



Detection and classification of falling in elderly people using customized deep learning algorithm

Bnar Abdulsalam Abdulrahman^{1*} and Aree Ali Mohammed²

¹Student affairs office, Komar University of Science and Technology, Kurdistan Region, Iraq

²Computer Science Department, College of Science, University of Sulaimani, Kurdistan Region, Iraq.

*Corresponding author's e-mail: bnar.abdulsalam@komar.edu.iq

Article info

Abstract

Original: 15 October 2020
Revised: 22 November 2020
Accepted: 29 November 2020
Published online: 20 June 2021

Key Words:

Biometrics
Fall detection
Fall classification
Customized deep learning
Accuracy
Detection

This research work proposes a fall detection system in elderly people based on customized human body detection using You Only Look Once (YOLO) version-3 algorithm. This scheme provides a high accuracy rate of classification for different cases (stand, sit, and fall). To design such system, IoT based-fall detection is implemented. The Raspberry pi is used to process the tested images captured from the live camera. Then, the YOLO detects the human body and classifies them into three categories. The proposed system has trained for 100 images for each class type based on positive samples. Moreover, in the test phase, static image, and live camera have been used to show the performance of the system in term of the accuracy detection. Finally, the accuracy rate is determined for different distances from the camera in order to improve the validity of the classification. Test results indicate that the proposed system is invariant for the light and environmental conditions and has a good accuracy when the live camera is used. The accuracy rates average of a live camera for different distances are 100%, 95%, and 90% for (stand, sit, and fall) classes respectively.

1. INTRODUCTION

There are many definitions of Internet of things, but at the basic level can be described as a network of devices that interact with each other through the exchange of data between devices (M2M), which allows you to collect and share data [1,2].

Improving the efficiency of healthcare infrastructure and biomedical systems is one of the most difficult tasks of modern society. Current patient monitoring, care, management and monitoring procedures are often carried out manually by the medical staff, which can lead to tragic errors in practice [3,4]. In addition, some possible examples are the automatic identification of people and their tracking of biomedical devices in hospitals, correct associations of medicines and patients, and monitoring of physiological parameters of patients in real time for early detection of clinical deterioration [3]. (IoT) have been widely recognized as a potential solution to reduce the drawbacks on health systems, and therefore have been at the center of many recent studies [1].

IoT have become a major stone in the development of a health surveillance system. The goal of an effective healthcare system for IoT is to provide remote monitoring of the patient's real-time health status[11]. It prevents critical cases of patients and improve the quality of life by using a smart Internet environment for things [5,12]. The IoT can not only ensure the safety and health of patients, but also improve the quality of medical care provided by doctors [13]. Internet healthcare things can increase patient involvement and satisfaction, allowing patients to spend more time communicating with their doctors [5].

Fall detection methods are classified into three main categories: wearable device based, ambient device based, and vision based. Wearable devices can be further divided into posture based and motion-based devices. Ambience devices can be further classified into presence and posture-based sensors. And the camera (vision) based systems can be further categorized into three classes as shape change, inactivity and

3D head motion. Most of the existing approaches share the same general framework. Data acquisition varies from one sensor to multiple sensors and from one fixed camera to multiple cameras and moving cameras [6]. According to the World Health Organization (WHO) report, falls are the second leading cause of unintentional or accidental death. One bad fall can mean a long hospital stay, permanent disability, painful rehabilitation, a loss of independent life, or worse. Falling can be considered as one of the most commonly mishap happens with elderly people, which might cause severe injuries [7]. Many fall detection systems based on wearable sensors have been proposed. Wearable device-based approaches depend on costumes with embedded sensors to detect the motion and location of the body of the subject [8]. Vision based fall detection system can be done effectively using fixed or moved cameras. Camera captures scene, foreground which is moving object in this case is detected and different features are extracted from that object. Extracted features are used to detect fall event [9]. The accuracy of any fall detection system relies on the technique (with or without sensors) and the type of classification. A vision-based system could be adopted for fall detection that can be adapted to the deep machine learning strategies. The deep network is exploited to differentiate humans (foreground) from the background. Adaptation is required to solve dynamic changes in the visual conditions (illumination, shadows, and background changes) which are very often for a real-life environment [10].

YOLO is viewed as an efficient deep learning method for real time object detection algorithm, which is considered as one of the most effective object detection algorithms that also enclosed many of the most state-of-the-art ideas coming out of the computer vision research community. It could be also used for customized object detection in many areas of research such as healthcare, safety-car driven, and human action recognition.

In this research work, customized human fall detection is proposed using the YOLO algorithm for both detection and classification. It first trains three different classes namely (stand, sit, and fall) of 100 images for each class. And then, it iterates for 50 epochs to detect the human body. The main contribution of this work is to use vision-based method through live camera or the webcam to detect and classify the human fall.

The rest of the paper is organized as follows: an overview of recent related works of the human fall detection has been presented in section 2. In section 3, a brief description of the proposed system has been introduced. Section 4 shows the performance evaluation of the system. Finally, the conclusions and some suggestions for future work are presented in Section 5.

2. RELATED WORK

F. Wu et al. [14] developed a new fall detection system based on a wearable sensor device (Accelerometer). The system monitors the movements of human body, recognizes a fall from normal daily activities by an effective quaternion algorithm, and automatically sends request for help to the caregivers with the patient's location. The system has also low power consumed hardware design and highly efficient algorithm which could extend the service time of the wearable device. Both the hardware and software designs are suitable for wearable and outdoor application.

K. Miguel et al. [15] proposed a novel low-cost fall detector for automated smart homes based on artificial vision algorithms. However, the system presented in this research work is still in progress, it is already able to reliably detect falls in controlled environments, while taking into account several common events found in real settings. The detector uses several algorithms (background subtraction, Kalman filtering and optical flow) as input to a machine learning algorithm with high detection accuracy. The conducted tests on more than 50 different fall videos have shown a detection ratio of greater than 96% .

Z. Zhang et al. [18] presented a survey of fall detection systems and algorithms which aim at automatically detecting cases where a human fall and may have been injured. Existing fall detection methods can be classified as using sensors, or being exclusively vision-based. This literature review focuses on vision-based methods. They also introduced five publicly available fall datasets. Three of them were recorded using Kinect cameras, one was collected by a single RGB camera and one was made with multiple calibrated RGB cameras. The overall fall detection system may contain additional modules, both to improve accuracy, and to include additional functionality, such as an acoustic module and a module sending an alert about the detected fall.

M. Yu et al. [17] suggested a new computer vision-based fall detection system, which can be applied for

the health-care of the elderly people group. The first step is to apply the background subtraction method for a recorded video stream to extract the human body contour. Extracted silhouettes corresponding to daily activities are applied to construct a convolutional neural network, which is applied for classification of different classes of human postures (e.g., bend, stand, lie and sit) and detection of a fall event (i.e., lying posture is detected in the floor region). The proposed system showed that the high classification rate is achieved.

K. Lu and E. Chu [16] designed and implemented an Image-based Fall Detection System (IFADS) for nursing homes, where public areas are usually covered with surveillance cameras. dissimilar to existing fall detection algorithms, one can essentially focused on falls that occur while sitting down and standing up from a chair, because the two activities together account for a higher proportion of falls than forward walking. IFADS first puts in an object detection algorithm to identify people in a video frame. Then, a posture recognition method is used to keep tracking the status of the people by checking the relative positions of the chair and the people. An alarm is activated when a fall is detected.

O. Seregin et al. [19] proposed an algorithm which is based on the skeleton features encoding on the sequence of consecutive frames and support vector machine classifier. An improved version of a cumulative sum method is used for combining the individual decisions on the consecutive frames. One-class classifier is used for detecting the low-quality skeletons. The 95.8% accuracy of the proposed fall detection procedure was obtained in the cross-validation procedure based on the removal of records of a particular person from the database (Leave-one-Person-out).

X. Kong et al. [20] constructed a data set consisting of various daily activities and fall events. They also studied the effect of camera/sensor height on fall-detection accuracy rate. Each activity in the data set is performed by eight contributors in eight directions and grabbed with the depth camera at five different heights. Numerous related works completely relied on human segmentation by using Kinect SDK, but this is not reliable enough. To overcome this problem, this study proposed Enhanced Tracking and Denoising Alex-Net (ETDA-Net) to improve tracking and denoising performance and classify fall and non-fall events. Experimental results indicated that fall-detection accuracy is affected by camera height, against which ETDA-Net is robust.

3. PROPOSED APPROACH

The proposed system is equipped with a webcam, Raspberry Pi V4, and a smart phone device as an alarm system for both doctors and patient’s relatives. Also, a connection to a server to perform the heavy workload image processing and the fall detection algorithm is needed. Figure 1 presents an overview of the system architecture.

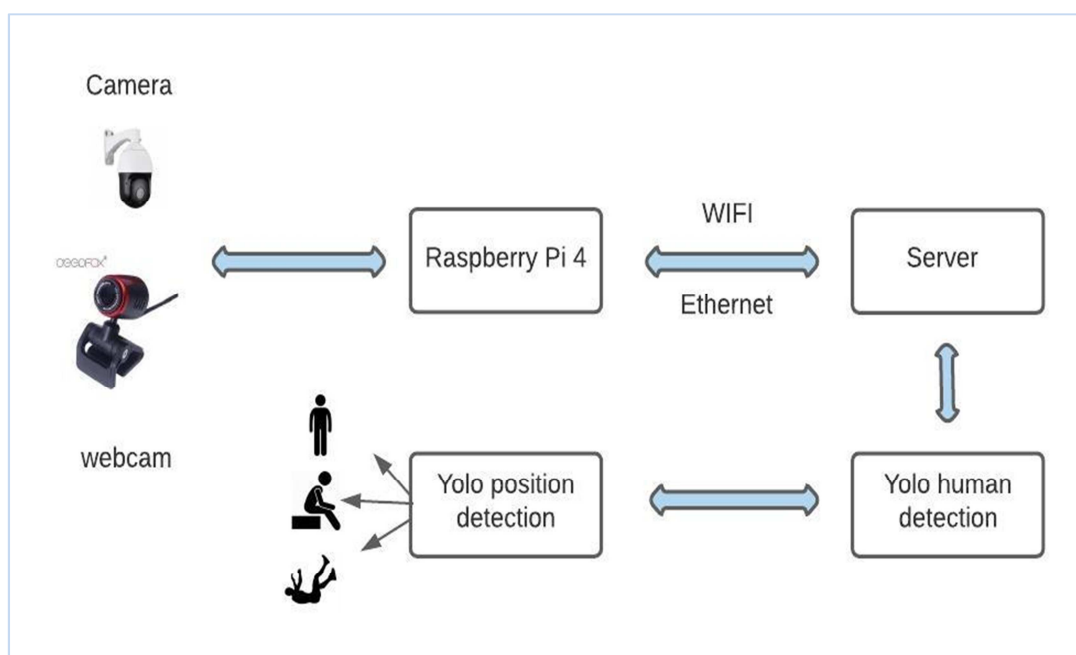


Fig. 1 System-architecture overview

3.1 Raspberry pi 4 specifications

Raspberry Pi 4 Model is the latest product in the common Raspberry Pi range of computers. It offers an increasing in processor speed, multimedia performance, memory, and connectivity compared to the prior-generation Raspberry Pi 3 Models. Figure 2 shows the specification of Raspberry pi4 compared with the raspberry pi3.

Raspberry Pi 4B vs 3B+		
Features/Specs	Raspberry Pi 4 Model B	Raspberry Pi 3 Model B+
Release Date	24th June 2019	14th March 2018
SoC Type (Processor)	Broadcom BCM2711 (with metal cover)	Broadcom BCM2837B0 (with metal cover)
Core Type	Cortex-A72 64-bit (ARMv8)	Cortex-A53 64-bit (ARMv8)
No. of Cores	Quad-Core	
GPU	VideoCore VI	VideoCore IV
Multimedia	H.265 decode (4Kp60) H.264 decode (1080p60) H.264 encode (1080p30) OpenGL ES 1.1, 2.0, 3.0 Graphics	H.264, MPEG-4 decode (1080p30) H.264 encode (1080p30) OpenGL ES 1.1, 2.0 Graphics
CPU Clock	1.5 GHz	1.4 GHz
Memory/OS storage	microSD	
RAM	LPDDR4: 1GB, 2GB, 4GB and 8GB options	LPDDR2 1GB
Ethernet	True Gigabit Ethernet	Gigabit over USB 2.0 (Max 300Mbps)
USB Port	2 x USB 3.0 + 2 x USB 2.0	4 x USB 2.0
HDMI	2 x micro HDMI support Dual Display	1 x full size HDMI
WiFi	802.11 b/g/n/ac (2.4GHz+5GHz & Shielded)	
Bluetooth	5.0 + BLE (Shielded)	4.2 + BLE (Shielded)
Antenna	PCB Antenna (Similar to Rpi Zero W)	
GPIO	40 pins (Fully backwards-compatible with previous boards)	
Operating System	Raspbian (> 24 June 2019)	Raspbian (> March 2018)
Dimension	85mm x 56mm	
Power Input	5V via USB Type C (upto 3A) 5V via GPIO header (upto 3A) Power over Ethernet, requires PoE HAT	5V via USB Micro B (upto 2.5A) 5V via GPIO header (upto 3A) Power over Ethernet, requires PoE HAT

Fig. 2 Raspberry pi4 versus pi3 specifications

3.2 Yolov3 architecture

In this paper, a real-time customized object detection system YOLOv3 for person detection is applied, which has proven to be a better replacement to other algorithms in terms of efficiency and accuracy. The YOLO network is basically taken an image or a frame from a camera and divided it into $S \times S$ grids. Each grid predicts B bounding boxes. In addition, each box provides a confidence score which reflects how likely the box contains an object. Bounding boxes with this parameter above a threshold value are selected and used to locate the object, a person in our case. The proposed approach deals with the fall detection challenge based on two steps, person detection and person position classifications (stand, sit, and fall). The person detection

algorithm is aimed to localize an elderly person in an image. Its output is the enclosed bounding box and the confidence score that reflect how likely it is that the boxes contain a person. In classification step, the detected person is estimated whether it is stand, sit, or fall. If any fall has been detected the system will directly send a notification to the doctor through SMS if there was no response in few seconds for second time it will resend the notification through email.

YOLOv3 is considered as an improvement over previous YOLO detection networks. Compared to prior versions, it makes multi-scale detection for feature extraction, stronger feature extractor network, and some modification in the loss function. Consequently, this network can detect many more object targets ranging from small to big objects. Also, YOLOv3 runs quite faster and makes real-time inference possible on GPU devices. YOLOv3 makes use of a new architecture of a feature extraction named Darknet-53 that originally has a 53-layer network trained in ImageNet. For the detection task, 53 more layers are stacked, giving us a fully convolutional underlying architecture of 106 layers for YOLO v3 and each followed by a sampling layer. This helps prevent the loss of low-level features that are often attributed to pooling. YOLO v3 uses a variant of Darknet, multiscale prediction, which means that it is detected on multiscale feature maps. For this reason, the accuracy of target detection is improved. The detail of its architecture is shown in Figure 3 [21].

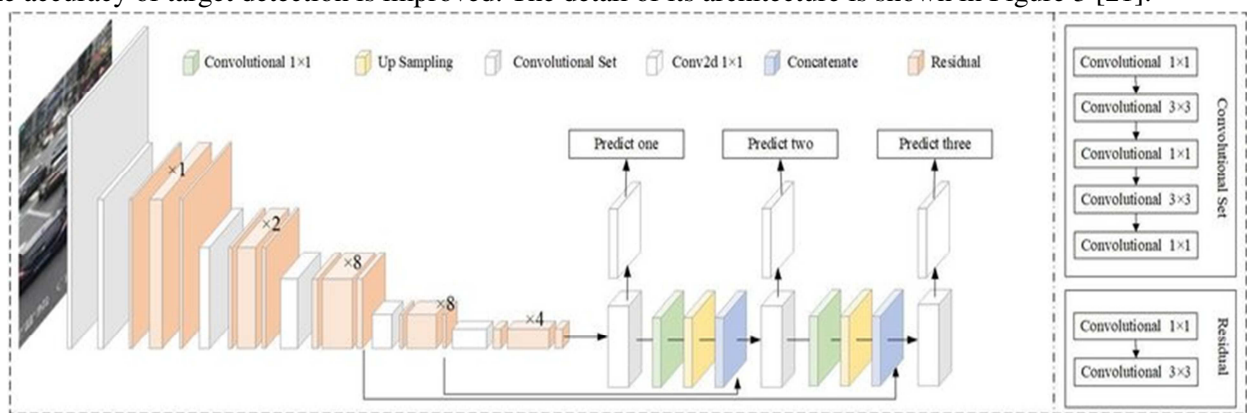


Fig. 3 Detailed YOLOv3 architecture [21]

To prepare YOLOv3 custom training data, the Keras implementation has used for feature extraction which is considered as a backbone of the network model. The necessary steps for training the proposed model are summarized as follows:

- Selecting up the environment, model architecture, and pre-fit weights.
- Loading in data via Image label.
- Determining the model configuration, such as (epochs to train, training batch sizes, the size of our training vs test set, and our learning rate).
- Initiating training.
- Using trained model for inference (predictions).

Figure 4 shows the flowchart of the proposed fall detection system.

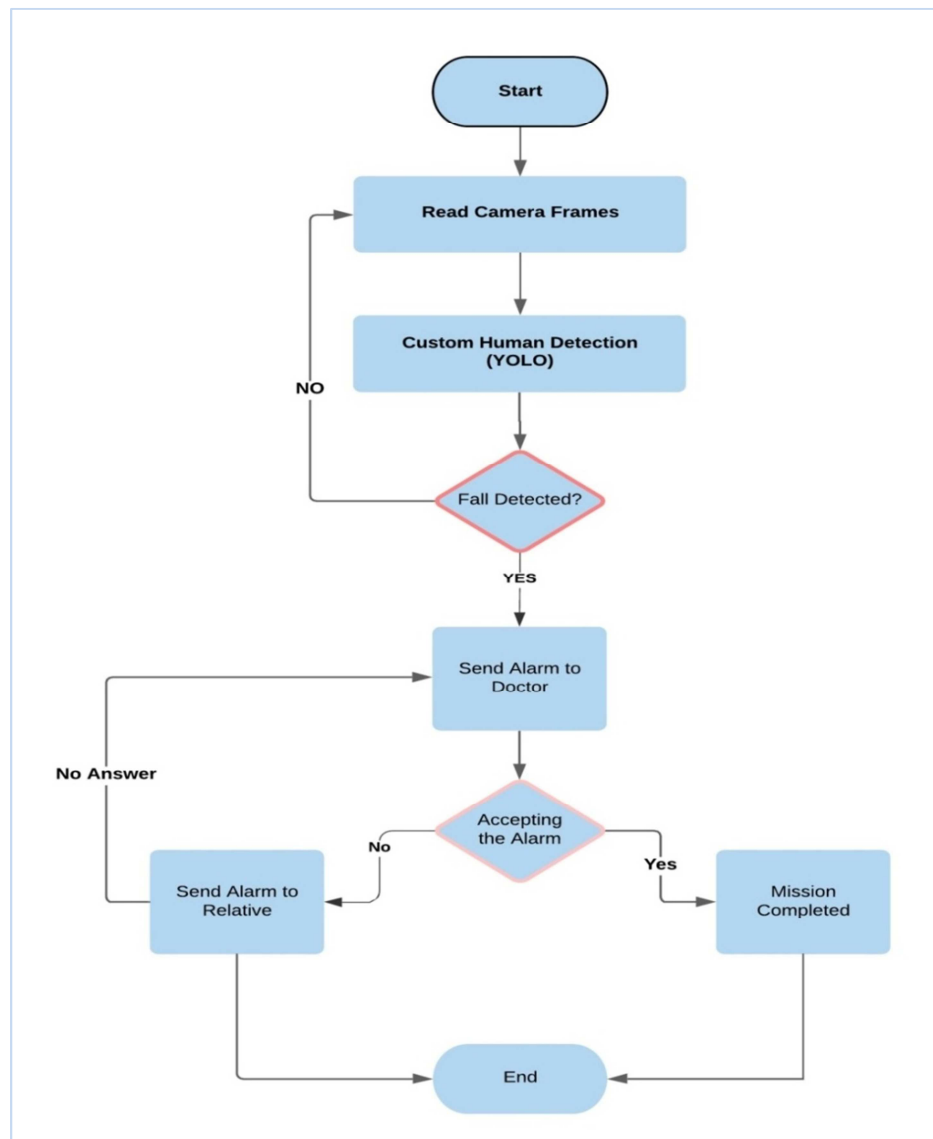


Fig. 4 Proposed fall detection flowchart

4. DATA COLLECTION AND PRE-PROCESSING

In this section the data collection is discussed in addition to the pre-processing applied to the data which has been collected. This step increase detection rate and decrease the computational time.

4.1 DATA COLLECTION

As showing in Figure 5, the data are collected image by image manually from the website (<https://www.freepik.com>) which is usually used by graphic designers since the quality of the pictures are very high. The photos that have been downloaded can be divided into three types: stand, sit and fallen people. The number of downloaded photos is 100-150 photos for each class. The photos contain different images that vary in terms of positions, environment, and illumination. Images in this study are manually selected from the above-mentioned website database sources as described in table 1.

Table 1. Data collection details

Ethnicity Class	ImageSource	No. of True Images	No. of False Images	Total
Stand	Freepik	110	10	120
Sit	Freepik	140	10	150
Fallen	Freepik	90	10	100
Total Image				370

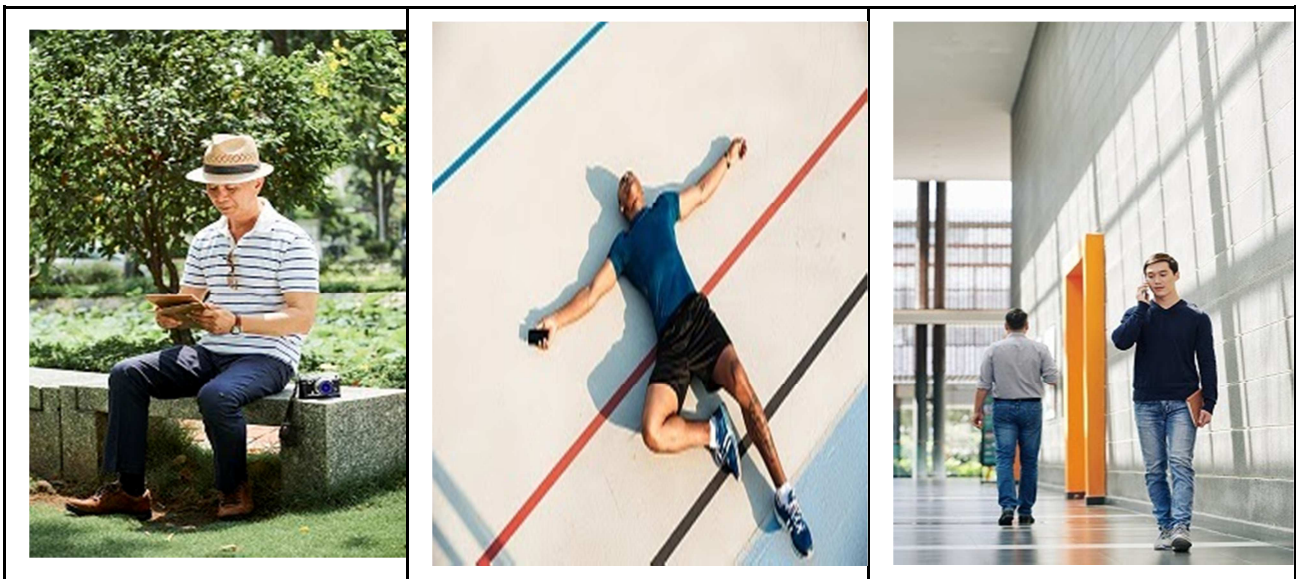


Fig.5 Images used for three classes (sit, fall, stand).

4.2 DATA PREPARATION

In data preparation (ImageLabing) tool has been used. As shown in Figure 6 the tool used to prepare the photos for the training. The object boundaries which are human selected by us then we chose the classes based on each picture (sit, stand, and fall). Then we saved it as an XML file as shown in figure 7.

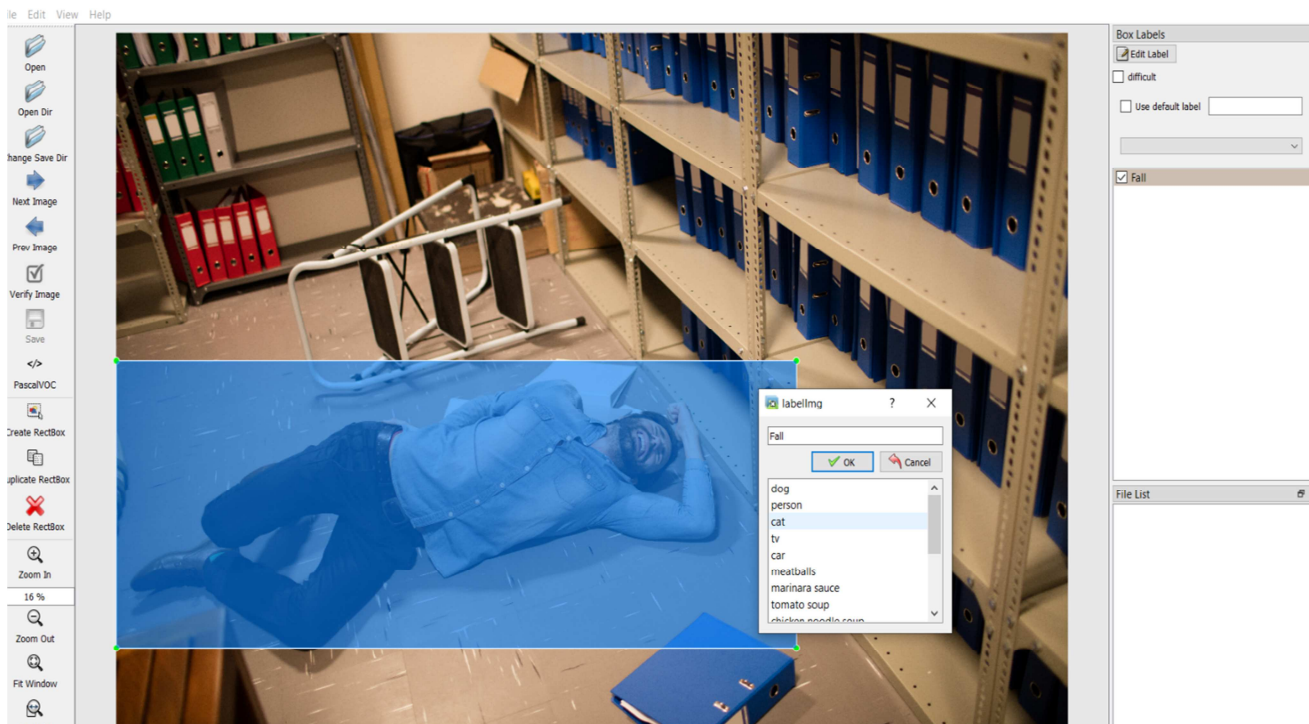


Fig.6 Image labeling tool (Fall example)

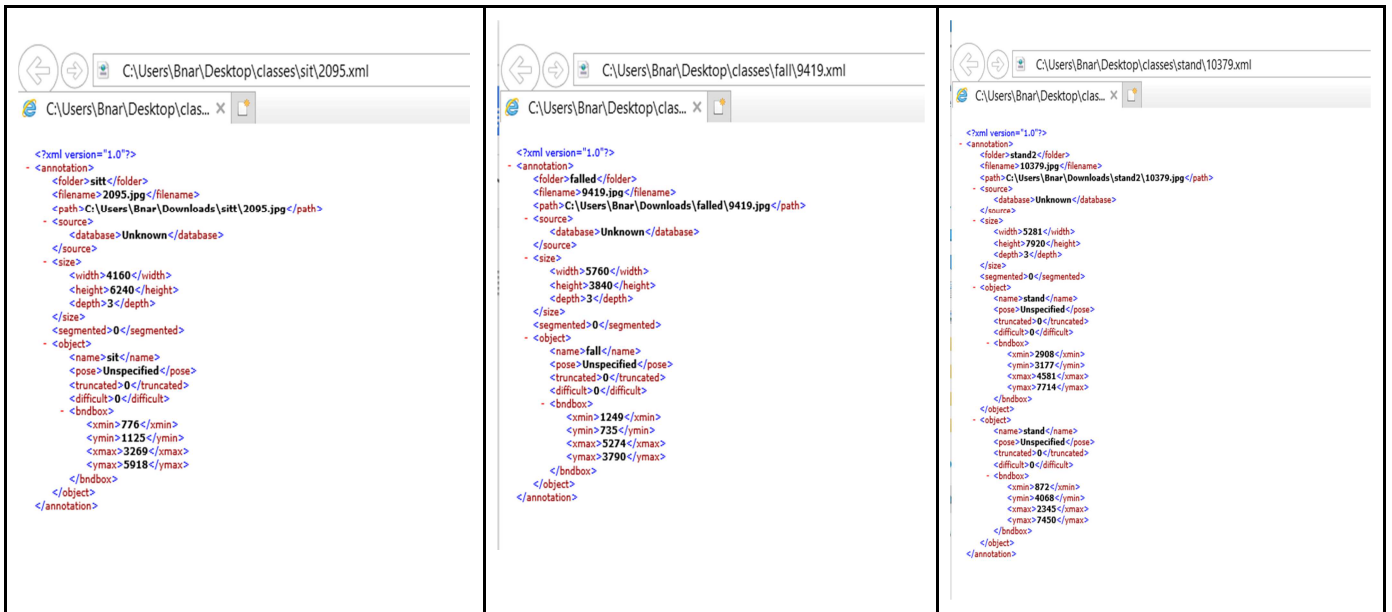


Fig.7 XML code for Images used in three classes (sit, fall, stand).

The above xml files have been used as an input for training our system through Yolov3. The below pseudo code identifies the steps and clarify where the xml files have been used.

Input: image or videos

Output: fall detection accuracy

Start Loop

Preprocess input data and Get XML files for each image.

Training the input data using Yolov3 (XML files are input data)

Test the model using Yolov3

If fall detected then send an alarm to the doctor

End loop

5. EXPERIMENTAL RESULTS

To evaluate the system’s performance, the test image data is collected from different sources (static image, video file, live camera). Some of the data is obtained from a well-known database especially video file and some others recorded in a house or a laboratory. Parameters used for performance evaluation are:

Sensitivity: presents the percentage of fall events detected.

Specificity: is the percentage of events without falls detected correctly.

Precision: is the percentage of fall alerts that represent actual falls.

Accuracy: presents the percentage of correctly detected events.

They are based on True Positive (TP) when a system properly detects a fall when fall has occurred. False Positive (FP) when a system detects a fall when no fall has occurred. True Negative (TN) when a system detects no fall when no fall has occurred. False Negative (FN) when a system detects no fall when a fall has occurred.

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

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

$$(3) \text{ Precision} = TP / (TP + FP)$$

$$(4) \text{ Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

In this subsection, the trained model is tested with 10 images of nine positive samples and one negative sample. Table 2 shows the number of TP, TN, FP, and FN for different class types (sit, stand, and fall) and the average accuracy detection for each class while in Figure 8, the results of three classes are presented.

Table 2. Accuracy test results for image

Test	Sit	Stand	Fall
TP	9	9	9
TN	1	1	1
FP	0	0	0
FN	0	0	0
Accuracy	100%	100%	100%

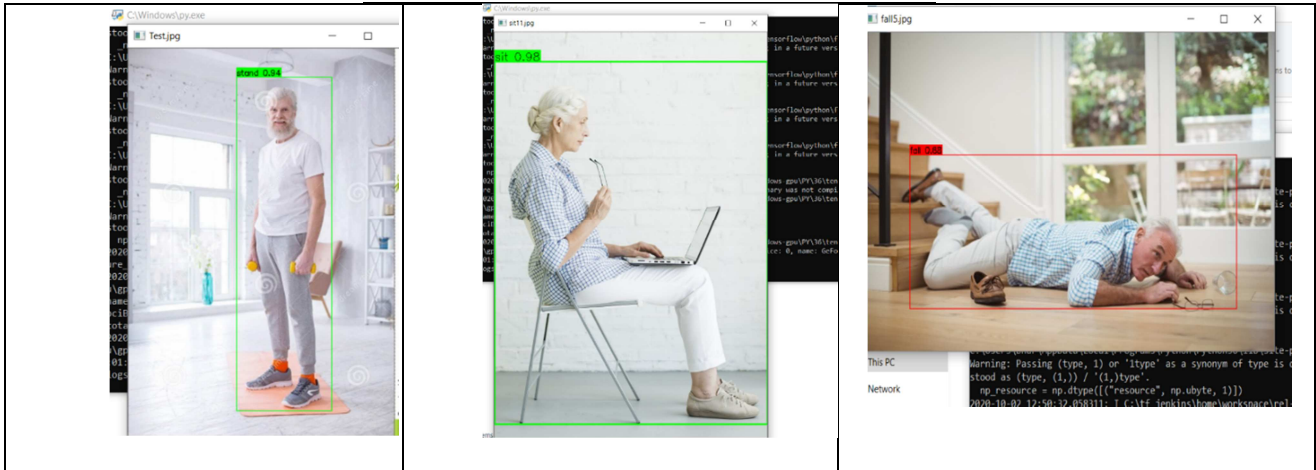
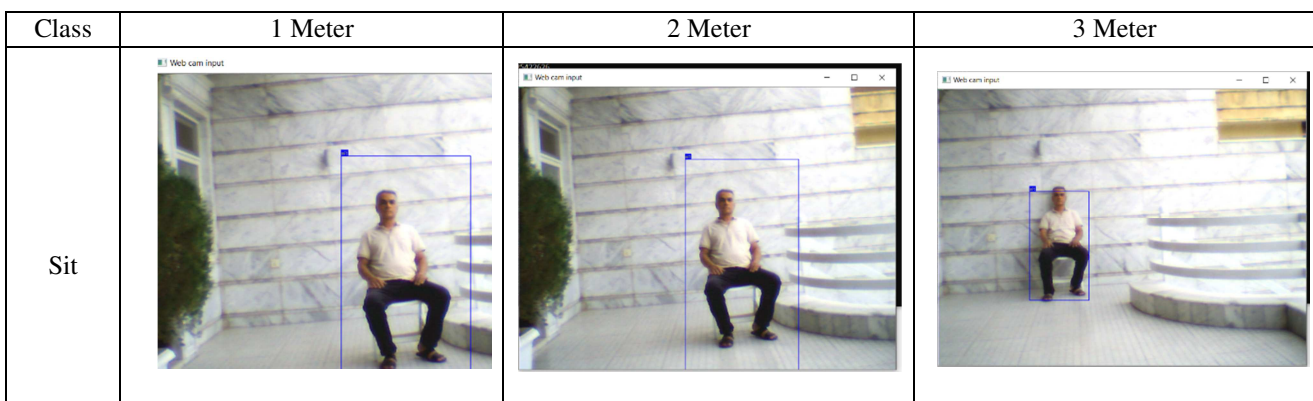


Fig.8 Image test results for three classes (stand, sit, and fall).

The webcam system has also tested for the sitting, standing, and falling classes for different distances (1,2, and 3 meters). In Table 3, the test results of performance parameters (TP, TN, FP, and FN) and the average accuracy of 10 trials for each class are shown. Figure 9 shows the accuracy detection for different distances.

Table 3. Accuracy test results for webcam

Test	Sit	Stand	Fall
TP	7	9	8
TN	1	1	1
FP	0	0	0
FN	2	0	1
Accuracy	80%	100%	90%



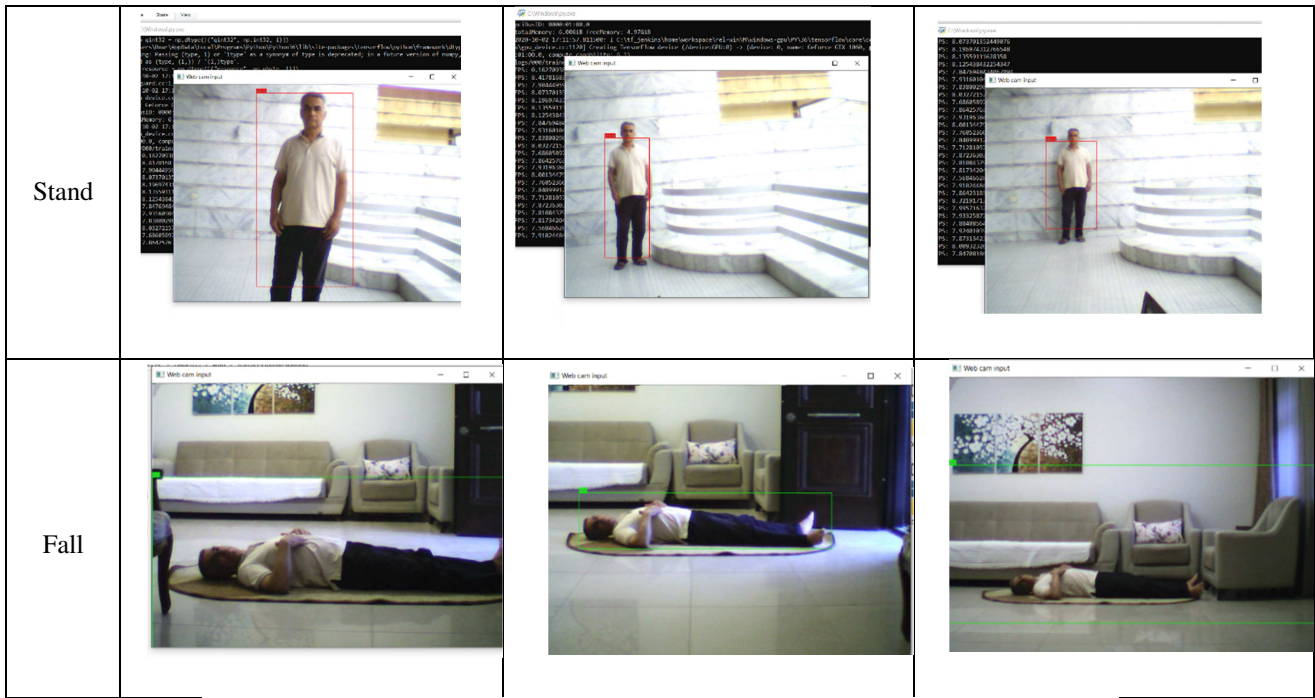


Fig.9 Webcam test for three classes (stand, sit, and fall) with three different distances.

Finally, the performance accuracy of the proposed system versus the webcam distance from the human body is shown in figure 10. As it can clearly be seen, the proposed system can perfectly detect and categorize the stand and fall classes for almost all distances whereas, the accuracy of sitting class is gradually decreased when the distance is increasing.

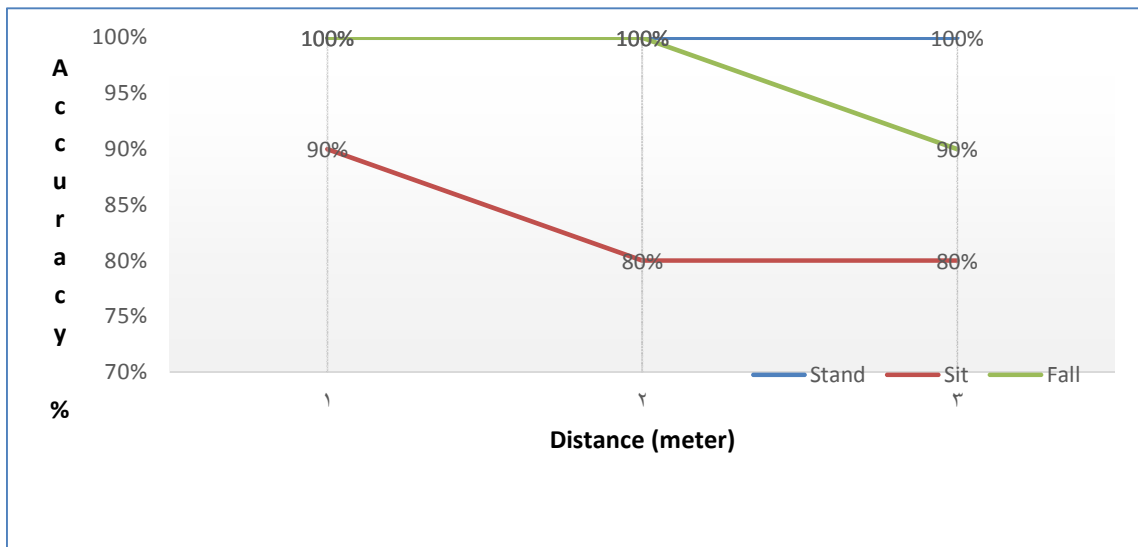


Fig.10 Accuracy versus distances for three classes (stand, sit, and fall).

As shown in table (4) our proposed method has been compared to four paper results from the references we have used in fall detection accuracy rate.

Table 4. Accuracy results comparison (Fall detection)

Related Researchs	Machine Learning Method	Accuracy Rate (for image)	Accuracy Rate (for video)
Paper [7]	SVM	99.98%	N/A
Paper [10]	Deep Adaptable Neural Network	N/A	93%
Paper [16]	IFADS	95.96%	N/A
Paper [17]	CNN	N/A	94.75%
Proposed Model	Yolov3	100%	90%

N/A – Not Available

6. CONCLUSIONS

In this paper, a customized human fall detection system is developed based on deep learning networks (YOLOv3). It mainly emphasizes on improving the accuracy rate detection for images and live camera as presented in Table 2, Figure 6 and Table 3, Figure 7 respectively. The system has trained with more than 350 images of different classes as indicated in Table 1. Test results show that the average accuracy has attained an optimal value 100% for all classes when the system is tested only with images while for the live webcam, it is decreasing for different distance values. Finally, the performance accuracy of the proposed system based on positive and negative samples versus the webcam distance from the human body is determined. As it can clearly be seen in figure 8, the proposed system can perfectly detect and categorize the stand 100% and fall 95% classes for almost all distances whereas, the accuracy of sitting class is gradually decreased when the distance is increasing. The reason for maintaining the limited distance from the live webcam is that the experimental area was small and for the large distance greater than 3 meters, the performance accuracy is considerably declined especially for sit and fall classes. To improve the accuracy rate, we will train our model with more datasets as a future work.

References

- [1] M. Pareraki et al. "*Biosensors and Internet of Things in smart healthcare applications: challenges and opportunities*". *Biosensors and Internet of Things in smart healthcare applications: challenges and opportunities*. Vol. 7, pp. 25-53. (2020).
- [2] P. Gope and T. Hwang. "*BSN-Care: A Secure IoT- Based Modern Healthcare System Using Body Sensor Network*". *IEEE Sensors Journal*. Vol. 16, no. 5, pp. 1368–1376. (2016).
- [3] R. K. Kodali, G. Swamy, and B. Lakshmi. "*An implementation of IoT for healthcare*". *IEEE Recent Adv. Intell. Comput. Syst. RAICS* Vol. 2015, pp. 411–416. (2016).
- [4] Koppa, Anant R, and Venugopalachar Sridhar. "A workflow solution for electronic health records to improve healthcare delivery efficiency in rural India," In *eHealth, Telemedicine, and Social Medicine, 2009. eTELEMED'09. International Conference on*, pp. 227–232, IEEE, 2009
- [5] S. Sivagami, D. Revathy, and L. Nithyabharathi. "*Smart Health Care System Implemented using IoT*". *Int. J. Contemp. Res. Comput. Sci. Technol*. Vol. 2, No. 3, pp. 641–646. (2016).
- [6] Mubashir M., Shao L., Seed L. "*A survey on fall detection: Principles and approaches*". *Neurocomputing*, pp. 144–152. (2013).
- [7] F. Hussain, M. Umair, M. Ehatisham, I. Pires, T. Valente, N. Garcia, and N. Pombo. "*An Efficient Machine Learning-based Elderly Fall Detection Algorithm*". *The Ninth International Conference on Sensor Device Technologies and Applications*, pp. 88-93. (2018).
- [8] T. Shany, S. J. Redmond, M. R. Narayanan, and N. H. Lovell. "*Sensors-based wearable systems for monitoring of human movement and falls*". *IEEE Sensors Journal*. Vol. 12, No. 3, pp. 658–670. (2012).
- [9] N. Joshi and S. Nalbalwar. "*A Fall Detection and Alert System for an Elderly using Computer Vision and Internet of Things*". *2nd IEEE International Conference On Recent Trends in Electronics Information & Communication Technology*, pp. 1276-1281. (2017).
- [10] D. Anastasios and D. Nikolaos. "*Adaptive Deep Learning for a Vision-based Fall Detection*". *Association for Computing Machinery Conference*. pp. 558-565, (2018).
- [11] B. Stephanie, W. Xiang, and I. Atkinson. "*Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities*". *IEEE Access*, Vol. 5, pp. 26521-26544. (2017).
- [12] M. Padmanabhan, P. Yuan, G. Chada and H. V. Nguyen. "*Physician-Friendly Machine Learning: A Case Study with Cardiovascular Disease Risk Prediction*". *Journal of Clinical Medicine*, Vol. 8, No. 10, pp. 1-13. (2019).
- [13] A. Yazar, A. Cetin, and F. Erden. "*Multi-sensor ambient assisted living system for fall detection*". *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (2014).

- [14] F. Wu, H. Zhao, Y. Zhao, and H. Zhong. "*Development of a Wearable-Sensor-Based Fall Detection System*". International Journal of Telemedicine and Applications. Vol. 2015, pp. 1-11. (2015).
- [15] K. Miguel, A. Brunete, M. Hernando, and E. Gambao. "Home Camera-Based Fall Detection System for the Elderly", Journal of Sensors, Vol. 17, No. 28, 2017.
- [16] K. Lu and E. Chu. "*An Image-Based Fall Detection System for the Elderly*". Journal of Applied Science., Vol. 8. (2018).
- [17] M. Yu, L. Gong, and S. Kollias. "*Computer Vision Based Fall Detection by a Convolutional Neural Network*". Proceedings of ACM International Conference on Multimodal Interaction. pp. 416-420. (2017).
- [18] Z. Zhang, C. Conly, and V. Athitsos. "*A Survey on Vision-based Fall Detection*". Proceedings of the ACM International Conference on Pervasive Technologies related to Assistive Environments. Article No. 46. (2015).
- [19] O. Seredin A. Kopylov, S. Huang , and D. Rodionov. "*A Skeleton Features-Based Fall Detection using Microsoft Kinect V2 with One Class-Classifier Outlier Removal*". Int. Workshop on Photogrammetric & Computer Vision Techniques for Video Surveillance, Biometrics and Biomedicine. Vol. XLII-2/W12, pp. 189-195. (2019).
- [20] X. Kong, L. Chen, Z. Wang, Y. Chen, L. Meng, and H. Tomiyama. "*Robust Self-Adaptation Fall-Detection System Based on Camera Height*". Journal of Sensors. Vol. 19, No. 3768. (2019).
- [21] Qi-Chao Mao , Hong-Mei Sun, Yan-Bo Liu and Rui-Sheng Jia. "*Mini-YOLOv3: Real-Time Object Detector for Embedded Applications*". 2941547, IEEE Access, DOI 10.1109/ACCESS.2019.