

## Real-Time Human Fall Detection and Emotion Recognition using Embedded Device and Deep Learning

Illuri Sreenidhi<sup>1</sup>, P. Satyanarayana<sup>2</sup>

<sup>1</sup>M.Tech student in Embedded Systems, KLEF, Vaddeswaram, Andhra Pradesh, India, sreenidhi2910@gmail.com

<sup>2</sup>Professor, Department of ECM, KLEF, Vaddeswaram, Andhra Pradesh, India satece@kluniversity.in

### ABSTRACT

The fall detection is one of the most discussed topic among the researchers because of its usage in home care and medical places for the patients. Because death due to falls is considered as most dangerous and injuries sustained from these incidents are pretty fatal and serious. Sometimes if the young age person falls he might suffer from minor injuries but we cannot say the same to the older aged person. Sometimes even though the person is suffered from the fall we can save the person by providing immediate care. Because there are some cases in which the death occurred by not responding quickly. Raspberry pi is used as a embedded device and Mog2 algorithm is used for video analysis. After detecting the fall the people are notified using internet based REST APIs like TWILIO. And for emotion recognition it is mostly used to detect the emotional state of the person. This system is mostly used in health care and marketing sectors to provide better results for the people. For emotion recognition system we use CNN (Convolutional Neural Network) since it considered as the best classifier for the video based recognition system. A data set with various emotions is trained and introduced to the system. For both these systems by using deep learning algorithm with embedded device and OpenCV we can provide better results and the performance of the systems are improved.

**Key words :** OpenCV, Raspberry Pi4, Deep Learning, Twilio, Mog2 Algorithm, CNN(Convolutional Neural Network), REST APIs.

### 1. INTRODUCTION

Presently, falls are regarded to be the most potentially fatal to humans regardless of age. According to WHO (World Health Organization) reports, falling is the fourth major cause of death throughout the world. From WHO survey it is confirmed that the death due to fall majorly occurs in the ages between 65-70. One of the review or study deduced that about

1.8 million of people around the world are suffering from fall related issues. This figure may rise to 15% by the year 2025. One of the fact in the world is about 90% of the elderly people or senior or retired citizens in the world live in the assistant homes. A dangerous fall can cause disaster or serious disability. Most of the falls happen accidentally in our routine activities. Occasionally these falls can be fatal, and others may be healed by appropriate treatment [8]. Falls are more frequent in aged people due to critical health problems. Falls typically occur whilst standing and sitting, and sometimes when walking. According to some research studies, falls are quite common in older people and patients living in residential homes, as well as in medical facilities [3]. The risk of falling grows with an increase in age. According to some studies there is a 40% increase in falls of older aged people and patients and they suffer severe injuries and sometimes critical treatment, but there are chances that they might suffer with minor injuries such as bruises, or sprains. But due to the calcium deficiency in such age the even a small or minor fall can lead to fracture or bones breaking. And these falls sometimes leads to brain injuries like concussion. Such accidents will always cause some serious psychological effects like PTSD for elders, and such stress in older age will lead to heart problems too. So it doesn't matter whether the fall is major or minor there will always be consequences. But such injuries can be sometimes fatal and theses can be avoided if we detect the fall immediately and the patients were treated with proper care this is only possible if we alert the medical personal fast [1]. This system of fall detection is the most common studied topic for the researchers just to provide better care for the people and to avoid any impossible situations as much as possible. There are two ways related to fall one is fall prevention and the other is fall detection. Fall prevention systems usually predicts the fall using posture and physical analysis of the person who might have a chance of falling and prevents them by alerting the concerned person [5]. Fall detection system detects the fall and alerts the authorities. The detection of the fall have different kind of approaches, such as sensor based, vision based, auxiliary equipment based, and now-a-days there are that can detect and predict the fall using

machine learning algorithms too. In every system the data from the various methods used are collected and analyzed using their respective algorithms [4]. These systems need to be trained beforehand so that when the system will run without any glitches. The algorithms provide the decision regarding the fall with help of data provided [2]. In this paper we will discuss the fall detection using vision based system and deep learning algorithm. The algorithm used for this video analysis is MoG2 algorithm which is totally based on the Gaussian mixture models. And after detecting the fall the authorities are alerted using internet rest APIs like TWILIO. Whereas, due to its popularity and high utilization in both commercial and academic potential, the Emotion Recognition System appears to be one of the most studied or researched topics in recent times. With the help of emotion recognition system we can know the emotional state of the patient and give the proper care. This system also helps to prevent the fall detection in some cases. If we can identify in which emotional state the patient is in it can really be helpful to provide a better care for patient. The accuracy of this system is considered to be 90% [13]. Mostly this system follows the facial expression of the person in the real time video. Sometimes the people may not identify the accurate state of emotion of the person but when we give the pre processed data to a system and by using machine learning algorithm the computer can identify the emotion using the face muscle movement and facial expressions. With the help of machine learning the dataset we trained and uploaded in to the systems algorithm is closely supervised and the prediction of emotion of the person is made easy [18]. Some of the algorithms used to predict the emotion state are Bayesian networks, Gaussian mixture models, Hidden Markov Models and Neural Networks [14]. But in this proposed system we use Deep Learning Algorithm like Convolutional Neural Network (CNN) and Computer Vision for the accurate prediction of emotion [16]. The dataset used in this system are provided with the series of images of various emotions and their facial expressions [12]. This dataset should be trained for the usage of system without any glitches or backsets. The CNN algorithm requires many parameters for their deployment of the system so we take the real time systems before they become unpredictable [10]. So for overcoming this problem we introduced a visualization system which is real time with back-propagation which is proposed by few researchers.

## 2. BACKGROUND STUDY

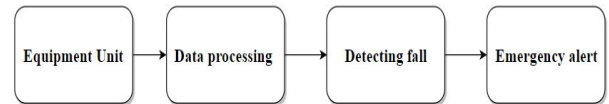
### 2.1 Fall Detection

Usually fall related systems are reviewed by the researchers based upon a few categories such as Data availability, using specific related category, on the perspective of energy efficiency and extensive survey on different methods used for fall detection or prevention [7]. Here the specific categories for the fall detection or prevention are Smartphone based,

multisensory based, Accelerometer based, and Vision based. Usually fall related systems are classified into two different ways like Device or Equipment based and Vision based. The basic fall detection process is given in fig below

#### 2.1.1. Device or Equipment based

The systems which are used for detecting the fall that are developed using devices like sensors, embedded devices or wearable devices which are mostly used to monitor the patients in the indoor environments are said to be described as



**Figure 1:** General fall detection process

Equipment based [1]. These devices are used to monitor the patient and their movements and then store the data. The sensors used in such devices are PIR, Accelerometers, Gyroscope, pressure sensors and Hybrid sensors. And basically all these sensors are used in Smartphone so tracking with phones are found as an easy process [6]. Whereas the ambient sensors like vibration, PIR, and pressure are placed in various places in the room of the patient and to alert the person and record the data of the patient a base station is placed to detect the fall. The sensors are divided into different groups based on the system requirement. The groups are Context based, Sensor fusion Based, Inertial Based, and RF based. The main work of these sensors or any sensors is to collect the data from the patient body like any sudden movements or unexpected disturbances [4]. In order to collect the data the sensors use Micro Electrical and Mechanical devices such as barometers, gyroscopes, accelerometers, and magnetometers. These devices were placed in the various parts of the patient's body to track the muscle movements and sudden movements of the body. The sensors like RF based work by using the Radar and radio signals or frequency and communicate using Wifi to evaluate the fall. Some sensors are fused with other sensors just to provide additional information and to improve the accuracy of the system so that it can provide better results for predicting the fall or sometimes detecting it. The Smartphone based devices collect the raw data which are used for accuracy of the fall without any disturbances created when the filtering process taken place, so these systems predict or detect the accurate falls. The Smartphone system are said to be 95% accurate [6]. The most common device used by many patients is the wearable device system. The smart watches were most common device used by almost all people irrespective of the age and health. Since they track every person closely and precisely, and even the alert system in these devices provides accurate results. But one drawback in this system is the battery, every once in while the watches should be charged or any such devices which work under the battery usage have this drawback. So if any fall occurs in this time it is difficult to know or alert. So for people

who are severely unable to move or unpredictable health issues such systems are totally avoided. The accuracy of the wearable devices are said to be 71% and for the ambient sensors it is 91%.

### 2.1.2. Vision based

The vision based system is mostly considered by the scientists and researchers because of the limitations created by the Device based systems. The main reason is due to cameras that are fixed in every household now-a-days; they are mostly used to track any movements or activities of the residents as well as the patients. The camera usually gives information about the person, such as actions, position and motion. Standard cameras are replaced with depth cameras, multiple cameras, and RGB cameras, and multiple cameras for more reliable and accurate performance [9]. The movements and activities of the person are captured by these cameras and analyze them by using the algorithms provided in the system to predict or detect the fall. There are a few algorithms which can detect the fall by analyzing the position and posture of the person. Some of the machine learning algorithms which are used in fall detection are HMM (Hidden Markov Models), SVM (Support Vector Machines), and Neural Networks like CNN (Convolutional Neural Networks), R-CNN, ANN. The vision based system are more reliable than Device based system. In this proposed system the camera is paired with embedded device to provide better results, and when the fall occurs the concerned authorities were notified by sending a message using internet based APIs.

### 2.2 Emotion Recognition

There are many research topics that discuss the effectiveness of the emotion recognition and its automated version. There are many proposals that use the emotion recognition in various methods like Image processing, Signal Processing, computer Vision, Bayesian Networks, Neural Networks, Gaussian mixture models, Machine Learning algorithms, Hidden Markov Models, and Supporting Vector Machines. Each and every system provides better accuracy than the other based on the human expressions, psychology and physiology of the facial movements based on the video analysis of the system [11]. Different emotions have different way of detection process. This technology basically uses the emotions that are formally used in the social connections and their way of reacting to the different conversations and their emotional state. All the existing systems are generally divided into categories based on their detecting techniques like Knowledge based methods, Hybrid methods, and calculative methods [18]. Knowledge based methods utilizes the main knowledge that are from the system based on the characteristics in order to detect the emotion. One of the main advantages of this technique is the large availability of the datasets and resources brought by the knowledge based systems. Usually this system produces the output based on the philosophical and linguistic rules which are used in the communication or conversation process [17]. But the main

drawback for this approach is the inability to use the rules that are provided by the linguistic process. In the hybrid methods they take the complex characteristics from both the knowledge based methods and calculative methods and utilize them by computing them [19]. They gain advantage by the usage of combining two techniques. In the calculative methods they mostly use the algorithms produced by the machine learning. Generally these algorithms provide better results in classification methods and the usage of large datasets makes them process easy and analyze them [13]. Some of the algorithms used are SVM, HMM, Deep Learning, CNN, ANN, LSTM(Long Short-term Memory), computer vision, NLP, Image Processing, Signal Processing, and video analysis.

## 3. METHODOLOGY

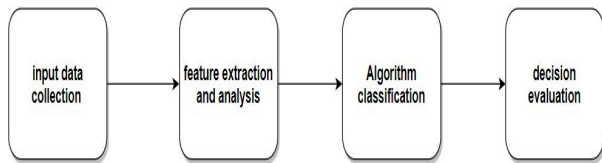
The proposed system for both fall detection and emotion recognition consists of Hardware and software description. The hardware description for both the systems is same since we use the same embedded device. Both the systems are processed with the help of OpenCV which is installed in the Raspberry pi.

### 3.1 Hardware Description

In this developed system we use an embedded device which is Raspberry pi4. It is a considered as a mini computer which runs with a single core processor. This device is used by connecting it with monitor or desktop or any LCD screen. It works on the ARM architecture Processor with both Graphic chip and Program Memory. The processor is BROADCOM BCM 2835 with the improved performance of the CPU which is Cortex-A53 cores since this project is processed better with the help of high deficiency monitors. This device has 40 GPIO pins, 4 USB ports, a SD card slot and type C power supply , and it supports both Bluetooth and Wi-Fi but there is also an Ethernet LAN port of 2.5 GHz and 5GHz for specific use. It can also be connected with both keyboard and mouse or a touch LCD display can also be connected if we are comfortable. But for this system for the easy identification and understanding the situations of the fall we are using a desktop computer. This device has a 4 GB RAM memory and also consists of audio and video jack; it has 2 micro HDMI slot for desktop connectivity and a pi camera supporting slot. It comes with the inbuilt cooling fan. Before installing the OpenCv in its memory it is installed with the Rasbian OS which is specially designed or developed for raspberry pi.

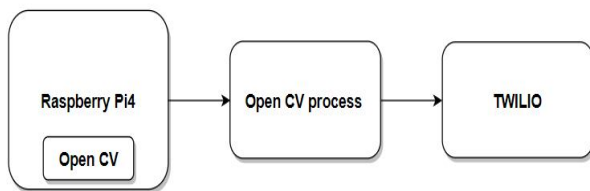
### 3.2 Fall Detection

The standard fall detection system includes input data and the collected data will be further processed for extraction and analysis, with the *help of algorithm that is used in the system*. Then the decision about the fall is predicted from the collected data after analyzing with the algorithm. This process is shown in the figure.2. The proposed method of fall detection system includes both video analysis and image classification.



**Figure 2:** general process of fall detection system

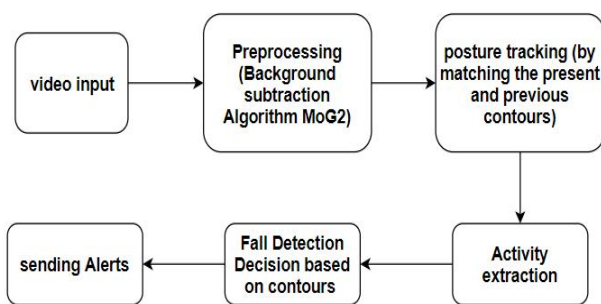
For this system we use both embedded device and machine learning algorithm for precise prediction of fall. For the embedded device we use the Raspberry pi 4 and Open CV is installed in the raspberry pi and in its database, using certain commands. This process is shown in figure .The alert notification is sent to the concerned person and authorities with the help of internet based messaging platforms like TWILIO. Even this messaging platform is to be installed in the raspberry. The input data for this system is collected from the cameras that were installed in the patient’s room. The algorithm used to process this video analyzes each and every frame and the decision is taken based on the data given.



**Figure 3:** Proposed system for Fall Detection and Alert system

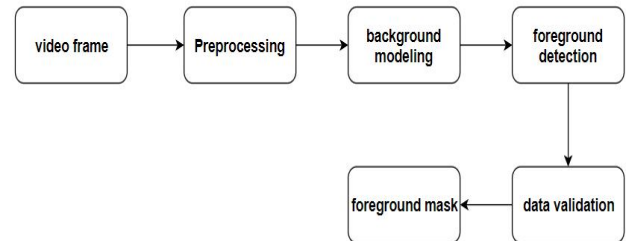
**3.2.1 Software description**

OpenCV is nothing but of an Open Source library in which some of the machine learning algorithm used packages are preinstalled for the better usage. Using this platform the development of any system which uses machine learning or deep learning algorithms are made easy due to the packages in its library. Object detection and face recognition applications primarily use this OpenCV. It is used for easy analyzing of image and video that was obtained from the monitoring cameras. This opencv is installed in the device using certain commands in the terminal of the raspberry pi and the process used in this system is shown in Figure.4.

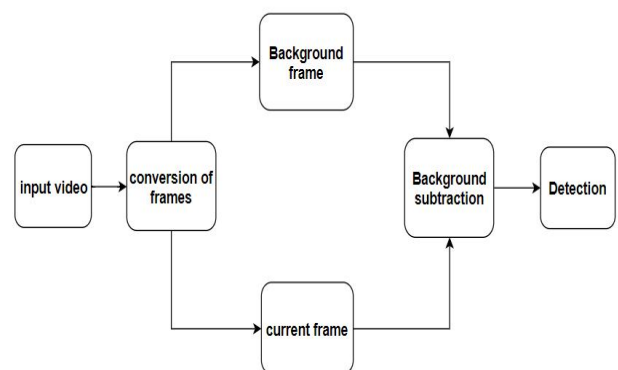


**Figure 4:** OpenCV process for the proposed system

In this fall detection we use a background subtraction MoG2 algorithm which is a deep learning algorithm. This is a Mixture of adaptive Gaussian algorithm which is used for eliminating the background of the image and concentrating on the required or important part of the image. It analyzes each and every frame and involves major steps of preprocessing techniques. This algorithm process is shown in figure.5.



**Figure 5:** The Mog2 algorithm process for video analysis  
 It is also said to be background and foreground subtraction algorithm. Background modeling is the important step in the Mog2 algorithm. It is strong against the environmental changes and sensitive enough to detect the person in the video. The technique used in this algorithm is mean and median. Median defines each and every pixel in the image which is classified from the video. For better concentration on the image it uses Kalman filter to remove the noise. It is capable of analyzing the video by frame to frame and detects the events caused by the person. This process is shown in Figure.6. And for alerting the concerned authorities we use the TWILIO which is a internet based messaging platform. At first twilio is installed in the raspberry pi using commands and a twilio account is created for the phone numbers or email id for which we need to get the alert notifications. And after this process we get a unique twilio number which is used to recharge the account from time to time, because at first twilio provides with the trail run account if we are satisfied with its performance we can develop the personal phone number since twilio provides a high security for our privacy the hacking of this system will be impossible.



**Figure 6:** the fall detection process used by the algorithm

### 3.3 Emotion Recognition

The embedded device used for this system is also Raspberry pi and OpenCV is booted in the SD card of the raspberry pi. We used raspberry pi 4 for better implementation of the algorithm and for better performance. Since we use some heavily processed algorithms they take much clock speed so pi4 is preferred.

#### 3.3.1 Software Description

The OpenCV is installed in the embedded device and the algorithm we use in the emotion recognition is a neural network. CNN (Convolutional Neural Network) are considered as the most accurate algorithm when it comes to detect and analyze from the given series of images or the video. At first we need to collect the dataset for various emotions like “angry”, “disgust”, “happy”, “sad”, “surprise”, and “neutral”. All these images are trained according to process of the algorithm and sent to the database. CNN usually has two important steps which are used in processing the image such as Convolution and Pooling. At first convolution means operation of two functions that produces a third function on expressing how the shape of one is modified by the other. But in programming it is a function that maps a tuple of sequences into sequence of tuples. CNN is also version of MLP (Multi Layer Perception). The three layers in the CNN are input, hidden and output layers. Input layer is where the series of input is given into the model and pixels are divided into many layers which features as the data this input image or the data is represented in the form of matrix. Hidden layer consists of multiple convolutional layers in which each frame is divided into many pixels with a range of (0,255). And the output of each layer is represented in the form of matrix. When all these matrices are multiplied with each other they give a single metric which is considered as output. All the images that were processed have 3 channels (R, G, B) with the depth, height and width of (3,255,255). The CNN observes the pattern of the input image. While a detection occurs it applies many layers for the image until the detection process is clearly occurred. In the detection human contours and expression are considered as the features. The algorithm for the proposed model is given in Figure.7.

In this CNN classification we can see an image or video which declassified as the sequence of the images are given as the input. Convolution 2D is a process in CNN in which a kernel (small matrix of weights) is multiplied with the given input the resulted output is a single pixel image. The hidden layers in the image are reduces and shifted in batch normalization layer. Pooling is nothing but of down sampling, it is the most important layer in the CNN. It is used to reduce the computation in each layer so that the detection process is improved with better speed. In max-pooling a kernel of size  $n*n$  is moved to each matrix and output is displayed by reducing the value. In Global Average Pooling the layers are reduced when the max-pooling occurs. These are the main steps in this proposed CNN classifier. These steps are repeated until the system recognizes certain requirement of the output.

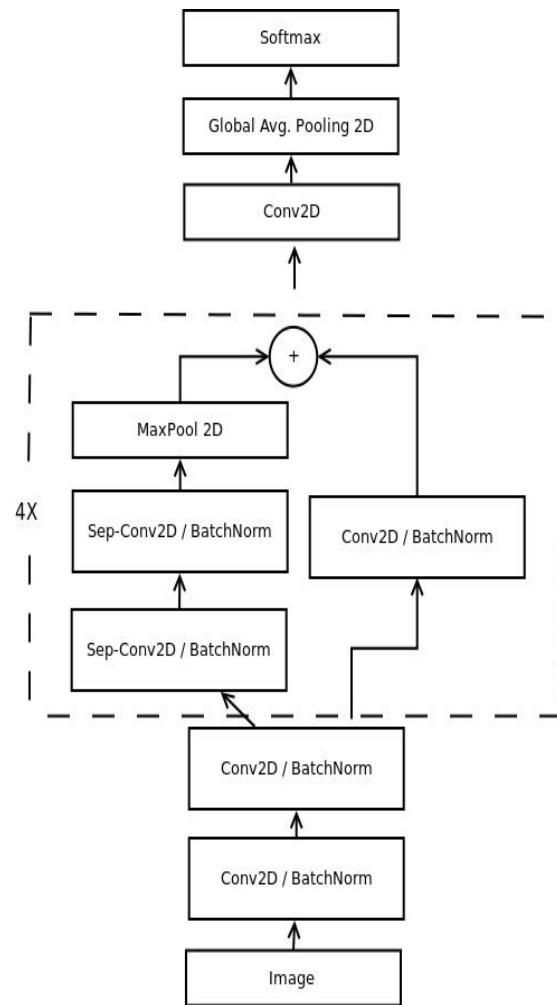
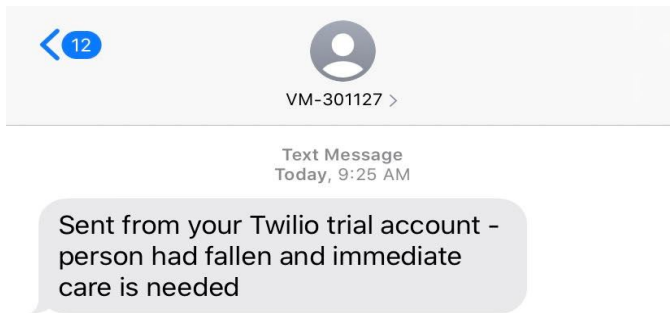


Figure 7: The CNN classification for the proposed system.

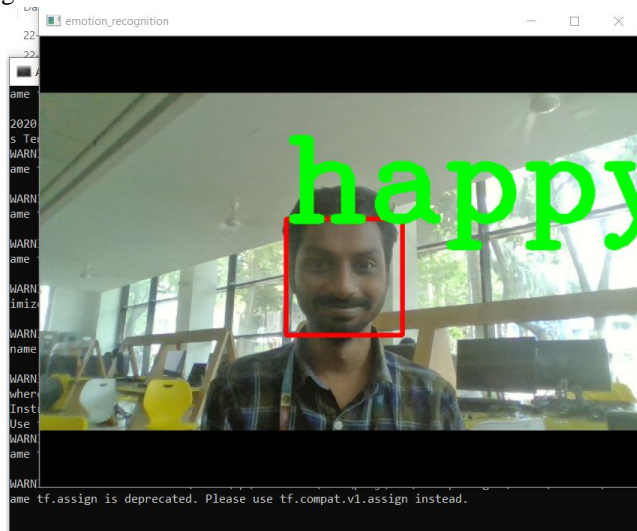
### 4. RESULTS AND DISCUSSION

The input data given for the fall detection system is processed by each frame and the Mog2 algorithm is implemented for better results and the background of the image is subtracted. The mog2 algorithm eliminated the noise and unwanted portions of the image and concentrates only on the person movements and reflexes in the given input video. Each frame is analyzed using the processing techniques of the algorithm. The algorithm concentrates on the person by drawing contours around him from these contours we can know the position and posture of the person. Because there is a different calculation of the person's height and width for each posture. If the height of the person is greater than the width then the person is standing. If it is vice-versa then the person is fallen. When the height and width is considered to be equal the person would likely to be in sitting position. This process takes place when algorithm process each and frame without fail. So when the algorithm identifies the person had fallen the system automatically sends an alert message for the care taker and authorities saying that the fall is detected and immediate care is needed shown in Figure.8.



**Figure 8:** Alert notification for the fall detection system

For both the systems the output is taken by implementing the system in real time. For the emotion recognition the dataset is introduced and trained with the system so that it system might produce the output within the required time. And the recognition of the emotion is deduced from the inputs of the different emotions provided by the classification of the facial expressions and movements in the facial muscles provided in the input. The facial movements like eyebrow movement, cheek movement, frowning and widening of eyes were considered to be different for each emotion. So the system process the input data and when the real time application is given it compares both the real time input and actual trained input and deduces the facial expression. When the lips are widened the output is displayed as the happy along the accuracy percentage. This system output is shown in the Figure.9.



**Figure 9:** The emotion recognition for the real-time data.

## 5. CONCLUSION AND FUTURE WORK

The proposed systems for both fall detection and emotion recognition are producing the output much effectively. These systems use less equipment and non complex algorithms so

that the prediction of the outputs is effective and fast. And the consumption of power is also considered to be less for these systems when compared to other proposed or existing systems. For fall detection we can use the monitoring camera or the CCTV which are installed in every household. But for emotion recognition system we can use any camera but the camera is placed near to the person so that the facial expressions for recognizing the emotion is clearly seen. But in future we can use high deficiency camera so that they can be mounted in any corner of the room and monitored for the emotion. So by using the high deficiency camera we can combine both the systems as one and with the help of emotion recognition we can develop better care for the persons. And by introducing the ECG and EEG of the patients we can also monitor the vitals of the patient and predict the persons health and prevent the fall or any unprepared circumstances.

## REFERENCES

1. P. Vallabh, R. Malekian, N. Ye and D. C. Bogatinoska, "Fall detection using machine learning algorithms," 2016 24th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, 2016, pp. 1-9. doi: 10.1109/SOFTCOM.2016.7772142.
2. W. Min, H. Cui, H. Rao, Z. Li and L. Yao, "Detection of Human Falls on Furniture Using Scene Analysis Based on Deep Learning and Activity Characteristics," in IEEE Access, vol. 6, pp. 9324-9335, 2018. doi: 10.1109/ACCESS.2018.2795239
3. E. Auvinet, F. Multon, A. Saint-Arnaud, J. Rousseau, and J. Meunier, "Fall detection with multiple cameras: An occlusion-resistant method based on 3-D silhouette vertical distribution," IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 2, pp. 290-300, Mar. 2001.
4. Noury et al., "Fall detection - Principles and Methods," 2007 29th Annual International Conference on IEEE Engineering in Medicine and Biology Society, Lyon, 2007, pp. 1663-1666. doi: 10.1109/IEMBS.2007.4352627
5. Ma, H. Wang, B. Xue, M. Zhou, B. Ji, and Y. Li, "Depth based human fall detection via shape features improved extreme learning machine," IEEE J. Biomed. Health Inform., vol. 18, no. 6, pp. 1915-1922, 2014.
6. Popoola, K. Wang, "Video-based abnormal human behavior recognition—A review", IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev., vol. 42, no. 6, pp. 865-878, Nov. 2012.
7. Kumar S. "Facial expression recognition: A review"; Proceedings of the National Conference on Cloud Computing and Big Data; Shanghai, China. 4-6 November 2015; pp. 159-162.
8. Xavier Glorot, Antoine Bordes, and Yoshua Bengio. "Deep sparse rectifier neural networks". In

- Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323, 2011.
9. L. Tong, Q. Song, Y. Ge, and M. Liu, **“HMM-based human fall detection and prediction method using tri-axial accelerometer,”** IEEE Sensors J., vol. 13, no. 5, pp. 1849–1856, May 2013.
  10. S. Wang, X. Wang, R. Chen, Y. Liu and S. Huang, **“Real-time Detection of Facial Expression Based on Improved Residual Convolutional Neural Network,”** 2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Dalian, China, 2019, pp. 1-4. doi: 10.1109/ICSPCC46631.2019.8960776.
  11. M. S. Bartlett, G. Littlewort, I. Fasel and J. R. Movellan, **“Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction.,”** 2003 Conference on Computer Vision and Pattern Recognition Workshop, Madison, Wisconsin, USA, 2003, pp. 53-53. doi: 10.1109/CVPRW.2003.10057.
  12. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; The IEEE Conference on **Computer Vision and Pattern Recognition (CVPR)**, 2016, pp. 770-778.
  13. H. Abanoz and Z. Çataltepe, **“Emotion recognition on static images using deep transfer learning and ensembling,”** 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4. doi: 10.1109/SIU.2018.8404346.
  14. H. Lee and K. Hong, **“A study on emotion recognition method and its application using face image,”** 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, 2017, pp. 370-372. doi: 10.1109/ICTC.2017.8191005.
  15. Samira Ebrahimi Kaou, Christopher Pal, Xavier Bouthillier, Pierre Froumenty, Caglar Gulcehre, Roland Memisevic, Pascal Vincent, Aaron Courville, Yoshua Bengio, **“Combining Modality Specific Deep Neutral Networks for Emotion Recognition in Video”**, ICMI, 2013.
  16. Christian Wallraven, Heinrich H. Bulthoff, Douglas W. Cunningham, Jan Fischer, and Dirk Bartz. **“Evaluation of real-world and computer-generated stylized facial expressions”**. In ACM Transactions on Applied Perception, volume 4, page 16, New York, NY, USA, 2007. ACM.  
<https://doi.org/10.1145/1278387.1278390>
  17. Stacy Marsella. **“Computationally Modeling Human Emotion. Communications of the ACM”**, Vol. 57 No. 12, Pages 56--67, 2015.
  18. Zhan C., Li W., Ogunbona P., Safaei F. **“A real-time facial expression recognition system for online games”**. Int. J. Comput. Games Technol. 2008;2008 doi: 10.1155/2008/542918.
  19. Bartlett MS, Littlewort G, Frank M, Lainscsek C, Fasel I, Movellan J (2005) **“Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior”**, IEEE Conf. Computer Vision and Pattern Recognition, pp. 568–573.  
<https://doi.org/10.1109/CVPR.2005.297>