Human fall detection using Modified EfficientNet

S.Sowmiya[1], D.Menaka[2]

[1,] NICHE,Kanyakumari,TamilNadu,India.,ORCID ID :0000-0001-6502-4095, Email: sowmiyabashir@gmail.com
[2,]NICHE,Kanyakumari,TamilNadu,India., ORCID ID :0000-0001-6539-0247,Email:menakaberita@gmail.com

Abstract:Falls among humans can have severe consequences, especially among elderly individuals. These incidents can lead to long-lasting physical impairments and even death. Falls pose a significant health issue that should not be underestimated.It is a social crave to prevent human falls but it is not fully possible because it is impractical to ensure full prevention.An effective approach to mitigate the effects of falls is the prompt detection of such incidents in real time. so that anemergency aid can quickly be provided which helps to reduce the severity of the fall ,results in abiding health consequences.With the increase in elderly population the demand for fall detection methods is highly increased.This study proposed is a non-invasive, vision-based system that does not require any wearable devices. This system can be implemented in various settings such as homes, hospitals, elderly care facilities, and rehabilitation centres.

Keywords: transformer; human fall detection, time series,IoT, ADL (Activities of daily living)

1. Introduction

Among the various human activity,fall is an important and abnormal activity which has to be identified with great priority.Standard supervised machine learning techniques cannot be applied to recognise this activity due to unavailability of sufficient training data.

Researches are advanced to identify ADL using various machine learning techniques,deep learning and using transformers.But the case of fall detection cannot be missed because it can impose health and safety risks on an individual especially in the case of elderly care or patient monitoring.Obtaining a dataset for fall detection can be challenging and risky as it involves capturing real falls, which can pose a danger to the individual. While data can be generated in controlled laboratory settings, it may not accurately reflect actual falls that occur in real-life scenarios. This presents a significant problem in ensuring that the fall detection system is reliable and effective. The difficulty in acquiring a representative dataset is a significant barrier in the development of robust fall detection systems.As the aging population grows, there is a growing interest in developing Ambient Assisted Living (AAL) systems to promote independent living. This trend is driven by the increasing old-age dependency ratio and the need to support individuals in their daily lives. AAL systems aim to enhance the quality of life and provide support to individuals, enabling them to live independently for as long as possible. This has led to increased attention and investment in the development of these systems to meet the needs of the aging population.Various schemes and programmes are introduced for the welfare of senior citizens and the area is of prime concern for the government.Research for AAL community mainly focus on the recognition of activities,behaviour and situations within an environment which is essential for giving the assistance and emergency care .Ambient Assisted Living offers a variety of services and technologies aimed at improving the lives of individuals and enabling them to age in place with dignity and independence. The solutions provided by Ambient Assisted Living help individuals maintain their health, safety, and social connections, which are crucial factors in promoting a fulfilling and active lifestyle. By fostering an environment that supports health and wellness, Ambient Assisted Living empowers individuals to stay active, engaged, and connected to the world around them for as long as possible.AAL also focus on techniques and solutions for helping people with disabilities, and for assisting caregivers and medical staffAmbient Assisted Living incorporates the use of advanced technology, such as smart devices, wireless networks, and connected nodes, to improve the quality of life for individuals. These solutions can range from smart phones and wearable sensors to non-wearable monitoring systems. The goal of AAL is to provide individuals with the tools they need to live a more self-sufficient and comfortable life.

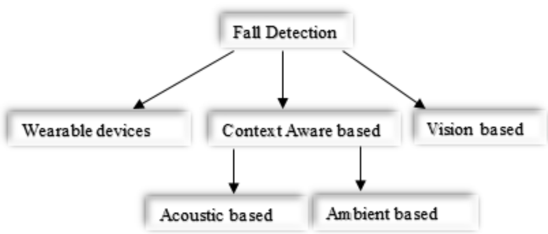

Fig 1: Classification of fall detection

Based on the nature of data collection, human fall-detection systems are classified into three extensive categories (Nizam et al., 2016)[1] a)wearable device based system, b) vision based system and c) context-aware based system (non-wearable device system) where it can be acoustic or ambient based system.While using wearable sensors ,it is more intrusive to the users at the same time context aware based system depends on the environment conditions.The vision based system is more robust and gives better performance and high precision.

Until recently most of the fall detection methods relied upon machine learning techniques like SVM, kNN, GMM etc where feature engineering is done manually.But with the advances in hardware ,utilising the power of GPU deep neural networks can be implemented which can automatically extract features .It works on big amount of data, the efficiency is much more and appropriate for resolving complex issues. The next phase of research in fall detection is in the avenue of deep learning techniques. Among them CNN is widely used technique .Also variations of the CNN ,RNN models, hybrid models including CNN with LSTM also experimented. Now it's the era of models with scaling of neural networks.Starting with the basic CNN models,several variants are proposed by researchers like AlexNet, ResNet, MobileNet, InceptionNet,EfficientNet.AlexNet is implemented in GPU while ResNet efficiently tackle vanishing gradient problem.MobileNet is a lighter version and InceptionNet can extract small and large features simultaneously.In EfficientNet the authors proposed compound scaling powered with squeeze and excitation network.

The main contributions and intuitions of this paper are summarized as follows:

- This study proposed a deepneural network model based on CNN called EfficientNet architecture for processing and detection of fall.
- Based on the compound scaling mechanism, MingxingTan et al. 2019 [15], EfficientNet architecture is designed .
- The proposed model is evaluated using publically available datasets for fall detection and achieved state-of-the-art prediction performance.

The remainder of this paper is structured as follows.In Section 2, we present a review of existing work in this research field. In Section 3, we present our approach followed by, results in Section 4, and conclusion and future scope in Section 5.

2. Related Work

Boyle and Karunanithi,[2]and work by Chen et al. (2011)[3]experimented threshold-based method for the classification between a fall and an ADL.Zhao et al [4] used tri axial gyroscope for fall detection using decision tree and achieved outstanding accuracy. Jokanović and Amin (2017) [5]with the help of range-Doppler RADARs using logistic regression withclassifier in DNN detected fall.In the field of Acoustic-based fall detection systems,most works are done using KNN and Gaussian mixture model with SVM.Li et al [6],Zhuang [7].For the case of vision based fall detection system,Rougier et al utilised GMM method using background subtraction and analysis of human shape deformation.[8].Also from the studies of Anderson et al [9] using Multi camera vision system with two level fuzzy logic classified fall and not fall.Salleh et al[10] using Nonlinear Auto Regression neural network (NARnet) with the help of Invensense sensor detected fall..

In fall detection,SVM is the commonly used ML algorithm followed by kNN. SVM fall detection is based on pattern recognition and since the pattern of a standing and falling person have a wide margin around the hyperplane, SVM can easily discriminate fall and not fall. SVM are a popular machine learning algorithm that have the ability to handle high dimensional data and are known for their robustness against overfitting. SVM use a concept called margin to separate the classes, which helps to reduce the risk of overfitting. Additionally, the use of regularization techniques in SVM can further control the overfitting issue. As a result, SVMs are often considered a good choice for complex and high dimensional data sets.

The k-Nearest Neighbor (kNN) algorithm is another popular machine learning method that is known for its ability to make real-time predictions. The classification results can be obtained very quickly, making it suitable for real-time applications such as fall detection. However, one of the main drawbacks of using kNN is that it can be computationally expensive, as a large amount of memory is required to store the training data. This can slow down the system, making it less efficient and less suited for use in resource-constrained environments. Despite this drawback, kNN is still widely used due to its ability to adapt to different data points and produce accurate results, making it a useful tool for many applications.Neural networks have the ability to learn and make predictions independently. They can identify patterns and relationships in data on their own, without relying solely on the input data. This allows them to continuously improve their performance over time, as they are exposed to more data.Most of the recent works on fall detection using vision based system used CNN techniques[16] followed by hybrid techniques[17].Chen.Y et al[18] proposed a fall detection method based on wireless channel state information (CSI). They used a method that uses continuous wavelet transform (CWT) to generate images and then uses transform learning of convolutional networks for classification. And a wavelet scattering network to automatically extract features and classify them using aLSTM network. Other DL techniques like RNN and Auto Encoder is also used.Many research works used Thermal and infrared sensors to protect the privacy of the users. Depth cameras use infrared sensors to capture the images and can work even in less light environments. So there is no concern for privacy.

3. Proposed Work

3.1. EfficientNet Architecture

EfficientNet is a innovative Convolutional Neural Network (CNN) architecture that seeks to achieve optimal balance between accuracy and computational efficiency. This is achieved by using a scaling method that uniformly scales all dimensions of width, depth, and resolution using a single compound scaling coefficient. This results in models that outperform traditional, individually scaled models in terms of accuracy and efficiency. EfficientNet uses a technique called compound coefficient to uniformly scale up models in three dimensions in a simple but effective manner. Unlike the usual method of randomly scaling up width, depth or resolution, compound scaling uniformly scales each dimension with a certain fixed set of scaling coefficients. Using the Auto ML and scaling method there are seven models of various dimensions in EfficientNet, which surpassed the state-of-the-art accuracy of conventional CNN, and have anoutstanding accuracy and efficiency with an order of magnitude fewer parameters and FLOPS.

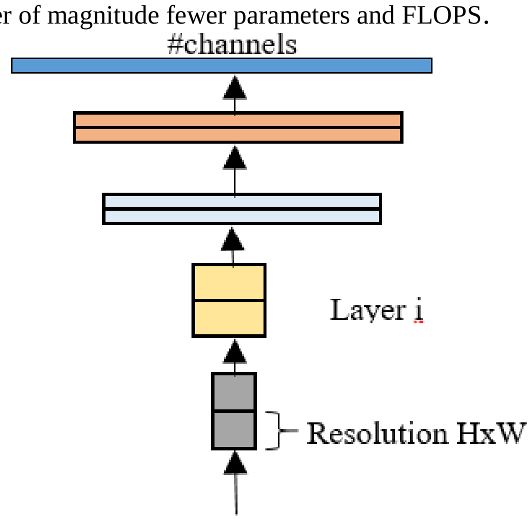

Fig 2: Baseline-EfficientNet

EfficientNet is developed by the neural architecture search using the AutoML MNAS framework, based on the baseline network. Fig 2 shows the baseline for EfficientNet.Conventional scaling as well as compound scaling are possible on this baseline architecture to provide different variations of the model .The network is fine-tuned for obtaining maximum accuracy but computational cost is high.When the network takes a lot of time to make predictions, it is penalized for slow inference time. The architecture is based on mobile inverted bottleneck convolution similar to MobileNet V2 and is much larger because of the increase in FLOPS. This baseline model is compound scaled up to attain the family of EfficientNets.

For developing the method of compound scaling, the research studies have been made to know the impacts that each scaling technique has on the model’s performance and efficiency. They figured that scaling single dimensions helps improve model performance. So it is proved that balancing the scale in all the three dimensions ;depth,width, and image resolution ,while considering the variable available resources best improve the overall model performance.

The compound scaling method used in EfficientNet is based on the concept of balancing the dimensions of width, depth, and resolution by scaling each dimension with a constant ratio. This results in a model that is more efficient and accurate compared to traditional models that only scale one dimension at a time.. The equations below show how it is achieved mathematically,

$$\text{Depth } d=\alpha^{\phi},\text{Width } w=\beta^{\phi},\text{Resolution } r=\gamma^{\phi}$$

$$\text{such that } \alpha.\beta^{2}.\gamma^{2}=2$$

$$\alpha\geq1,\ \beta\geq1,\ \gamma\geq1$$

The underlying reasoning behind using larger networks for larger input images is the need to expand the receptive field. This is because a larger image requires the network to have a wider field of view to capture all the details and patterns within the image. Additionally, as the size of the image increases, it becomes necessary to have more channels in the network to be able to detect finer patterns within the image. This helps to maintain the accuracy of the network even when working with larger images.. The compound scaling technique also helped improve the model efficiency and accuracy of previous CNN models such as MobileNet and ResNet by around 1.4% and 0.7% ImageNet accuracy, respectively, compared to otherrandom scaling methods.

The performance of EfficientNet compared to other network architectures. EfficientNet B7, the largest among the seven EfficientNet models, has achieved state-of-the-art results on the ImageNet and CIFAR-100 datasets. This model achieved a top-1 accuracy of 84.4% and a top-5 accuracy of 97.3% on ImageNet. These results demonstrate the exceptional performance capabilities of EfficientNet B7.At the same time, the model size was 8.4 times smaller and 6.1 times faster than the best CNN counterpart. It obtained 91.7% accuracy on the CIFAR-100 data set and 98.8% accuracy on the Flowers dataset.EfficientNet model provides better Class Activation Maps (CAM),since the scaling model focus more on the relevant regions with more object details, thus provides an enhanced explainability of the model.

3.2.Fall detection workflow

Figure 3 presents the block diagram of the proposed fall detection process with various stages, such as Training ,Feature extraction, Cross validation and classification. The various stages are typically carried out in a sequential manner to effectively detect fall events.The image data set is divided into train (90%) and validation (10%) and the model is built on the proposed method. Finally the model is verified using the test dataset.Also the performance of the model is compared with existing works.

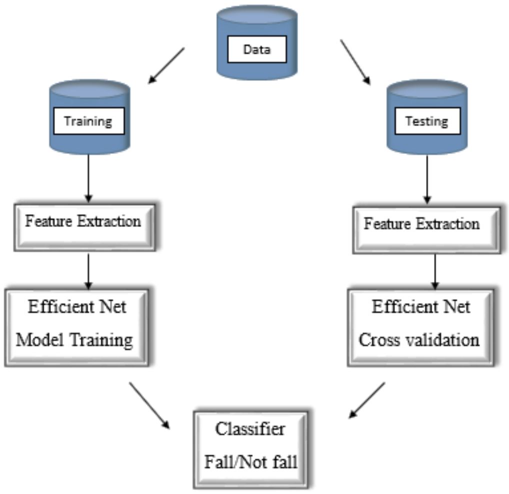

Fig 3: Fall detection process

Fig 4 shows the system framework for human fall detection.The various surveillance cameras inside the room monitors the person and the data is send to the care-taker of the family. If any fall is detected an emergency

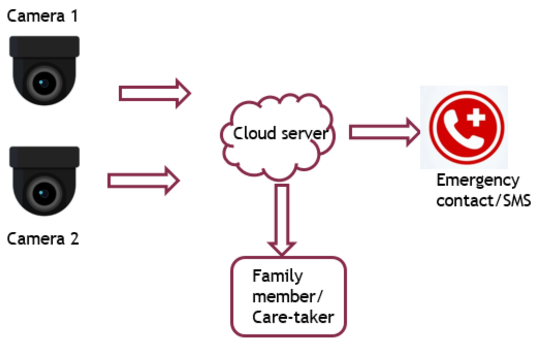

Fig.4.System framework

call/SMS will be send the responsible person.Thus giving alert immediately emergency aid can be given and the consequence of fall can be reduced.

4. Experimental results and discussion

The dataset used here is a standard dataset which is publically available taken from Kaggle [11]. The dataset

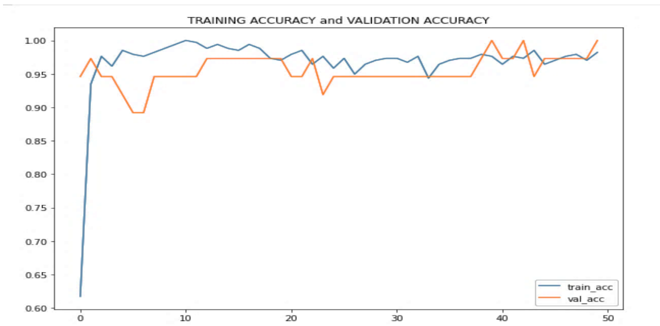

Fig 5:Training,Validation accuracy plot

containstraunans test datafor preparing the neural network model. Images directories consist of two subdirectories train which is used for training and val for validation. Another directory named label consists of two subdirectories train and val.In this directory text files with labels of that particular image is included.Labels file corresponding to the image file names consists ofone class label and four bounding box

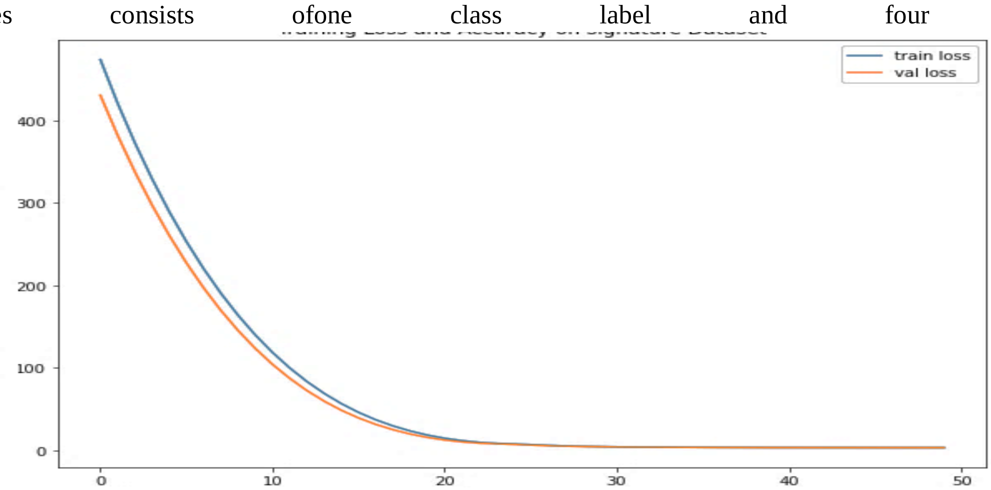

Fig 6:Training ,Validation loss plot

Author	Method	Accuracy
Ge et al. (2017) [12]	SVM	90.50%
de Miguel et al. (2017) [13]	kNN	SE - 96% SP - 97%
Zhao et al. (2018) [14]	Large Margin based kNN	SE - 86% SP - 98%
Proposed method	EfficientNet	97.92%

Table 1: Comparison

values for that particular image.In this dataset the labels are Fall Detected, Walking, and Sitting.In first phase we train the model using the dataset.

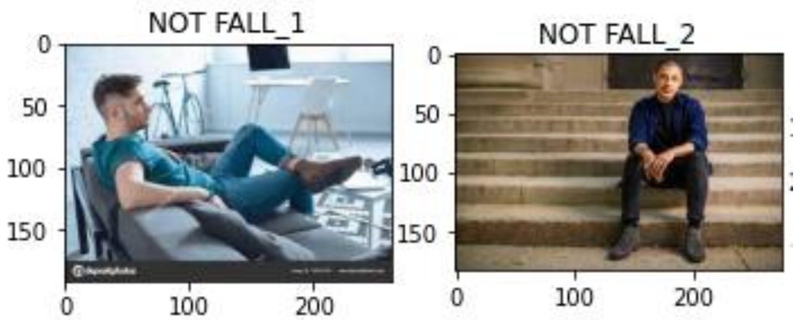

Fig 7:Not Fall images

In the second phase, we produce the input to the modified EfficientNet model for final prediction. 50 epochs were experimented with a learning rate of 0.00125 with Adam optimiser.Six blocks of EfficientNetB0 base model are designed for the proposed model which increases the efficiency.

4.1.Performance evaluation metrics

In fall detection, two evaluation metrics are given priority: Sensitivity (SE) and the Specificity (SP) .The SE (or recall) corresponds to how many relevant elements are actually selected. It gives a measure of fall detection probability which means percentage value of falls(True Positives TP) that have actually been detected. On the other hand The SP corresponds to the probability of not fall detection which indicates the percentage value of non-relevant elements (True Negatives TN) selected, i.e., how many events classified as non-falls are actually non-falls.

$$Sensitivity = \frac{TP}{(TP + FN)}$$

$$Specificity = \frac{TN}{(TN + FP)}$$

And accuracy is given by

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

The model achieved an accuracy of 97.92% which is comparable with other latest deep neural network models.Plot of accuracy and loss for both validation and training is shown in Fig 5 and Fig 6.Also the test results showing fall and not fall images are given in Fig 7 and Fig 8.BesidesTable1 presents the comparison of our proposed method with the existing state of art models.

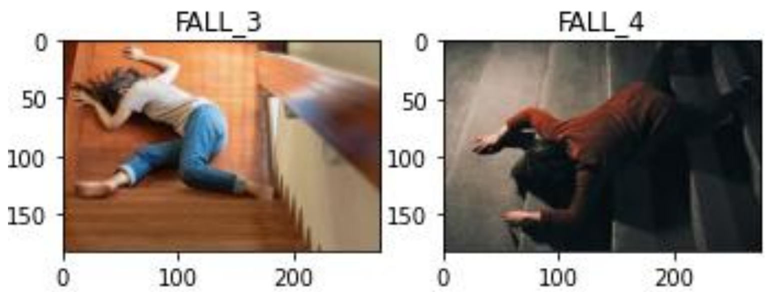

Fig 8: Fall images

5.Conclusion and Future scope

In this study, we introduced a human fall detection method using images and the EfficientNet architecture to differentiate between fall and non-fall events. Additionally, it is important to analyze the pre-fall and post-fall activities for a comprehensive understanding of the fall event. To accomplish this, activity recognition should be performed for a 10-15 minute period before and after the fall. This analysis can help evaluate the accuracy and severity of the detected fall and determine the person's activities after the fall. The use of GPS location, speed, and weather data can provide additional insights into the circumstances surrounding the fall and support the development of more effective fall prevention strategies in the future.

6.References

[1] Nizam, Y., Mohd, M.N.H., & Jamil, M.M.A. (2016). A study on human fall detection systems: daily activity classification and sensing techniques. *International Journal of Integrated Engineering*, *8*(1), 35-43.

[2] Boyle, J., &Karunanithi, M. (2008, August). Simulated fall detection via accelerometers. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 1274-1277). IEEE. Vancouver, BC, Canada.

[3] Chen, D., Feng, W., Zhang, Y., Li, X., & Wang, T. (2011, December). A wearable wireless fall detection system with accelerators. In *2011 IEEE International Conference on Robotics and Biomimetics*(pp. 2259-2263). IEEE. Phuket, Thailand.

[4] Zhao, S., Li, W., Niu, W., Gravina, R., &Fortino, G. (2018b). Recognition of human fall events based on single tri-axial gyroscope. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)* (pp. 1-6). IEEE. Zhuhai, China.

[5] Jokanovic, B., Amin, M., Ahmad, F., &Boashash, B. (2016 b, May). Radar fall detection using principal component analysis. In *Radar Sensor Technology XX* (Vol. 9829, p. 982919). International Society for Optics and Photonics. SPIE Defense + Security, 2016, Baltimore, Maryland, United States.

[6] [Li, Y., Ho, K.C., &Popescu, M. (2012). A microphone array system for automatic fall detection. *IEEE Transactions on Biomedical Engineering*, *59*(5), 1291-1301.

[7] Zhuang, X., Huang, J., Potamianos, G., & Hasegawa-Johnson, M. (2009, April). Acoustic fall detection using Gaussian mixture models and GMM supervectors. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 69-72). IEEE. Taipei, Taiwan.

[8] Rougier, C., Meunier, J., St-Arnaud, A., & Rousseau, J. (2011). Robust video surveillance for fall detection based on human shape deformation. *IEEE Transactions on Circuits and Systems for Video Technology*, *21*(5), 611-622.

[9] Anderson, D., Luke, R.H., Keller, J. M., Skubic, M., Rantz, M., &Aud, M. (2009). Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer Vision and Image Understanding*, *113*(1), 80-89.

[10] MdSalleh, S., mohdyusoff, ainulhusna, ngadimon, K. and Koh , C.Z. 2020. Neural Network Algorithm-based Fall Detection Modelling. *International Journal of Integrated Engineering*. 12, 3 (Feb. 2020), 138–150.

[11] https://www.kaggle.com/datasets/uttejkumarkandagatla/fall-detection-dataset

[12] Ge, C., Gu, I.Y.H., & Yang, J. (2017, September). Human fall detection using segment-level CNN features and sparse dictionary learning. In *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1-6). IEEE. Tokyo, Japan.

[13] de Miguel, K., Brunete, A., Hernando, M., &Gambao, E. (2017). Home camera-based fall detection system for the elderly. *Sensors*, *17*(12), 2864.

[14] Zhao, F., Cao, Z., Xiao, Y., Mao, J., & Yuan, J. (2018a). Real-time detection of fall from bed using a single depth camera. *IEEE Transactions on Automation Science and Engineering*, *16*(3), 1018-1032.

[15] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *International conference on machine learning*. PMLR, 2019.

[16] X. Cai, X. Liu, M. An, G. Han, Vision-based fall detection using dense block with multi-channel convolutional fusion strategy, IEEE Access 9 (2021) 18318–18325

[17] A. Apicella, L. Snidaro, Deep neural networks for real-time remote fall detection, in: International Conference on Pattern Recognition, Springer, 2021, pp. 188–201.

[18] Chen, Y., Wei, Y., Pang, D., Xue, G. (2022). A Deep Learning Approach Based on Continuous Wavelet Transform Towards Fall Detection. In: Wang, L., Segal, M., Chen, J., Qiu, T. (eds) Wireless Algorithms, Systems, and Applications. WASA 2022. Lecture Notes in Computer Science, vol 13472. Springer, Cham.