Computer Systems Science & Engineering
DOI:10.32604/csse.2022.020361
Article

Tech Science Press

Vision Based Real Time Monitoring System for Elderly Fall Event Detection Using Deep Learning

G. Anitha[1,*] and S. Baghavathi Priya[2]

[1]Department of Information Technology, Rajalakshmi Engineering College, Chennai, 602105, India
[2]Department of Computer Science and Engineering, Rajalakshmi Engineering College, Chennai, 602105, India
*Corresponding Author: G. Anitha. Email: anitha.g@rajalakshmi.edu.in
Received: 21 May 2021; Accepted: 13 July 2021

Abstract: Human fall detection plays a vital part in the design of sensor based alarming system, aid physical therapists not only to lessen after fall effect and also to save human life. Accurate and timely identification can offer quick medical services to the injured people and prevent from serious consequences. Several vision-based approaches have been developed by the placement of cameras in diverse everyday environments. At present times, deep learning (DL) models particularly convolutional neural networks (CNNs) have gained much importance in the fall detection tasks. With this motivation, this paper presents a new vision based elderly fall event detection using deep learning (VEFED-DL) model. The proposed VEFED-DL model involves different stages of operations namely preprocessing, feature extraction, classification, and parameter optimization. Primarily, the digital video camera is used to capture the RGB color images and the video is extracted into a set of frames. For improving the image quality and eliminate noise, the frames are processed in three levels namely resizing, augmentation, and min–max based normalization. Besides, MobileNet model is applied as a feature extractor to derive the spatial features that exist in the preprocessed frames. In addition, the extracted spatial features are then fed into the gated recurrent unit (GRU) to extract the temporal dependencies of the human movements. Finally, a group teaching optimization algorithm (GTOA) with stacked autoencoder (SAE) is used as a binary classification model to determine the existence of fall or non-fall events. The GTOA is employed for the parameter optimization of the SAE model in such a way that the detection performance can be enhanced. In order to assess the fall detection performance of the presented VEFED-DL model, a set of simulations take place on the UR fall detection dataset and multiple cameras fall dataset. The experimental outcomes highlighted the superior performance of the presented method over the recent methods.

Keywords: Computer vision; elderly people; fall detection; deep learning; metaheuristics; object detection; parameter optimization

1 Introduction

In past decades, the security of elders living alone turned into a growing problem for society [1]. Falling event is a popular and significant risk among elder people since the older persons are deteriorating in loss of balance, physical function, and slower sensory response. Generally, these falls create loss of mobility, injury, and other health issues. Thus, effective recognition of fall events is vital for security in solitary indoor scenes. They could not alert any person for aid, specifically when they are unconscious and have sustained severe injuries [2]. It was stated that fifty percent of the elders lay on the floor for several hours when falling in 6 months afterward the accident. Additionally, fear of falling and not getting immediate medicinal support could limit the activity of elder's leads to helplessness, social isolation, and depression [3]. Alternatively, long-term nursery care at home is costly [4]. Thus, the fall is the main risk for elder person, with substantial physical, financial, and emotional effects. It is considered the main health concern particularly for people who live alone. Due to this, it is important for developing intelligent scheme that could manually detect falls, help elder people to live independently and securely in their homes. Simultaneously, it would decrease the cost and the time needed for caretakers for avoiding intervene in its dangerous consequence.

Recently, researchers on of fall detection system have improved remarkably with rapid growth of novel smart technologies and sensors which could extract various types of data from the platform. Based on type of sensors utilized, fall detection approaches are separated into 3 major classes [5], such as (i) ambient sensor based, (ii) wearable sensor based, and vision based (iii). The wearable sensor-based method utilizes particular device related to human body to detect other variations of his activity. Several wearable sensors are utilized for detecting falls like gyroscope, smart watch, accelerometer, or combination of them. Unfortunately, elder persons often forget to wear them mainly with cognitive impairment [6]. An ambient sensor-based system exploits pressure/vibration devices set up in the floor/bed to examine the vibration and sounds. These sensors are inexpensive and do not distract elder people. Their major limitations are lower fall detection accurateness and tend to create more false alarms. Lastly, vision-based method uses several cameras for detecting falls. These cameras are set up in daily environment and provide high data regarding peoples and their events. Fall detection is highly complex in real time environments because of pose changes, occlusions, illumination, light changing, shadows, and so on. To face these uncontrollable situations, numerous fall detection approaches were introduced. Vision based device methods are gradually utilized in various situations. Various vision sensors have been employed in fall detection task, comprises infrared sensors, RGB cameras, and depth sensors [7]. Amongst them, RGB cameras are inexpensive and easier to setup, as surveillance systems are established in daily lives.

Currently, DL has presented a higher classifier accuracy than handcrafted features on several computer vision tasks like fall detection, object detection, and action recognition. A DL scheme is depending upon depth data is introduced in [8] for distinguishing human from background and handle dynamic variation of environments like illumination, shadows. PCAnet utilized for extracting features from colour images and later employed SVM to categorize events. [9] proposed a CNN for detecting falls depending upon optical flow displacement and attained reasonable outcomes. Reference [10] fed CNN by a single integrated subtracted depth image and RGB. These approaches are depending upon single modality colour, motion/shape. However, the combination of these modalities has been established effective at performing various processes on multi modal data. For instance, [11] integrated RGB and motion features extracted using CNN for detecting events and established an optimum efficiency was attained related to other networks with one modality.

This paper presents a new vision based elderly fall event detection using deep learning (VEFED-DL) model. At the initial stage, the frames are processed in three levels namely resizing, augmentation, and min–max based normalization. In addition, MobileNet based spatial feature extraction and gated recurrent unit (GRU) based temporal feature extraction processes are involved. Lastly, a group teaching

optimization algorithm (GTOA) with stacked autoencoder (SAE) is employed as a classifier to compute the occurrence of fall/non-fall events. For examining the fall detection efficiency of the presented VEFED-DL model, a series of experimentation is carried out on the UR fall detection dataset and multiple cameras fall dataset.

2 Related Works

In Satpathy et al. [12], a double check technique for detecting falls in elderly people by an IMUL sensor and RGB camera is presented. The IMUL sensor is an integration of an IMU sensor (gyroscope and accelerometer) and an ultra-wideband signal based location sensor; the RGB sensor is fixed on a robot. The presented technique includes confirming and detecting the fall of an elderly person by IMUL sensor and an RGB image, correspondingly. If a significant fall happens, the single position data is coordinated with the movement information. In recognition, due to the serial nature of IMU data, a DL method named RNN is trained for fall classification. Lee et al. [13] proposed a framework for classifying fall events from other indoor natural events of human beings. Primarily, a 2DCNN module is presented for extracting features from video frames. Later, gated recurrent unit (GRU) network detects temporal dependencies of human motion. Lastly, sigmoid classification is utilized as binary classifier for human fall event detection.

Sultana et al. [14] aimed to identify fall event recognition in complex backgrounds depending upon visual data. Contrasting to most traditional background subtraction approaches that are based on background modeling. Chen et al. [15] proposed weighted multi stream DCNN that utilize high multimodal data given by RGB D cameras. This technique identifies manually fall events and transmits a request to the care takers. As compared to the initial feature that lacks the context data regarding prior and succeeding frames, the next modality describes the human shape variation. Later human silhouette, background subtraction, and individual detection is extracted and stacked for defining HBMI. The final 2 modalities are utilized for discriminating the motion data. Khief et al. [16] examined the technical features of FDS depending upon wearable device and AI methods DL for implementing an efficient method for detecting on-line falls. The presented classification is depending upon RNN module with fundamental LSTM blocks.

Musci et al. [17] defines the implementation of a fall detection system for elderly households living alone utilize lower resolution thermal sensor arrays. This technique has designed Bi-LSTM, LSTM, and GRU; the last stated method attained an optimum result at 93% in accurateness. The outcomes attained goal to be a useful tool for preventing accidents and physicians for managing the information. In Taramasco et al. [18], a 3D CNN based technique for detecting falls is established by video kinematic data for training an automated feature extractor and avoid the necessity for huge fall datasets of DL solution. The 2D CNN can encode spatial data, and applied 3D convolution can extract movement feature from temporal series that is essential for detecting fall. For locating the ROI in every frame, LSTM based spatial visual attention system is combined. In Lu et al. [19], automated fall detection utilizing DL is modeled by RGB images collected from single camera source. Mainly, it defines the sensitive detail that is conquered in the original image and guarantees privacy, extensively consider for protection and safety. Several researches are performed by real-world falling datasets. Reddy et al. [20] proposed a vision-based solution using CNN for deciding when a series of frames comprises an individual fall. For modeling the video movement and create the scheme scenario independently, they utilize optical flow images as input to the network succeeded by a new 3 phase training.

3 The Proposed Fall Detection Framework

The system architecture of the proposed fall detection framework is displayed in Fig. 1. Generally, elderly people are monitored by the use of digital cameras that are kept in the room. It helps to provide details about the surrounding environment in case of fall event occurred. The video cameras generate videos of dynamic length and then frames are generated from the captured videos. Then, the sequential frames are fed into the MobileNet model to extract the spatial key features. Afterwards, the extracted features are given into the GRU model to extract the temporal features. Finally, the outcome of the GRU is fed into the GTOA-SAE model to determine the class labels. The detailed working of these modules is given in the succeeding sections.

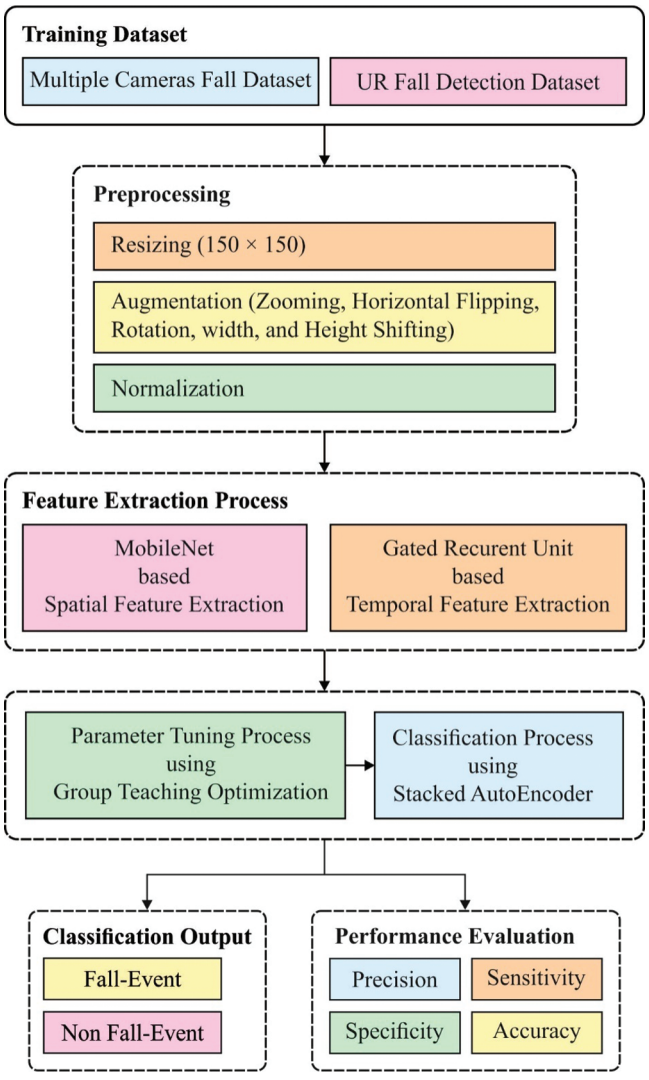

Figure 1: The working process of VEFED-DL model

3.1 Data Pre-Processing

At the initial stage, the frames are pre-processed to improve the characteristics of the image, eliminate the noise artefacts, and enhance particular set of features. Here, the frames are processed in three major levels such as resizing, augmentation and normalization. For reducing the computation cost, the resizing of frames

take place to 150 × 150. Followed by, the frames are augmented where the frames are transformed at every training epoch. To augment the frames, different processes such as zooming, horizontal flipping, rotation, width, and height shifting. Finally, normalization process is applied for improved generalization of the model.

3.2 MobileNet Based Spatial Feature Extraction

CNN generally contains pooling layer, full connection (FC) layer, and convolutional layer. Initially, the features are extracted by more than one pooling and convolution layer [21]. Next, every feature map from final convolution layer is converted to 1D vector for FC. Lastly, the output layer categorizes the input images. The network adjusts the weight variables by backpropagation (BP) and minimizes the square variance among classification outcomes and predictable outputs. The convolutional layer comprises numerous convolutional filters, extract distinct features from the image by convolution function. Later, the non-linear features map is attained with the help of activation function. The pooling layer, so called subsampling layer, is behindhand the convolutional layer. It executes down sampling function, with certain value as output in a specific sub region [22]. Simultaneously, the flexibility of the network to the modifications of image translation and rotation is raised. The Popular pooling functions are highest pooling and average pooling. The structure is depending upon pooling and convolutional layers could enhance the strength of the network module. The CNN could attain deeper multi-layer convolutions. Using amount of layer improving, the features attained by learning turn into additional global.

MobileNet contains high precision, small structure, and lesser computation, that is utilized to embedded devices and mobile terminals. Depending upon depth wise separable convolution, MobileNets utilize 2 global hyper parameters for keeping a balance among accuracy and efficiency. The essential concept of MobileNet is the decompositions of convolution kernel. With depth wise separable convolutional, the typical convolution is decomposed to depth wise convolution and a point wise convolutional with 1 × 1 convolutional kernel. The depth wise convolutional filter performs convolution to every channel, and the 1 × 1 convolution is utilized for combining the outcomes of depth wise convolutional layers. Like this, N standard convolutional kernels are substituted by M depth wise convolutional kernel and N point wise convolution kernel. A typical convolutional filter integrates the inputs to a novel set of output when the depth wise separable convolutional divisions the input into 2 layers, one for filtered and another for merged.

3.3 GRU Based Temporal Feature Extraction

Amongst every kind of human movement, it is needed to consider the temporal features along with the spatial ones to categorize the fall events. The RNN extracts the temporal features by keeping essential details from the past. But, at this operation, it faces vanishing and exploding gradient problems. Therefore, GRU model is employed for resolving the issue of RNN by the use of update and reset gates. The RNN presents the feedback relationship among the hidden layer units, thus the network could maintain the learned data to the present moment and define the last output result of the network along with input of present moment. The efficiency of the RNN in resolving timing problems are authenticated in several study fields and produced various stimulating outcomes, like machine translation, image description, and speech recognition. Additionally, RNN is utilized for predicting the events connected to time sequences, like stock prediction. But, in the training procedure, RNN overcomes the problem of vanishing gradient that results in network failure to normally converge and RNN could not face the effect of long term dependency.

Various network architectures to improve the RNN have been presented, and extensively utilized and efficient structure is the LSTM [23]. A cell comprises output gate, input gate, and forget gate, correspondingly. The advanced study literature and result show that LSTM is an efficient method to solve

long term dependency, and the problem of vanishing gradient is alleviated by gating technique.Since the common variant of LSTM, gated recurrent unit (GRU) simplified the gated structure in LSTM cells and utilizes reset gate and upgrade gate for replacing 3 gates in LSTM, where the reset gate defines the novel input data with the prior memory, and the upgrade gate determines the prior information should be saved for the present time step. The GRU have attained suitable outcomes in various study field, and related to LSTM, GRU could save more trained time and computing resource. Representation for unfolded structure of GRU is given in Fig. 2. The fundamental estimation procedure of GRU is given in the subsequent formula.

(a) Candidate State

$$\bar{h}_t = g\left(W_{fh}x_t + W_{rh}(h_{t-1} \odot r_t) + \varphi_h\right) \tag{1}$$

(b) Reset Gate

$$r_t = f\left(W_{fr}x_t + W_{rr}h_{t-1} + \varphi_r\right) \tag{2}$$

(c) Update Gate

$$z_t = f\left(W_{fz}x_t + W_{rz}h_{t-1} + \varphi_z\right) \tag{3}$$

(d) Current State

$$h_t = (1 - z_t) \odot \bar{h}_t + z_t \odot h_{t-1} \tag{4}$$

where r_t and z_t denotes output vector of reset and upgrade gates at present time step t, Let h_t and $\bar{h}_t$ denotes state and candidate states vector at present time step t, correspondingly $\varphi_h \in R^{n\times 1}$, $\varphi_r \in R^{n\times 1}$, and $\varphi_z \in R^{n\times 1}$ indicates bias vectors. $W_{fh} \in R^{n\times m}$, $W_{fr} \in R^{n\times m}$, and $W_{fz} \in R^{n\times m}$ denotes weight matrices of the feed forward connection, correspondingly Also, $W_{rh} \in R^{n\times n}$, $W_{rr} \in R^{n\times n}$, and $W_{rz} \in R^{n\times n}$ denotes weight matrix of the recurring connection, correspondingly. Particularly, weight sharing method is employed for distinct time steps t.$\odot$ denotes operational rule of elementwise multiplication among vectors. g and f denotes neuronal activation function, whereas g denotes tanh function and f indicates sigmoid function.

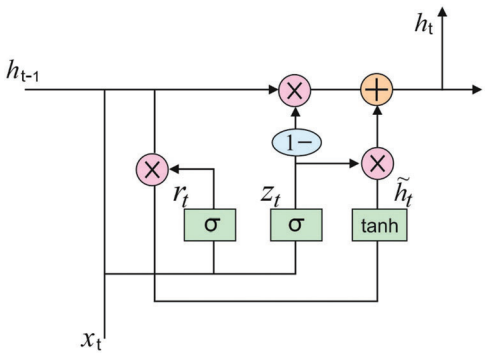

Figure 2: The structure of GRU

Because of the multi objective classification included in the research, SoftMax classifier is utilized as last output layer of GRU in this study:

$$softMax(y_i) = \frac{e^{y_i}}{\sum_{i=1}^{I} e^{y_i}} \tag{5}$$

whereas y_i denotes ith element of GRU's output vector and fulfills $\sum_{i=1}^{I} softmax(y_i) = 1$. I represents dimension of the output vector.

3.4 SAE Based Classification Process

The AE is a NN for unsupervised feature learning [24]. The illustrative layers of the AE are consisting of a decoder and encoder, which consist of succeeding non-linear AE functions:

$$f(x) = e_f(W_e x_i + b_e) \tag{6}$$

$$g(x) = d_f(W_d f(X_i) + b_d) \tag{7}$$

whereas $f(x)$ and $g(x)$ denotes decoder and encoder functions, correspondingly; W_e and W_d represents weight matrix, whereas b_e and b_d indicates bias vector. For activation function, sigmoid function is used by decoding and encoding layers, represented by Eqs. (8) and (9):

$$e_f = \frac{1}{1 + e^{-(W_e x_i + b_e)}} \tag{8}$$

$$d_f = \frac{1}{1 + e^{-(W_d f(x_i) + b_d)}} \tag{9}$$

The encoder layer converts higher dimension data to lower dimension data when decoding recovers the lower dimension data and becomes higher dimension data which is similar to the original input structure. Particularly, the AE has benefits in feature extraction and dimension reduction of non-linear data. Therefore, this research presented a variant AE network through SAE. Thorough data and scientific equations of the SAE are illuminated in the succeeding section. Fig. 3 illustrates the architecture of SAE.

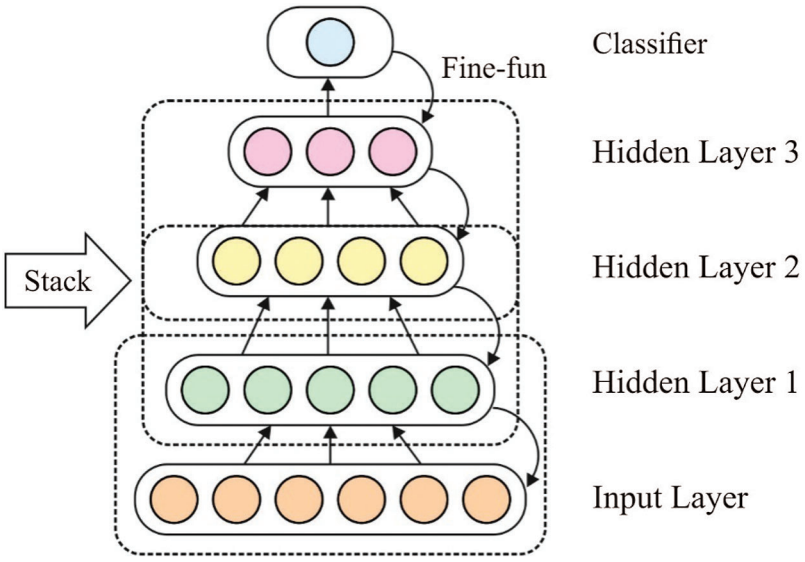

Figure 3: The architecture of SAE

The essential standard of SAE is equivalent to the original AE network. SAE is an alternate to the fundamental AE network while handling complex feature data of the hyperspectral image cube. In contrast to the AE that has a single hidden layer, SAE contains many encoded and decoding layers, denoted in subsequent formula:

$$f_k(\times) = e_{k,f}\left(W_{k,e^x k,i} + b_{k,e}\right) \tag{10}$$

$$g_t(\times) = d_{t,f}\left(W_{t,d} f\left(x_{t,i}\right) + b_{t,d}\right) \tag{11}$$

whereas $f_k(x)$ and $g_t(x)$ indicates encoder and decoder function in kth and tth layers, correspondingly $W_{k,e}$ and $W_{t,d}$ represents weight matrix in the kth and tth layer, when $b_{k,e}$ and $b_{t,d}$ denotes bias vectors. For optimizing SAE network, the error among input and output information must be reduced.

$$Y = \min\left(\frac{\Sigma_{i=1}^{N}(g(x) - I_o)^2}{N}\right) \tag{12}$$

where Y indicates cost function, N represents amount of nodes, $g(x)$ denotes recreated input, and I_o denotes original input. For training the SAE network, the BP error derivative updates the network variables in the AE layers in the network by function in Eq. (13)

$$\delta = \frac{\partial Y}{\partial A_f} \tag{13}$$

whereas A_f denotes AE function.

3.5 Parameter Tuning Process

To optimally adjust the parameters involved in the SAE model, GTOA is employed to improve the fall detection rate. In GTOA, the knowledge of the entire class (c) is enhanced, that is the fundamental concept after the presented method, viz. GTOA. For implementing GTOA optimization, a simple group of teaching modules was proposed according to the succeeding rules.

a) The variance among students has the ability to be acquiescent of knowledge. The great challenge for the teacher for developing the teaching strategy based on the variances in capability for accepting knowledge.
b) The quality of an optimum teacher is to focus more on students have a poor ability for accepting knowledge.
c) With self-learning, or interaction with classmates, a student could enhance their knowledge in the free time.
d) For improving the knowledge of students, a good teacher allocation technique is useful.

For representing the knowledge of entire classes, a normal distribution function is utilized by Eq. (14).

$$f(x) = \frac{1}{\sqrt{2\pi}\delta} exp^{(x-\mu)^2} \tag{14}$$

whereas x denotes value for normal distribution function is needed. μ indicates mean and δ denotes standard deviation [25]. An outstanding group is a set of students with an optimum capability for understanding the knowledge, where the group with poor ability for grasping knowledge so-called average group.

In the teacher stage, students learn from teacher that is the additional rules defined previously. In GTOA the teacher makes separate strategies for average and outstanding groups.

Teacher phase I:

The teacher emphases enhancing the knowledge of entire classes because of the good capability of students for accepting knowledge. Student belongs to an outstanding group could develop the knowledge using Eq. (6).

$$X_{teacher\_j}^{t+1} = X_j^t + a \times (T^t - F \times (B \times M^t + c \times X_j^t)) \tag{15}$$

$$M^t = \frac{1}{N}\sum_{j=1}^{N} X_j^t \tag{16}$$

$$b + c = 1 \tag{17}$$

whereas an amount of students is denoted as N, X_j indicates knowledge of every student, T denotes teacher knowledge, M denotes mean group knowledge. Teacher factor denotes $F, X_{teacher\_j}$ indicates knowledge of student j learning to a teacher. Arbitrariness is presented as, b, and c from the range zero and one.

Teacher phase II:

Because of weaker knowledge accepting capability, depending upon additional rules, the teacher focuses more on average groups. An average group of students could attain knowledge *via* Eq. (18).

$$X_{ieacher,j}^{t+1} = X_j^t + 2 \times d \times \left(T^t - X_j^t\right) \tag{18}$$

whereas d denotes arbitrariness $(0, 1)$. Eq. (19) addresses the problems when a student could not attain knowledge from the teacher phase.

$$X_{teacher,\,j}^{t+1} = \begin{cases} X_{teacher,\quad j}^t f(X_{ieacher,j}^{t+1}) < f(X_j^t) \\ X_j^t, \quad f(X_{teacher,\,j}^{t+1}) \geq f(X_j^t) \end{cases} \tag{19}$$

In spare time student could attain knowledge by self-learning, or with interaction fellow students, that is denoted arithmetically by Eq. (20). The student phase relations to the 3rd rule by involving student phases I and II.

$$X_{teacher,\,j}^{t+} = \begin{cases} X_{teacher,\,j}^t + e \times (X_{teacher,\,j}^{t+1} - X_{teacher,\,k}^{t+1}) + g \times (X_{teacher,\,j}^{t+1} - X_j^t), & f(X_{teacher,\,j}^{t+1}) < f(X_{teacher,\,k}^{t+1}) \\ X_{teacher,\,j}^t - e \times (X_{teacher,\,j}^{t+1} - X_{teacher,\,k}^{t+1}) - g \times (X_{teacher,\,j}^{t+1} - X_j^t), & f(X_{teacher,\,j}^{t+1}) \geq f(X_{teacher,\,k}^{t+1}) \end{cases} \tag{20}$$

whereas e and g denotes 2 arbitraries $(0, 1)$, $X_{student,j}^{t+1}$ indicates knowledge of the student i, and $X_{teacher,\,j}^{t+1}$ represents knowledge of student j learned to teacher. Students could not attain knowledge from this phase. He/she could be tackled in Eq. (21).

$$X_j^{t+1} = \begin{cases} X_{teacher,\,j}^t f(X_{teacher,\,j}^{t+1}) < f(X_{student,\,j}^t) \\ X_{student,\,j}^t f(X_{teacher,\,j}^{t+1}) \geq f(X_{student,\,j}^t) \end{cases} \tag{21}$$

For enhancing the student knowledge, an optimum teacher allocation technique is critical and determined using 4th rule. Stimulated by hunting performance of grey wolves, the top 3 optimum students are chosen, as given by Eq. (22).

$$T = \begin{cases} x^t_{first} f(X^t_{\text{first}}) \leq f(\frac{X^t_{first}+X^t_{second}+X^t_{third}}{3}) \\ \frac{X^t_{first}+X^t_{second}+X^t_{third}}{3} > f(\frac{X^t_{first}+X^t_{second}+X^t_{third}}{3}) \end{cases} \quad (22)$$

whereas $X^t_{first}, X^t_{second}, f^{irst}$ and X^t_{third} denotes top 3 students, correspondingly

4 Performance Validation

The proposed VEFED-DL model is simulated using Python 3.6.5 tool. The proposed VEFED-DL model is tested using multiple camera fall (MCF) dataset and UR Fall Detection (URFD) dataset. The MCF dataset is commonly used for the classification of fall events. It comprises a set of 192 videos in which 96 videos signify fall and 96 indicate regular events. This dataset is captured under 24 distinct situations under the presence of nine actions such as crouching, walking, falling, etc. The videos have a frame rate of 30 fps with a frame size of 720 × 480. Fig. 4 showcases the sample test images from the input MFC dataset.

Figure 4: Sample images

Besides, the URFD dataset is collected using 2 Kinect sensors and includes frontal and overhead video series. The frontal series encompasses a total of 314 frames, including 74 frames with fall events and 240 frames with no-fall events. The overhead series has 302 frames, comprising 75 and 227 fall and non-fall events respectively. Few test images from the URFD dataset are illustrated in Fig. 5.

Tab. 1 and Fig. 6 offer a brief result analysis of the VEFED-DL model on the classification of fall and non-fall events interms of different measures. From the table, it is demonstrated that the VEFED-DL model has classified the frames as 'Fall-Event' with the precision of 99.98%, sensitivity of 100%, specificity of 99.86%, accuracy of 99.91%, and F-measure of 99.96%. In addition, the VEFED-DL model has categorized the frames as 'Non-Fall Event' with the precision of 99.96%, sensitivity of 100%, specificity of 99.89%, accuracy of 99.92%, and F-measure of 99.98%. Moreover, the VEFED-DL model has accomplished effectual outcomes with an average precision of 99.97%, sensitivity of 100%, specificity of 99.88%, accuracy of 99.92%, and F-measure of 99.97%.

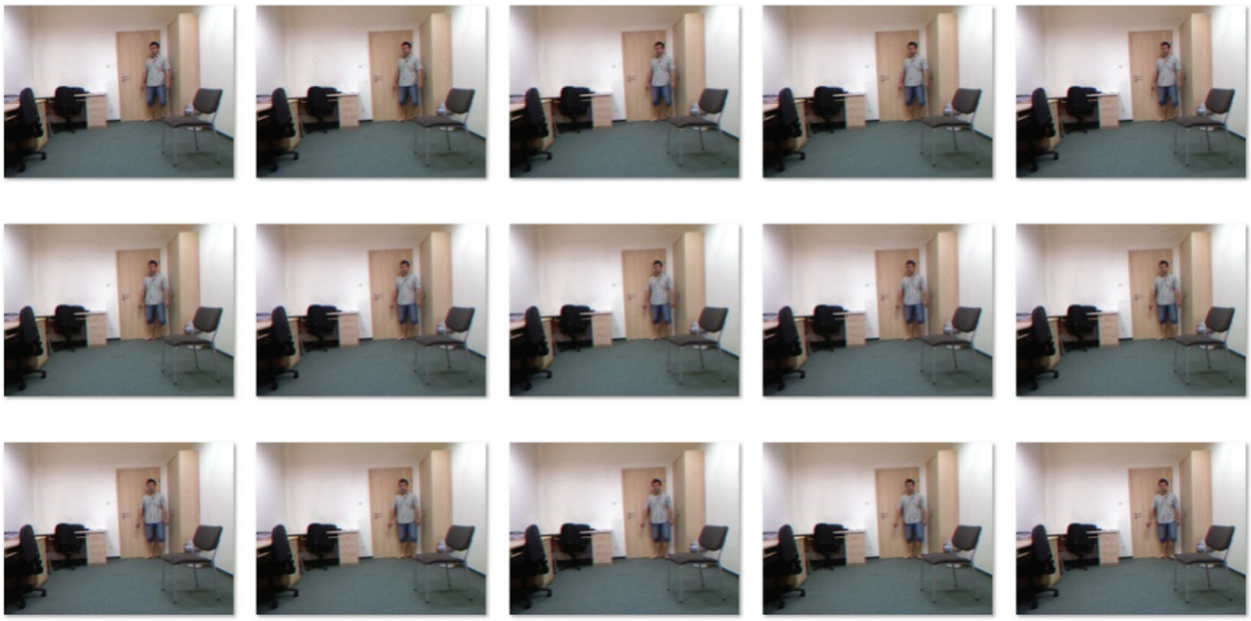

Figure 5: Test image of URFD dataset

Table 1: Results of proposed VEFED-DL method on MCF dataset with respect to distinct measures

Labels	Precision	Sensitivity	Specificity	Accuracy	F-score
Fall-event	99.98	100	99.86	99.91	99.96
Non-fall event	99.96	100	99.89	99.92	99.98
Average	99.97	100.00	99.88	99.92	99.97

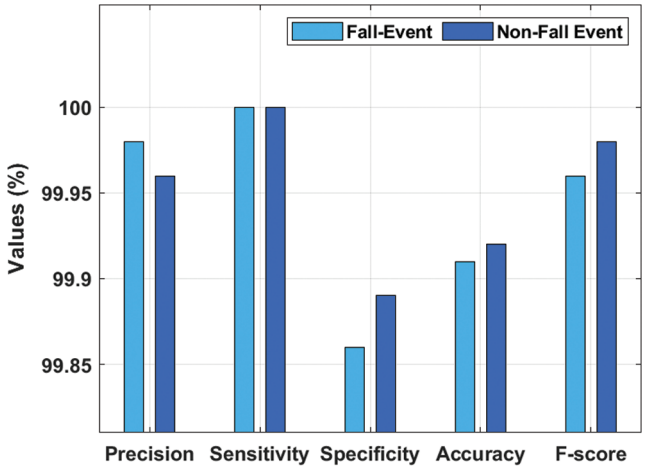

Figure 6: Result analysis of VEFED-DL model on MCF dataset

Tab. 2 and Fig. 7 provide a brief outcome analysis of the VEFED-DL method on the classification of fall and non-fall events with respect to distinct measures. From the table, it can be portrayed that the VEFED-DL method has classified the frames as 'Fall-Event' with the precision of 99.99%, sensitivity of 99.95%, specificity of 99.98%, accuracy of 99.97%, and F-measure of 99.95%. Also, the VEFED-DL model has categorized the frames as 'Non-Fall Event' with the precision of 100%, sensitivity of 100%, specificity of 99.97%, accuracy of 99.99%, and F-measure of 99.98%. Furthermore, the VEFED-DL model has accomplished effectual results with an average precision of 100%, sensitivity of 99.98%, specificity of 99.98%, accuracy of 99.98%, and F-measure of 99.97%.

Table 2: Results of proposed VEFED-DL method on UR fall detection dataset in terms of different measures

Labels	Precision	Sensitivity	Specificity	Accuracy	F-score
Fall-event	99.99	99.95	99.98	99.97	99.95
Non-fall event	100	100	99.97	99.99	99.98
Average	100.00	99.98	99.98	99.98	99.97

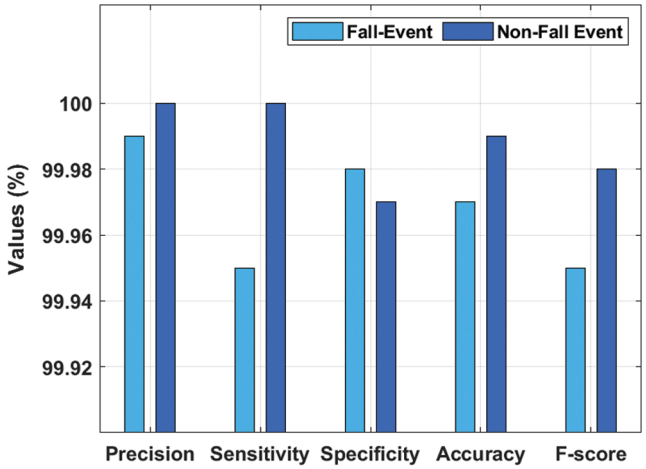

Figure 7: Result analysis of VEFED-DL model on UR fall detection dataset

Fig. 8 demonstrates the ROC analysis of the proposed VEFED-DL model on the applied MCF dataset. From the figure, it is clear that the VEFED-DL model has gained effective performance with a higher ROC of 99.9808.

Fig. 9 exhibits the ROC analysis of the presented VEFED-DL technique on the applied UR Fall Detection dataset. From the figure, it can be revealed that the VEFED-DL technique has gained effective performance with a maximum ROC of 99.9972.

A brief comparative study of the VEFED-DL model with existing methods on the applied MCF dataset is given in Tab. 3 and Fig. 10. From the obtained values, it is apparent that the PCANet model has accomplished insignificant outcomes with a sensitivity of 89.20% and specificity of 90.30%. Besides, the COF model has gained slightly enhanced performance with a sensitivity of 93.70% and specificity of 92.00%. Followed by, the VBFD-CNN model has demonstrated a certainly increased outcome with the

sensitivity of 99% and specificity of 96%. Moreover, the MSDCN model has resulted in a near optimal performance increased outcome with a sensitivity of 99.70% and specificity of 99.80%. However, the proposed VEFED-DL model has showcased superior outcomes with a sensitivity of 100% and specificity of 99.88%.

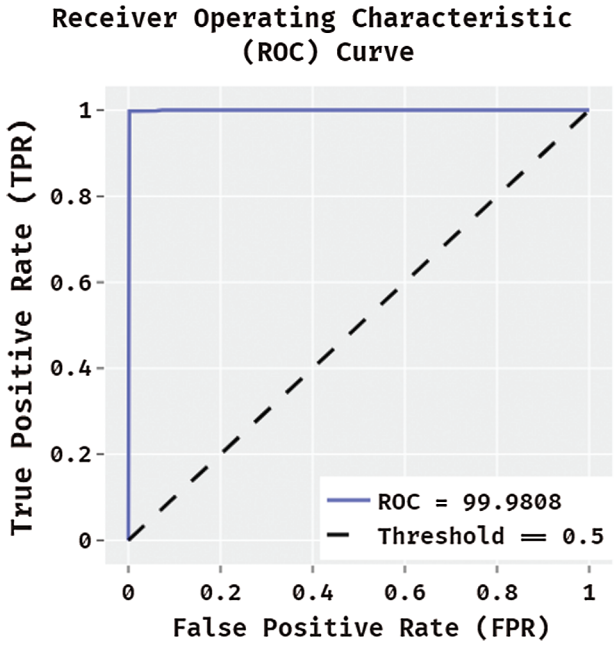

Figure 8: ROC analysis of VEFED-DL model on MCF dataset

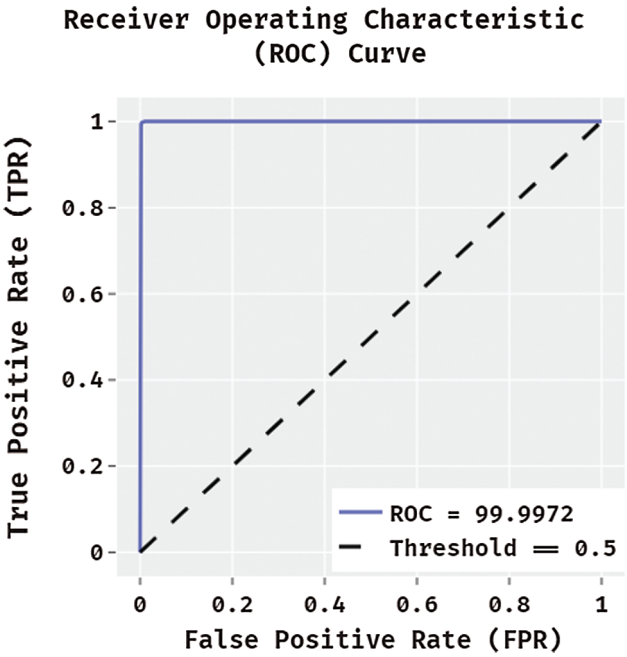

Figure 9: ROC analysis of VEFED-DL model on UR fall detection dataset

Table 3: Results of existing with proposed VEFED-DL method on MCF dataset in terms of sensitivity and specificity

Methods	Sensitivity	Specificity
VEFED-DL	100.00	99.88
MSDCN	99.70	99.80
VBFD-CNN	99.00	96.00
PCANet	89.20	90.30
COF	93.70	92.00

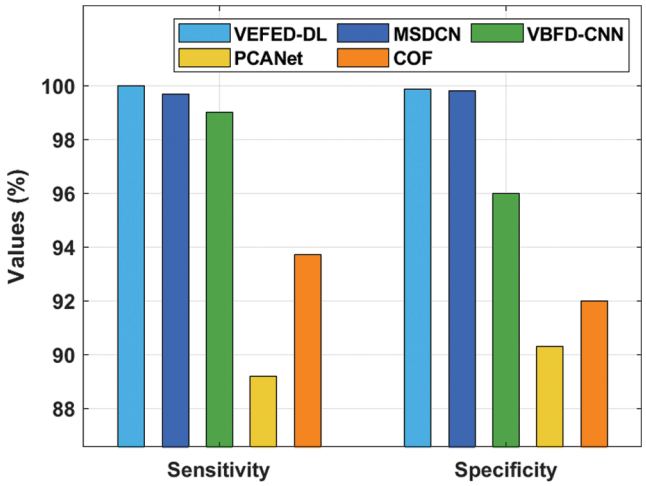

Figure 10: Sensitivity and specificity analysis of VEFED-DL model on MCF dataset

A brief comparative analysis of the VEFED-DL model with state-of-art techniques on the applied UR Fall Detection dataset is provided in Tab. 4 and Fig. 11 [15]. From the attained values, it can be revealed that the FRL-DS technique has accomplished insignificant results with a sensitivity of 96% and specificity of 82%. Similarly, the DM-WA approach has gained somewhat improved performance with a sensitivity of 100% and specificity of 80%. Eventually, the VBFD-CNN method has outperformed a certainly improved result with the sensitivity of 100% and specificity of 92%. In addition, the MSDCN technique has resulted in a near better performance increased outcome with the sensitivity of 100% and specificity of 95%. But, the presented VEFED-DL methodology has demonstrated higher result with a sensitivity of 99.98% and specificity of 99.98%.

Table 4: Results of existing with proposed VEFED-DL method on UR fall detection dataset in terms of sensitivity and specificity

Methods	Sensitivity	Specificity
VEFED-DL	99.98	99.98
MSDCN	100	95
VBFD-CNN	100	92
DM-WA	100	80
FRL-DS	96	82

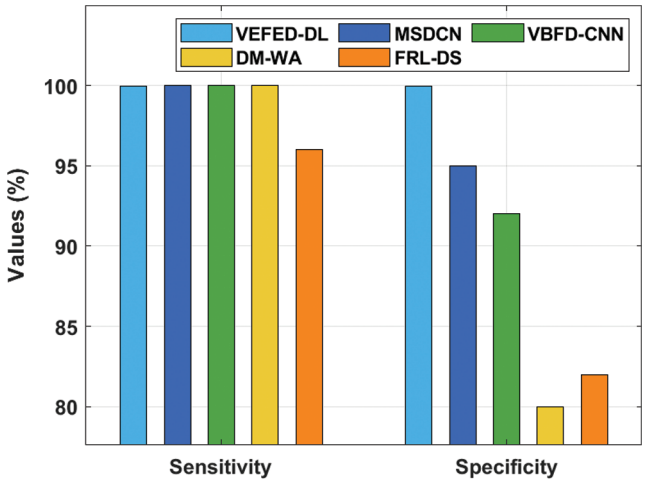

Figure 11: Sensitivity and specificity analysis of VEFED-DL model on UR fall detection dataset

In order to further ensure the goodness of the VEFED-DL model, an accuracy analysis with recent techniques is provided in Tab. 5. From the achieved outcomes, it is showcased that the 2DCNN-LSTM model has attained lowest performance with an accuracy of 0.89. At the same time, 1DCNN-GRU and 2DCNN-Bi-LSTM models have demonstrated improved performance with the accuracy of 0.943 and 0.955. Likewise, the VGG-16, VGG-19, Xception, 3DCNN-LSTM, and CIHFE-DL models have offered moderately closer performance with the accuracy of 0.98, 0.98, 0.99, 0.99, and 0.998 respectively. However, the proposed VEFED-DL model has resulted in higher accuracy of 0.999.

Table 5: Results of existing with proposed VEFED-DL method on MCF dataset in terms of accuracy

Methods	Accuracy (%)
VGG-16	0.980
VGG-19	0.980
Xception model	0.990
1DCNN-GRU	0.943
2DCNN-Bi-LSTM	0.955
3DCNN-LSTM	0.990
2DCNN-LSTM	0.890
CIHFE-DL	0.998
VEFED-DL	0.999

To further confirm the goodness of the VEFED-DL technique, an accuracy analysis with existing models is provided in Tab. 6 [13]. From the attained results, it can be portrayed that the 2DCNN-LSTM method has attained worse performance with an accuracy of 0.88. Likewise, 1DCNN-GRU and 2DCNN-Bi-LSTM techniques have showcased increased performance with the accuracy of 0.927 and 0.95. Similarly, the 3DCNN-LSTM, VGG-16, VGG-19, Xception, and CIHFE-DL approaches have offered moderately closer performance with the accuracy of 0.975, 0.976, 0.98, 0.98, and 0.98 correspondingly. But, the

presented VEFED-DL methodology has resulted in a maximum accuracy of 1.000. From the above-mentioned tables and figures, it is apparent that the proposed VEFED-DL model is found to be an effective fall detection tool to assist elderly people in real time.

Table 6: Results of existing with proposed VEFED-DL method on UR fall detection dataset in terms of accuracy

Methods	Accuracy (%)
VGG-16	0.976
VGG-19	0.980
Xception Model	0.980
1DCNN-GRU	0.927
2DCNN-Bi-LSTM	0.950
3DCNN-LSTM	0.975
2DCNN-LSTM	0.880
CIHFE-DL	0.980
VEFED-DL	1.000

5 Conclusion

This paper has developed a novel VEFED-DL model for fall detection in elderly people. The proposed VEFED-DL model involves different stages of operations such as pre-processing, MobileNet based spatial feature extraction, GRU based temporal feature extraction, SAE based classification, and GTOA based parameter tuning. In the proposed VEFED-DL model, the GTOA is employed for the parameter optimization of the SAE model in such a way that the detection performance can be enhanced. For examining the fall detection efficiency of the presented VEFED-DL model, a series of experimentation is carried out on the UR fall detection dataset and multiple cameras fall dataset. The experimental analysis has shown that the proposed VEFED-DL model outperformed the existing methods interms of several performance measures. As a part of future work, the performance of the VEFED-DL model can be enhanced by the use of advanced deep learning architectures with class attention layer.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. R. Lord and J. Dayhew, "Visual risk factors for falls in older people," *Journal of the American Geriatrics Society*, vol. 49, no. 5, pp. 508–515, 2001.

[2] B. M. Kistler, J. Khubchandani, G. Jakubowicz, K. Wilund and J. Sosnoff, "Falls and fall-related injuries among us adults aged 65 or older with chronic kidney disease," *Preventing Chronic Disease*, vol. 15, pp. 170518, 2018.

[3] G. Allali, E. I. Ayers, R. Holtzer and J. Verghese, "The role of postural instability/gait difficulty and fear of falling in predicting falls in non-demented older adults," *Archives of Gerontology and Geriatrics*, vol. 69, pp. 15–20, 2017.

[4] E. R. Burns, J. A. Stevens and R. Lee, "The direct costs of fatal and non-fatal falls among older adults—United States," *Journal of Safety Research*, vol. 58, no. 7, pp. 99–103, 2016.

[5] M. Mubashir, L. Shao and L. Seed, "A survey on fall detection: Principles and approaches," *Neurocomputing*, vol. 100, no. 1, pp. 144–152, 2013.

[6] M. Wortmann, "Dementia: A global health priority-highlights from an ADI and World Health Organization report," *Alzheimer's Research & Therapy*, vol. 4, no. 5, pp. 40, 2012.

[7] J. Han, L. Shao, D. Xu and J. Shotton, "Enhanced computer vision with microsoft kinect sensor: A review," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1318–1334, 2013.

[8] A. Doulamis and N. Doulamis, "Adaptive deep learning for a vision-based fall detection," in *Proc. of the 11th Pervasive Technologies Related to Assistive Environments Conf.*, Corfu Greece, pp. 558–565, 2018.

[9] A. Núñez-Marcos, G. Azkune and I. Arganda-Carreras, "Vision-based fall detection with convolutional neural networks," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–16, 2017.

[10] S. Neelakandan and D. Paulraj, "A gradient boosted decision tree-based sentiment classification of twitter data," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 18, no. 4, pp. 1–21, 2020.

[11] S. Neelakandan and D. Paulraj, "An automated exploring and learning model for data prediction using balanced CA-SVM," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, 2020.

[12] S. Satpathy, P. Mohan, S. Das and S. Debbarma, "A new healthcare diagnosis system using an IoT-based fuzzy classifier with FPGA," *The Journal of Supercomputing*, vol. 76, no. 8, pp. 5849–5861, 2020.

[13] D.-W. Lee, K. Jun, K. Naheem and M. S. Kim, "Deep neural network-based double-check method for fall detection using IMU-L sensor and RGB camera data," *IEEE Access*, vol. 9, pp. 48064–48079, 2021.

[14] A. Sultana, K. Deb, P. K. Dhar and T. Koshiba, "Classification of indoor human fall events using deep learning," *Entropy*, vol. 23, no. 3, pp. 328, 2021.

[15] Y. Chen, W. Li, L. Wang, J. Hu and M. Ye, "Vision-based fall event detection in complex background using attention guided bi-directional LSTM," *IEEE Access*, vol. 8, pp. 161337–161348, 2020.

[16] C. Khraief, F. Benzarti and H. Amiri, "Elderly fall detection based on multi-stream deep convolutional networks," *Multimedia Tools Applications*, vol. 79, no. 27–28, pp. 19537–19560, 2020.

[17] M. Musci, D. De Martini, N. Blago, T. Facchinetti and M. Piastra, "Online fall detection using recurrent neural networks on smart wearable devices," *IEEE Transactions on Emerging Topics in Computing*, vol. 2, no. 2, pp. 1, 2020.

[18] C. Taramasco, Y. Lazo, T. Rodenas, P. Fuentes, F. Martínez *et al.,* "System design for emergency alert triggered by falls using convolutional neural networks," *Journal of Medical Systems*, vol. 44, no. 2, pp. 50, 2020.

[19] N. Lu, Y. Wu, L. Feng and J. Song, "Deep learning for fall detection: Three-dimensional CNN combined with LSTM on video kinematic data," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 314–323, 2019.

[20] G. P. Reddy and M. K. Geetha, "Video based fall detection using deep convolutional neural network," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 2, pp. 5542–5551, 2020.

[21] A. Núñez-Marcos, G. Azkune and I. Arganda-Carreras, "Vision-based fall detection with convolutional neural networks," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–16, 2017.

[22] W. Wang, Y. Hu, T. Zou, H. Liu, J. Wang *et al.,* "A new image classification approach via improved mobilenet models with local receptive field expansion in shallow layers," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–10, 2020.

[23] B. Yan and G. Han, "LA-GRU: Building combined intrusion detection model based on imbalanced learning and gated recurrent unit neural network," *Security and Communication Networks*, vol. 2018, no. 1, pp. 1–13, 2018.

[24] J. Pyo, H. Duan, M. Ligaray, M. Kim, S. Baek *et al.,* "An integrative remote sensing application of stacked autoencoder for atmospheric correction and cyanobacteria estimation using hyperspectral imagery," *Remote Sensing*, vol. 12, no. 7, pp. 1073, 2020.

[25] M. H. Zafar, T. Al-shahrani, N. M. Khan, A. F. Mirza, M. Mansoor *et al.,* "Group teaching optimization algorithm based MPPT control of PV systems under partial shading and complex partial shading," *Electronics*, vol. 9, no. 11, pp. 1962, 2020.