Arabic Command Based Human Computer Interaction

Hani S. Hassan1, Jammila Harbi S2 and Maisa’a Abid Ali Kodher1
1Computer science Department, University of Technology, Iraq
2Computer science Department, Mustansiriya University, Iraq
1Hani.co81@gmail.com
2dr.jameelahharbi@gmail.com
3 masiaa.Ali2015@gmail.com

Abstract. This paper means to build up a strategy for helping debilitated individuals old and amputee, by using unimodal human computer interaction; real time voice command and interaction between patient and computer where these mixes give a promising answer for help the incapacitated individuals. The fundamental goal of the work is to design a project that deals easily with a patient that uses Arabic commands. The project is based on training a dataset that collected from different persons that have ages from 12-58 years old, unknown word and background noise. The propose system can works with different direction in Arabic language ("ﺍﺫﻫﺐ","ﻳﻤﻴﻦ","ﻳﺴﺎﺭ") and with a multispeed that based on pulse width modulation (PWM). These functions based on code written in MATLAB environment. The project uses ATmega328microcontroller.

1. Introduction
Audio handling frameworks have been a piece of numerous individuals’ lives since the development of the phonograph during the 1870s. The subsequent series of developments started by that troublesome innovation have finished inevitably in the present versatile sound gadgets, for example, Apple's iPod, and the omnipresent MP3 (or also compacted) sound records that populate them. These might be tuned in to on compact gadgets, PCs, as soundtracks going with Blu-beam movies and DVDs, and in countless different spots [1]. Human PC interface/cooperation (HCI) may be called as Man-Machine interfacing or association. HCI was naturally spoken to with the emerging of PC. HCI term contains three parts (human, computer and interaction between them) [2]. The quality in the plan of HCI relies upon usefulness and ease of use. Usefulness of a framework is a lot of administrations or activities that is given to its people [3]. These days, robots interface intimately with a mix of people in their ordinary condition in region recreation, social insurance, nursing, stimulation, help impaired individuals. The quantity of individuals with handicaps has altogether expanded in these days [4]. the voice orders have been contemplated in a previous couple of years and becomes one of the real fields in HCI. Rakhi, et al. [5] (2013), proposed design "Automatic Wheelchair using Gesture Recognition". Many of the needs of humans with disabilities, human who cannot use their arms to strength a manual wheelchair this work uses head movements and infrared sensor integrated with wheelchair. In reality, this design inappropriate for amputee. Patil, et al. [6] (2014), planned the “Wheelchair Using Finger Operation with Image Processing Algorithms”. They present structure for helping incapacitated patient who have capacity to move fingers. This work depends on image handling and Vegard RISC processor (AVR) microcontroller for controlled wheelchair. They drive their wheelchair through finger activity utilizing image preparing by image location and acknowledgment dependent on transformation of finger picture from RGB to HSV. Undoubtedly, the procedure demonstrates getting pixels of nose district for face following. This plan suits some debilitated individuals who can move their fingers and does not oblige people who have experienced a mishap in the advantages and fingers and lower appendages because of this plan depend on image processing and this proposal suite it for amputee because of it depend on speech recognition by using discriminative model. Buvanswari et, at. [7] (2015), planned "Eye Scrutinized Wheelchair for People Affected with Tetraplegia ". His work presents a wheel situate system for the all-inclusive community that is affected by Tetraplegia. This work dependent on image handling in two stages. In first stage, iris acknowledgment by utilizing Gaussian haze work. In second stage, ascertain the focal point of understudy and send data to discover course. The wheelchair which is picked here uses the
Baughman's computation and depends upon the method for the eye to send requests to the microcontroller to control the wheel situate every which way and stops the movement. Additionally, Arduino board gets orders from PC for controlling wheelchair course. Hani Saeed Hassan et, at. [8] (2017) planned, "Human Computer Interface for Wheelchair Movement" the connection among human and PC by distinguishing a human face, following it and as needs be setting advanced stick of Arduino UNO stop, low or high. By utilizing MATLAB, face recognition calculation and CAM Shift calculation have been created with the method of thresholding. Also, proposed system work with stop, low and high. Amiel Hartman et, at.[9](2019) planned "Human-Machine Interface for a Smart Wheelchair". His work presents a wheelchair. It portrays the reconciliation of equipment and programming with sensor innovation and PC handling to build up the cutting-edge savvy wheelchair. This design is a PC group configuration to test elite processing for savvy wheelchair activity and human association. The LabVIEW bunch is produced for ongoing self-ruling way arranging and sensor information handling. Four little structure factor PCs are associated over a Gigabit Ethernet neighborhood to shape the PC group. At last, this work required four PCs are connected over Gigabit Ethernet thus, in contrast to proposed system that is single PC and robust its environment noise.

2. Overview

2.1. Human computer interaction

The human PC cooperation (HCI) framework comprises of a few sections that are fundamental to the plan of such frameworks. The human part ought to consistently think about what people ponder and require what physical ability confinements they may have how frameworks are debatable, what they discover appealing and fun when they use PCs. At the point when people cooperate with PCs, they check the gathering lifetime experience. Originators must choose how to make the items engaging without scattering clients from their obligations. The PC part comprises of PC or microcomputer, for example, Arduino Atmel ATmega328 for unique reason and advanced gadget communicate with a human. The interface part is a point where two articles meet. Communication happens among people and PCs. The human connection incorporates both programming and equipment. Individuals should utilize PCs or implanted gadgets which are diverse for various purposes, so they need to communicate with these gadgets. Scientists have fabricated various interfaces and techniques for those software engineers and fashioners to locate a sensible harmony between what can be modified inside the calendar and spending plan, and would be perfect for people [10]. In pattern recognition, there are two primary sorts of scientific model discriminative model and generative model the qualification between them Depends on probability distribution. In Discriminative models, straightforwardly register the likelihood of a yield given an information. In the other hand the generative model gives joint likelihood dissemination of the output and the input [11].

Speech recognition comprises a broad field, and is typically arranged by a couple of key clear expressions: Automatic speech recognition (ASR), Continuous speech recognition and Natural language processing (NLP). In A continuous speech recognition, depics a speech recognition framework that can perceive nonstop sentences of discourse. In principle, this would not require a client to delay when talking, and would incorporate correspondence and interpretation frameworks. The option is a discrete word acknowledgment framework, utilized fundamentally for taking care of vocal directions, that perceives single words delimited by stops [12]. Machine learning (ML) has become an adult and has reformed a few fields in registering and past, including human-PC communication (HCI). Human-subject investigations have been embracing Machine learning procedures for over 10 years, for instance for movement acknowledgment and wearable registering. There now additionally exists plenty of utilization areas where in ML methodologies are advancing intuitive registering research. Here we wish to feature a portion of the traps that HCI specialists ought to stay away from while utilizing Machine learning systems in their examination [13]. Machine learning is a subfield of artificial intelligence (AI), began to take off during the 1950s after the British mathematician Alan Turing distributed a progressive paper about the plausibility of contriving machines that think and learn. His well-known Turing Test surveys a machine's knowledge by verifying that if an individual can't recognize a machine from a person, the machine has genuine insight. Today, AI gives PCs the capacity to gain from named models and perceptions of information and to adjust when presented to new information rather than being unequivocally customized for each assignment. Specialists are creating PC projects to assemble models that distinguish designs, draw associations, and make expectations from
information to build educated choices about what to do straightaway. The consequences of AI are clear all over, from Facebook's personalization of every part's NewsFeed to discourse acknowledgment frameworks like money related market apparatuses, Siri, email spam filtration, proposal motors, for example, Netflix Amazon and language interpretation administrations [14]. The deep learning strategies have been as of late proposed as a plausible system to perform start to finish grouping models, wherein preprocessing and highlight extraction steps can be forget about. Specifically, novel techniques for dissecting discourse sign utilizing deep learning methodologies have been proposed, either as highlight extraction draws near as category models that utilization conventional acoustic descriptors [15]. There are various hotspots for huge information, which may incorporate utilizing existing data, acquiring information directing a battle to record the necessary information, etc. The absolute most intriguing information is what as of now exists and has been caught – unintentionally or something else – alongside other data [16]. Convolutional Neural Network has had weighty outcomes over the previous decade in an assortment of fields identified with design pattern recognition; from voice, handling to image discover. The most advantageous part of CNNs is lessening the quantity of parameters in Artificial Neural Networks (ANN). This accomplishment has provoked the two scientists and engineers to approach bigger models so as to illuminate complex undertakings. To acquire good recognition of CNN must be began with elements Convolution, Stride, Padding and Feature of CNNs [17]. By using a parametric rectified linear unit, it improves exactness at insignificant extra computational expense equation (1) shown how parametric rectified linear unit works, any input value less than zero it multiplied by scaler learned [18].

\[
\sigma(z) = \begin{cases} 
  z, & x > 0 \\
  0, & x \leq 0 
\end{cases} \tag{1}
\]

\(\sigma(z)\) activates the neurons of the layer.

In the training option there are many algorithms to perform this process such as derived from adaptive moment estimation (Adam), root mean square propagation (Rmsprop) [19] and the stochastic gradient descent with momentum (SGDM), the last algorithm used in the proposed system. The SGDM algorithm apple to sway along the way of steepest plummet towards the optimum. it update is by the equation (2) [20]

\[
\theta_{t+1} = \theta_t - \alpha \nabla E(\theta_t) + \gamma(\theta_t - \theta_{t-1}) \tag{2}
\]

\(\gamma\): decides the commitment of the past previous gradient step to the present cycle.
\(a\): \(a > 0\) is the learning rate.
\(\theta\): parameter vector.
\(E(\theta)\): loss function.

Voice upgrade is the route toward improving the nature of the voice signal by decreasing the establishment fuss and other unwanted sounds from the voice signal. voice signal quality is much of the time weighed by its clarity, lucidity, and appeal. voice improvement is a preliminary framework in the talk dealing with an area, including discourse blend, discourse coding, discourse recognition, and discourse investigation. voice sign recorded in a consistent circumstance may contain unwanted sound, for instance, playing uproarious statement by people, sound of fan, climate control system, etc. These are considered under the class of racket. To the group of spectator’s individuals, these impediments are significantly unpalatable and should be diminished in the solicitation to improve the quality and rationality of talk signal. Also, the talking sign getting ready counts to depend on the doubt that the voice sign is free from establishment uproar. The proximity of establishment uproar in voice sign will cut down the introduction of the talk dealing with structure basically [21].

3. Proposed system

The proposed framework consists of two parts recognizing direction of wheelchair and navigation of wheelchair that is shown in Figure 1. The first part consists of four stages gathering dataset, preprocessing of dataset, training dataset and recognizing voice command. Firstly, gathering real dataset from different humans (man and woman) with different ages (12-58) years old that consist of an isolated word of voice command recorded in Arabic language ("اذهب","يمن","يسار","قف") by using MATLAB each dataset has a properties such length 1 second, .wav format, Bit rate 256 Kbps, sampling rate 16000, Nbit 16 and number of channel 1. In another hand, we recorded a noise environment and unknown voice command for enhancing recognize. The background noise datasets with the different length but the
other properties like the isolated word command but the unknown voice command datasets properties like the isolated word voice command properties. All datasets equal 4500 files that are consist of isolated word voice command, background noise and unknown voice command. Secondly, preprocessing of dataset (recorded of isolated word command) by using Butterworth (bandpass filter 500-5000 Hz) and wiener filters for acquiring a pure waveform of isolated Arabic words commands ("ﺍﺫﻫﺐ","ﻳﻤﻴﻦ","ﻳﺴﺎﺭ"ﻭ"ﻗﻒ").

Thirdly, preparing neural systems is most effortless when the contributions to the system have a sensibly smooth conveyance and are standardized. After gathering all the datasets must be handle it before training such as separated (isolated word command) into training and validation and prepare the speech waveform for effectiveness training by transform the waveform into log-mel spectrograms and then obtain data with a smoother distribution by compute the spectrum and normalize it to the logarithm of the spectrograms with make an equivalent clips number (4000) of background noise. Compute spectrum of background noise and split spectrum between training and validation datasets. An expansion of the efficient size of the training data and help keep the net system from overfitting, creates more image datastore during input and training. Define a neural network geometry by using convolutional 2D layers, batch normalization layers and max pooling layers. Our neural network have size 24178.7236KB and the network geometry consists of 10 layers with 40 filters and using interesting nonlinear activation functions, parametric rectified linear unit at training time it used to multiply input value that less than zero by scaler learned. The proposed training network used 50 epochs for training and select SGDM algorithm with momentum 0.95. Proposed system used the shuffle (validation and training data) at each epoch before training and learning rate every number of epoch by projection learning rate during training that based on multiplied with factor (0.1) at each (20) number of epoch. At the end, the training of datasets is end with training error 4.971% and validation error 23.7729%. Lastly, recognizing voice command that classification a voice command depend it on training datasets at the pervious stage. In this stage specify the sampling rate to (16e3) and using a microphone to input voice from the current patient that use a wheelchair. Also, indicate parameters for the spilling spectrogram calculations and introduce a initialize a buffer for input voice that’s mean grab the classification label and configure the buffer for input the human voice. Figure 2 shows the flow chart of recognizing direction of proposed system. The second part, navigation of wheelchair based on detection voice command in Arabic language such as ("ﺍﺫﻫﺐ","ﻳﻤﻴﻦ","ﻳﺴﺎﺭ"ﻭ"ﻗﻒ") that acquired. The command is sent to Arduino via serial port then the Arduino translate the command into function and it supply the different voltage to the two-wheel drive motors based on technique pulse width modulation (PWM). Such function, when the command is "ﺍﺫﻫﺐ" the wheelchair goes to forward.

Figure 1. Diagram of proposed system
4. Result and Analysis
Subsequent to building up the entire framework, proposed wheelchair project is made and testing it. The entire framework based on input microphone and training datasets with different persons (record of isolated word command) with preprocessing, unknown words and background noise. All these datasets training by using deep learning for recognize voice command ("اذهب","يمين","يسار"و"قف") and then communication with Arduino uno and two-wheel drive motors. All these processes depict in. Figure 3 shows the preprocessing datasets (isolated word command) by using Butterworth bandpass filter (500-5000) Hz. Figure 4 that shows draw the spectrum and waveforms of some training examples such as ("اذهب"و"قف") and unknown word. Figure 5 shows the smoothing of data distribution of the datasets. Figure 6 shows the different class labels in the training and validation sets. Figure 7 shows the confusion matrix of accuracy and review for each class the largest confusion between orders ("اذهب","يمن","يسار"و"قف").Figure 8 shows the wheelchair prototype that consist of Arduino Uno, electronic circuit and two wheel drive(M1 and M2). It should be mentioned that all hardware component of prototype works with 5 DC voltage.

Figure 2. Flow chart of recognizing direction
Figure 3. Preprocessing of waveform dataset, $a(a)$, $b(b)$, $c(c)$, and $d(d)$, filtered by Butterworth.
Figure 4. Plot of some training examples. a: waveform and b: spectrum

Figure 5. Histogram of distribution of pixel value
Table 1 shows the precision of movement of wheelchair when the patient talks via microphone by using one of the commands ("أذهب", "يمين", "يسار", "وقف"). All accuracy of the commands is calculated by using the equation (3).

\[
\text{Precision} = \frac{(\text{true movement} + \text{true detect})}{(\text{false detect} + \text{false movement} + \text{true detect} + \text{true movement})} \times 100 \% \quad \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \ cd
5. Conclusion
The aim of our proposed system is to propose HCI system project wheelchair for helping elderly, disabled people and amputee. This proposal uses recognize in real-time Arabic speech and this system depends on the discriminative model, training recorded Arabic voice orders real dataset with varied background noise and unknown words with filter to enhance recognition. Trained a real dataset by using a deep learning and the validation accuracy result is 76.23% and the high precision is forward direction have 85% Our system robust against background noise, recorded different background noise from different places such as people talking, mall, restatement, market, street,…etc Voice commands of the patient via a microphone that interpreted into signal via a microcontroller, this microcontroller must be upheld by Pulse Width Modulation (PWM) to create diverse voltage D.C dependent on voice orders.

References
[1] McLoughlin, Ian. 2016 “Speech and Audio Processing”, 1st ed. Padstow Cornwall, TJ International Ltd.1-6p.
[2] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, Mo Nours Arab. 2008. “Human-Computer Interaction” International Journal on Smart Sensing and Intelligent Systems, vol.1, no.1.
[3] Dov Te’eni, Jane M Carey, Ping Zhang, 2007. Human Computer Interaction: Developing Effective Organizational Information Systems”, 1st ed. Hoboken, Wiley.
[4] Chhabria, S. A. Dharaskar, 2011. “Multimodal Interface for Disable Persons” International Journal of Computer Science and Communication, vol. 2, no.1, p. 223-225.
[5] Rakhi A. Kalantri, D.K. Chitre, 2013. “Automatic Wheelchair using Gesture Recognition” International Journal of Engineering and Advanced Technology, vol.2, no.6.
[6] Chhaya G.Patil, Sayali K.Gharge, Sonal V.Modhave, Y.S.Angal, 2014. Design of “Wheelchair Using Finger Operation With Image Processing Algorithm” International Journal of Research in Engineering and Technology, vol.3, no.2, p. 232.
[7] B.Buvanswari, T.Kalpalatha Reddy, 2015. “Eye Scrutinized Wheelchair for People Affected with Tetraplegia” International Journal of Computer Science, Engineering and Information Technology, vol.5, no.2.
[8] Hani S. 2017. Multimedia Based Human Computer Interaction MSc Thesis University of technology, Iraq.
[9] Amiel Hartman, Vidya K. Nandikolla, 2019. Human-Machine Interface for a Smart Wheelchair, journal of Robotics, 4837058.
[10] Zahra, K. 2013. Techniques for Real-time Multi-person Face Tracking for Human-robot Dialogue. M.Sc. Dissertation, Blekinge Institute of Technology, School of Engineering.
[11] Ch Chen, 2016. Handbook of Pattern Recognition and Computer Vision, 5th ed. World scientific publishing Co. pte.Ltd.U.S.A. 27p.
[12] Ian McLoughlin, 2009, Applied Speech and Audio Processing, 1st ed. Cambridge University Press, 174-176p.
[13] Vassilis Kostakos, Mirco Musolesi, 2017, Avoiding Pitfalls When Using Machine Learning in HCI Studies, ACM, 24(4):5556, Doi.org/10.1145/3085556.
[14] Masoud Hemmatpour, Renato Ferrero, Bartolomeo Montrucchio, Maurizio Rebaudengo, 2019, A Review on Fall Prediction and Prevention System for Personal Devices: Evaluation and Experimental Results, Advances in Human-Computer Interaction. Doi.org/10.1155/2019/9610567.
[15] Fu, J., Mao, Q., Tu, J., Zhan, Y., 2019, Multimodal shared features learning for emotion recognition by enhanced sparse local discriminative canonical correlation analysis. Multimed. Springer.25(5), doi.org/10.1007/s00530-017-0547-8.
[16] Ian Vince McLoughlin, 2016, *Speech and Audio Processing*, 1st ed. Cambridge University Press, 223-225p.

[17] Saad Albawi, Tareq A. 2017, Understanding of a Convolutional Neural Network, The International Conference on Engineering and Technology, 2017, At Antalya, Turkey.

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, 2016, Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, arxiv.org/abs/1502.01852.

[19] Diederik P. Kingma, Jimmy Lei Ba, 2017, adam: an explanation for stochastic optimization, conference paper at the 3rd International Conference for Learning Representations, San Diego.

[20] A. Yashwanth, Shaik Shammer, R. Sairam, G. Chamundeeswari, 2019, “A novel approach for indoor-outdoor scene classification using transfer learning” International Journal of Advance Research, Ideas and Innovations in Technology, vol. 5, no. 2, p. 1759.

[21] Shajeez K U, Sachin Kumar S, Pravena D, Soman K P, 2012, “Speech Enhancement based on Savitzky–Golay Smoothing Filter” International Journal of Computer Applications, vol. 21, no. (0975 – 8887), p. 112-6.