Human Fall Detection in Poor Lighting Conditions Using CNN-Based Model

Md. Sabir Hossain[1][a] and Md. Mahfuzur Rahman[1,2][b]

[1]*Department of Information and Computer Science, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, 31261, Saudi Arabia*

[2]*Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS), King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, 31261, Saudi Arabia*

{*g202314790, mdmahfuzur.rahman*}*@kfupm.edu.sa*

Keywords: Elderly Care, Fall Detection, CNN, Low Lighting, Deep Learning.

Abstract: Human fall detection for elderly care has become a crucial field of research as it can cause serious injuries and impact the quality of life. In this article, we present a deep learning-based approach for human fall detection in low-lighting conditions using a convolutional neural network (CNN). We trained and evaluated our model on multiple datasets, both annotated for fall detection. The proposed architecture captures and analyzes the falls-related features effectively, even in achieving a significant amount of precision, recall, and F1-scores for human fall detection. Moreover, our proposed architecture outperforms (91% accuracy) several state-of-the-art models, including ResNet50, InceptionV3, MobileNet, XceptionNet, VGG16, VGG19, and DenseNet. With a reliable human fall detection architecture, this research significantly contributes to enhancing safety measures for elderly individuals.

1 INTRODUCTION

Elderly people fall is a burning issue as it often leads to serious health injuries, even death. World Health Organization (WHO) reported that deaths among people aged ≥65 are because of fall-related injuries (Ageing and (AAH), 2008). According to the report, approximately 28-35% of people aged ≥65 fall each year, but increasing to 32-42% for those aged ≥70. These phenomena affect individual's health as well as create challenges for healthcare domain. Again, according to World Bank Data, the total amount of aged (over 65) population is 10%. This large amount of aged people all over the world puts an immense need to address fall-related health issues. Frontier health industries can play an important role in human fall detection and leverage opportunities to take timely action. Early fall detection can reduce health injuries in a great context and improve the quality of living for elderly people.

Existing fall detection techniques depend on sensors such as accelerometers and gyroscopes (Chen et al., 2022; Lian et al., 2021; Gomes et al., 2022). These approaches generally ask the users to use different wearables to collect data. Though these techniques achieved some success but lack in many cases. Many users are not comfortable wearing device constantly. Moreover, in dynamic environments, sensor-based technologies perform very poorly. In this research, we aim to develop a vision-based human fall detection model using deep learning methods. Our approach uses neural networks to analyze the visual data to learn the fall-related features. We aim to improve the accuracy of the fall detection model as well as enhance its reliability. Additionally, we also intend to deal with low-light environment challenges while capturing images. Our final objective is to develop an accurate and robust human fall detection system to provide reliable healthcare services.

The rest of the paper is organized as follows. The section 2 discusses the related research on human fall detection. Our proposed CNN-based architecture with overall methodology is described in section 3. After presenting the results and comparisons with other existing state-of-the arts models in section 4, we conclude the paper in section 5. Additionally, future research directions are added in the section 5.

[a] https://orcid.org/0000-0003-4545-6872
[b] https://orcid.org/0000-0002-2871-9119

Hossain, M. S. and Rahman, M. M.
Human Fall Detection in Poor Lighting Conditions Using CNN-Based Model.
DOI: 10.5220/0013502200003938

In *Proceedings of the 11th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2025)*, pages 414-420
ISBN: 978-989-758-743-6; ISSN: 2184-4984

2 LITERATURE REVIEW

Elderly fall detection has drawn significant attention from the research community. It has made an enormous impact on health and quality of life. Researchers explored the use of sensors, wearable devices, and cutting-edge technologies like machine learning, computer vision, etc. to detect human falls. This literature review focuses on the application of deep learning techniques in fall detection among elderly people.

Alam et al. presented a comprehensive review on vision-based human fall detection systems (Alam et al., 2022). They classified existing techniques into various deep learning models, including CNN, LSTM, auto-encoders, MLP, and hybrid techniques. Moreover, they described architectures and evaluation metrics such as accuracy, sensitivity, and specificity. This study also analyzed different benchmark datasets and measured performances of different techniques on them. They also identified different limitations such as lack of real fall data, privacy-preserving issue, and detection in low-lighting or occlusions.

B. Luo proposed elderly fall detection for smart home environments using Yolo networks (Bo, 2023). The author demonstrates less memory usage and superior accuracy (95%) with the Yolov5 network over other networks. The proposed method is sensitive to the camera field of view (FOV) for accurate detection. The author suggested using other sensors with vision sensors to improve the existing accuracy. Additionally, the author suggested the integration of other sensors with vision sensors further enhance the accuracy of fall detection systems. X. Zi et al. also offered a detection technique in poor lighting condition scenario (Zi et al., 2023).

In contrast, X. Kan et al. laid a lightweight approach named CGNS-YOLO for human fall detection by integrating the GSConv module and the GDCN module with YOLOv5 network (Kan et al., 2023). They also reduced the proposed model's size and discarded less pertinent information by incorporating a normalization-based attention module (NAM). They achieved 1.2% enhancement in detection accuracy compared with the conventional YOLOv5s framework. This paper also considered challenging environments like different lighting conditions, and occlusions in their research. However, this research requires validation regarding the different lighting conditions of fall detection.

Gunale et al. presented a novel way of using CNN to detect falls to assist elderly people (Gunale et al., 2023). They combined multiple datasets to achieve generalization and used CNN for automatic feature extraction. They performed both qualitative and quantitative analysis and received 97.93% accuracy. However, the sensitivity value for combined datasets (URFD, MCFD, FDD and SDU) is very low (64.46%) compared to other state-of-the-art architecture, indicating a performance constraint. Their predictive model also suffered from the scarcity of appropriate data to predict correctly. A few recent studies also focused on using deep learning networks in human fall detection (Hoang et al., 2023; Alanazi and Muhammad, 2022). Adrián Núñez-Marcos and Ignacio Arganda-Carreras proposed a video fall detection system using transformer-based model (Núñez-Marcos and Arganda-Carreras, 2024). Their suggested model determines whether or not a fall has occurred based on a video clip. It uses a sliding window style in a video stream to sound an alarm as soon as it detects a fall.

In brief, the above articles discussed the application of YOLO, Transformer, and Convolutional Neural Networks to detect human falls for elderly care in different challenging environment. However, further research is required to address the challenges and improve real world fall detection performance.

3 METHODOLOGY

The proposed methodology for detecting human falls in low lighting conditions consists of several steps; pre-processing, model training, and prediction using our CNN model and other pre-trained models [Figure 1]. Initially, the images are annotated with the bounding box to locate the human subject. Later, the images are resized to a standard dimension to make them consistent throughout the dataset. For model training, we have extracted the illumination-invariant features using our proposed CNN architecture. Then, the trained CNN model predicts based on the test image to classify as a fall or non-fall scenario. Furthermore, we have compared our CNN model's performance with state-of-the-art (SOTA) pre-trained models trained on ImageNet namely ResNet50 (He et al., 2016), InceptionV3 (Szegedy et al., 2016), MobileNet (Howard et al., 2017), XceptionNet (Chollet, 2017), VGG16 (Simonyan and Zisserman, 2015), VGG19 (Simonyan and Zisserman, 2015) and DenseNet (Huang et al., 2017) to ensure optimal detection capacity in low-lighting environments. By designing CNN-based architecture adaptive to different low-lighting conditions, this proposed approach provides a robust and reliable human fall detection and contributes to the safety and independent living of elderly people.

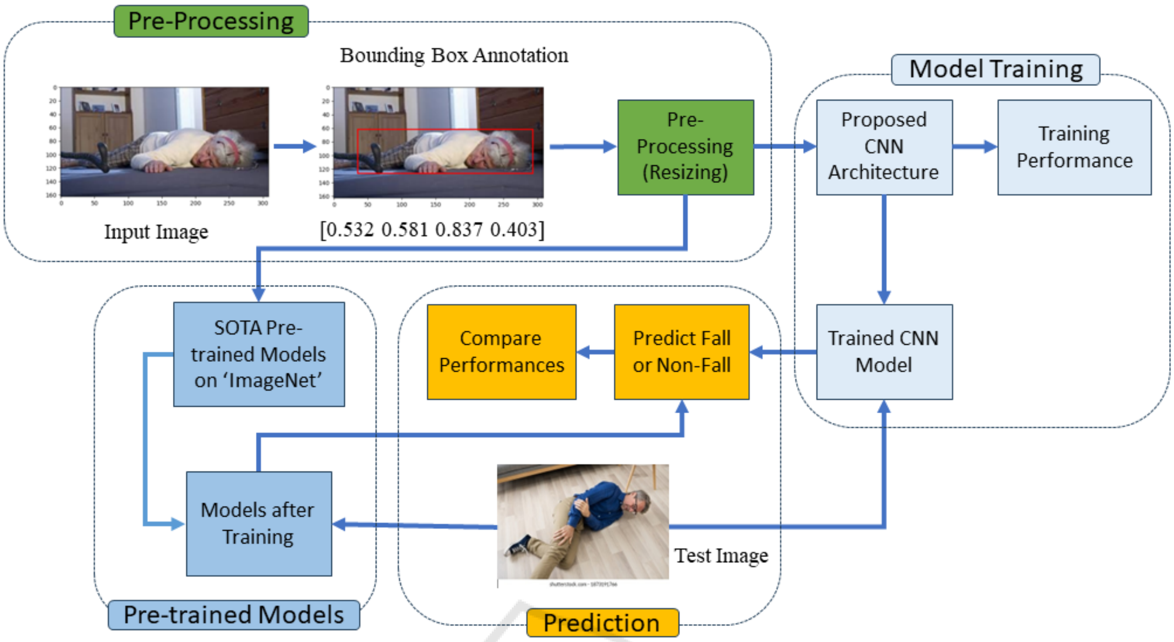

Figure 1: Proposed Methodology for Human Fall Detection.

Figure 2: Image samples from a) Fe2i Dataset and, b) Fall Detection Dataset.

3.1 Data Collection and Annotation

In experimentation, we have used two datasets namely Fe2i Fall Detection Dataset (Xing et al., 2023) and Fall Detection dataset (Kandagatla, 2022). Fe2i dataset is an annotated version of the original video dataset, and contains 2996 sample images, each annotated with relevant information for fall detection. The original Le2i incorporated multiple scenarios like living room, coffee room, office, and classroom. The scenarios were recorded in different lighting conditions. Each image annotation contains class labels and the bounding box coordinates. The annotations contain two classes: fall and upright. The annotations were in XML format. We have converted it into text format (.txt) to prepare for training and testing the CNN model.

On the other hand, the Fall Detection dataset consists of 374 images for training and 111 images for validation. The images were labeled using the Make Sense website (MakeSense.AI, 2024). After uploading the images, a bounding box needs to be drawn. Then, a class label from three (fall, sit, and walk) is assigned. The class labels with bounding box coordinates were exported as text files for each image. The annotations process provides information about the activities as well as enables training and testing processes. Samples from both datasets are shown in Figure 2.

3.2 Proposed CNN Architecture

Our proposed CNN architecture captures and analyzes features related to falls effectively including low lighting conditions. The architecture includes several layers that contribute to detecting fall detection accurately. The proposed CNN architecture is depicted in Figure 3. A brief description of each component and its functionalities provided below:

The input layer takes the images as input of size 150 x 15 pixels with three color channels (Red, Green, and Blue). The previously defined fixed size confirms the compatibility with other layers of the model. In the first convolutional layer, 32 filters with kernel size

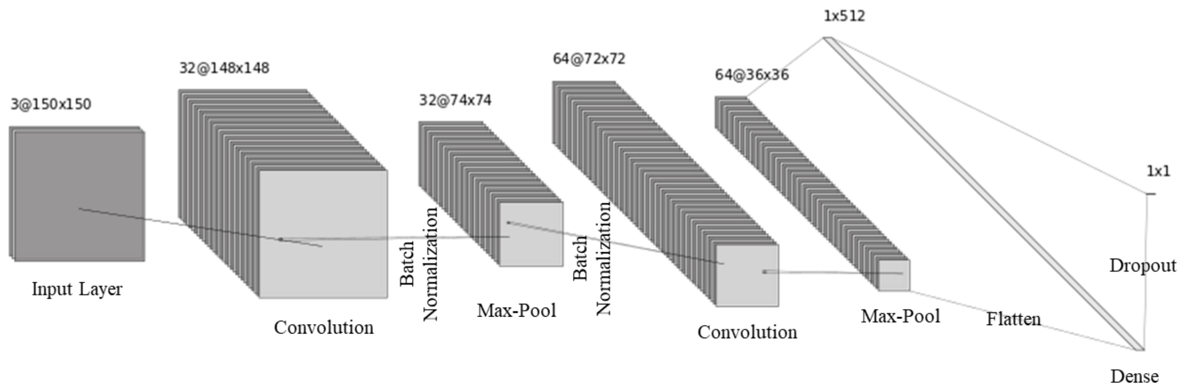

Figure 3: Proposed CNN Architecture to Detect Human Fall in Low Lighting Conditions.

3x3 are applied to the preprocessed images. The target of this component is to identify the crucial features prevalent in images like edges and textures to understand human shapes and movements. We have introduced the ReLU activation function to add non-linearity to the model to learn complex patterns in the data. We used batch normalization after each convolutional layer to normalize the activations. It results in fast and more stable training of the network. This step is also effective for varying characteristic images due to low lighting conditions as it decreases the internal covariate shift.

The first max pooling layers after convolutional layers minimize feature maps by reducing spatial dimensions. However, this step keeps the most significant information. As a result, the computational complexity decreases, and the model becomes more robust to variations in input. Higher-level features are learned in the subsequent convolutional layers from the feature maps generated in immediate layers. In these layers, the more abstract representations of the input images are extracted which are significant for detecting human falls in diverse environments and lighting conditions. At the end of the architecture, dense layers are introduced to learn complex patterns from the flattened feature maps. There are 512 neurons in dense layers and the ReLU activation function is applied to capture complicated relationships in the data. A dropout layer after dense layers prevents overfitting by randomly deactivating a fraction of neurons during training. It improves model generalization to unseen data and extends detection accuracy. The single neuron with a sigmoid activation function in the final layer generates a probability score between 0 and 1. A closer value to 1 indicates the likelihood of detecting a human fall. We have provided layer-wise detailed structure in Table 1 to confirm reproducibility. It will facilitate the researchers to benchmark and further enhance its performance for human fall detection in real-world applications.

In brief, our proposed model is capable of detecting human falls in low-lighting conditions effectively by extracting and analyzing visual features from input images. Several layers including convolutional, batch normalization, activation functions, and dropout regularization are introduced so that the model learns and generalizes to different environments to develop a suitable real-world fall detection technique.

Table 1: Layer-wise details of the proposed CNN model.

Layer	Kernel/Units	Out. Shape	Parameters
Input	150×150×3	150×150×3	-
Conv2D	3×3, 32 filters	148×148×32	896
BN	-	148×148×32	128
ReLU	-	148×148×32	0
MaxPool	2×2	74×74×32	0
Conv2D	3×3, 64 filters	72×72×64	18,496
BN	-	72×72×64	256
ReLU	-	72×72×64	0
MaxPool	2×2	36×36×64	0
Flatten	-	82944	0
Dense	512 neurons	512	42,467,840
BN	-	512	2,048
ReLU	-	512	0
Dropout	0.5	512	0
Output	1 neuron (Sigmoid)	1	513

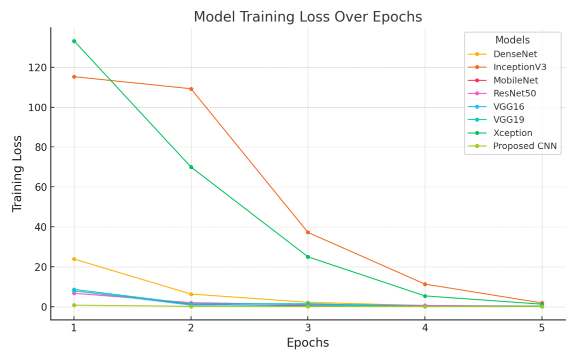

Figure 4: Model loss for each model for different epochs.

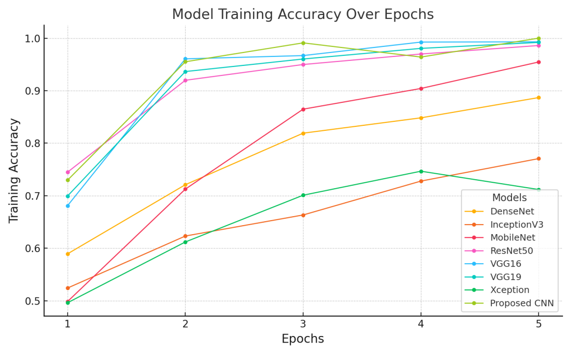

Figure 5: Model accuracy for each model for different epochs.

4 RESULTS AND ANALYSIS

4.1 Training Loss and Accuracy Results

The training loss and accuracy statistics for different models laid an idea about the superiority of our proposed model. For both cases, our proposed model shows faster convergence [Figure 4] and higher accuracy [Figure 5] compared to state-of-the-art models like ResNet50, InceptionV3, MobileNet, Xception, VGG16, VGG19, and DenseNet.

Figure 4 represents that our proposed CNN model efficiently minimizes classification errors. The higher loss values for InceptionV3 (115.27) and XceptionNet (133.25) in the initial epoch indicate that these models did not learn the relevant features for falling detection. Furthermore, ResNet50, VGG16, and DenseNet demonstrated significantly higher loss values compared to our proposed CNN model. Although these models finally reduced their losses over time, they are still higher. On the other hand, our proposed model gradually reached a final loss of 0.0397, indicating its ability to learn fall-related features efficiently, optimize parameter updates effectively, and generalize the training data. Thus, our proposed model can detect human fall under low-light conditions.

The training accuracy depicted in Figure 5 also supports the effectiveness of the proposed CNN model in fall detection. Few models like ResNet50 (0.986), VGG16 (0.993), and VGG19 (0.992) achieved significant amounts accuracy in later epochs but didn't achieve highest accuracy (1.00). However, other models, InceptionV3 and XceptionNet did not reach optimal accuracy levels, with 0.7707 and 0.7118, respectively. In contrast, our proposed model achieved maximum accuracy showing it's effectiveness in human fall detection in challenging situation like low lighting.

4.2 Model Performance

The performance of our proposed deep learning-based model for human fall detection in low lighting conditions is presented in Table 2. We use precision, recall, and F1-score as evaluation metrics to assess the effectiveness of our model.

Our model achieved a precision score of 0.90 for detecting class 0 (fall) and 0.93 for class 1 (non-fall). Precision indicates the percentage of true positive detection over total positively identified instances including true positives and false positives. With the high precision score, false positive detection is minimized. Our proposed model's precision score indicates that it can detect falls and non-falls with minimal errors, making it reliable for practical applications.

Additionally, the recall values of our proposed model are also significant, 0.88 for class 0 and 0.94 for class 1. Recall value is measured by the ratio of the true positive detection out of all actual positive instances (true positive and false negative). The recall value is very crucial to detect the actual falls and non-falls to ensure the safety of elderly people. Our proposed model achieved high recall scores which indicates the effectiveness of our model in detecting the majority of falls, even in challenging low-lighting conditions.

The F1-score presents a balanced measure of the model's performance by calculating the harmonic mean of precision and recall. Our model performed in a balanced manner and achieved F1-scores of 0.89 for class 0 and 0.94 for class 1. The high F1-scores are an indication of the model's overall effectiveness in accurately and reliably detecting human falls. Moreover, our model achieved 0.92 overall accuracy, which indicates that the model correctly identifies falls and non-falls.

In brief, the performance measurements of our proposed deep learning-based model show its effectiveness in human fall detection, even in low-lighting conditions. The high precision, recall, F1-scores, and overall accuracy denote that our model provides a reliable solution for enhancing the safety and well-being of elderly individuals.

Table 2: Precision, recall, and F1-score for each class.

Class	Precision	Recall	F1-Score
0	0.90	0.88	0.89
1	0.93	0.94	0.94
Accuracy	0.92 (111 samples)		
Macro Avg	0.91	0.91	0.91
Weighted Avg	0.92	0.92	0.92

4.3 Comparative Analysis

The Table 3 presents the comparative analysis of various state-of-the-art models used in human fall detection. Our CNN-based proposed model outperforms other notable models listed in the table, achieving an accuracy of 0.9099.

ResNet50 is one of the promising models in computer vision, achieved 0.8834 accuracy. ResNet50 is a popular option for many image processing tasks, specially classification, due to its deep architecture and residual connections. Though it's a robust model, it performed approximately 2.65% less compared to our proposed model accuracy. Another widely used model, InceptionV3, performed the poorest among the compared models with an accuracy of 0.4577. In spite of its efficiency and accuracy in various applications, this poor performance indicates that InceptionV3 might not be well-suited for human fall detection, especially in low-lighting conditions.

MobileNet, another popular model designed for mobile and embedded vision applications, achieved an accuracy of 0.7460. Due to the presence of low lighting conditions, it is compared to other models (ResNet50, VGG16, VGG19, and DenseNet) especially, 16% less than our proposed model. XceptionNet, a model created to improve efficiency by expanding on the Inception module achieved an accuracy of 0.6469. While its unique design is commendable the results suggest that XceptionNet may not excel much in this task when compared to our suggested method. Interestingly, VGG16, VGG19, and DenseNet, these three models achieved an identical accuracy of 0.8434. These models are capable to capture image details. Though these models outperformed than InceptionV3 MobileNet and XceptionNet, they still fall short of our proposed model by, approximately 6.65%.

The results demonstrated that the CNN-based architecture we proposed performed better than all other state-of-the-art models tested. The higher accuracy rate (0.9099) denotes its effectiveness in detecting human falls, in challenging low-light settings. The exceptional performance is achieved due to the customized design and fine-tuning of the CNN network, which probably improves its capability to recognize and understand the characteristics of falls.

Table 3: Comparison with the state-of-the-art models with the proposed model.

Model	Accuracy
ResNet50	0.8834
InceptionV3	0.4577
MobileNet	0.7460
XceptionNet	0.6469
VGG16	0.8434
VGG19	0.8434
DenseNet	0.8434
Proposed Network	0.9099

5 CONCLUSIONS

In this research, we have developed an accurate and reliable CNN-based human fall detection model to enhance elderly care through early and reliable fall detection. By achieving 91% overall accuracy our proposed model outperforms most state-of-the-art (SOTA) models. Furthermore, our proposed CNN-based customized model performs accurately in different challenging situations including low lighting. Our proposed CNN-based model is a notable contribution in the advancement in the area of elderly care. Detecting human falls at the earliest possible time may avoid severe injuries and contribute to the life of elderly individuals, leveraging a safe environment at living places.

Though our proposed model has competing results, there are still a few areas where researchers can contribute in the future. One of them is validating the model in in diverse environments and varying lighting conditions. In future research, we can consider amalgamation of deep learning-based models with other sensors, such as accelerometers and gyroscopes. It will improve the accuracy and reliability of fall detection systems. This diverse method may leverage more comprehensive data which will lead to better detection capabilities. Another future research direction might be real-world deployment and continuous improvement based on user feedback. As fall detection in real-world scenarios is crucial, it can be refined by constant monitoring and providing input from real-world deployments. By contributing to the above-mentioned areas, we can improve the accuracy, reliability, and applicability of the fall detection model. Finally, the enhanced model can be an indispensable tool in elderly care and other safety-critical applications.

ACKNOWLEDGMENT

The authors would like to thank Dr. Saeed Anwar for his comments on the project presentation which improved the quality of the manuscript.

6 FUNDING STATEMENT

This research was fully funded by the Interdisciplinary Research Center for Intelligent Secure Systems (IRC-ISS) through the Deanship of Research, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia. The IRC-ISS research grant #INSS2516 supported the required data and computing resources for this experiment.

7 DATA AND CODE AVAILABILITY

Both datasets used in this study are publicly available on Kaggle, provided in (Xing et al., 2023), (Kandagatla, 2022). The code for this study is available from the authors upon reasonable request.

REFERENCES

Ageing and (AAH), H. (2008). Who global report on falls prevention in older age. https://www.who.int/publications/i/item/9789241563536. Accessed: Mar. 03, 2024.

Alam, E., Sufian, A., Dutta, P., and Leo, M. (2022). Vision-based human fall detection systems using deep learning: A review. *Computers in biology and medicine*, 146:105626.

Alanazi, T. and Muhammad, G. (2022). Human fall detection using 3d multi-stream convolutional neural networks with fusion. *Diagnostics*, 12(12).

Bo, L. (2023). Human fall detection for smart home caring using yolo networks. *International Journal of Advanced Computer Science and Applications*, 14(4).

Chen, Z., Wang, Y., and Yang, W. (2022). Video based fall detection using human poses. In *CCF Conference on Big Data*, pages 283–296. Springer.

Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1251–1258.

Gomes, M. E. N., Macêdo, D., Zanchettin, C., de Mattos-Neto, P. S. G., and Oliveira, A. (2022). Multi-human fall detection and localization in videos. *Computer Vision and Image Understanding*, 220.

Gunale, K. G., Mukherji, P., and Motade, S. N. (2023). Convolutional neural network-based fall detection for the elderly person monitoring. *Journal of Advances in Information Technology*, 14(6):1169–1176.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Hoang, V. H., Lee, J. W., Piran, M. J., and Park, C. S. (2023). Advances in skeleton-based fall detection in rgb videos: From handcrafted to deep learning approaches. *IEEE Access*, 11:92322–92352.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Huang, G., Liu, Z., Maaten, L. V. D., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708.

Kan, X., Zhu, S., Zhang, Y., and Qian, C. (2023). A lightweight human fall detection network. *Sensors*, 23(22).

Kandagatla, U. K. (2022). Fall detection dataset. https://www.kaggle.com/datasets/uttejkumarkandagatla/fall-detection-dataset. Accessed: Dec. 20, 2024.

Lian, J., Yuan, X., Li, M., and Tzeng, N. F. (2021). Fall detection via inaudible acoustic sensing. *Proc ACM Interact Mob Wearable Ubiquitous Technol*, 5(3).

MakeSense.AI (2024). Makesense.ai - free annotation tool for ai and machine learning. https://www.makesense.ai/. Accessed: 2024-12-20.

Núñez-Marcos, A. and Arganda-Carreras, I. (2024). Transformer-based fall detection in videos. *Engineering Applications of Artificial Intelligence*, 132:107937.

Simonyan, K. and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826.

Xing, Z., Chaturvedi, K., Braytee, A., Li, J., and Prasad, M. (2023). Fe2i fall detection dataset (annotated). https://www.kaggle.com/datasets/starxz/fe2i-fall-detection-datasetannotated. Accessed: Dec. 21, 2024.

Zi, X., Chaturvedi, K., Braytee, A., Li, J., and Prasad, M. (2023). Detecting human falls in poor lighting: Object detection and tracking approach for indoor safety. *Electronics (Switzerland)*, 12(5).