A Novel Concept of the Rehabilitation Training Coach Robot for Patients with Disability

Seung-Ho Han, Han-Gyu Kim and Ho-Jin Choi
School of Computing, Korea Advanced Institute of Science and Techology (KAIST), Daejeon, Korea
Email: {seunghohan,kimhangyu,hojinc}@kaist.ac.kr

***Abstract*—This paper proposes the rehabilitation treatment coach robot which will help at-home patients do their rehabilitation exercises at home without any professional trainers. The coach robot is designed to be cheap enough for patients to afford it. The robot suggests the rehabilitation program and corrects the posture of the patients during the exercise. The deep neural network is used for posture correction. Besides, the voice interface is applied for convenient interaction between robot and patients during the exercise. The emergency detection module is adopted which will inform doctors when emergency happens on patients. The emergency detection will be implemented using deep neural network on voice input and video input simultaneously. The detailed data collection plan for training deep neural network and performance evaluation plan are also provided in the paper.**

***Keywords*-rehabilitation training; coach robot; neural networks; exercise pose correction; emergency detection**

I. Introduction

The rehabilitation treatment plays important role in helping patients recover from various physical or neurological diseases, such as multiple sclerosis [1], [2], Parkinsons disease [3], [4] and stroke [5]. A patient should do proper exercise provided by the rehabilitation program for quick recovery.

Improper exercise is usually helpless, or even causes deterioration of the patient's health condition. Thus, for a proper treatment on patients, physical trainers are necessary. However, there are insufficient physical trainers to care all patients, because to become a certificated trainer is a long and difficult process. Besides, physical trainers often cost so much that patients cannot afford them. In addition, for at-home patients, personal physical trainer is not efficient because the trainer will cost most of time on the way to visit each patient.

In order to solve the problem described above, we propose a rehabilitation treatment coach robot. This robot plays the role of the personal trainer for rehabilitation treatment of the patient. In order to archive such purpose, the proposed robot should be able to suggest the training program for patients and correct the posture of patients during the rehabilitation exercise. Besides, we also propose to add the emergency detection function to the robot, with which the robot may call the doctor when serious emergency happens during the exercise, for example, patient's falling over.

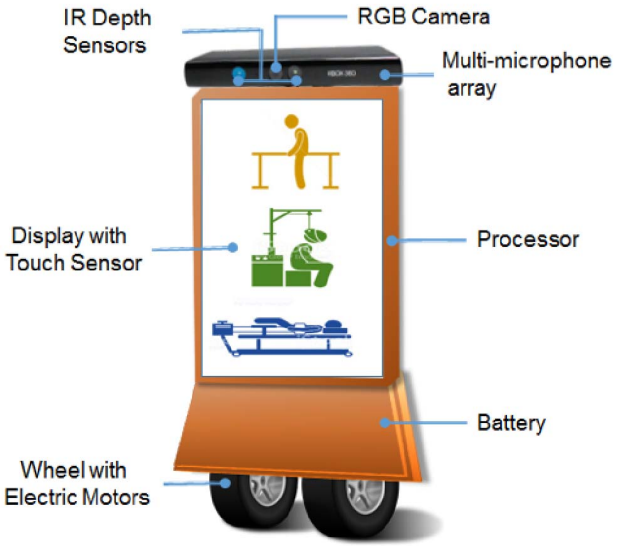

Figure 1. The design concept of the rehabilitation treatment coach robot

Most of the developed healthcare robots cost too much money that an individual cannot afford it [6]. The rehabilitation coach robot should be cheap enough that at-home patients may afford it. We will design a robot with simple hardware so that the price of the robot does not become too expensive. Besides, for convenience of the patients, we will integrate simple interface based on touch screen and voice recognition into the robot. The details of the proposed robot is described in the following section.

II. Method

A. Overview of the Robot Design

The most important factor we considered for designing the coach robot is the price, as it should be guaranteed that every patient can afford the robot. To make the coach robot as cheap as possible, we use basic hardwares to design the robot.

For the robots movement, simple wheels with electric motors are selected instead of robot feet with articular muscles, as robot feet are difficult to control and cost too much. Microsoft Kinect [7] is included in the robot design, as Kinect has various useful functions, such as video cameras

2375-0324/17 $31.00 © 2017 IEEE
DOI 10.1109/MDM.2017.65

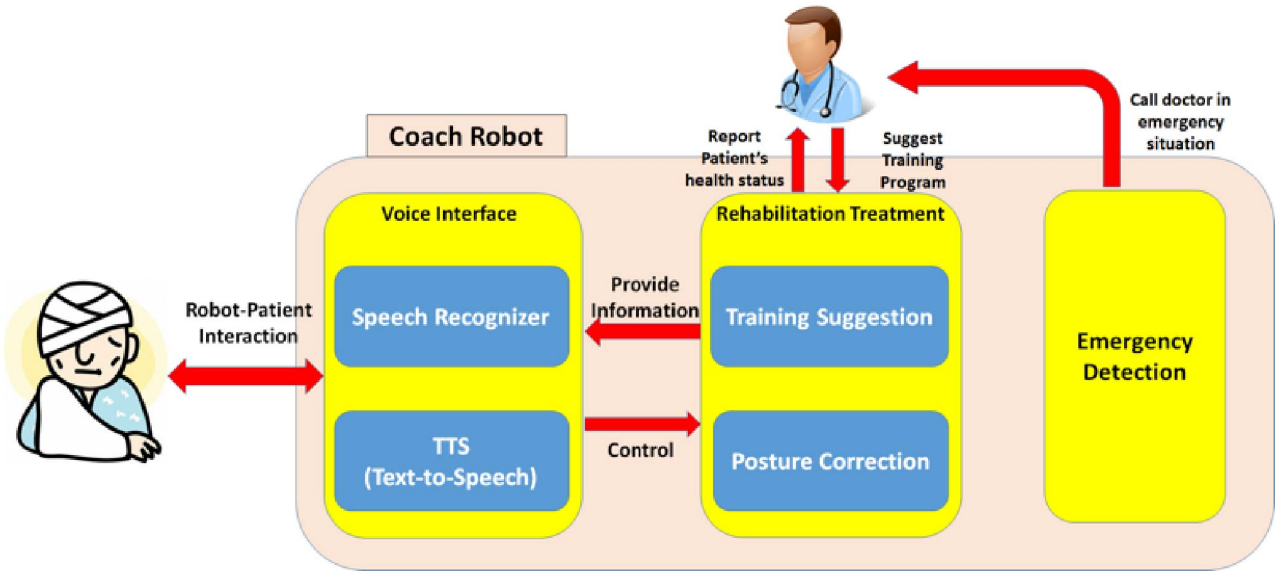

Figure 2. The functions of the coach robot

for recognizing patients, 3-D depth sensors for finding out the posture of the patients during exercising, and multi-microphone array for speech recognition. The monitor with touch sensor attached on it is also included in the robot for convenient robot control with touch interface. The speaker should also be included for the voice interface-based robot control. One processor should be integrated into the robot for image processing and voice recognition. The processor does not have to be extremely powerful as our proposed robot does not executes heavy computations. Thus, cheap processor is enough for our robot. The whole design concept of the above mentioned coach robot is shown in Figure 1.

For convenient control of the robot, the touch interface and voice interface is used simultaneously. The touch interface may be simply realized by designing and programming touch menus. The voice interface is more important compared to touch interface because it is inconvenient for patients to touch the screen during the physical exercise. Thus, we design to integrate the voice interface into our proposed coach robot. However, implementing the voice interface is much more difficult than implementing touch interface. The details of the implementation of the voice interface is described in the following subsections.

As mentioned in the introduction section, the main function of our proposed robot is to proper rehabilitation coaching to the patients. In order to achieve such purpose, two modules should be implemented: one is rehabilitation training program suggestion; the other is exercise posture correction. Besides, we will also implement emergency detection module to deal with the emergency during the rehabilitation exercising. The details for these modules are also explained in the following subsections. The above functions of the rehabilitation treatment coach robot is described in Figure 2.

B. Voice Interface Using Deep Neural Network

The voice interface is composed of two modules: speech recognition and speech synthesis. We will implement speech recognition and synthesis system of our own instead of using already-implemented application programming interface (API), such as Google speech API [8], because implementing task specific recognizer and synthesizer helps improving performance.

The traditional speech recognition system uses hidden markov models (HMM) with Gaussian mixture models (GMM) as the observation probability function [9]. Mel-frequency cepstrum coefficients (MFCC) are used as the audio feature vectors [10]. Recently, deep learning algorithms are proven to be effective in pattern recognition. The state-of-the-art speech recognition system replaces GMM with deep neural network (DNN) in the HMM-based recognizer [11]. Such speech recognizer based on HMM-DNN algorithm may be easily trained using open source speech recognition toolbox called Kaldi [12]. We will use Kaldi to implement the speech recognizer that is specialized in recognizing the commands that will be used to control our rehabilitation treatment coach robot.

In order to interact with patients, the text to speech (TTS) module is also needed. TTS should be close to the natural human voice so that patients feel comfortable while communicating with the robot. However, implementing TTS close to a human is a very difficult task. Recently, the Google Deepmind proposed deep recurrent neural network (RNN) based TTS called WaveNet [13]. WaveNet produces speech that is very close to the natural human voice. In our proposed

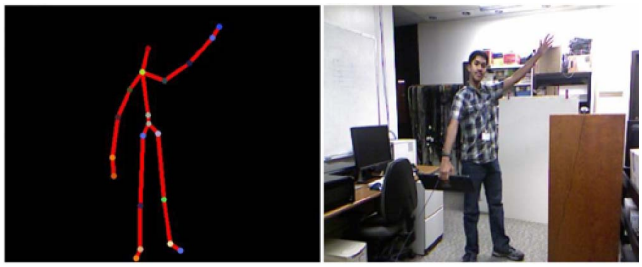

Figure 3. The original video and the extracted skeleton from the Kinect API

robot, we will adapt the WaveNet algorithm in implementing our own TTS.

C. Rehabilitation Training Program Suggestion

The main function of the proposed coach robot is to offer the service that a rehabilitation trainer may provide. For such purpose, the robot should be able to provide the rehabilitation training program suggestion to the patients.

Our coach robot will communicate with the doctors and provide the doctors training suggestion to the patients. The robot will check how well the patient has completed the training programs. Then, the robot will send a report about the patients health condition to the doctor periodically to obtain new training programs which fits the current status of the patient.

We will not implement the automatic suggestion system using machine learning methods, because the automatic suggestion necessarily accompanies the risk of the wrong suggestions. Even only one wrong suggestion may be greatly harmful to a patient. By providing training programs according to the suggestions of the doctors, the patients may achieve the most effective rehabilitation.

D. Rehabilitation Exercise Posture Correction

In order to achieve the main function of the robot of playing the role of a rehabilitation trainer, the most important module is the rehabilitation exercise posture correction module. Such module recognizes the posture of the patient, and when there exist any mistake in exercising posture, the robot will find out the mistake and guide the patient to perform proper exercise with correct posture.

In order to implement the posture correction, the posture should be recognized first. The skeleton of a target person may be easily obtained by Microsoft Kinect using API provided by Microsoft [14], [15]. The extracted skeleton is shown in Figure 3. The skeleton contains the information of the posture of the patient.

We will use deep learning algorithm to implement the pose correction module. The deep learning is used because it shows good performance in many areas such as image

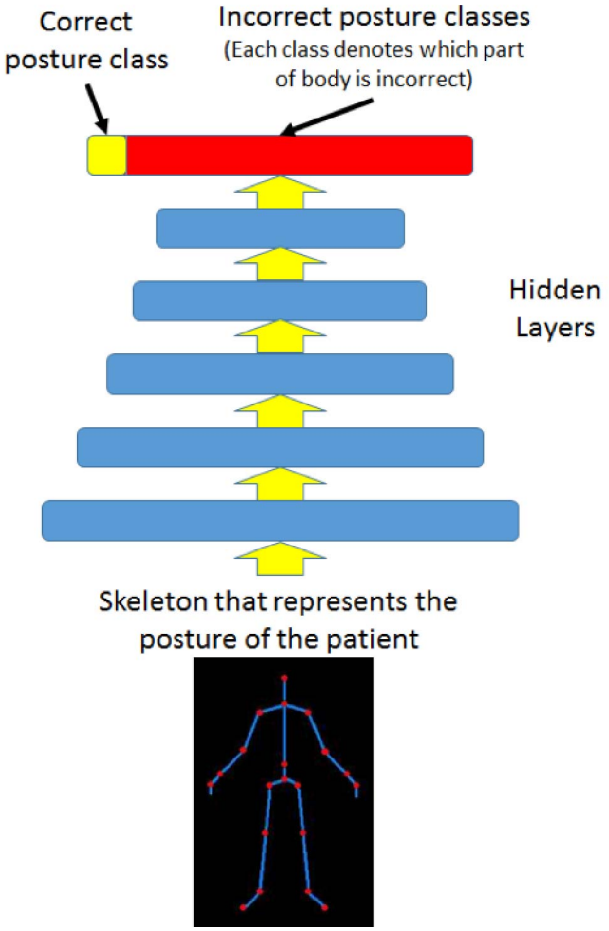

Figure 4. The DNN structure for posture correction

recognition [16] and natural language processing [17]. Besides, researches which are similar to the posture correction is already done using deep neural network and obtained good performance. Such researches include deep learning-based hand posture recognition [18], [19] and human activity recognition [20].

In order to implement a good DNN-based correction algorithm, a large size of data composed of correct cases and incorrect cases should be collected. The correct postures should be collected from trainers and the incorrect postures will be collected from the patients or ordinary persons. The incorrect postures will be categorized into several classes where each class is constructed according to the incorrect part of the posture.

With the dataset described above, we will train the DNN that can find out the incorrect posture of a patient. The DNN receives the skeleton of a person as input, and exports the class of the posture as the output. The output class may either be correct or the incorrect part of the posture. Several fully-connected hidden layers will be adopted in our DNN. The

number of hidden layers and the number of nodes in each hidden layer will be decided with performance evaluation experiments. Such DNN structure for posture correction is shown in Figure 4. Several techniques such as dropout [21] or batch normalization [22] will also be considered for make the best performance for our posture correction algorithm.

After the training of the DNN, in the inference step, the skeleton of a patient is extracted via Kinect. The skeleton is passed on to the DNN as the input. The DNN exports the incorrect part of the skeleton if the patients posture is not correct. The TTS module will create speech using the output of the DNN. Finally, the patient is informed of his/her incorrect posture with voice interface of the coach robot.

E. Emergency Detection

The rehabilitation exercises accompanies the risk of the emergency occurrence such as falling over, as the patients who does the exercises experiences inconvenience in moving their bodies. Even slight falling over may cause crucial result as the patients probably cannot overcome the emergency themselves. Therefore, we design to add the function of emergency detection into the coach robot. With the emergency detection module, the patients may do their rehabilitation exercises without anxiety, because the robot will call doctors or nurses for help when there happens emergency situation.

We will use both voice input from microphones and video input from cameras for emergency detection. The deep neural network technique will be adopted to analyse the input signals and find out the emergency situation. The deep neural network is used because, as mentioned above, deep neural network shows good performance in various areas.

The traditional emergency detection was based on specific sensors that is attached on human bodies [23], [24]. Such methods are not suitable for our robot as we will not use any extra sensors for the convenience of the patients. There have been several researches on emergency detection using audio sound. The basic idea of the audio-based emergency detection is to find out the voice that is accompanied by the emergency situation, such as scream or shout [25], [26]. There also have been several researches on detecting emergency from video streams [27], [28]. Besides, the emergency detection that fuses the audio signal and video signal is being researched actively [29]. In our system, we will use deep learning algorithm to improve the detection performance both for audio and video. We will adopt the fusion method proposed in [29] to gather the emergency detection results of the two different input signal together.

III. Experiment Plan

A. Data Collection Plan

The deep neural network needs a large size of dataset for training. The quality of the collected data is also important for high performance of DNN. We will collect datasets respectively for speech recognizer, speech synthesizer, posture correction and emergency detection modules.

For the speech recognizer, we will collect the speech from at least 50 persons of different sex and ages. Each person should pronounce all the commands that may be used to control the coach robot for at least 5 times. It will be much easier to collect the dataset for speech synthesize because the voice data of only one person needed. However, this person should pronounce each utterances for at least 10 times. The robot will seem friendly if the synthesized voice is listenable. Therefore, we will hire a voice actor to record the corpus for speech synthesizer.

For posture correction, there should be large number of correct cases as well as large number of incorrect cases. We will collect the correct postures for all possible exercises from the professional physical trainers. The incorrect postures will be collected from the patients and ordinary persons. The collected incorrection examples should be labeled by which part the posture is incorrect. Such labeling operation will be done manually by physical trainers. Only the labeled data may be used in training the neural network.

We will also collect the data for emergency detection. As it is hard to collect the data from real emergency situation, we will design some emergency scenarios and collect the dataset from those scenarios. Some possible scenarios are falling over, muscle cramps, screaming and shouting. More scenarios will be designed under the consultation of nurses and caregivers before collecting training data.

B. Evaluation Plan

Before launch the coach robot for the real world application, is should be evaluated to prove that the robot indeed help patients in taking rehabilitation exercises. There was no such type of robot before. Thus, we should design our own evaluation criteria for the coach robot. We will conduct two types of evaluation, which are objective evaluation and subjective evaluation.

The objective evaluation is implemented with some objective criteria such as recognition accuracy. We will evaluate the performance of the speech recognizer, posture correction and emergency detection modules with objective criteria. The speech recognizer will be evaluated using voice command recognition accuracy. The posture correction module will be evaluated using the accuracy of finding out the incorrect posture. The emergency detection algorithm will be estimated with false alarm rate which is generally used to evaluation detection system, indicating the proportion of the erroneous detecting of the target.

After the whole robot is designed, the subjective evaluation will be implemented in order to evaluate the overall performance of the robot, as the objective evaluation is difficult to be adopted in the whole system evaluation. The subjective evaluation will be conducted on two groups. One

group is composed of experts including trainers, nurses and doctors, and the other group is composed of patients.The experts are asked to evaluate how professional the coach robot is in assisting rehabilitation treatment. The patients will evaluate the convenience of using the coach robot. Besides, we will compare the health status of the patients with our coach robots and those without our coach robots, in order to prove the effectiveness of our proposed robot.

IV. Discussion

The voice interface of the proposed robot is only able to recognize and synthesize pre-defined commands. In the above part of the paper, applying natural language technique such as chatbot to our robot is not considered, because it is a complex algorithm and we need a powerful processor for real-time natural language processing, which may greatly increase the price of the robot. However, the robot which can talk naturally like human may be more friendly to the patients. Such robot will help improve mental stability of a patient who is having a hard rehabilitation training or communicate with a patient who is too sick to go out. Therefore, although is it not in our plan yet, we will try to integrate simple dialogue system into our robot.

V. Conclusions

In this paper, we propose the coach robot that helps outpatients with having rehabilitation training. The robot is designed with cheap parts, in order to make the patients can afford it. The main function of the coach robot is to suggest rehabilitation training program to patients, and correct the posture of the patients during the rehabilitation exercise. The robot will also adopt voice interface for convenient control, and emergency detection module for calling doctors in the emergency situation during the exercises. We plan to collect large size of datasets for training all modules adopted in the robot. We also provide the performance evaluation plan of the implemented coach robot.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government, Ministry of Science, ICT & Future Planning (MSIP)(No.2010-0028631)

References

[1] J. Freeman, D. Langdon, J. Hobart, and A. Thompson, "The impact of inpatient rehabilitation on progressive multiple sclerosis," *Annals of neurology*, vol. 42, no. 2, pp. 236–244, 1997.

[2] F. Patti, M. R. Ciancio, M. Cacopardo, E. Reggio, T. Fiorilla, F. Palermo, A. Reggio, and A. J. Thompson, "Effects of a short outpatient rehabilitation treatment on disability of multiple sclerosis patients," *Journal of neurology*, vol. 250, no. 7, pp. 861–866, 2003.

[3] G. Frazzitta, G. Bertotti, G. Riboldazzi, M. Turla, D. Uccellini, N. Boveri, G. Guaglio, M. Perini, C. Comi, P. Balbi *et al.*, "Effectiveness of intensive inpatient rehabilitation treatment on disease progression in parkinsonian patients: a randomized controlled trial with 1-year follow-up," *Neurorehabilitation and neural repair*, vol. 26, no. 2, pp. 144–150, 2012.

[4] G. Frazzitta, M. Morelli, G. Bertotti, G. Felicetti, G. Pezzoli, and R. Maestri, "Intensive rehabilitation treatment in parkinsonian patients with dyskinesias: a preliminary study with 6-month followup," *Parkinsons Disease*, vol. 2012, 2012.

[5] A. Henderson, N. Korner-Bitensky, and M. Levin, "Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery," *Topics in stroke rehabilitation*, vol. 14, no. 2, pp. 52–61, 2007.

[6] H. Iwata and S. Sugano, "Design of human symbiotic robot twendy-one," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 580–586.

[7] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[8] Y. Tian, Y. C. Liu, A. Bhosale, L. S. Huang, P. Tague, and C. Jackson, "All your screens are belong to us: Attacks exploiting the html5 screen sharing api," in *Security and Privacy (SP), 2014 IEEE Symposium on*. IEEE, 2014, pp. 34–48.

[9] L. R. Rabiner and B.-H. Juang, "Fundamentals of speech recognition," 1993.

[10] W. Han, C.-F. Chan, C.-S. Choy, and K.-P. Pun, "An efficient mfcc extraction method in speech recognition," in *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*. IEEE, 2006, pp. 4–pp.

[11] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 30–42, 2012.

[12] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," in *IEEE 2011 workshop on automatic speech recognition and understanding*, no. EPFL-CONF-192584. IEEE Signal Processing Society, 2011.

[13] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *CoRR abs/1609.03499*, 2016.

[14] Y. Chang, E.-S. Aziz, Z. Zhang, M. Zhang, S. Esche, and C. Chassapis, "A platform for mechanical assembly education using the microsoft kinect," in *ASME 2014 International Mechanical Engineering Congress and Exposition*. American Society of Mechanical Engineers, 2014, pp. V005T05A019–V005T05A019.

[15] M. A. Livingston, J. Sebastian, Z. Ai, and J. W. Decker, "Performance measurements for the microsoft kinect skeleton," in *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*. IEEE, 2012, pp. 119–120.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[17] A. Kumar, O. Irsoy, J. Su, J. Bradbury, R. English, B. Pierce, P. Ondruska, I. Gulrajani, and R. Socher, "Ask me anything: Dynamic memory networks for natural language processing," *CoRR, abs/1506.07285*, 2015.

[18] P. Barros, S. Magg, C. Weber, and S. Wermter, "A multichannel convolutional neural network for hand posture recognition," in *International Conference on Artificial Neural Networks*. Springer, 2014, pp. 403–410.

[19] A. Tang, K. Lu, Y. Wang, J. Huang, and H. Li, "A real-time hand posture recognition system using deep neural networks," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 6, no. 2, p. 21, 2015.

[20] K. Wang, X. Wang, L. Lin, M. Wang, and W. Zuo, "3d human activity recognition with reconfigurable convolutional neural networks," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 97–106.

[21] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[23] M. Cheon, J. Lee, W. Lee, C.-H. Hyun, and M. Park, "Intelligent emergency stop algorithm for a manipulator using a new regression method," *Sensors*, vol. 12, no. 6, pp. 7451–7467, 2012.

[24] Y. J. Yi and Y. S. Yu, "Emergency monitoring system based on a newly-developed fall detection algorithm," *Journal of information and communication convergence engineering*, vol. 11, no. 3, pp. 199–206, 2013.

[25] E. Principi, D. Fuselli, S. Squartini, M. Bonifazi, and F. Piazza, "A speech-based system for in-home emergency detection and remote assistance," in *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[26] E. Principi, S. Squartini, R. Bonfigli, G. Ferroni, and F. Piazza, "An integrated system for voice command recognition and emergency detection based on audio signals," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5668–5683, 2015.

[27] E. L. Andrade, R. B. Fisher, and S. Blunsden, "Detection of emergency events in crowded scenes," 2006.

[28] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Fall detection from human shape and motion history using video surveillance," in *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*, vol. 2. IEEE, 2007, pp. 875–880.

[29] G. Galatas, S. Ferdous, and F. Makedon, "Multi-modal person localization and emergency detection using the kinect," *Int. J. Adv. Res. Artif. Intell*, vol. 2, no. 1, pp. 41–46, 2013.