Detecting human fall using internet of things devices for healthcare applications

Zakaria Benhaili, Youssef Balouki, Lahcen Moumoun
Laboratory of Mathematics, Computer Science and Engineering Sciences, Faculty of Sciences and Techniques,
Hassan First University of Settat, Settat, Morocco

Article Info

Article history:

Received Dec 7, 2023
Revised Jul 4, 2024
Accepted Jul 26, 2024

Keywords:

Activity recognition
Deep learning
Fall detection
Internet of things
Sensors

ABSTRACT

Falls pose a significant threat to unintentional injuries, particularly impacting the independence of older individuals. Existing detection methods suffer from drawbacks, including inaccuracies, wearer discomfort, complex setup, resource-intensive computation, and limitations in detecting falls outside a specific setting. In response, our innovative fall detection system integrates with a pneumatic solution, analyzing fundamental human activities like running, walking, and sitting, both indoors and outdoors. This approach combines wearable sensors with a vision-based solution, utilizing a smart belt with embedded accelerometer and gyroscope, alongside wall-installed cameras in a smart house. The system triggers an airbag and sends an emergency alarm upon fall detection. To achieve this, we propose FallMixer a lightweight deep learning model, combined with 'you only look once' version 8 (YOLOv8) algorithm, fine-tuned on a collected video dataset to enable real-time detection. We found that the models result in competitive performance, as demonstrated on SisFall, UCI human activity recognition (HAR), and mobile health (MHEALTH) datasets with a remarkable mean average precision. Subsequently, we assess the hardware performance of our solution on edge devices.

Corresponding Author:

Zakaria Benhaili
Laboratory of Mathematics, Computer Science and Engineering Sciences
Faculty of Sciences and Techniques, Hassan First University of Settat
Settat 26000, Morocco
Email: z.benhaili@uhp.ac.ma

1. INTRODUCTION

Human activity recognition (HAR) has emerged as a crucial area of research within ambient and context-aware computing [1], [2]. It has a substantial impact in different fields, such as intelligent surveillance systems [3], healthcare [4], [5], human-computer interaction [6], and eldercare assistance [7]. While HAR has broad applications, one crucial area where it plays a vital role is in fall detection for elderly persons. The detection of falls holds paramount importance, particularly for those aged 65 and above, as it represents a significant public health concern and stands as the leading cause of injury-related fatalities within this demographic. The implementation of real-time fall detection devices holds immense value, offering a means to promptly summon assistance and thereby preventing injuries. Additionally, such devices contribute to an enhanced quality of life for seniors and provide caregivers with greater help. HAR frameworks enable the detection, classification, and understanding of specific body movement or activity through data collected from diverse sensors. In literature researchers detect the activities through two main methods: vision-based, which analyzes video or image data captured by cameras, and wearable sensor-based, which leverages data from non-intrusive, embedded sensors in smart devices like smartphones and smartwatches.

However, building models that generalize well across various activities and sensors remains a challenging task. Human activity signals and images can significantly vary among individuals and even for the same individual performing an activity at separate times. Additionally, different activities may exhibit similar signal patterns, further complicating activity classification. Traditionally, researchers have used handcrafted feature extraction methods combined with supervised machine learning techniques like k-nearest neighbors (KNN) [8], support vector machines (SVM) [9], decision trees (DT) [10], and ensemble approaches for classification [11]. Nevertheless, these approaches require domain expertise, rigorous data pre-processing, and might lack flexibility due to the difficulty of establishing spatial and temporal relationships among handcrafted features. Recent deep learning techniques have gained popularity, especially in areas like natural language processing [12], image recognition and classification [13]. Precisely multilayer perceptron (MLP) mixers [14], a novel architecture that has attracted considerable interest in natural language processing and computer vision. In contrast to the popular transformer [15], that relies on self-attention mechanisms, MLP-mixers use MLPs for processing input data. This alternative approach provides several advantages, such as simplicity and computational efficiency, leading to faster training times and reduced memory consumption, which is suitable for internet of things (IoT) and edge devices. Given these promising characteristics, we investigate the deployment of this class of network for time series data, and their suitability for tasks involving long-range dependencies.

We present a system design aimed at ensuring the safety of elderly individuals within their homes and minimizing the risk of injury in the event of a fall outdoors. To detect movements, we employ two types of sensors: accelerometers and gyroscopes. These sensors help interpret body motions and identify potential falls. In addition, we use cameras strategically installed on the walls. To accomplish this objective, we introduce FallMixer, a deep neural network designed to exploit sensory data and extract temporal features from it. For real-time fall detection, we utilized YOLOv8, a state-of-the-art object detection system. Where for training and testing we employed three datasets from wearable sensors, alongside a collected video dataset featuring instances of falls. To implement this solution, we propose embedding the system into a smart belt or smart jacket. This wearable device will help to detect pre-fall actions using the capabilities of FallMixer. In the event of a potential fall, the system can trigger the protection mechanism, such as deploying an airbag, to safeguard sensitive bones and minimize the risk of injury. The integration of visual detection further contributes to reducing false positives and negatives, ensuring the arrival of the alert to the medical personnel.

In this research, our contribution lies in presenting a resilient approach for fall detection for elderly people within a household, leveraging deep learning algorithms. In contrast to prior studies that concentrate on single-modality detection, our novel approach adopts a dual-sensor strategy to augment detection capabilities. Our proposed algorithm exhibits the ability to identify a diverse range of actions, surpassing the performance of existing methodologies, particularly in fall detection. Additionally, we explore the practicality of implementing our approach on edge devices, highlighting its potential for real-world applications.

The remainder of the paper unfolds in a structured manner. Section 2 provides an overview of methodologies proposed by other researchers in the field. The specifics of our proposed method are outlined in section 3. Section 4 presents the experimental setup, detailing the configuration and parameters used for evaluation. Results and discussions are comprehensively covered in section 5, and last, we conclude the paper.

2. RELATED WORKS

Several noteworthy studies have contributed to advancing the accuracy and efficiency of detecting falls and related activities. Sengül *et al.* [16] explored deep learning-based fall detection using smartwatches, with a primary emphasis on healthcare applications. Their method involved collecting gyroscope and accelerometer using a smartwatch, augmented by an interpolation technique. This approach highlighted efficacy in accurately identifying instances of falling. This augmentation technique has advanced the accuracy of bidirectional long short-term memory (Bi-LSTM) algorithm to reach 99%, collected data have submitted the creation of 38 features. In another work [17], a fall detection system, using optimized convolutional neural networks (CNN) and wearable IoT sensor data, was developed. Data from 14 individuals and six sensors were processed to extract distinctive features, which were then reduced using multilinear principal component analysis. An 8-layer, AlexNet-based CNN significantly advanced fall detection. Another approach, presented by Lee *et al.* [18], uses foot plantar pressure and acceleration data. Their innovative system effectively distinguished between various activities of daily living (ADL tasks) and diverse types of falls. By utilizing DTs and a threshold technique, the method achieved a 95% accuracy rate in recognizing fall activities at an average speed of 317 milliseconds. Luo *et al.* [19] introduced a binary convolutional network tailored for real-time HAR. Their focus on improving fall detection using wearable sensor data, particularly for mobile devices, led them to incorporate dilated convolutions to effectively capture features from time series data.

Wearables, being battery-operated with limited noninvasive applicability, have led other researchers to adopt a visual approach for fall detection. They investigate human posture estimation using various types of cameras. Interestingly, most of these studies use various types of CNN or YOLO algorithms for real-time detection. Hasan *et al.* [20] use a two-layer long short-term memory (LSTM) on video data, utilizing the OpenPose algorithm for 2D body pose estimation, and employing sequential frames for fall identification. Feng *et al.* [21] addressed challenges in complex environments for pedestrian detection, by integrating YOLOv3, Deep-Sort for object tracking, and VGG-16 with an attention-LSTM network for fall detection. This showcased adaptability in intricate scenarios. Fei *et al.* [22] approached fall detection by using the Flow-position Net model, combining optical flow and human pose data. Notably, their system demonstrated good accuracy on two open datasets URFD and Le2i, highlighting it is potential for reliable fall detection. Research by Kan *et al.* [23], the YOLO network undergoes modifications and enhancements in conjunction with other modules to attain a commendable mean average precision (mAP) and guides the development of a lightweight fall detection network.

The preceding methodologies have primarily concentrated on fall detection through various sensor modalities but have not prioritized the anticipation of a fall event, initiating a warning signal in advance of a fall occurrence holds the potential for lifesaving interventions. Furthermore, a refined approach entails distinguishing between indoor and outdoor fall detection, enhancing the system's resilience by considering the distinctive environmental attributes of each setting. Considering these considerations, we put forth our approach, which centers on the detection of falls through the integrated use of both wearable sensors and surveillance cameras.

3. METHOD

In this section, we present our approach for activity detection, with a specific focus on identifying falls. Leveraging wearable sensors seamlessly integrated into smartphones, our system promptly recognizes fall patterns, triggering an alert to initiate a protective mechanism. This mechanism takes the form of a belt airbag designed to safeguard vulnerable bones. Introducing FallMixer, a neural network, we evaluate its performance through implementation on a smartphone. The analysis could also be done on cloud servers to further enhance detection capabilities. We incorporate surveillance cameras tailored for smart home scenarios. Then employing YOLO [24], we train the model using a collected dataset, enabling it to effectively track and detect falls with precision. This comprehensive system not only leverages wearable technology but also integrates sophisticated camera-based surveillance, ensuring a robust and reliable approach to fall detection and protection. Figure 1 presents the details of the solution.

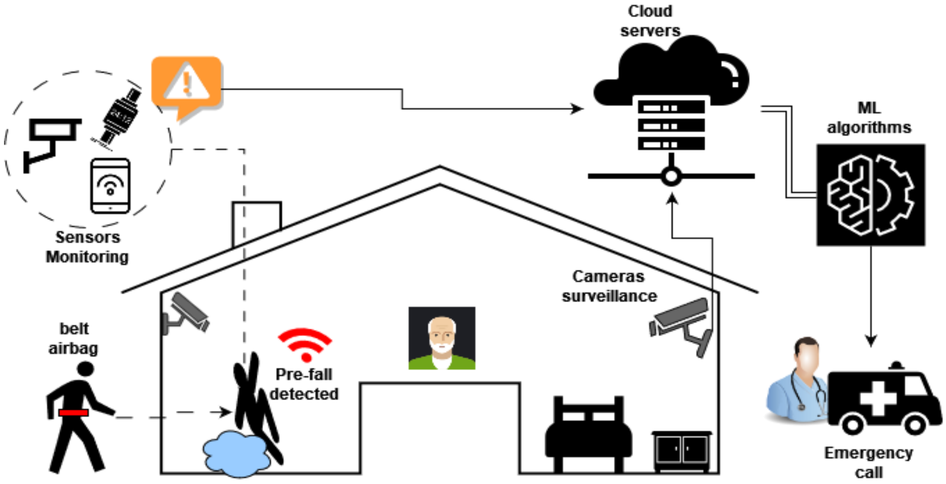

Figure 1. Fall detection system in smart home using camera and wearable sensors

3.1. FallMixer architecture

The proposed architecture presented in Figure 2 is inspired from the paper [25] we named it "FallMixer" it is designed to address the challenge of efficiently processing time series data, using lightweight operations. The model utilizes a combination of convolutional and mixer layers to extract relevant features and facilitate information flow. Architecture's input: it is a one-dimensional time series data, which is reshaped to make it compatible with convolutional layers. Like the original architecture we also used the patch embedding layer, specifically we applied a 2D convolution operation with a patch size equal to the kernel size, to capture local information from the input sequence. The output is then passed through a Gaussian error linear unit (GELU) activation function followed by batch normalization to introduce non-linearity and stabilize training. The core of the architecture is the FallMixer layer composed of a series of depthwise convolutional layers. Each FallMixer consists of three main components.

Convolution layer that uses a depthwise convolution with a kernel size of two to identify spatial relationships between neighboring patches and to capture dependencies within local patches of the input sequence. Skip connection that keeps data from earlier layers in the model. This promotes information flow and reduces the vanishing gradient, which facilitates the model's ability to pick up pertinent representations. Pointwise convolution that takes the output of the skip connection and passes it through a 1×1 convolutional layer to aggregate information and increases the model's capacity to capture more complex patterns in the data. A global average pooling layer and a dense layer are used in the architecture's output.

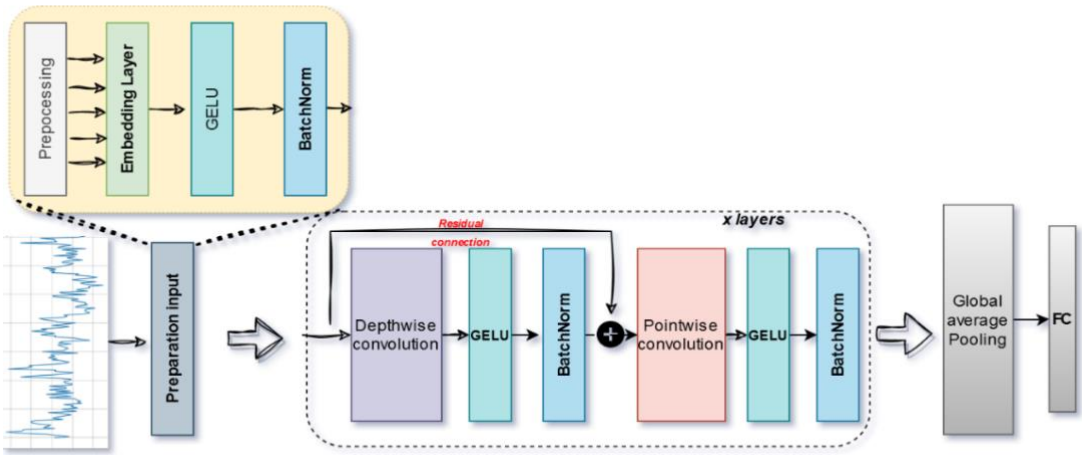

Figure 2. The architecture of FallMixer used for activity recognition

3.2. Real time fall detection algorithm

In this study, we utilize the YOLOv8 network, the latest iteration in the YOLO family of algorithms. The YOLOv8 model comprises four main network components: the neck-end for feature fusion, the backbone network for extracting image features, the input for data enhancement, and the decoupled header output, which separates the classification from the detection header. With numerous improvements built on the YOLOv5 framework, YOLOv8 surpasses YOLOv5 in terms of speed and accuracy. Additionally, it provides a unified framework for training models that manage tasks such as image classification, object recognition, and instance segmentation.

4. EXPERIMENT SETUP

This section offers a comprehensive overview of the datasets and configurations utilized in the experiments conducted in this study. To evaluate our models across various settings, we selected four datasets, three publicly available and one collected by our team. Each dataset is distinguished by unique traits and environments. We provide detailed insights into the dataset partitioning training configurations employed in our study, aiming for reproducibility across all analyses.

4.1. Datasets description

UCI HAR [26] dataset was gathered using an iPhone 6s placed in the front pocket of the participants. A total of 24 participants, including 10 women and 14 men, with diverse ages, weights, and heights, engaged in six activities across 15 trials under consistent environmental conditions. The activities performed were walking, jogging, sitting, standing, stairs down, and stairs up. Each accelerometer and gyroscope axis in the dataset contains 1,304,950 samples. The data was collected at a sampling rate of 50 Hz, resulting in a cumulative recording time of 435 minutes (about 7 and a half hours). Mobile health (MHEALTH) [27] dataset 10 participants perform static and dynamic activities, captured using various sensors, including three accelerometer sensors, an electrocardiogram sensor, two gyroscope sensors, and three magnetometer sensors. The position of all these sensors is on the left ankle, the chest, and the right wrist. MHEALTH encompasses twelve varied actions such as sitting, cycling, and climbing stairs.

SisFall [28] selected for testing our model, the SisFall dataset sets itself apart from other publicly available datasets. It involves data from fifteen healthy, independent elderly individuals and outshines competitors with its inclusion of more participants, diverse activity types, and a larger number of recordings. Comprising 2,706 activities of daily living and 1,798 falls, this dataset addresses the scarcity of datasets realistically simulating activities and falls by elderly individuals in the domain of elderly fall detection research. Specifically, it introduces up to thirty-four activities (falls and activities of daily living) performed by thirty-eight participants equipped with a wearable device fixed to their waist. The information for those three datasets is organized in Table 1 to facilitate a straightforward comparison.

Video dataset in our visual approach, we collected a dataset sourced from YouTube and Kaggle, comprising various scenarios depicting individuals experiencing falls. It is important to highlight that this dataset is limited. To mitigate this limitation, we applied data augmentation techniques to expand the dataset's diversity. The dataset was categorized into two distinct classes: fall and not fall.

Table 1. Datasets recording details

Variable	Sensors	Activities	Frequency (Hz)	Samples
UCI HAR	2	6	50	10,299
MHEALTH	3	12	50	120
SisFall	2	2	200	4,505 used

4.2. Data partitioning

We approached dataset partitioning uniformly. Only the UCI dataset had a pre-defined split into training and test sets, and we maintained consistent test partitions for comparative analysis. Our preprocessing methodology exclusively involved the use of raw signals post-down sampling, distinguishing our approach from the rest of studies on UCI HAR. These studies often either applied their own preprocessing techniques to raw data or utilized provided hand-crafted features. Hand-crafted features were employed for the SisFall dataset, and we adopted the same setup for the MHEALTH dataset. Multiple data transformations were applied to enhance the quality of the image dataset. Across all datasets, we adhered to a standardized split of 30% for testing and 70% for training.

5. RESULTS AND DISCUSSION

5.1. Evaluation measures

We provide a concise overview of the evaluation criteria employed in this study. Precision: is presented as the proportion of correctly predicted positive observations relative to the total predicted positive observations. It is calculated using the terms true positive (TP) and false positive (FP).

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

Recall (sensitivity): recall, also referred to as sensitivity, is the ratio of correctly identified positive observations to the total number of observations in the corresponding actual class.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

F1 score: it serves as a harmonic mean of both 'precision' and 'recall.' Consequently, it incorporates information about false positives and false negatives to yield a more comprehensive assessment. While not as straightforward as accuracy, F1 is often more informative, especially in scenarios where there is an imbalance in class distribution.

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

We also employ mAP, a metric that offers a more comprehensive assessment than accuracy alone for evaluating a model's precision and recall trade-off in object detection tasks.

5.2. Classification results

Training configuration in our learning environment, we set up the training process for the FallMixer model, which is tailored for a classification task. We trained the model using the Adam optimizer, setting the learning rate to 0.001, and employed the categorical cross-entropy loss function. To mitigate overfitting, we incorporated early stopping with a patience of 500 epochs. Additionally, a callback was implemented to save the optimal model based on validation accuracy. The training process extends up to a maximum of 1,000 epochs, utilizing a batch size of 512. This setup ensures the model's efficient learning and helps us achieve accurate classification results, making it suitable for handling sequential data. During the experiments, we thoroughly evaluated the FallMixer model with varying configurations, exploring different numbers of filters and depths while keeping the kernel size and patch size constant. This systematic analysis enabled us to identify the optimal settings for the number of filters and depth, maximizing the model's feature capturing capabilities and achieving accurate classifications. By maintaining constant kernel and patch size, we ensured a fair comparison and focused on understanding the impact of other hyperparameters on performance. In subjecting our work to a comparative analysis with various approaches, all adhering to the same data splitting configuration, a notable observation becomes known, FallMixer emerges as particularly noteworthy for attaining superior accuracy. Table 2 highlights the performance of different architectures, and amidst them. The compared methodologies in the table employed methods like fusion of LSTM and CNN architectures or ensemble methods. However, consistently, FallMixer exhibits enhanced performance in terms of accuracy. This outcome not only underscores the efficacy of our approach. This significance is further exemplified in the context of the MHEALTH dataset, as illustrated in Table 3. Here, the utilization of only an accelerometer with FallMixer demonstrates commendable activity detection capabilities, and underscoring the versatility and effectiveness of FallMixer, particularly in scenarios involving diverse combinations of wearable sensors.

Table 2. Previous works on UCI HAR

Architecture	Accuracy
MobileHART [29]	97.67%
FallMixer	97.56%
ViT [29]	93.66%
CNN LSTM [30]	92.79%±0.34

Table 3. Previous works on MHEALTH dataset

Architecture	FallMixer	EkVN [31]	MEMM [32]
Accuracy	97.71%	67%	90.91%

Fall detection results: Following the comprehensive testing of FallMixer across various activities, our focus shifted to evaluating its performance using the SisFall dataset with a specific emphasis on the detection of a singular activity, namely falls in daily living. The results show that it can detect the fall with an accuracy close to 99%. The confusion matrix is presented in Table 4.

Table 4. Confusion matrix of SisFall dataset

	Predicted fall	Predicted ADL
True fall	556	0
True ADL	0	345

5.3. Hardware performance

We chose the Tensorflow Lite model maker library for its user-friendly interface in mobile model development, seamless integration with TensorFlow templates, and the ability to create efficient, lightweight TFLite models suitable for mobile deployment. We moved our experimental models to an iPhone 11 Pro Max with a Hexa-core CPU, 4 GB RAM, and an Apple A13 Bionic Chipset (7 nm+) with a 4-core graphics Apple GPU. Over 1,000 inferences were performed using four CPU threads, the inferences were exclusively conducted on the device's CPU, and multiple measures were taken during the process.

Analysis: the outcomes of the experiments are detailed in Table 5. It highlights a substantial decrease in the average inference time for FallMixer in comparison to ViT and HART, suggesting its superior efficiency in making predictions on the device. While our Model exhibits a higher memory footprint than HART but a lower one than ViT, it features the smallest model size, indicating a reduced need for storage space. This proves advantageous, particularly for devices with limited storage capacity. However, the optimal architecture choice depends on specific application requirements, considering factors like available memory and the desired balance between speed and model complexity.

Table 5. Inferential duration and memory usage throughout 1000 inferences on the device

Network	Average inference time (ms)	Memory (MB)	Model size (MB)
ViT [29]	8.213±1.518	32.07	15.22
HART [29]	5.376±1.104	12.74	5.9
Our model	2.46 (mean)	25.5	3

5.4. Fall detection using cameras

By utilizing YOLO's detection capabilities and fine-tuning it on our dataset, the obtained precision is equal to 97.2%, the recall is 82.6% and the mAP is equal to 89.6%. Human falls are shown in Figure 3. In the image, we observe successful fall detection, with the bounding box effectively encompassing the individual. This outcome is encouraging, particularly given the use of a modest dataset, and a limited training time.

Figure 2. Indoor and outdoor fall detection using surveillance cameras and YOLO algorithm

5.5. Discussion

This study introduces a novel approach to fall detection by integrating wearable sensors with accelerometers and gyroscopes, along with visual sensors utilizing cameras. By combining these two modalities, the proposed algorithms achieved remarkable accuracy rates on various datasets, surpassing previous methods. Notably, FallMixer attained 97.56% accuracy on the UCI HAR dataset, 97.71% on MHEALTH, and 99% on Sisfall dataset, while Yolov8 demonstrated 97.2% precision on a custom fall video dataset. This advancement outperforms traditional techniques like LSTM and convolution, as well as transformer-based algorithms, showcasing the potential for attention-free architectures in extracting temporal properties from sensor data. Despite these promising results, further research is essential to optimize the integration of visual and wearable sensor data, address implementation challenges such as incorporating airbag protection for bones and explore deployment options in both device and cloud environments. These findings suggest exciting prospects for enhancing detection accuracy and reliability using hybrid sensing, particularly in smart homes and daily living activity monitoring for older adults.

6. CONCLUSION

This investigation has focused on the development of a fall detection system tailored to the specific needs of senior citizens, capable of being deployed in both indoor and outdoor settings. The proposed network can detect movements, particularly falls, by utilizing sensors built into smartphones or other IoT devices. The use of a video-based approach enhanced the detection of the fall, while also providing the ability to monitor and analyze various other movements, promoting a more active lifestyle. The proposed FallMixer model showcases a significant level of reliability and accuracy in recognizing basic movements like walking and sitting. It achieves an accuracy of 97.71% on MHEALTH and 97.67% on UCI HAR. Moreover, it reaches a 99% accuracy in detecting falls in the Sisfall dataset, utilizing information from accelerometer and gyroscope sensors. The incorporation of YOLOv8 finetuned aids in achieving a precision of 97.2% for detecting two classes, namely Fall and NoFall, on a collected video dataset. In the future, we are thinking about improving the system by adding a belt with an airbag that is intended to shield important bones in the event of a fall.

Real-time detection could be performed using a Raspberry Pi with a camera, and the sensors could be integrated into a smartwatch. This allows the suggested system to be reliably deployed in a smart home context, reducing worries among older residents while also improving their safety and well-being.

REFERENCES

[1] Y. Asim, M. A. Azam, M. E. -Ul-Haq, U. Naeem, and A. Khalid, "Context-aware human activity recognition (CAHAR) in-the-wild using smartphone accelerometer," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 4361–4371, 2020, doi: 10.1109/JSEN.2020.2964278.

[2] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, "GCHAR: An efficient group-based context—aware human activity recognition on smartphone," *Journal of Parallel and Distributed Computing*, vol. 118, pp. 67–80, 2018, doi: doi: 10.1016/j.jpdc.2017.05.007.

[3] A. Hussain, S. U. Khan, N. Khan, M. Shabaz, and S. W. Baik, "AI-driven behavior biometrics framework for robust human activity recognition in surveillance systems," *Engineering Applications of Artificial Intelligence*, vol. 127, 2024, doi: doi: 10.1016/j.engappai.2023.107218.

[4] A. Subasi, K. Khateeb, T. Brahimi, and A. Sarirete, "Chapter 5 - Human activity recognition using machine learning methods in a smart healthcare environment," in *Innovation in Health Informatics*, Academic Press, 2020, pp. 123–144, doi: 10.1016/B978-0-12-819043-2.00005-8.

[5] R. Liu, A. A. Ramli, H. Zhang, E. Henricson, and X. Liu, "An overview of human activity recognition using wearable sensors: healthcare and artificial intelligence," in *Internet of Things–ICIOT 2021*, Cham: Springer International Publishing, 2022, pp. 1–14, doi: 10.1007/978-3-030-96068-1_1.

[6] R. Zhang, C. Jiang, S. Wu, Q. Zhou, X. Jing, and J. Mu, "Wi-Fi sensing for joint gesture recognition and human identification from few samples in human-computer interaction," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 7, pp. 2193–2205, 2022, doi: 10.1109/JSAC.2022.3155526.

[7] L. Schrader *et al.*, "Advanced sensing and human activity recognition in early intervention and rehabilitation of elderly people," *Journal of Population Ageing*, vol. 13, no. 2, pp. 139–165, Jun. 2020, doi: 10.1007/s12062-020-09260-z.

[8] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE (OTM 2003)*, Catania, Italy, 2003, pp. 986–996, doi: 10.1007/978-3-540-39964-3_62.

[9] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998, doi: 10.1023/A:1009715923555.

[10] R. A. Berk, "Classification and regression trees (CART)," in *Statistical Learning from a Regression Perspective*, New York: Springer, 2008, pp. 1–65, doi: 10.1007/978-0-387-77501-2_3.

[11] L. Rokach, "Ensemble methods for classifiers," in *Data Mining and Knowledge Discovery Handbook*, Boston, MA: Springer US, 2005, pp. 957–980, doi: 10.1007/0-387-25465-X_45.

[12] I. Lauriola, A. Lavelli, and F. Aiolli, "An introduction to deep learning in natural language processing: Models, techniques, and tools," *Neurocomputing*, vol. 470, pp. 443–456, 2022, doi: 10.1016/j.neucom.2021.05.103.

[13] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv-Computer Science*, pp. 1-17, 2020.

[14] I. O. Tolstikhin *et al.*, "MLP-mixer: An all-mlp architecture for vision," *Advances in neural information processing systems*, vol. 34, pp. 24261–24272, 2021.

[15] A. Vaswani *et al.*, "Attention is All you Need," in *Advances in Neural Information Processing Systems*, 2017.

[16] G. Şengül, M. Karakaya, S. Misra, O. O. Abayomi-Alli, and R. Damaševičius, "Deep learning based fall detection using smartwatches for healthcare applications," *Biomedical Signal Processing and Control*, vol. 71, 2022, doi: 10.1016/j.bspc.2021.103242.

[17] A. Alarifi and A. Alwadain, "Killer heuristic optimized convolution neural network-based fall detection with wearable IoT sensor devices," *Measurement*, vol. 167, 2021, doi: 10.1016/j.measurement.2020.108258.

[18] C. M. Lee, J. Park, S. Park, and C. H. Kim, "Fall-detection algorithm using plantar pressure and acceleration data," *International Journal of Precision Engineering and Manufacturing*, vol. 21, no. 4, pp. 725–737, Apr. 2020, doi: 10.1007/s12541-019-00268-w.

[19] F. Luo, S. Khan, Y. Huang, and K. Wu, "Binarized neural network for edge intelligence of sensor-based human activity recognition," *IEEE Transactions on Mobile Computing*, vol. 22, no. 3, pp. 1356–1368, 2023, doi: 10.1109/TMC.2021.3109940.

[20] M. M. Hasan, M. S. Islam, and S. Abdullah, "Robust pose-based human fall detection using recurrent neural network," in *2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON)*, 2019, pp. 48–51, doi: 10.1109/RAAICON48939.2019.23.

[21] Q. Feng, C. Gao, L. Wang, Y. Zhao, T. Song, and Q. Li, "Spatio-temporal fall event detection in complex scenes using attention guided LSTM," *Pattern Recognition Letters*, vol. 130, pp. 242–249, 2020, doi: 10.1016/j.patrec.2018.08.031.

[22] K. Fei, C. Wang, J. Zhang, Y. Liu, X. Xie, and Z. Tu, "Flow-pose Net: an effective two-stream network for fall detection," *The Visual Computer*, vol. 39, no. 6, pp. 2305–2320, Feb. 2022, doi: 10.1007/s00371-022-02416-2.

[23] X. Kan, S. Zhu, Y. Zhang, and C. Qian, "A lightweight human fall detection network," *Sensors*, vol. 23, no. 22, 2023, doi: 10.3390/s23229069.

[24] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023, doi: 10.3390/make5040083.

[25] A. Trockman and J. Z. Kolter, "Patches are all you need?," *arXiv-Computer Science*, pp. 1-16, 2022, doi: 10.48550/arXiv.2201.09792.

[26] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruge*, Bruges, Belgium, pp. 437-442, Apr. 2013.

[27] O. Banos *et al.*, "mHealthDroid: A novel framework for agile development of mobile health applications," in *Ambient Assisted Living and Daily Activities*, Cham: Springer International Publishing, 2014, pp. 91–98, doi: 10.1007/978-3-319-13105-4_14.

[28] A. Sucerquia, J. D. López, and J. F. Vargas-Bonilla, "SisFall: A fall and movement dataset," *Sensors*, vol. 17, no. 1, 2017, doi: 10.3390/s17010198.

[29] E. K. Sannara, F. Portet, and P. Lalanda, "Lightweight transformers for human activity recognition on mobile devices," *arXiv-Computer Science*, pp. 1-20, 2022, doi: 10.48550/arXiv.2209.11750.
[30] R. Mutegeki and D. S. Han, "A CNN-LSTM approach to human activity recognition," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 362–366, doi: 10.1109/ICAIIC48513.2020.9065078.
[31] K. D. Garcia *et al.*, "An ensemble of autonomous auto-encoders for human activity recognition," *Neurocomputing*, vol. 439, pp. 271–280, 2021, doi: 10.1016/j.neucom.2020.01.125.
[32] Y.-J. Kim, B.-N. Kang, and D. Kim, "Hidden Markov model ensemble for activity recognition using tri-axis accelerometer," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 2015, pp. 3036–3041, doi: 10.1109/SMC.2015.528.

BIOGRAPHIES OF AUTHORS

Zakaria Benhaili received the engineering degree from the National School of Applied Sciences at Sultan Moulay Slimane University, Khouribga, Morocco, currently a Ph.D. student in Faculty of Sciences and Technologies at Hassan 1st University. His main research area includes internet of things, deep learning, and it is applications in human activity recognition, pattern recognition, and smart homes. He can be contacted at email: z.benhaili@uhp.ac.ma.

Youssef Balouki Laboratory of Mathematics, Computer and Engineering Sciences, Faculty of Sciences and Techniques, Hassan I University, Settat, Morocco. He is currently a professor of computer science with the Faculty of Science and Technologies, Hassan I University, Settat, Morocco. He has conducted many Ph.D. theses and has written fifty scientific articles in the domain of artificial intelligence, software engineering, model-driven development, data mining, and formal methods. He can be contacted at email: balouki.youssef@gmail.com.

Lahcen Moumoun received the advanced graduate diploma in structural calculation from the National School of Electricity and Mechanics of Casablanca from Hassan Second University, Casablanca, Morocco, in 1995, and the Ph.D. degree in image processing from in Faculty of Sciences and Technologies at Hassan 1st University, Morocco, in 2011. He has been a professor of computer science engineering with Hassan 1st University, since 2014. He has authored or co-authored more than 100 refereed journal and conference papers, 10 book chapters, and three edited books with Elsevier and Springer. His research interests include pattern recognition, computer vision, big data, and deep learning. He can be contacted at email: lahcen.moumoun@uhp.ac.ma.