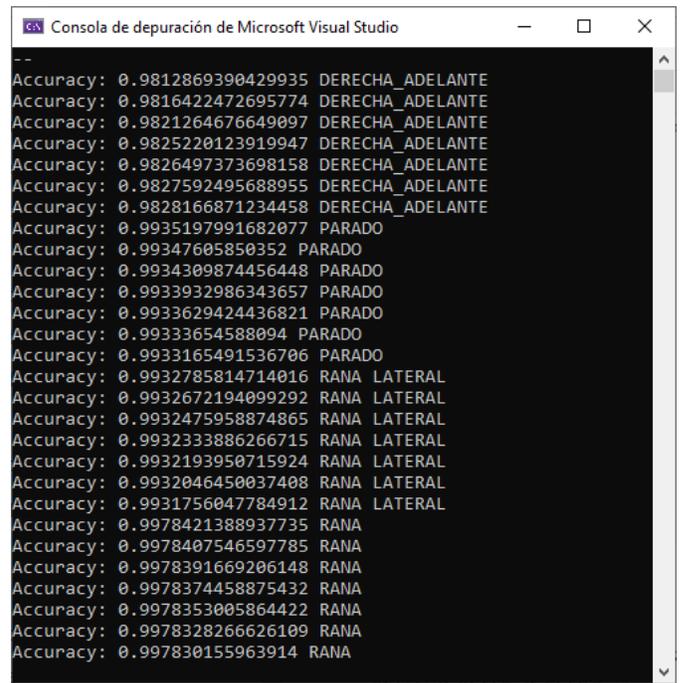


Fig. 9. Sample of Tests Performed on Saved Objects.

TABLE IV. RESULTS OF THE RECOGNITION OF 4 POSITIONS FOR 28 SAMPLES (7 FOR EACH POSTURE)

Sample / Posture	A	B	C	D
Sample 1	0.9813	0.9935	0.9933	0.9970
Sample 2	0.9816	0.9935	0.9933	0.9970
Sample 3	0.9821	0.9934	0.9932	0.9970
Sample 4	0.9825	0.9934	0.9932	0.9970
Sample 5	0.9826	0.9934	0.9932	0.9970
Sample 6	0.9828	0.9933	0.9932	0.9970
Sample 7	0.9828	0.9933	0.9932	0.9970
Average accuracy %	0.9822	0.9934	0.9932	0.9970

However, Choubik and Mahmoudi also show other clustering-based learning techniques such as Support Vector Machines (SVM) or k-nearest neighbors (KNN), in the case of SVM, accuracies of up to 100% were obtained using a linear kernel function, while with KNN accuracies were also obtained of 100%, although SVM showed higher precision rates as the amount of data increased more. Le and Nguyen [21] used SVM in the same way, obtaining a precision of 100%, testing with different combinations of angles formed by the joints of the body.

V. CONCLUSIONS

In this work, we proposed one more way to take the data provided by Kinect to convert it into an entry that represents as much information as possible about the body joints, all in order to recognize postures. The entrances are represented by ten angles made up of different groups of three-body joints.

The developed ontology allowed to broadly identify the most outstanding classes or entities of the application, as well as their relationships with other parts of the system.

The neural network trained with the Back-propagation algorithm allowed to give good predictions of the postures based on the angles of the joints, and the more data have, the more significant the recognition precision.

Despite having obtained high classification precision such as 99.70%, there is still room for improvement, considering that Clustering techniques are usually better with data of many dimensions, we may consider extending this work in the future using SVM as learning.

REFERENCES

- [1] R. S. Valtanen, A. Arshi, B. V. Kelley, P. D. Fabricant, and K. J. Jones, "Articular Cartilage Repair of the Pediatric and Adolescent Knee with Regard to Minimal Clinically Important Difference: A Systematic Review," *Cartilage*, 2020.
- [2] T. Junge, L. Runge, B. Juul-Kristensen, and N. Wedderkopp, "Risk factors for knee injuries in children 8 to 15 years: The CHAMPS study DK," *Med. Sci. Sports Exerc.*, 2016.
- [3] G. Aktas and G. Baltaci, "Does kinesiotaping increase knee muscles strength and functional performance?," *Isokinet. Exerc. Sci.*, 2011.
- [4] C. Sinpithakkul, W. Kusakunniran, S. Bovonsunthonchai, and P. Wattananon, "Game-based Enhancement for Rehabilitation Based on Action Recognition Using Kinect," in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2019.
- [5] G. F. Muñoz, J. G. Casero, and R. A. M. Cárdenas, "Usability study of a Kinect-based rehabilitation tool for the upper limbs," *Adv. Intell. Syst. Comput.*, vol. 931, pp. 755–763, 2019.
- [6] A. Vitali, D. Regazzoni, C. Rizzi, and F. Maffioletti, "A new approach for medical assessment of patient's injured shoulder," in *Proceedings of the ASME Design Engineering Technical Conference*, 2019.
- [7] R. J. Nogales and M. Benalcázar, "Reconocimiento de Gestos de la Mano en Tiempo Real Usando Leap Motion Controller y Machine Learning," in *Conference Proceedings*, 2019, vol. 3, no. 1, pp. 823–835.
- [8] F. Cary, O. Postolache, and P. S. Girão, "Kinect based system and Artificial Neural Networks classifiers for physiotherapy assessment," in *IEEE MeMeA 2014 - IEEE International Symposium on Medical Measurements and Applications, Proceedings*, 2014.
- [9] C. Youness and M. Abdelhak, "Machine Learning for Real Time Poses Classification Using Kinect Skeleton Data," in *Proceedings - Computer Graphics, Imaging and Visualization: New Techniques and Trends, CGiV 2016*, 2016.
- [10] A. Da Gama, P. Fallavollita, V. Teichrieb, and N. Navab, "Motor Rehabilitation Using Kinect: A Systematic Review," *Games for Health Journal*, vol. 4, no. 2. Mary Ann Liebert Inc., pp. 123–135, 01-Apr-2015.
- [11] M. Morando, S. Ponte, E. Ferrara, and S. Dellepiane, "Definition of motion and biophysical indicators for home-based rehabilitation through serious games," *Inf.*, 2018.
- [12] S. T. L. Pöhlmann, E. F. Harkness, C. J. Taylor, and S. M. Astley, "Evaluation of Kinect 3D Sensor for Healthcare Imaging," *J. Med. Biol. Eng.*, 2016.
- [13] A. D. Calin, H. F. Pop, and R. F. Boian, "Improving movement analysis in physical therapy systems based on Kinect interaction," in *HCI 2017: Digital Make Believe - Proceedings of the 31st International BCS Human Computer Interaction Conference, HCI 2017*, 2017.
- [14] D. Leightley, J. Darby, B. Li, J. S. McPhee, and M. H. Yap, "Human activity recognition for physical rehabilitation," in *Proceedings - 2013 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2013*, 2013.
- [15] G. Paraskevopoulos, E. Spyrou, and D. Sgouropoulos, "A real-time approach for gesture recognition using the Kinect sensor," in *Proceedings of the 9th Hellenic Conference on Artificial Intelligence*, 2016, pp. 1–4.
- [16] M. Linna, J. Kannala, and E. Rahtu, "Real-time human pose estimation with convolutional neural networks," in *VISIGRAPP 2018 - Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2018.
- [17] M. Morando, M. Trombini, and S. Dellepiane, "Application of SVM for evaluation of training performance in exergames for motion rehabilitation," in *ACM International Conference Proceeding Series*, 2019.
- [18] J. Ressiman, E. Rasmussen-Barr, and W. J. A. Grooten, "Reliability and validity of a novel Kinect-based software program for measuring a single leg squat," *BMC Sports Sci. Med. Rehabil.*, 2020.
- [19] Aforge.net, "AForge.NET :: Framework," *AForge.NET Framework*, 2020. [Online]. Available: <http://www.aforgenet.com/framework/>. [Accessed: 25-Jul-2020].
- [20] N. B. Pham, V. T. Phan, V. B. Nguyen, and V. D. Nguyen, "Game-based virtual rehabilitation system for upper extremity using low-cost camera," in *BMEiCON 2015 - 8th Biomedical Engineering International Conference*, 2016.
- [21] T. L. Le, M. Q. Nguyen, and T. T. M. Nguyen, "Human posture recognition using human skeleton provided by Kinect," in *2013 International Conference on Computing, Management and Telecommunications, ComManTel 2013*, 2013.
- [22] Y. Li, "Hand gesture recognition using Kinect," in *ICSESS 2012 - Proceedings of 2012 IEEE 3rd International Conference on Software Engineering and Service Science*, 2012.
- [23] M. Peris and K. Fukui, "Both-hand gesture recognition based on KOMSM with volume subspaces for robot teleoperation," in *Proceedings - 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, CYBER 2012*, 2012.
- [24] J. M. K. Westlund and C. Breazeal, "Transparency, teleoperation, and children's understanding of social robots," in *ACM/IEEE International Conference on Human-Robot Interaction*, 2016.
- [25] R. Saini, P. Kumar, B. Kaur, P. P. Roy, D. P. Dogra, and K. C. Santosh, "Kinect sensor-based interaction monitoring system using the BLSTM neural network in healthcare," *Int. J. Mach. Learn. Cybern.*, 2019.