Joint Positions Detection for the Elderly Exercises using Backpropagation Neural Network

Porawat Visutsak[1,*], Teeradon Chaiwong[1], Khemmarin Jirawangkaewworrawut[1], and Mohamed Daoudi[2]

[1]Department of Computer and Information Science, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand
[2]IMT Lille Douai, Univ. Lille, CNRS, UMR 9189 CRIStAL, F-59000 Lille, France
*Corresponding author. E-mail: porawatv@kmutnb.ac.th

ABSTRACT

The inspiration of this work is based on the limitation of the elderly's movement. The lack of monitoring while the elderly is doing the exercises may cause the injury. This system can prevent the injury of the elderly by real time detecting and alerting. The video clips of the elderly exercises used in this system were created by the sport science specialists from Thai Health Promotion Foundation (THPF). There are 2 modes of exercise: sit mode and standing mode, the elderly may prompt to see the demonstration of each exercise from the video clip on the GUI. The Kinect extracts the joints data during the exercise, and the system will alert if the wrong posture is detected. Our method used the backpropagation neural network for training and testing models. The 16 exercises of the elderly were recorded and extracted to 8,300 frames in total as the training data, e.g. we generated the model for the exercise no. 1 by using 500 frames, our 16 models used 25 hidden nodes in the second layer and 2 outputs. In the real time classification, the system yields the results of 0.79, 0.77 and 0.79 of precision, recall and F1 score, respectively. The user interface is quite simple and easy to use for the elderly.

Keywords: backpropagation, elderly, exercise, kinect, motion detection, neural network

INTRODUCTION

Exercise has countless benefits for those of all ages, especially in the elderly. The usual exercise reduces the risk of chronic diseases, lowers the chance of injury, and can even improve the mental health. Many researchers found that when the elderly walk three or more times a week, the occurrence of dementia was 35% lower than those elderly who were not involved in any type of physical activity (The Importance of Exercise for Seniors, 2014). Unfortunately, the elderly's muscle mass begins to decrease; therefore, this is the limitation of the elderly's movement. Leaving the elderly do the exercise on their own may cause the injury (Visutsak & Daoudi, 2017; Singh et al., 2018).

Article history:
Received 13 February 2021; Received in revised from 09 June 2021;
Accepted 10 June 2021; Available online 21 June 2021.

This paper presents “Joint Positions Detection for the Elderly Exercises using Backpropagation Neural Network (BPNN)”. By the real time monitoring, the system can prevent the injury of the elderly by detecting the elderly’s poses while they are doing their exercises. The system will alert to the elderly whether they do the wrong poses to prevent the injury of muscle and joints. Unfortunately, persuading the elderly to use this kind of technology is exceedingly difficult, since the elderly may resist or refuse to use the assisted devices. Therefore, the challenges of this work are 1) the application must be user-friendly for the elderly and 2) the application should help the elderly enjoy. Therefore, we gathered the user requirements from the caregivers and the elderly, and we designed the graphics user interface (GUI) to suit the user needs. To ensure that the GUI is fit for the requirements, we test the GUI prototype to the caregivers and the elderly until the final GUI meet their needs. In order to help the elderly enjoy while they are using this application, we selected the video clips of the elderly exercises created by the sport science specialists from Thai Health Promotion Foundation (THPF). The exercises consist of 2 set: the sit mode consists of 12 exercises, the elderly can do these exercises by sitting on the chair and follow the instruction on the video clips; the standing mode consists of 4 exercises, the elderly can do these exercises using the chair to assist while they are moving the lower parts of their body, e.g. raising the right/left knee. The GUI of this work show in Figure 7-12, and our work was selected as the show case in the National Science and Technology Fair 2018 Thailand (NST2018) as seen in Figure 14. The system is easy to use in the daily life, by attaching the Kinect camera with the laptop or PC. The system will show the demonstration of the exercises to the elderly; upon doing the exercises, the system will show the alert sign (✖) to the screen if the elderly do the wrong pose, as well as the correct sign (✓) when the elderly perform the correct pose. This paper consists of 6 parts: introduction, a review of literature, system framework, implementation, results, and future works.

REVIEW OF LITERATURE

In this paper, the algorithm used for training and testing model for classifying the activities is Backpropagation Neural Network. Neural network models is the artificial software created based on the biological neurons and synapses to create a system that can be used in forecasting tasks e.g. forecasting the rainfall in the harvesting period of fruits and predicting stock market prices (Leung, & Haykin, 1991). The basic idea is to use a set of training data (with known inputs and outputs) to fine tune the neural network, in order to solve the set of numeric constants (weights and biases) that result in the best fit of the training data. Therefore, the trained neural network model, using the best fitting constants, can make predictions on new data inputs with unknown outputs (McCaffrey, 2017; McCaffrey, 2012). Backpropagation uses mathematics to determine the direction and magnitude of neural network errors on the training data, and then modifies the weights and biases; the calculation is repeated until some stopping condition is reached.

We investigated the method of segmentation of human body from the background (Roh, et al., 2007); the method used Continuous Dynamic Programming (CDP) to extract the 2D of human shape from a low-quality video (a sport video). We also investigated the depth feature which is the importance feature of Kinect (Xia, Chen, & Aggarwal, 2011; Shotton et al., 2011; Patsadu, Nukoolkit, & Watanapa, 2012; Liu, et al., 2013). The recent works of using Kinect were also reviewed (Del Rio Guerra, & Martin-Gutierrez, 2020; Klishkovskaia, et al., 2020; Ma, Liu, & Cai, 2020; Fang, et al., 2014; Sengto, & Leauhatong, 2012). Del Rio Guerra, & Martin-Gutierrez (2020) used Kinect sensor to measure the gestures set. The 20 full-body gestures were analyzed and used as the requirements specification. This useful information can help the UX designer for designing the universal gesture based HCI (Human Computer Interaction) application to help individuals with Down Syndrome in daily living. Klishkovskaia, et al. (2020) proposed human posture detection with MATLAB. This work used the displacement of vector lengths and angles of 25 Kinect joints to analyze 13 basic exercises. This work also proved that the computational complexity was less than the deep learning-based algorithms. The analysis of daily activities using the assessment of upper body gesture was presented by Ma, Liu, & Cai (2020); the functional movement of shoulder and elbow was trained using the recurrent neural network. By comparison to the marker-based three-dimensional motion capture system (3DMC), this work yielded more accuracy in less laboratory space usage. The activity recognition in smart home using Backpropagation (BP) was also investigated by Fang, et al. (2014). This work used motion sensors embedded in each room to capture inter-class distance as the feature set, the activity classification was based on Activities of Daily Living (ADLs). The results of this work were compared with Naïve Bayes (NB) classifier and Hidden Markov Model (HMM). The fall detection technique using Back-propagation algorithm was reviewed (Sengto, & Leauhatong, 2012). This work used tri-axial accelerometer embedded in wearable wrist device to classify the ADLs and four falls (front fall, back fall, left fall, and right fall). The construction of BP of this work comprised of the following parameters: 300 input neurons, 90 hidden neurons (1 hidden layer), and 8 output neurons. The sophisticated works of deep learning-based sensor-based activity recognition were also shown in Su, Tong, & Ji (2014), Wang, et al. (2019), and Ordóñez, & Roggen (2016). Su, Tong, & Ji (2014) and Wang, et al. (2019) compared many algorithms in activity recognition such as Gaussian Mixture Models, Artificial Neural Networks (ANN), Rule-Based classifier, Support Vector Machine (SVM), and fuzzy inference system. Ordóñez, & Roggen (2016) used the Long-short-term memory recurrent (LSTM) together with Convolutional neural networks (CNN) to recognize the multimodal signal captured from wearable device and iBeacon embedded in the house.

The Kinect is the depth camera using the infrared to extract the human joints into the low-level feature. The Kinect camera can be captured 3D video with the frame rate 200 F/S, and the highest resolution is 1,280x960 pixels (Microsoft Developer Network, Tracking Users with Kinect Skeletal Tracking, 2014). Figure 1 shows the Kinect camera and human joints data captured from Kinect.

Figure 1 Kinect camera and human joints extracted from Kinect

The distance in x, y, z is the major feature captured from Kinect, Figure 2 shows the sensor direction. The distance in z direction is the depth distance between the human body to the Kinect (the distance in z direction is always positive and the magnitude of z would be increase if the object is far away from the camera). The appropriate distance between the object to the Kinect is 1.8-3.0 m.

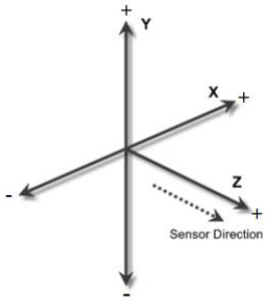


Figure 2 The Kinect direction in x, y, z

SYSTEM FRAMEWORK

The block diagram of our method shows in Figure 3. The method consists of 4 steps: 1) Video data captured from Kinect; 2) The depth data will be extracted in the feature extraction and converted to CSV (we called these chunk of data as “floor plan”); 3) The floor plan will be trained in BPNN as the motion classification; and in the last step (the real time exercise); the BPNN will be used to detect the elderly exercises. The system will alert if the elderly’s position is in the wrong pose to prevent the injury of muscle and joints. The class diagram of the left arm lift and the right arm lift in the exercise series is also shown in the appendix.

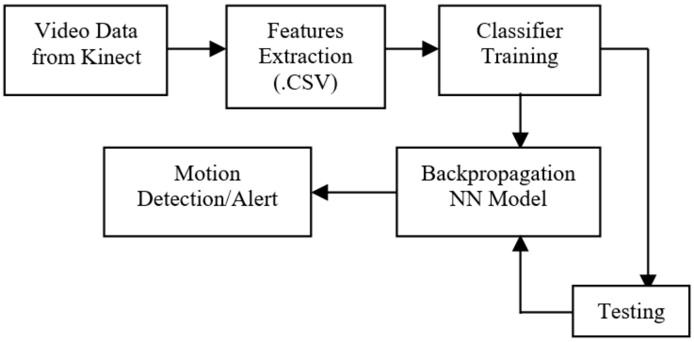
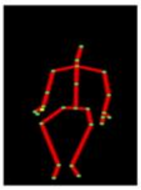
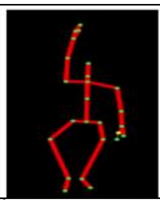
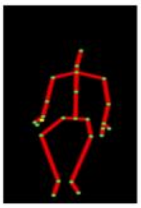
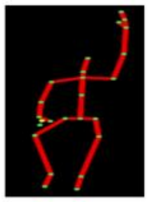
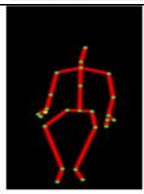
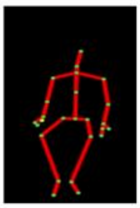
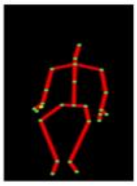
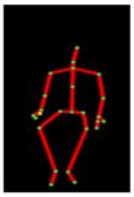
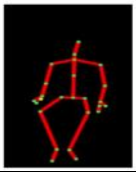
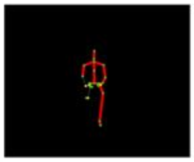
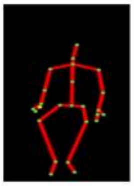
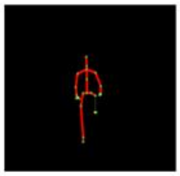
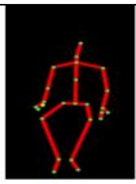
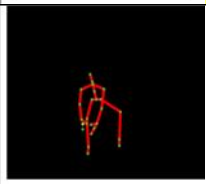
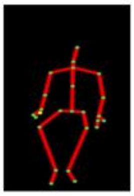
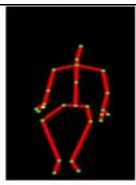
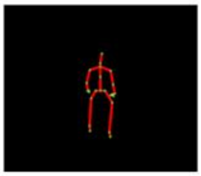
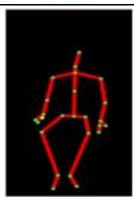
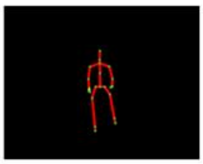


Figure 3 The system framework

In the data preparation, we trained the following of the elderly exercises using Kinect. The 16 exercises can be divided into 2 major exercise modes: the sit mode (exercise no. 1-12) and the standing mode (exercise no. 13-16) (Figure 4).

Exercises (Sit mode)				
1. Raise left hand				
2. Raise right hand				
3. Raise both hands				
4. Raise left hand to the front				

5. Raise right hand to the front				
6. Raise both hands to the front				
7. Raise left leg to the front				
8. Raise right leg to the front				
9. Raise left leg to the back				
10. Raise right leg to the back				
11. Left kick				
12. Right kick				

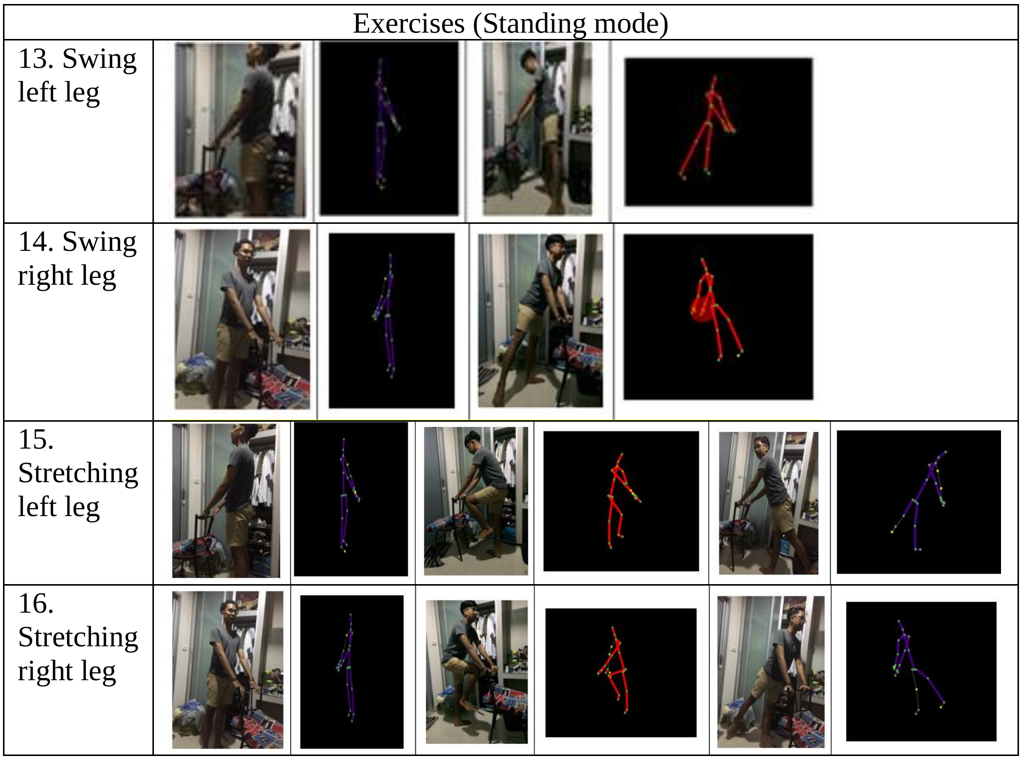


Figure 4 The 16 exercises and Kinect extraction

Table 1. shows the number of frames used in the data preparation. The total frames of the 16 exercises = 8,300 frames.

Table 1 The training data according to the exercises

Exercise No.	The training set (frames)
1	500
2	500
3	500
4	500
5	500
6	500
7	500
8	500
9	800
10	800
11	500
12	500
13	500
14	500
15	350
16	350

IMPLEMENTATION

The GUI design is shown in Figure 5. As we mentioned earlier, the system must be easy to use by the caregiver and the elderly, the GUI was designed using the low-fidelity concept to turn our ideas into testable artifacts and used to collect and analyze feedback until we meet all requirements from the caregiver and the elderly.

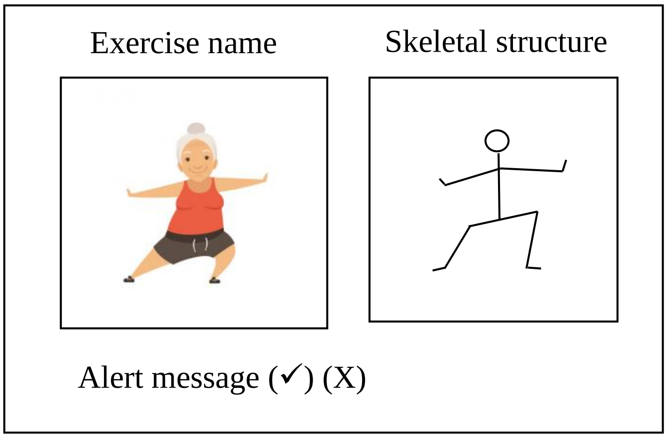


Figure 5 The GUI design

We used the backpropagation algorithm in C# for training neural networks; the original source code was developed by Breast Cancer Neural Network (2014). Based on the training frames, the model will be generated by determining the direction and magnitude of neural network errors, and then modifies the weights and biases. Like the other optimization techniques, the stopping rules is set and the calculation will be repeated until it reaches the condition (Visutsak, 2021). Figure 6 shows a generic model of BPNN; $x_1 \rightarrow x_n$ are nodes in an input layer (initial data for the neural network), this model has 1 hidden layer: node $1 \rightarrow m$ (intermediate layer between input and output layers and place where all the computation is done), and an output layer: $y_1 \rightarrow y_l$ (produce the result for given inputs). Our model used 25 hidden nodes in the second layer, and 2 outputs (one model per one exercise).

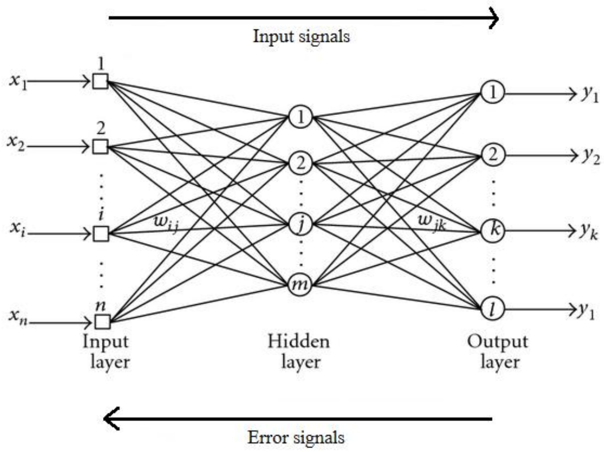


Figure 6 A generic model of backpropagation neural networks

The system can be divided to 5 parts:

1) Feature Selection: The floor plan calculation and the converting of the exercise data to CSV file. In this step, we created the data.csv file for storing the signals captured from Kinect. The depth data from Kinect will be read according to the joints. The set of joints = {hip left, hip right, knee left, knee right, neck, shoulder left, shoulder right, spine base, spine mid, spine shoulder, thumb left, thumb right, wrist left, wrist right, foot left, foot right}.

2) The NN settings: In this step, we created the input columns to fit the data in the csv file. We set the input nodes = the length of the input columns. We also set the number of hidden layers = 1 with the hidden nodes = 25, and the number of outputs = 2, respectively.

3) The data normalization: We computed the mean of the collected data. All data in each row will be subtracted by the mean value and the result will be squared. This normalization is used to eliminate those data that are extremely high or low.

4) The NN weight computing in CSV file (for the online mode): The model will be constructed as the results of this step.

5) Testing: The unknown exercise was performed, and the output of the classification was shown.

RESULTS

We show the screen captured of the system in the following figures:

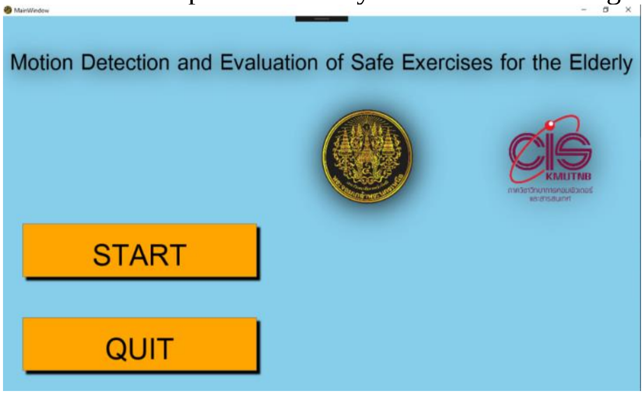


Figure 7 The start screen

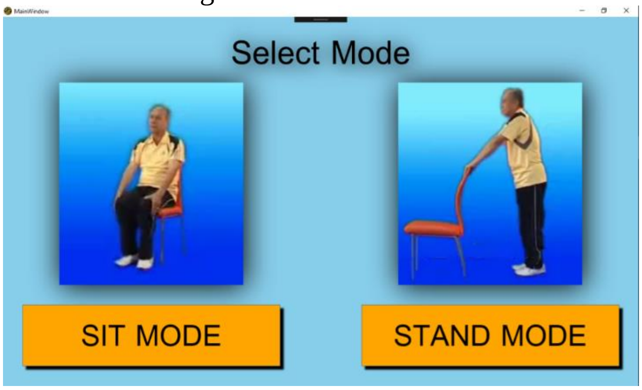


Figure 8 The 2 modes of exercise: sit mode and standing mode

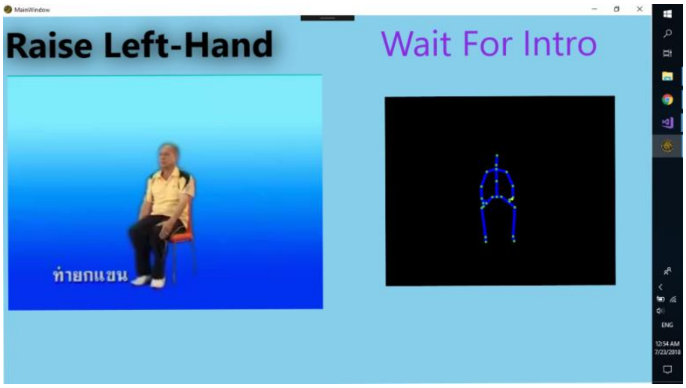


Figure 9 The user may prompt to see the demonstration of each exercise from the video clip shown on the left-hand side of the screen

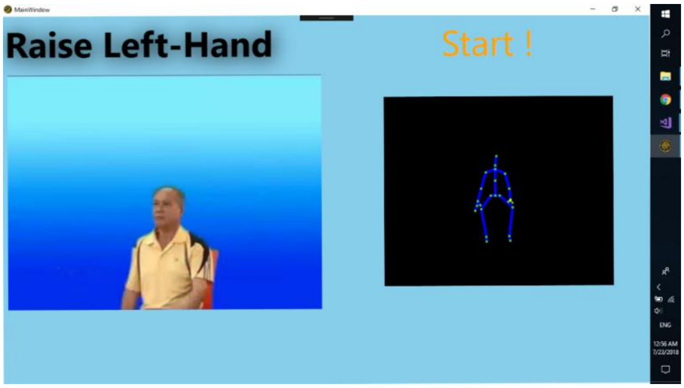


Figure 10 The user does the exercise in front of the Kinect camera

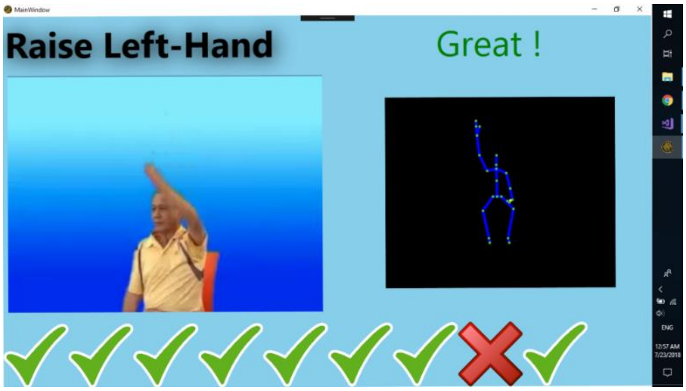


Figure 11 The user repeats the exercise and the system will show the results at the bottom of the screen (✓ indicates the correct pose, while × indicates the wrong pose)

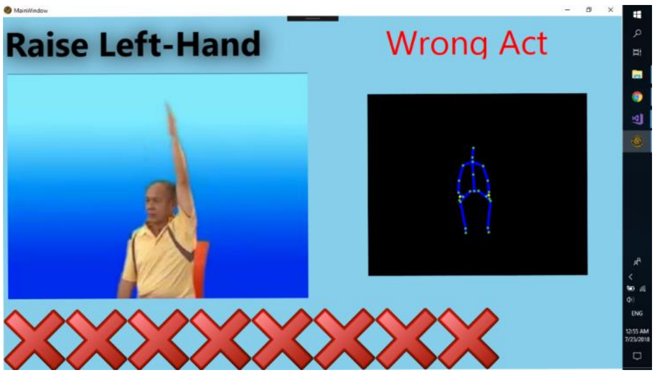


Figure 12 The system will alert if the wrong posture was detected

We used the following equations for the system evaluation:

$$\text{Sensitivity} = TP / (TP+FN) \tag{1}$$

$$\text{Specificity} = TN / (TN+FP) \tag{2}$$

$$\text{Accuracy} = TP+TN / (TN+TP+FN+FP) \tag{3}$$

Where TP is True Positive.
FP is False Positive.
TN is True Negative.
FN is False Negative.
Sensitivity is the ability to detect the correct posture.
Specificity is the ability to detect the wrong posture.
Accuracy is the overall of detection.

Table 2 shows an example of how to evaluate the system, the system was tested by letting the elderly do their exercises 10 repetitions per exercise, e.g. the elderly do 10 repetitions of exercise no. 3; exercise no.3 is in a sit mode; the elderly do the exercise by raising their arms over their head while they are sitting on the chair (see Figure 4). The result shows that the user does the right poses of exercise no.3 = 9 reps and the system also detects that they are the right poses. Whereas, the user does the wrong pose = 1 but the system detects that it is a right pose. Therefore, the accuracy of this test = 90%.

The test results of exercise No.1 is shown in table 2 with the accuracy = 100%.

Table 2 The Test Results of Exercise No. 3

Number of doing the exercise = 10		User:	
		Number of doing the wrong pose	Number of doing the correct pose
Program:	Number of detecting the wrong pose		
	Number of detecting the correct pose	1	9

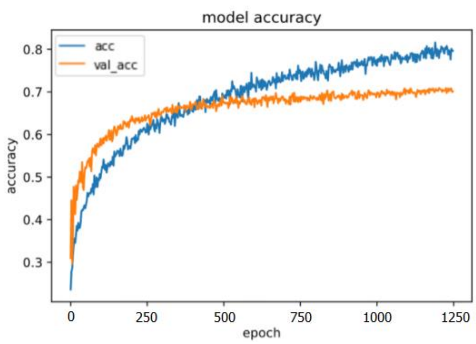
Table 3 shows classifying results based on the proposed model, the evaluation results (precision, recall and F1 score) are also shown. The proposed model can classify all 16 exercises. It gives 0.79 and 0.77 of average precision and recall, respectively and it gives 0.79 of average F1 score. It can be concluded that our BPNN model can classify class of basic activities with high accuracy.

Table 3 Classification Results Based on Precision, Recall, and F1 Score

Exercise No.	Precision	Recall	F1 Score
1	0.90	0.85	0.87
2	0.81	0.83	0.82
3	0.90	0.90	0.90
4	0.80	0.80	0.84
5	0.82	0.83	0.86
6	0.80	0.78	0.81
7	0.88	0.82	0.86
8	0.82	0.80	0.82
9	0.80	0.80	0.82
10	0.80	0.78	0.82
11	0.81	0.79	0.80
12	0.72	0.74	0.78
13	0.72	0.74	0.70
14	0.70	0.68	0.70
15	0.68	0.62	0.68
16	0.70	0.63	0.64
Average	**0.79**	**0.77**	**0.79**

CONCLUSIONS

The accurate result and the model loss from the proposed model are illustrated in Figure 13. The model can maintain the accuracy ≈ 0.7 when reaches 1,250 epochs, whereas the model loss ≈ 1.0. The system was exhibited as the show case in the National Science and Technology Fair 2018 Thailand (NST2018) during 16-26 August 2018 (Figure 14). We also collected the questionnaires from the users to use these feedbacks for the future development.



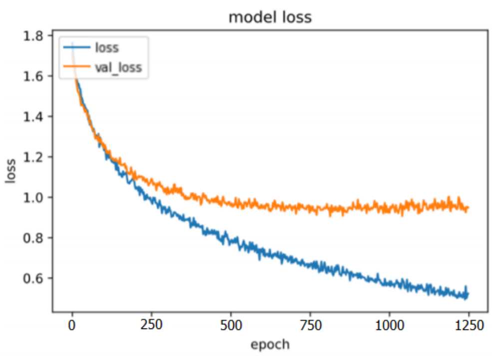


Figure 13 Detection results based on accuracy and model loss

Figure 14 The demonstration of the system at NST2018

DISCUSSION and FUTURE WORKS

The major problem of our experiment is the training set in exercise no. 15 and 16. Referred to Table 1, although the total of trained data = 8,300 frames but the unsuitable data in exercise no. 15 and 16 can affected the overall accuracy since we can separated the video clip of exercise no. 15 and 16 into 350 frames/exercise (the length of exercise no. 15 and 16 is only 1 minute, compare to the other exercises which the length = 2 minutes in average). Since the input of BPNN is a vector not exactly a series of image frames and the insufficient data is also the major problem of doing the experiment. To solve these problems, the data augmentation can be used to replicate the data that would be enough for training the neural network (Preechasuk, et al., 2019), unfortunately this solution is suit for the object recognition and classification of still image not the data in video format.

We have planned to develop the classification method by using Convolutional Neural Networks (CNNs) - Deep Learning. CNNs are another version of multilayer perceptrons but they take advantage over multilayer perceptrons since CNNs have hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme (Preechasuk, et al., 2019). We also have planned to develop the modified classification method (CNNs) with the Tai Chi exercises by using 2 Kinect cameras as the 2nd phase of our project (Motion Detection and Evaluation of Safe Tai Chi Exercises for the Elderly using multiple Kinect cameras).

ACKNOWLEDGEMENTS

This research was funded by the King Mongkut's University of Technology North Bangkok. Contract no. KMUTNB-GOV-60-53.

We would like to thank the Junior Research Fellowship Program 2017 (the Franco-Thai Junior Research Fellowship Program 2017) and Ambassade de France en Thailande for travel grant during the study at CRIStAL laboratory, Lille 1 University, Lille, France.

REFERENCES

Breast Cancer Neural Network. (2014). https://github.com/sebastianbk/BreastCancerNeuralNetwork (accessed on 6 July 2018).

Del Rio Guerra, M.S. & Martin-Gutierrez, J. (2020). Evaluation of Full-Body Gestures Performed by Individuals with Down Syndrome: Proposal for Designing User Interfaces for All Based on Kinect Sensor. Sensors, 20, 3930.

Fang, H., He, L., Si, H., Liu, P. & Xie, X. (2014). Human activity recognition based on feature selection in smart home using back-propagation algorithm. ISA transactions, 53(5), 1629-1638.

The Importance of Exercise for Seniors. (2014) https://www.asccare.com/importance-exercise-seniors/ (accessed on 6 July 2018).

Klishkovskaia, T., Aksenov, A., Sinitca, A., Zamansky, A., Markelov, O.A. & Kaplun, D. (2020). Development of Classification Algorithms for the Detection of Postures Using Non-Marker-Based Motion Capture Systems. Appl. Sci., 10, 4028.

Leung, H. & Haykin, S. (1991). The Complex Backpropagation Algorithm. IEEE Transactions on Signal Processing, vol. 39, no. 9, pp. 2101-2104.

Liu, Y., Zhang, Z., Li, A. & Wang, M. (2013). View independent human posture identification using Kinect. Biomedical Engineering and Informatics (BMEI), Proceedings of 5th International Conference on, October 16-18, Chongqing, China, pp. 1590-1593.

Ma, Y., Liu, D. & Cai, L. (2020). Deep Learning-Based Upper Limb Functional Assessment Using a Single Kinect v2 Sensor. Sensors, 20, 1903.

McCaffrey, J. (2012). Coding Neural Network Back-Propagation. https://jamesmccaffrey.wordpress.com/2012/11/20/coding-neural-network-back-propagation/.

McCaffrey, J. (2017). Deep Neural Network Training. https://docs.microsoft.com/en-us/archive/msdn-magazine/2017/september/test-run-deep-neural network-training.

Microsoft Developer Network, Tracking Users with Kinect Skeletal Tracking. (2014). https://msdn.microsoft.com/en-us/library/jj131025.aspx (accessed on 6 July 2018).

Ordóñez, F. J. & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. Sensors, 16(1), 115.

Patsadu, O., Nukoolkit, C. & Watanapa, B. (2012). Human Gesture Recognition using Kinect Camera. Computer Science and Software Engineering (JCSSE), Proceedings of International Joint Conference on, May 30–June 1, Bangkok, Thailand, pp. 28-32.

Preechasuk, J., Chaowalit, O, Pensiri, F. & Visutsak, P. (2019). Image Analysis of Mushroom Types Classification by Convolution Neural Networks, Proceedings of the 2nd Artificial Intelligence and Cloud Computing Conference (AICCC 2019). Association for Computing Machinery, New York, NY, USA, pp. 82–88.

Roh, M. C., Kim, T. Y., Park, J. & Lee, S. W. (2007). Accurate Object Contour Tracking based on Boundary Edge Selection. Pattern Recognition, 40(3), 931-943.

Sengto, A. & Leauhatong, T. (2012). Human falling detection algorithm using back propagation neural network. In The 5th Biomedical Engineering International Conference (pp. 1-5). IEEE.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A. & Blake, A. (2011). Real-Time Human Pose Recognition in Parts from Single Depth Images, Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, June 20-25, Providence, Rhode Island, USA, pp. 1297-1304.

Singh, D., Psychoula, I., Kropf, J., Hanke, S. & Holzinger, A. (2018). Users' Perceptions and Attitudes Towards Smart Home Technologies, Proceedings of International Conference on Smart Homes and Health Telematics, pp. 203-214.

Su, X., Tong, H. & Ji, P. (2014). Activity recognition with smartphone sensors. Tsinghua science and technology, 19(3), 235-249.

Visutsak, P. & Daoudi M. (2017). The Smart Home for the Elderly: Perceptions, Technologies, and Psychological Accessibilities: The Requirements Analysis for the Elderly in Thailand, Proceedings of XXVI International Conference on Information, Communication and Automation Technologies (ICAT), Sarajevo, Bosnia and Herzegovina, pp. 1-6.

Visutsak, P. (2021). Activity Classification Using Backpropagation Neural Networks for the Daily Lives of the Elderly. International Journal of Machine Learning and Computing vol. 11, no. 3, pp. 188-193.

Wang, J., Chen, Y., Hao, S., Peng, X. & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. Pattern Recognition Letters, 119, 3-11.

Xia, L., Chen, C. & Aggarwal, J. K. (2011). Human Detection using Depth Information by Kinect, Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference on, June 20-25, Colorado Springs, USA, pp. 15-22.