IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 10, 2019, accepted May 31, 2019, date of publication June 5, 2019, date of current version June 21, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921257

A Socially Assistive Robot for Elderly Exercise Promotion

ESTER MARTINEZ-MARTIN, (Senior Member, IEEE), AND MIGUEL CAZORLA, (Senior Member, IEEE)

RoViT, Department of Computer Science and Artificial Intelligence, University of Alicante, 03690 Alicante, Spain

Corresponding author: Ester Martinez-Martin (ester@ua.es)

This work was supported in part by the Spanish Government under Grant TIN2016-76515-R, and in part by Feder Funds.

ABSTRACT The population ageing phenomenon leads to an unceasing need for home-based healthcare systems to continuously monitor the elderly's cognitive and physical health. In this sense, physical activity may be beneficial in preserving cognition in elder life as well as in providing clinicians and therapists with the indicative of elderly's health condition. Nevertheless, current systems fail to promote and monitor the elderly's physical activity in their living environments. This paper is aimed at providing a socially assistive robot solution for this task. Since robot acceptance depends to a great extent on its robustness in performing tasks, we have focused on exercise recognition due to its great importance for both clinicians and elderly. For that, two different tasks were carried out. First, an image dataset for physical exercise recognition has been generated. Then, a comparative analysis of several deep learning techniques is presented. This paper reveals a great performance in the exercise recognition of CNN-LSTM with an exercise recognition accuracy of 99.87%.

INDEX TERMS Assistive robotics, elderly healthcare, human action recognition.

I. INTRODUCTION

Globally, mortality rates have fallen significantly leading to considerable changes in the age distribution in societies. This fact may result in a great increase in age-related chronic diseases [1]. That can have a heavy impact on health and social care systems. For that reason, it is required to move health and care practices from hospitals to home, merging treatment and care with prevention [2]–[4].

In this context, Robotics can play an important role with respect to healthcare support and independent life, avoiding (or at least delaying to the limit) being institutionalised in sheltered homes, or nursery homes when problems related to ageing appear. In this regard, the literature in eldercare Robotics features different kinds of robots based on the covered senior needs such as therapy (PARO [5]), rehabilitation (ZORA [6]), telehealth (GiraffPlus [7]), company (MARIO [8]), activity planning (Aido [9], Pearl [10]), navigation aid (Pearl [10]) or entertainment (Manzai robots [11]).

Going a step further, socially assistive robotics adds a social capability in elderly assistance. So, these robots are intended to socially engage people through human-robot interactions while performing daily tasks and supporting human activities and mobility. An example is the EU supported project HOBBIT [12] which is intended to develop a socially assistive robot allowing elderly people to stay longer at home thanks to its implemented functionalities (i.e. searching and bringing objects, transporting small items, keeping floors clutterfree, warnings about reminders, and emergency detection). In a similar way, the EU-funded project ENRICHME (ENabling Robot and assisted living environment for Independent Care and Health Monitoring of the Elderly) [13] proposed an integrated platform for Ambient Assisted Living (AAL) with a mobile service robot for the provision of advanced user services to enhance their quality of life such as health monitoring, complementary care and social support, or cognitive and physical activities to remain active and independent for longer. On its behalf, the EU-funded project RAMCIP (Robotic Assistant for MCI Patients at home) [14] has aimed to develop a service robot with high-level cognitive functions to assist people with mild cognitive impairments and Alzheimer's disease in their daily life. For that and taking into account the patient's loss of cognitive functions and behavioral abilities, a proactive assistance has been implemented. So, RAMCIP supports daily activities' performance (i.e. cooking, eating and taking

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Liu.

medication) through human activity recognition and patient-robot communication.

Moreover, apart from assistance, elderly people require technological solutions promoting active ageing. In particular, this paper is focused on exercise promotion due to its great benefits in preserving motor and cognition at senior life [15]. Along this line, Matsusaka *et al.* [16] developed *TAIZO*, a small humanoid robot acting as an exercise trainer. Activated by a human demonstrator via voice or keypad input, *TAIZO* explains the asked arm exercises while performing them. Note that *TAIZO* is just an actuator (i.e. an exercise demonstrator) considering that no perception sensors have been included. Therefore, its main handicaps are the lack of autonomy, user feedback, active guidance or personalised training.

Gadde *et al.* [17] used the humanoid robot *RoboPhilo* to show and monitor the user's physician-prescribed exercise program. The interaction starts with the user's presence what triggers the user's consent request. After the expected gesture (i.e. the user's hand wave), *RoboPhilo* starts mimicking the first exercise in the scheduled routine while verbally explains the different body movements. Thereafter, the robot monitors the user's performance with the purpose of analysing their form and timing. So, the user is congratulated when their performance is correct, otherwise exercise repetition is required. At the end of the interaction cycle, a verbal feedback about the overall user's performance is provided. This system has several limitations. For instance, the user should be positioned at the centred of the video frame at the optimal distance (determined by the user's height). Moreover, a good performance is only obtained under good light conditions (enough light is necessary for accurate hand/face detection).

On their behalf, Ngee Ann Polytechnic school developed *Robocoach Xuan*, a trainer humanoid robot encouraging physical exercise in classes [18]. In this case, the robot conducts simple workout sessions for seniors. In addition, thanks to its motion sensors, *Robocoach Xuan* is able to evaluate senior movements and provide one-on-one feedback about their performance. However, only seated exercises are considered, what considerably restricts *Robocoach Xuan*'s applicability.

More recently, Gorer *et al.* [19] presented an autonomous exercise tutor for elderly people. In its first stage, a therapist teaches the exercise performance to a user and, when the user is repeating the exercise, their movements are learnt by a NAO robot. In the second stage, the NAO robot behaves as an exercise tutor by repeating the learnt movements with their corresponding verbal explanation when available. Then, the robot observes the user's motions and compares them with the stored performance for their evaluation. This process is repeated until the exercise program is completed. The session is concluded after providing the user with an overall performance score. The main drawbacks of this system include the loss of user's progress since each user is only identified by name. In addition, the provided feedback omits the user's physical capabilities because a specific feedback is associated with a certain score range. Moreover, the exercise program was designed for an average, healthy elderly person, making it inappropriate for the majority of elder community. Finally, the user's performance cannot be studied when the person is seated because their skeleton data is unrealiable.

An analysis of the literature reveals that no appropriate system is available for assisting elderly in doing physical exercise at home. This paper introduces an assistive robot aiding seniors in their daily physical activities and promoting their practice at home. In particular, this paper is focused on the exercise recognition process due to its great importance to give appropriate and relevant feedback as well as to report elderly's health condition in terms of loss of limb mobility.

So, this paper is organised as follows: Section II describes our socially assistive robot; Section III analyses the exercise recognition problem; Section IV presents and discusses the experimental results; and, finally, some conclusions are presented in Section V.

II. SYSTEM DESCRIPTION

A. ROBOT PLATFORM

As pointed out by Fasola and Mataric [20] in their study about the design and effectiveness of socially assistive robots aimed to motivate and engage elderly users in physical exercise, a physical robot is perceived as more helpful, more socially attractive, and as having greater social presence than a virtual robot. In addition, given that attitudes towards robots in elderly care are systematically sceptical, the proposed robot platform should be friendly, intuitive, interactive and proactively assistive. These requirements result in the use of a Pepper robot [21], an endearing human-shaped robot with high levels of acceptance in elder community (see Figure 1).

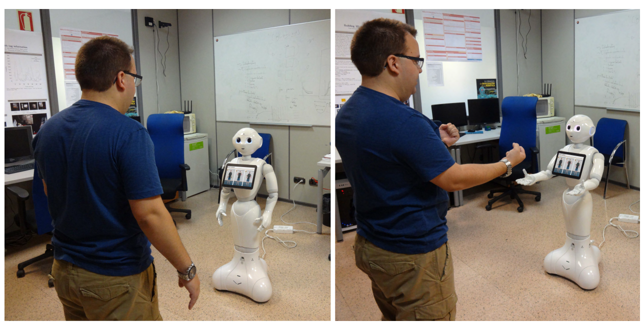

FIGURE 1. Our socially assistive robot interacting with a user.

B. OUR ARCHITECTURE

Our proposal consists in an interactive robot platform designed for assisting the elderly in their daily physical activities at home. As illustrated in Figure 2, the interaction starts when Pepper reminds the elderly user their scheduled daily physical exercise session (set according to the elderly preferences and therapist recommendations).

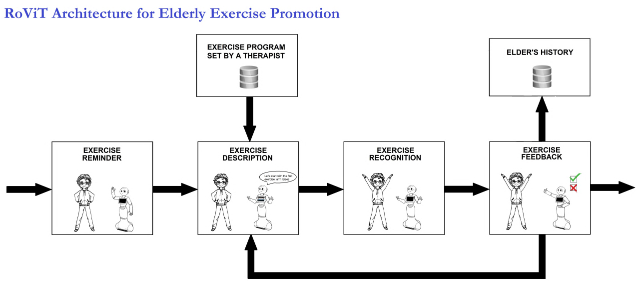

FIGURE 2. RoViT socially assistive robot architecture.

When the user is ready, Pepper explains the first therapy exercise to be done in both verbal and visual ways. Then, Pepper continuously monitors the elderly movements.

After that, Pepper gives relevant feedback to the elderly about the exercise performance. At the same time, the elder's performance is stored with the purpose of being later analysed by the therapist. In this way, the specialist can follow the user's evolution as well as early detect any loss of limb mobility.

This process is repeated until all the recommended exercises for the current session are done.

III. HUMAN EXERCISE RECOGNITION

A key issue of exercise promotion is exercise recognition for two main reasons: providing therapists and clinicians with information about how well the elderly performs physical activities (what is an indicative of the elderly's health condition); and, on the other hand, being able to give relevant feedback to the elderly doing exercise.

Given the importance of limb positions in the exercise recognition process, wearable sensors could make it easy if it were not for their inaccuracy and their cumbersome usability.

In this context, image analysis could be a solution so that elderly people could keep their autonomy and independence while being monitored.

A review of the traditional computer vision approaches brings to light the need of working under certain conditions to achieve good results, especially in terms of background characteristics and/or illumination conditions. However, this cannot be guaranteed in living environments.

Instead of processing standard visual input, an alternative is to extract high-dimensional visual features describing the interest points and combine them into a fixed-sized level description with the aim of classifying them to get a final prediction about the exercise done.

In regard to human representation, the literature is broadly divided into two groups: representations based on local features (i.e. keypoints in the spatio-temporal space); and, skeleton-based representations, where a small number of representative joints encode the whole body configuration. However, the high computational cost, the lack of information for textureless regions and the inability of representing multiple individuals in the same scene make approaches based on local features inappropriate for the task at hand.

On their behalf, skeleton-based approaches provide the system with an estimation of limb position, what is crucial for exercise recognition. Thus, a human representation based on 3D skeleton information with a high frame rate is required. Despite the wide research on this area, we have chosen Openpose [22] to develop this work due to its real-time execution and its adaptability to identify multiple people in the image. Basically, it is a bottom-up approach for real-time multi-person pose estimation. So, this two-branch multi-stage Convolutional Neural Network (CNN) outputs a 18-keypoint body skeleton for all the people in the image. For that, human parts are predicted by means of confidence maps and combined through joint associations such that an articulated system of rigid segments connected by joints is generated for each person. Note that a human skeleton is built from the output human joints. The reason lies in making the learning of joint relationship easier and more accurate.

The next step is to identify the exercise(s) done. It is worthy to take into account that exercise recognition involves capturing spatio-temporal context across frames. One of the most widely used approaches nowadays is Recurrent Neural Networks (RNNs) since they model the temporal dependencies between the visual data. However, the design of architectures capturing this information include multiple options which are not trivial. Although RNNs have been successfully applied to temporal data, it has been proved that ResNets can improve the performance on the training set, which is a prerequisite to do well on the validation and test sets. This results in an

FIGURE 3. Exercise recognition flowchart when multiple people do physical exercise.

FIGURE 4. **Some human poses representing the key limb positions for some physical exercises. Each limb is shown with its corresponding skeleton representation overlapped in the taken frames.**

easy way for a residual block to learn the identity function by adding skipped connections. Therefore, the combination of both architectures joins their advantages leading to a structure able to improve the movement recognition and, as a consequence, the exercise recognition in a more efficient way.

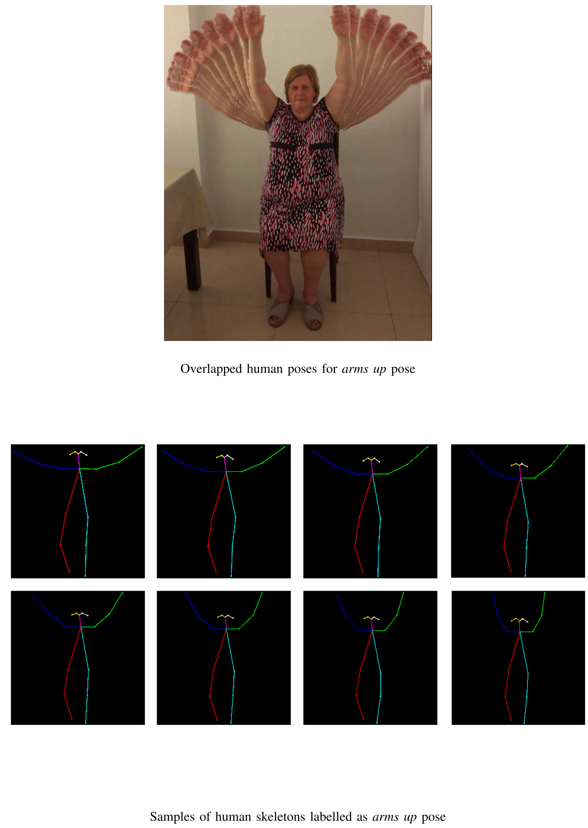

FIGURE 5. **Some human poses representing the same key limb position *arms up* with the aim of overcoming the elderly physical limitations when doing physical exercise.**

In addition, this combination allows the system to exploit spatio-temporal information, which is crucial to achieve the required accuracy. A comparative analysis of different RNNs (and CNN-RNN) architectures is presented and discussed.

Notice that, for simplicity, only one person is illustrated in the images. However, this procedure can be applied to multi-person session. In this case, the *Openpose* output is used to generate different skeletons such that all the skeletons are processed one-by-one in the neural network architecture. In this way, an exercise evaluation is performed for each person in the image (see Figure 3).

IV. EXPERIMENTAL RESULTS

A. PHYSICAL EXERCISES

Physical activity can be defined as any repetitive, structured activity helping to improve or maintain the fitness or condition of the body. Obviously, the daily amount of physical activity varies according to age, gender and physical condition. In this sense, the British National Health Security (NHS) has published online an exercise handbook containing the recommended exercises for older adults [23]. This includes written notes and pictures of the various exercise poses. In addition, this guidebook is divided into four exercise groups: sitting, flexibility, strength and balance. So, a daily elder's exercise program would include a number of exercises

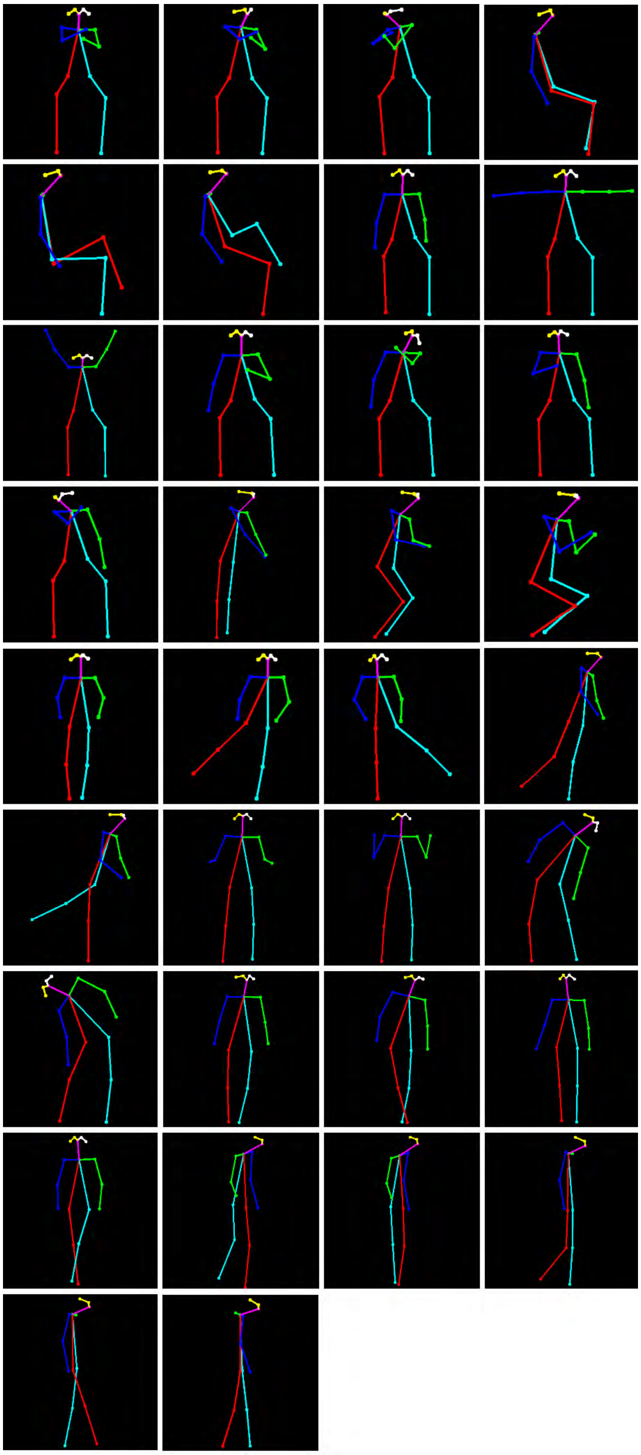

FIGURE 6. The 34 considered human poses for exercise recognition.

from each group based on their age and their physical limits.

Note that a subset of the 24 proposed exercises have been considered in our experiments: four *sitting* exercises (*upper body twist*, *hip marching*, *arm raises*, *neck stretch*); two *flexibility* exercises (*neck stretch*, *sideways bend*); four *strength* exercises (*mini squats*, *sideways leg lift*, *leg extension*, *bicep curls*); and two *balance* exercises (*simple grapevine*, *heel to toe walk*). Note that each exercise has been repeated for each side (left and right) and, consequently, the twelve above-mentioned exercises result in a total of 24 exercises. It is noteworthy that there are some exercises like *neck rotation*, that appears in two different groups: sitting and flexibility. In addition, there are three more exercises that were discarded for this study since the limb motion in the skeleton representation is visually imperceptible.

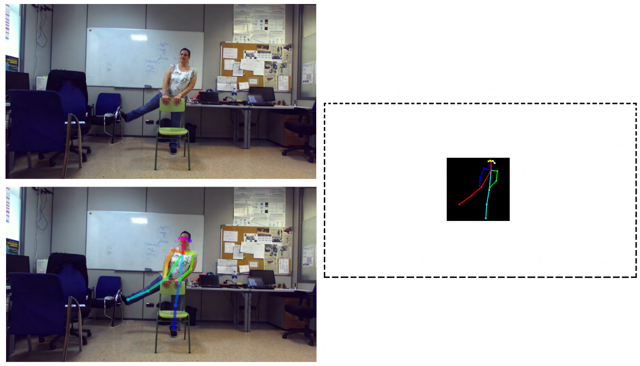

FIGURE 7. Image dataset generation: top left image represents the taken frame (640×480×3); bottom left image corresponds to the image generated by the Openpose algorithm [22] (640×480×3); and the right image is the obtained image after isolating the human skeleton and accordingly cropping the image (224×224×3).

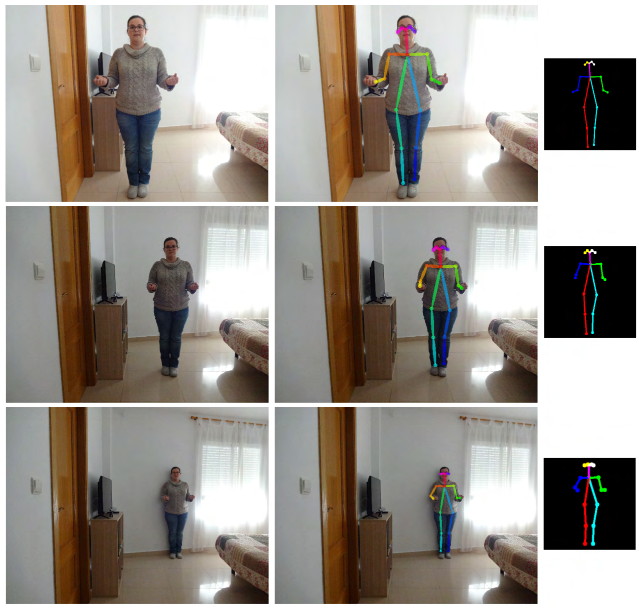

FIGURE 8. The human skeleton 224×224×3 images generated from the visual input varying the user's distance to the robot and/or the user's position within the image.

For proper exercise recognition, each physical exercise has been divided into a series of human poses representing the key limb positions (some examples are illustrated in Figure 4).

It is worth noting that elderly people may present several physical limitations. For this reason, their movements, and consequently, their limb positions can be restricted. This results in the fact that some end positions are unattainable.

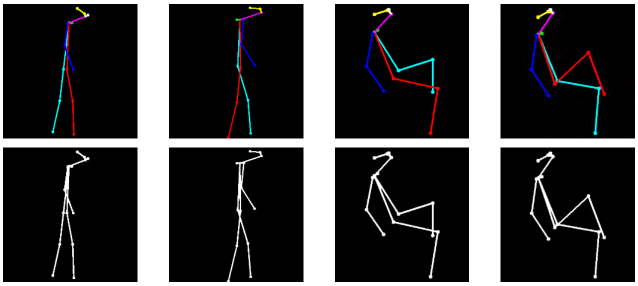

FIGURE 9. Samples illustrating the need of coloured skeleton with the purpose of improving the exercise recognition results.

With the purpose of overcoming this issue, most of the defined human poses cover more than one limb position. So, for instance, the human pose corresponding to the limb position *arms up* includes all the arm positions ranging from 100 to 180 degrees as illustrated in Figure 5. Consequently, 34 key limb positions have been defined to cover all the considered physical exercises (see Figure 6).

B. EXPERIMENTAL DATA

The image dataset was generated in two stages. Firstly, some university staff, who was not in the expected age group, was contacted to participate in the early system's evaluation. So, in addition to check the robot's functionality, this group of volunteers performed all the 24 physical exercises as proposed by the robot, while being recorded by the Pepper's top camera (SOC Image Sensor OV5640). Then, the robot was tested with a group of elderly people who, following the robot instructions, performed all the physical exercises under study. These performances were also recorded by the Pepper's top camera.

Each video recording of the 11 participants has been divided in images at a rate of 15 frames per second. Given their resolution (640 $\times$ 480 $\times$ 3) and aiming to avoid background overfitting, the images are processed to obtain 224$\times$224$\times$3 images only containing the human skeleton as shown in Figure 7. For that, each image was cropped to the skeleton bounding box. After that, the resulting image is resized to 224$\times$224$\times$3. With this procedure the problems derived from variable distance to the robot or user's position within the input image are automatically resolved (see Figure 8).

Thus, our image dataset is composed of 224$\times$224$\times$3 images of coloured skeletons on a black background generated from the visual input of Pepper's top camera. Note that each limb is represented with a different colour. The reason lies in the lack of 3*D* information about the limb pose. Consequently, some human poses could be easily confused whether the skeleton was represented by just one colour as shown in Figure 9.

C. EXPERIMENTAL ANALYSIS

As above-mentioned, an infinity RNNs architectures could fit the exercise recognition task. For that reason, a comparative analysis has been carried out. To begin with, the popular Long Short Term Memory networks (LSTMs) [24] are used. LSTMs are explicitly designed to avoid the long-term dependency problem. Given that a temporal analysis is performed, the length of the video sequence corresponding to any physical exercise plays a main role in its proper recognition. This variability is due to two main reasons. The first one is the different number of human poses involved in a physical exercise. So, for example, the *sideways bend* exercise requires three human poses for each direction (*stand up - direction bended - stand up*), while the *bicep curls* requires five (*rest position - half curled arms - curled arms - half curled arms - rest position*). The second reason is the elderly's mobility what by far influences the exercise duration. As a solution, the input sequences were truncated and padded so that all the input data

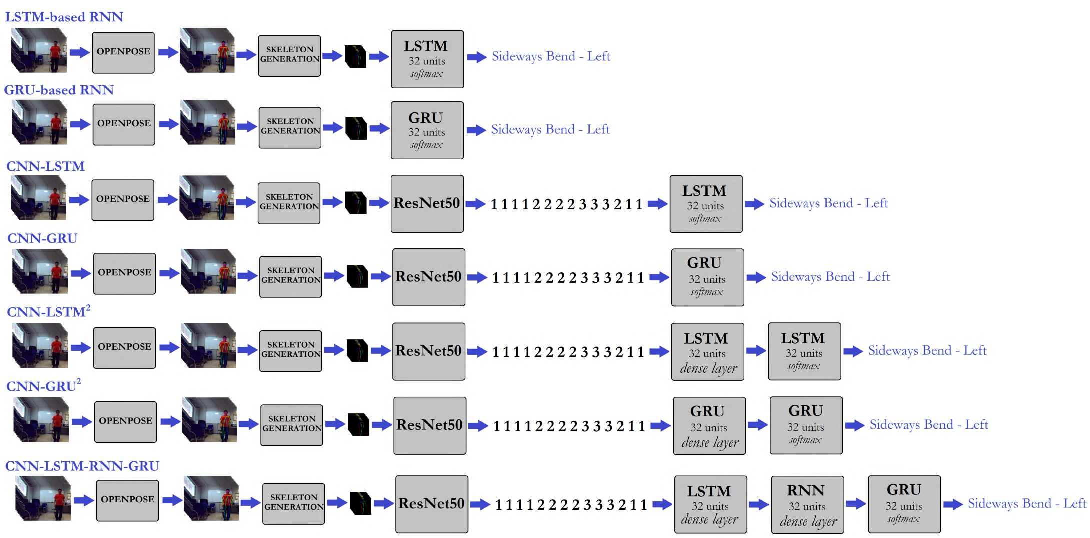

FIGURE 10. Deep learning architectures analysed and compared in this work.

TABLE 1. Confusion matrix corresponding to LSTM-based RNN test evaluation with 150 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
0.16	0.75	0.00	0.03	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.41	0.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.10	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.03	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.02	0.02	0.00	0.00	0.57	0.00	0.07	0.21	0.00	0.02	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.15	0.00	0.00	0.00	0.53	0.03	0.15	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.04	0.11	0.00	0.00	0.00	0.29	0.18	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.17	0.00	0.00	0.17	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.03	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.41	0.19	0.13	0.00	0.00	0.09	0.00	0.00	0.00	0.06	0.00	0.00	0.03	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.25	0.54	0.00	0.00	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.06	0.19	0.00	0.00	0.00	0.00	0.00	0.63	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.58	0.00	0.08	0.33	0.00	0.00	0.00	0.00
0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.44	0.00	0.00	0.19	0.00	0.00	0.00	0.00
0.25	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 2. Comparison of the implemented deep learning techniques in terms of number of parameters to be set.

Deep Learning Technique	No. parameters
LSTM-based RNN	4822176
GRU-based RNN	3616896
CNN-LSTM (Resnet50 ➔ LSTM)	23600160 + 5144
CNN-GRU (Resnet50 ➔ GRU)	23600160 + 4056
CNN-LSTM^2 (Resnet50 ➔ LSTM ➔ LSTM)	23600160 + 13464
CNN-GRU^2 (Resnet50 ➔ GRU ➔ GRU)	23600160 + 10296
CNN-LSTM-RNN-GRU (Resnet50 ➔ LSTM ➔ RNN ➔ GRU)	23600160 + 13464

TABLE 3. Comparison of the implemented deep learning techniques in terms of accuracy.

Deep Learning Technique	Accuracy (%)
LSTM-based RNN	37.42
GRU-based RNN	35.97
CNN-LSTM (Resnet50 ➔ LSTM)	99.87
CNN-GRU (Resnet50 ➔ GRU)	89.09
CNN-LSTM^2 (Resnet50 ➔ LSTM ➔ LSTM)	4.17
CNN-GRU^2 (Resnet50 ➔ GRU ➔ GRU)	87.96
CNN-LSTM-RNN-GRU (Resnet50 ➔ LSTM ➔ RNN ➔ GRU)	73.77

has the same length for a correct processing. Thus, the shorter input sequences were completed with zero values (i.e. black images in the case of image input and 0 in the case of human pose input). In this way, zero values are learnt as *no information* elements. Specifically, the video sequence length was experimentally set to 60 frames. So, chunks of 60 coloured skeleton images feed into a RNN consisting of a LSTM with 32 units and a Dense layer with *softmax* activation for exercise classification (see Figure 10). A 150-epoch running results in an exercise recognition accuracy of 37.42 %. Most of the image sequences were erroneously recognise as the same exercise in the opposite direction (i.e. left side instead right side and viceversa) as illustrated in Table 1. In addition, almost any video sequence corresponding to the *heel-to-toe-walk* was properly identified. These bad results could be a consequence of the great amount of parameters to be set (see Table 2) making the epochs to be trained insufficient.

A total of 39537 frames were obtained such that 24545 frames are used for training and 14992 frames for test. All these coloured skeleton images have been manually labelled for both CNN and RNN validation. Note that, with the purpose of being able to evaluate the patient's health status, two types of RNN sequences has been considered: (1) sequences representing complete physical exercises; and, (2) sequences representing partial performance of a physical exercise. That is, with the aim of determining the completeness of each activity, different subsequences of that activity were evaluated. So, the completeness of an activity is given by the performed human poses. For instance, in the case of the arm raises exercise, the exercise is 100 % complete if five poses are achieved in the right order: down arms - extended arms - rose arms - extended arms - down arms. On the contrary, when the person performs the sequence down arms - extended arms - down arms, the system outputs the right exercise (i.e. arm raises exercise), but with a score of 60 %. In this way, a therapist can monitor and analyse the elderly's health status based on their daily performance. Therefore, partial exercises belong to a different category from the one corresponding to the whole exercise. By simplicity, all the categories belonging to the same physical exercise appeared under the same category in the following confusion matrices.

Secondly, a RNN based on Gated Recurrent Unit (GRU) [25]. GRU can be considered as a variation of

TABLE 4. Confusion matrix corresponding to GRU-based RNN test evaluation with 150 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
0.77	0.19	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.68	0.26	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.33	0.60	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.05	0.71	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.09	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.21	0.32	0.00	0.03	0.00	0.29	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.21	0.54	0.00	0.00	0.00	0.04	0.18	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.17	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.15	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.58	0.00	0.00	0.04	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.04	0.07	0.00	0.00	0.11	0.00	0.75	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.03	0.00	0.00	0.03	0.00	0.03	0.00	0.14	0.00	0.00	0.69	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.07	0.00	0.69	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.04	0.00	0.00	0.04	0.17	0.17	0.00	0.17	0.00	0.00	0.00	0.21	0.00	0.00	0.08	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.08	0.23	0.00	0.00	0.08	0.00	0.23	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.64	0.36	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 5. Confusion matrix corresponding to ResNet50 test evaluation with 100 epochs when 32 human poses are considered.

0.85	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.85	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.31	0.00	0.64	0.00	0.00	0.00	0.02	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.87	0.07	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.01	0.94	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.07	0.00	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.01	0.00	0.00	0.00	0.00	0.00	0.88	0.10	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.97	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.20	0.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.48	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.12	0.23	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.11	0.00	0.02	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.02	0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.95	0.00	0.02	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00
0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.16	0.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.02	0.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.73	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.77	0.18	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.97	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.86	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.02	0.00	0.00	0.20	0.09	0.58
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.02	0.00	0.06	0.83	0.04

the LSTM. It is aimed to solve the vanishing gradient problem that comes with a standard RNN. In particular, the implemented RNN is composed of a GRU with 32 units and a Dense layer with *softmax* activation for exercise classification. Again, 150 epochs were necessary to achieve an exercise recognition with a poor accuracy of 35.97 % (see Table 3 for a comparison). As illustrated in Table 4, a common error is to confuse the same exercise but with opposite directions. That is, for instance, the *upper body twist-right* exercise is misclassified as its analogous the *upper body twist-left* exercise. This could be due to the skeleton thinness. Additionally, any sequence corresponding to the *heel-to-toe-walk* was correctly identified.

The bad results obtained with the proposed RNN architectures can be attributed to a slow (or lack of) convergence. So, with the aim of achieving better network convergence, a CNN could be used as starting point. Thus, the CNN layers are used for feature extraction on input data, while the RNN layers interpret those features across time steps. In this way, the resulting CNN-RNN architecture is both spatially and temporally deep. Along this line, a series of experiments was also carried out.

Similarly to RNN, there are countless CNN architectures fitting the task at hand. However, among all them, a 50-layer Residual Network (*ResNet50* [26]) was used in all our CNN-RNN implementations due to its successful

TABLE 6. Confusion matrix corresponding to CNN-LSTM training evaluation with 50 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.002	0.998	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.001	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.995	0.000	0.000	0.005	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	0.001	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.996	0.000	0.000	0.004
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000	0.997	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

TABLE 7. Confusion matrix corresponding to CNN-LSTM test evaluation with 50 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.002	0.998	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.993	0.000	0.000	0.007	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.993	0.000	0.000	0.007	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.991	0.000	0.000	0.009
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.002	0.993

results in visual recognition tasks. For that, this network has been trained with our image dataset to properly identify each human pose. In this case, 34 classes were defined according to each considered human pose (see Figure 6). After 100 epochs of training, an accuracy of 84.76 % was obtained. As shown in Table 5, the most confusing human pose corresponds to the person walking (*heel-to-toe-walk* exercise). In fact, almost any sample was correctly recognised as that human pose. Another human pose commonly misclassified is when a person holds their shoulder down with the opposite arm.

Once the CNN model has been trained (and fixed), the RNN model is added. As previously, the beginning implementation involves LSTMs. Thus, the initial CNN-RNN architecture (referenced as CNN-LSTM) comes from the combination of *ResNet50* with a LSTM with 32 units with a Dense layer (with 24 units) on the output. In this case, the approach converged at 50 epochs with an exercise accuracy of 99.87 %. As observed in Tables 6 and 7, the worst recognition result corresponds to the sequences describing the *heel-to-toe-walk*. The reason could lie in the classification errors coming from the *ResNet50* human pose recognition.

TABLE 8. Confusion matrix corresponding to CNN-GRU training evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
0.958	0.042	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.057	0.943	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.989	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.111	0.889	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.998		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.001	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.001	0.999	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.985	0.000	0.000	0.015	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.984	0.012	0.000	0.000	0.000	0.003	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.216	0.756	0.000	0.000	0.000	0.019	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.895	0.105	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.090	0.905	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.977	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.017	0.000	0.000	0.006	0.867	0.013	0.002	0.055	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.012	0.000	0.000	0.000	0.119	0.713	0.025	0.110	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.009	0.000	0.000	0.000	0.040	0.016	0.613	0.304	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.015	0.000	0.000	0.000	0.016	0.008	0.006	0.948	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.598	0.000	0.000	0.401	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.985	0.000	0.000	0.015	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.988	0.000	0.000	0.012
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.082	0.000	0.000	0.918	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.112	0.000	0.000	0.888	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.230	0.000	0.000	0.770

TABLE 9. Confusion matrix corresponding to CNN-GRU test evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
0.947	0.053	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.062	0.938	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.988	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.120	0.880	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.995	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.005	0.995	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.002	0.998	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.986	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.979	0.016	0.000	0.000	0.000	0.002	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.233	0.733	0.000	0.000	0.000	0.023	0.002	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.859	0.136	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000	0.120	0.869	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.961	0.030	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.046	0.012	0.000	0.000	0.009	0.820	0.037	0.005	0.071	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.030	0.009	0.000	0.000	0.000	0.092	0.733	0.018	0.118	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.016	0.000	0.000	0.000	0.035	0.016	0.631	0.293	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021	0.005	0.000	0.000	0.000	0.023	0.005	0.007	0.940	0.000	0.000	0.000	0.000	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.578	0.000	0.000	0.417	0.005	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.977	0.000	0.000	0.023	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.993	0.000	0.000	0.007
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.090	0.000	0.000	0.910	0.000	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.111	0.000	0.000	0.889	0.000
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.217	0.000	0.002	0.781

The next implementation makes use of GRUs. So, the CNN-GRU starts with ResNet50 followed by a GRU with 32 units and a Dense layer for exercise recognition. For comparative issues, 100 epochs were also used. After them, the CNN-GRU architecture shows a classification accuracy of 89.09 %. As illustrated in Tables 8 and 9, the worst recognised exercise is the *heel-to-toe-walk* because of the slight difference between the defined human poses. As previously, there is some confusion between the two directions for the same exercise. That is, the left performance is continuously misclassified as its right counterpart.

Going a step further, like CNN layers, RNN layers could be combined with the purpose of improving the exercise recognition. Firstly, two layers of the same type are combined. Thus, we started with two LSTMs with 32 units fed by the ResNet50 output. Far from improving the performance of the *CNN-LSTM* approach, the resulting accuracy of *CNN-LSTM*^{2} for exercise recognition is only 4.17 %. As illustrated in Table 10, all the exercise sequences were classified as *mini-squats* exercise. One reason for this result could be the higher number of parameters to be set (see Table 2). For that reason, more epochs are required to be able to get a better accuracy.

TABLE 10. Confusion matrix corresponding to CNN-LSTM^2 test evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE 11. Confusion matrix corresponding to CNN-GRU^2 training evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.09	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.97	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.05	0.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.91	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.88	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.93	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.26	0.67	0.01	0.02	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.07	0.26	0.59	0.06	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.06	0.01	0.09	0.83	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.67	0.00	0.00	0.33	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.93	0.00	0.00	0.06	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.88	0.00	0.05	0.06
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.94	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.00	0.00	0.94	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.28	0.63

Similarly, when two consecutive GRU layers (with 32 units each one) are used for the RNN model, the CNN-GRU^2 accuracy decreases about 1 % with respect to CNN-GRU (87.96 % versus 89.09 %, respectively) (see Table 3). So, adding a GRU deteriorates exercise recognition apart from increasing the time required for training and, therefore, for adding a new physical exercise to the system. As shown in Tables 11 and 12, the CNN-GRU^2 fails in the recognition

TABLE 12. Confusion matrix corresponding to CNN-GRU^2 test evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.09	0.91	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.97	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.06	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.89	0.10	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.00	0.00	0.92	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.31	0.63	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.04	0.30	0.59	0.05	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.06	0.01	0.10	0.80	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.66	0.00	0.00	0.34	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94	0.00	0.00	0.06	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.82	0.00	0.06	0.10
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.92	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.95	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.31	0.60

TABLE 13. Confusion matrix corresponding to CNN-LSTM-RNN-GRU training evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.97	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.61	0.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.30	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.17	0.00	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.13	0.00	0.87	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.79	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.07	0.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.00	0.00	0.66	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.04	0.86	0.05	0.02	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.05	0.90	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.96	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.98	0.00	0.00	0.01	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.98	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.99

of the *heel-to-toe-walk* exercise, especially when stopping the movement, due to the similarity in the movement and the *ResNet50* misclassification. In addition, a confusion between the *sideways-bend* and *simple-grapevine* exercises as well as between the *sideways-leg-lift* and *sideways-bend*/*simple-grapevine* exercises (especially during training) can be also observed. Some errors are also made between different directions of the same exercise.

TABLE 14. Confusion matrix corresponding to CNN-LSTM-RNN-GRU test evaluation with 100 epochs when 24 exercises are considered.

Upper Body Twist - Left	Upper Body Twist - Right	Hip Marching - Left	Hip Marching - Right	Arm Raises	Neck Stretch - Left	Neck Stretch - Right	Sit to Stand	Mini Squats	Sideways leg lift - Left	Sideways leg lift - Right	Leg extension - Right	Leg extension - Left	Bicep curls	Sideways bend - Left	Sideways bend - Right	Simple grapevine - Left	Simple grapevine - Right	Heel to toe walk - Starting Right	Heel to toe walk - Left Step	Heel to toe walk - Right Step	Heel to toe walk - Starting Left	Heel to toe walk - Stop Right	Heel to toe walk - Stop Left
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.64	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.29	0.00	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.19	0.00	0.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.14	0.00	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.75	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.07	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.64	0.32	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.03	0.86	0.06	0.03	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.93	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.98	0.00	0.00	0.02	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.97	0.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.97

Finally, a RNN combining different kinds of layers was studied. In particular, the architecture can be summarised as a ResNet50 layer connected to a LSTM feeding a Simple RNN layer linked to a GRU layer ended with a Dense layer. Note that the Simple RNN layer corresponds to a predefined fully-connected RNN where the output is to be fed back to input. The complexity in the design implies a great number of parameters to be set (see Table 2). As abovementioned, this fact could require a higher number of epochs to get good results. With 100 epochs, the *CNN-LSTM-RNN-GRU* approach obtains an accuracy in exercise recognition of 73.77%. Again, the different directions of the same kind of exercise (e.g. *upper-body-twist*-left versus *upper-body-twist*-right) are misclassified. In addition, all the sequences corresponding to the *sideways-bend*-left exercise were erroneously recognised during the training, although the result improves for the test sequences.

All the approaches have been implemented in Keras with TensorFlow as backend and run on a GeForce GTX 1080 Ti. All of them used the efficient gradient descendant algorithm adam [27] for optimization due to its broader adoption for deep learning applications in computer vision, while accuracy was set as the metric since it is considered as a classification problem.

V. CONCLUSIONS

The rapidly increasing ageing population demands technological solutions for healthcare, assistance and rehabilitation, while promoting independent living, active ageing and ageing in place. In this context, a main concern is cognitive and physical decline. A good practice is to do physical exercise that preserves cognition while improving person's overall health enhancing their quality of life. In this sense, socially assistive robots could assist older people in their daily physical routines. However, a review of the literature reveals that no appropriate system is available for assisting elderly in doing exercise at home.

To overcome this social issue, this paper presents a socially assistive robot provided with social and therapeutic capabilities to engage elderly to do daily physical exercise. Specifically, exercise recognition is the core of this work due to its importance for early detection of the elderly's health decline and for elderly motivation. With that aim, several deep learning techniques have been analysed.

The first step was the generation of an appropriate image dataset. For that, eleven people were recorded while interacting with our assistive robot and doing the recommended physical exercises. The video sequences were then divided into frames containing the different human poses defining each considered physical exercise. Then, with the purpose of avoiding background overfitting and errors due to differences in human bodies, a technique to extract human skeletons is used (in particular, Openpose). After that, images were cropped according to the information of interest, that is, the human skeleton. Next, a total of 39537 $224\times224\times3$ images were obtained such that 24545 images were used for training and 14992 images for test. After that, all those images were manually labelled for both CNN and RNN validation.

Once the experimental data was ready, seven deep learning approaches were implemented in Keras (with TensorFlow as backend) and tested: *LSTM-based RNN*, *GRU-based RNN*, *CNN-LSTM*, *CNN-GRU*, *CNN-LSTM*^{2}, *CNN-GRU*^{2}, and *CNN-LSTM-RNN-GRU*. The experimental results show that a pure RNN architecture needs a higher number of epochs to be able to converge that those integrating a CNN as starting point. However, adding more RNN layers deteriorates the accuracy as in the case of *CNN-LSTM*^{2}. The reason lies in the increase in the parameters amount to be set since this fact requires more samples and more epochs to get a good result.

Apart from design issues, the experimental results highlight the confusion of physical exercise when different directions are considered. That is, the video sequence corresponding to one exercise in one direction and in the opposite one are commonly misclassified (e.g. *upper-body-twist* to the left side versus *upper-body-twist* to the right side). Despite colouring the human limbs in different tones precisely to overcome this issue, most of the implemented deep learning techniques were not able to properly distinguish them. As a solution, thicker skeletons or more distinguishable colours could be used. This solution could be also used to overcome the *heel-to-toe-walk* misclassification, widely extended among the implemented (CNN-) RNN architectures.

Despite all the described issues, the best architecture for exercise recognition according to our experimental set-up is *CNN-LSTM* with an accuracy of 99.87 %.

As future work, we plan to evaluate its performance with other types of physical exercises and activities like eating. Moreover, different metrics have been also analysed to better evaluate the proposed architectures. In addition, aiming at helping clinicians to know the elderly's health status at any time, we plan to equip the proposed system with the ability to generate a statistical summary about the elderly's performance for each activity. Moreover, a comparative analysis of deep learning techniques with respect to genetic algorithms in terms of accuracy and execution time are also planned. Finally, the multi-sensory integration in living environments will be studied.

ACKNOWLEDGMENT

Experiments were made possible by a generous hardware donation from NVIDIA

REFERENCES

[1] (2017). U. N.-I. on Aging. *Why Population Aging Matters: A Global Perspective*. Accessed: May 2019. [Online]. Available: https://www.nia.nih.gov/sites/default/files/2017-06/WPAM.pdf

[2] A. Peine, I. Rollwagen, and L. Neven, "The rise of the 'innosumer'—Rethinking older technology users," *Technol. Forecasting Social Change*, vol. 82, pp. 199–214, Feb. 2014.

[3] A. Sanchis, V. JuliánJuan, M. Corchado, H. Billhardt, and C. Carrascosa, "Using natural interfaces for human-agent immersion," in *Proc. Commun. Comput. Inf. Sci.* Cham, Switzerland: Springer, 2014, pp. 358–367.

[4] J. A. Rincon, A. Costa, C. Carrascosa, P. Novais, and V. Julian, "EMERALD—Exercise monitoring emotional assistant," *Sensors*, vol. 19, no. 8, p. 1953, Apr. 2019.

[5] K. Wada, T. Shibata, T. Saito, and K. Tanie, "Effects of robot assisted activity to elderly people who stay at a health service facility for the aged," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Las Vegas, NV, USA, Oct. 2003, pp. 2847–2852.

[6] (2017). *Zorabots*. Accessed: May 2019. [Online]. Available: http://zorarobotics.be/index.php/en/

[7] *Giraffplus*. Accessed: May 2019. [Online]. Available: http://www.giraffplus.eu/

[8] *Mario project*. Accessed: May 2019. [Online]. Available:http://www.mario-project.eu/portal/

[9] Aido. *The Next Generation Home Robot*. Accessed: May 2019. [Online]. Available: https://www.startengine.com/aido

[10] M. Pollack, L. Brown, D. Colbry, C. Orosz, B. Peinter, S. Ramakrishnan, S. Engberg, J. T. Matthews, J. Dunbar-Jacob, C. E. McCarthy, S. Thrun, M. Montemerlo, J. Pineau, and N. Roy, "Pearl: A mobile robotic assistant for the elderly," in *Proc. AAAI Workshop Autom. Eldercare*, Aug. 2002, pp. 1–7.

[11] U. Tomohiro, A. Satoshi, A. Kazuhiro, M. Ryo, K. Tatsuya, and N. Akiyo, "Scalable component-based manzai robots as automated funny content generators," *J. Robot. Mechatronics*, vol. 28, no. 6, pp. 862–869, Dec. 2016.

[12] (2011). *Hobbit—The Mutual Care Robot*. Accessed: May 2019. [Online]. Available: http://hobbit.acin.tuwien.ac.at/index.html

[13] (2015). *Enrichme*. Accessed: May 2019. [Online]. Available: http://www.enrichme.eu/

[14] (2015). *Ramcip—Robotic Assistant for MCI Patients at Home*. Accessed: May 2019. [Online]. Available: http://www.ramcip-project.eu/

[15] L. Bherer, K. I. Erickson, and T. Liu-Ambrose, "A review of the effects of physical activity and exercise on cognitive and brain functions in older adults," *J. Aging Res.*, vol. 2013, Jul. 2013. Art. no. 657508.

[16] Y. Matsusaka, H. Fujii, T. Okano, and I. Hara, "Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration," in *Proc. IEEE Int. Symp. Robot Hum. Interact. Commun.*, Sep. /Oct. 2009, pp. 472–477.

[17] P. Gadde, H. Kharrazi, H. Patel, and K. F. MacDorman, "Toward monitoring and increasing exercise adherence in older adults by robotic intervention: A proof of concept study," *J. Robot.*, vol. 2011, Aug. 2011, Art. no. 438514.

[18] T. Wong. (2015). *Singapore Robot Exercise Coach for the Elderly*. Accessed: May 2019. [Online]. Available: https://www.bbc.co.uk/news/av/world-asia-35149344/singapore-s-robot-exercise-coach-for-the-elderly

[19] B. Görer, A. A. Salah, and H. L. Akin, "An autonomous robotic exercise tutor for elderly people," *Auto. Robots*, vol. 41, no. 3, pp. 657–678, Mar. 2016.

[20] J. Fasola and M. J. Matarić, "A socially assistive robot exercise coach for the elderly," *J. Hum.-Robot Interact.*, vol. 2, no. 2, pp. 3–32, Jun. 2013.

[21] *Pepper*. Accessed: May 2019. [Online]. Available: https://www.softbankrobotics.com/emea/en/pepper

[22] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7291–7299.

[23] *NHS Choices—Exercises for Older People*. Accessed: May 2019. [Online]. Available: https://www.nhs.uk/Tools/Documents/NHS_ExercisesForOlderPeople.pdf

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, Nov. 1997.

[25] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translat," Jun. 2014, *arXiv:1406.1078*. [Online]. Available: https://arxiv.org/abs/1406.1078

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Dec. 2015, *arXiv:1512.03385*. [Online]. Available: https://arxiv.org/abs/1512.03385

[27] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

ESTER MARTINEZ-MARTIN is a Computer Science Engineer since 2004. In addition, she received the master's degree in mobile and video games programming Jaume-I University, in 2017, the master's degree for Secondary Education, Vocational Training and Language Teaching, from Jaume-I University, in 2011, and the Ph.D. degree in engineering (Robotics) from Jaume-I University, in 2011.

She is currently an Associate Professor with the University of Alicante. She has published nine JCR-indexed articles and articles in scopus-indexed journals, a research book (Springer), several book chapters, three research books (as a co-editor), and more than 25 articles in congresses, both national and international. She has participated in several research projects of community, national, and European nature. She has also done four research stays in prestigious foreign centres including Università degli Studi di Genova (Pr. Silvio Sabatini), Sungkyunkwan University (Pr. Suthan Lee), Universidade do Minho (Pr. Paulo Novais), and the Technische Universitat (TU) Wien (Pr. Markus Vincze), all of which are financed in competitive public calls. Her research interest includes the use of vision in robotic tasks such as object detection and action recognition. In addition, she is an AERFAI Member. In addition, she has been the Organization Chair of the IEEE-RAS Summer School on Experimental Methodology, Performance Evaluation and Benchmarking in Robotics, in 2015, 13th International Conference on the Simulation of Adaptive Behavior (SAB 2014), and the 12th International AERFAI/UJI Robotics School on Perceptual Robotics for Humanoids (IURS 2012). She was also co-organizer of some tutorials in international conferences including HRI, in 2017, ICINCO, in 2014, IAS-13, in 2014, and the IEEE RO-MAN, in 2013. She has been a member of the programme committee and organization committee in several national and international conferences. She is a regular reviewer of JCR journals and international congresses.

MIGUEL CAZORLA is currently a Computer Engineer and the Ph.D. in computer engineering from the University of Alicante, in 1995 and 2000, respectively.

In 1995, he joined the University of Alicante, as a Professor, going through all the figures (associate, assistant, head of school, and the holder of University). Since 2017, he has been a Full Professor. He has completed several stays at foreign institutions (Carnegie Mellon University, University of Sydney, and University of Edinburgh). He has published over 50 articles indexed in JCR (with over 20 in Q1) and over 100 publications in national and international conferences. He has directed 11 Ph.D. theses. He is a Principal Investigator of several national projects (CICYT, Challenges), and he has made multiple transfer contracts to the company. He is a member of different program committees of national and international conferences. His line of research has always been centered on computational vision. From the beginning, he has applied these skills to try to solve robotic tasks. Almost since its inception in research has worked in the processing of 3D data. In recent years, he has diversified his lines to apply deep learning techniques to different areas (medical image, object recognition, depth estimation, and identification of traffic objects). All his research in recent years have been focused on social robotics: applications of all these techniques to help dependents.