Medicine in Novel Technology and Devices 20 (2023) 100272

Contents lists available at ScienceDirect

Medicine in Novel Technology and Devices

journal homepage: www.journals.elsevier.com/medicine-in-novel-technology-and-devices/

# A computer vision-approach for activity recognition and residential monitoring of elderly people

Sudhir Gaikwad [a,*], Shripad Bhatlawande [a], Swati Shilaskar [a], Anjali Solanke [b]

[a] E&TC Department, Vishwakarma Institute of Technology, Pune, Maharashtra, India
[b] E&TC Department, Marathwada Mitramandal's College of Engineering, Pune, Maharashtra, India

ARTICLE INFO

*Keywords:*
Assisted living systems
Human activity recognition
Fall detection
Key joint angle
Landmarks

ABSTRACT

In this study, we explore a human activity recognition (HAR) system using computer vision for assisted living systems (ALS). Most existing HAR systems are implemented using wired or wireless sensor networks. These systems have limitations such as cost, power issues, weight, and the inability of the elderly to wear and carry them comfortably. These issues could be overcome by a computer vision based HAR system. But such systems require a highly memory-consuming image dataset. Training such a dataset takes a long time. The proposed computer-vision-based system overcomes the shortcomings of existing systems. The authors have used key-joint angles, distances between the key joints, and slopes between the key joints to create a numerical dataset instead of an image dataset. All these parameters in the dataset are recorded via real-time event simulation. The data set has 780,000 calculated feature values from 20,000 images. This dataset is used to train and detect five different human postures. These are sitting, standing, walking, lying, and falling. The implementation encompasses four distinct algorithms: the decision tree (DT), random forest (RF), support vector machine (SVM), and an ensemble approach. Remarkably, the ensemble technique exhibited exceptional performance metrics with 99 % accuracy, 98 % precision, 97 % recall, and an F1 score of 99 %.

## 1. Introduction

Referring to both a United Nations report and the Indian census of 2011, it becomes evident that a significant demographic transition is taking place across the globe. This phenomenon affects the global population, including India, and is marked by a notable increase in the elderly population [1–3]. Improved medical facilities, increased public health awareness, and a lower birth rate are some of the factors contributing to this trend. However, this demographic transition presents new challenges, such as loneliness, inadequate healthcare, and social isolation, that impact the elderly population [4,5].

Addressing the needs of this ageing demographic poses significant challenges to societies and integrated health systems worldwide, given the woefully inadequate infrastructure to support the growing number of individuals facing these issues. As a response to these challenges, the field of study underscores the importance of Assisted Living Systems (ALS). Among various ALS approaches, Human Activity Recognition (HAR) emerges as a promising solution to provide assistance for elderly individuals [5,6]. HAR enables monitoring and analysis of both normal and abnormal activities of the elderly within ALS environments [7].

The human activity recognition for assisted living systems encompasses a wide array of research articles published in IEEE journals and conferences. The significance of human activity recognition in this context is evident from the extensive research efforts in the area [7–9]. The surveyed papers showcase various techniques and sensor modalities employed to develop efficient and context-aware systems that support elderly care and enhance the quality of life for individuals in assisted living environments. The Artificial intelligence-based approaches have gained substantial attention in recent years, demonstrating their potential in achieving higher accuracy and robustness for human activity recognition [10]. Environmental sensors and vision-based systems are also vital components of human activity recognition systems [11,12]. These studies emphasize the importance of context-awareness and the integration of multiple sensor types to enhance recognition capabilities. Furthermore, comparative studies and survey research presented in some papers analyze the performance of various recognition techniques [13,14]. These papers help researchers and practitioners choose suitable methods based on specific application requirements and constraints.

* Corresponding author.
*E-mail addresses:* sg22jn@gmail.com (S. Gaikwad), shripad.bhatlawande@vit.edu (S. Bhatlawande), swati.shilaskar@vit.edu (S. Shilaskar), anjali.solanke1@gmail.com (A. Solanke).

https://doi.org/10.1016/j.medntd.2023.100272
Received 10 April 2023; Received in revised form 23 October 2023; Accepted 26 October 2023

Real-time implementation of human activity recognition has also been addressed in the literature [15–17]. These studies focus on minimizing latency and achieving quick responses to support assisted living residents effectively. The fusion of data from different sources is another significant aspect explored in some papers [17,18]. These studies highlight the potential of combining audio and video data to improve recognition accuracy, providing more detailed and reliable activity information. While the surveyed papers demonstrate the progress made in human activity recognition for assisted living systems, challenges and opportunities remain [19–21]. Issues like Accuracy, cost, power, weight, and adaptability to elderly users need to be addressed for the wider adoption of these systems in real-world scenarios.

The research findings highlight the importance of detecting human activities within assisted living systems. The diverse array of techniques and sensor modalities showcased in the reviewed papers reflects the growing interest and commitment of researchers in this field. As technology continues to advance, further research in this area promises to bring about even more efficient, context-aware, and personalized solutions for assisted living environments. Given the range of methods at hand, this study seeks to make a meaningful addition to the realms of assisted living systems (ALS) and human activity recognition (HAR) by introducing a computer vision driven HAR system. The paper provides a comprehensive exploration of feature compilation analysis and its consequential impact on system accuracy. Furthermore, it conducts a qualitative analysis of implemented algorithms and addresses the identified issues present in the existing body of literature.

## 2. Methodology

This paper presents a computer vision system for the residential monitoring of elderly people. It detects normal and abnormal living patterns in elderly people. Normal living pattern detection includes sitting, standing, walking, and lying activity recognition. Abnormal living pattern detection includes fall detection. Here, fall detection is considered a trigger event to initiate an alarm. As shown in Fig. 1, The implemented computer vision system is a wall-mounted system. It includes a Jetson Nano computing board configured with a 12-megapixel web camera and a sound alarm. To facilitate the experimental process, the researchers selected a single hall within a residence for the elderly, with dimensions of 12 feet in length, 15 feet in width, and 10 feet in height. A single web camera is installed in the hall, and it is enough in this case. However, if the living space is larger or has different sections, having multiple cameras could ensure that all the details are captured effectively. So, it's important to consider the size and layout of each living space when deciding how many cameras to use for optimal performance.

The authors have tested SVM, DT, RF, and Ensemble approaches to implement the computer vision system. Based on the better results, an ensemble approach was used in the final system.

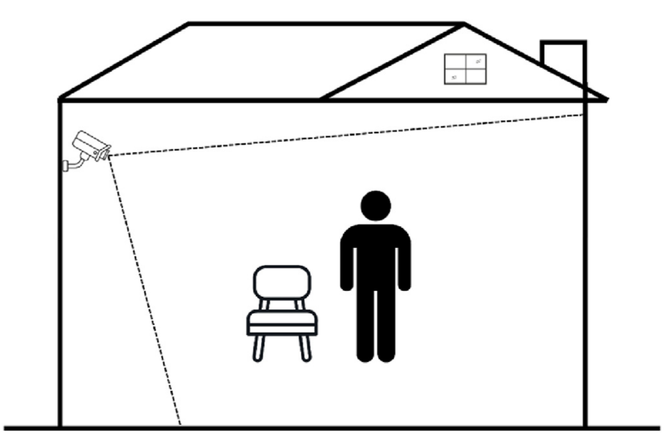

**Fig. 1.** A computer vision system setup.

### 2.1. Dataset creation

#### 2.1.1. Participants

There are a total of four elderly males (ages: 60–65 years; body mass index: 24–28 kg/m2; height: 171.0–186 cm; weight: 65.5–80 kg) are participating as participants in the research work. Handicapped and bedridden people are avoided as participants. To provide variety in the dataset, participants with different body shapes and heights are purposefully selected. Participants consent was taken for experimentation purposes. To ensure accurate monitoring and recognition, participants wear clothing that snugly follows the body contours. Loose or baggy clothing is avoided, as it can obscure crucial body features and movements, thereby making it more challenging for the model to accurately identify activities.

#### 2.1.2. Procedures

All participants' normal activities, like sitting, standing, lying, and walking, and abnormal activities, like falling, are video recorded using a 12-megapixel web camera from different angles. The video resolution of the Web camera is 640 × 480, with a frame rate of 30 fps. The webcam's distance from the people participating ranges from 180 cm to 300 cm. The camera's height ranges from 120 cm to 190 cm. Video recordings of each participant for all activities are converted into frames. For each activity, 1000 images per person are recorded. As per Eq. (1), the total number of images is 20,000.

$$C\ (n) = c\ (p) * c\ (a) * c\ (im) \tag{1}$$

Where,

C (n) = Total count of images,

c (p) = Count of participants,

c (a) = Count of activities,

c (im) = Count of images per activity.

The sample images are shown in Fig. 2.

The majority of existing HAR systems use large datasets with many images, which consume massive amounts of memory. To avoid these issues, the authors have used their own numerical dataset. The dataset is created using features like key joint angles, the Euclidean distance between the key joints, and the slope between the key joints. The authors have calculated these features using Cartesian coordinates of landmarks in the human skeleton. The deep machine learning framework BlazePose is used to detect landmarks. The kNN algorithm in BlazePose returns 33 landmarks of the human skeleton in a frame. These landmarks are shown in Fig. 3. Each landmark contains its own Cartesian coordinates. The authors have used Cartesian coordinates to calculate the following features.

#### 2.1.3. Key joint angle

The *a* and *b* coordinates of three landmarks are required for key joint angle calculation:

If (*a1, b1*), (*a2, b2*) and (*a3, b3*) are coordinates of landmark 1, 2 and 3 respectively. Then using these landmarks and co-ordinates, the key joint angle can be calculated as:

$$(\tan^{-1}(b3\text{-}b2,\ a3\text{-}a2)\ \text{-}\ \tan^{-1}(b1\text{-}b2,\ a1\text{-}a2)) \tag{2}$$

#### 2.1.4. Euclidean distance between the key joints

The *a*, *b* and *c* co-ordinates of two landmarks are required for the calculation of distance between the two key joints: *i. a1, b1, c1 = landmark1, ii. a2, b2, c2 = landmark2.*

Using these landmarks and co-ordinates, the distance between the two key joints is:

$$\sqrt{((c2-c1)^2 + (a2-a1)^2 + (b2-b1)^2)} \tag{3}$$


**Fig. 2.** (a) Standing (b) walking (c) sitting (d) lying.

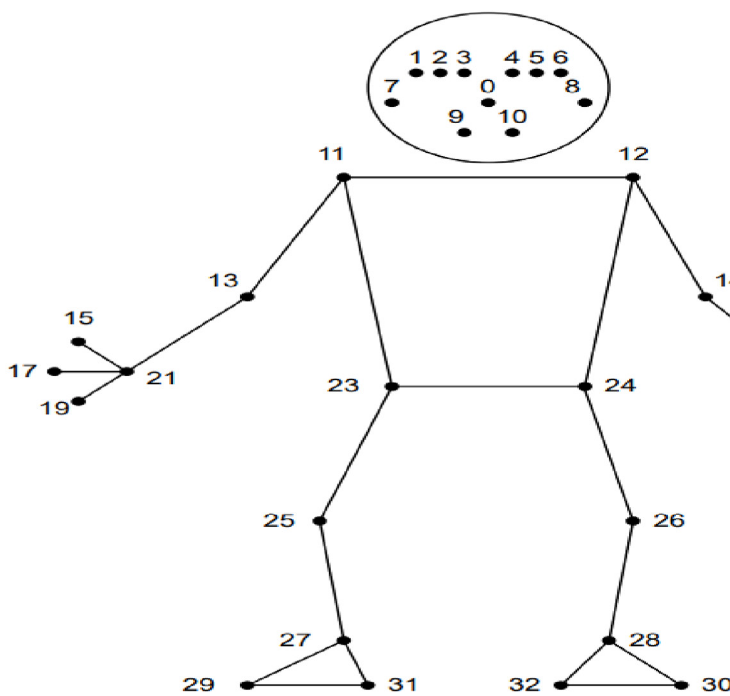


**Fig. 3.** Human skeleton with 33 landmarks.

### 2.1.5. *Slope between the key joints*

For calculating the slope between the two landmarks, respective *a* and *b* coordinates have been used. The formula for the slope is given in Eq. (4).

$$(b2 - b1) \ / \ (a2 - a1) \tag{4}$$

From Eqs. (2)–(4), the features in Table 1 are calculated and recorded to create the dataset. By combining the outcomes of Eqs. (2)–(4), a feature vector of 1*39 is created for each image in the dataset. There are total of 20,000 feature vectors in the dataset. The total number of feature values can be calculated from Eq. (5).

$$\mathrm{f(n)} = \mathrm{fv(n)} * \mathrm{l(fv)} \tag{5}$$

where,

**Table 1**
Selected features for the dataset creation (Numbers are with reference to Fig. 3).

| Parameter | Features |
|---|---|
| Euclidean distance (19 nos.) | 21-22,21–12,11-21,12–22,24-22,23–22,23-24,27–21,28-22,24–28,23-27,28–27,2-31,5–32,13-28,14–27,0-21,0–22,25-26,31–32. |
| Key-joint angle (12 nos.) | 12-14-22,11-13-21,12-14-24,11-13-23,13-23-25,12-24-26,24-26-28,23-25-27,11-27-28,12-28-27,11-22-21,12-21-22. (Middle number: specific key joint angle with reference to other two landmarks) |
| Slope between the key joints (8 nos.) | 24-28, 23–27,12-24,11–23,21-28,22–27,24-27,23-28 |

f(n) = Total number of feature values
fv(n) = Total number of feature vectors l (fv) = length of feature

**Table 2**
Dataset properties.

| Dataset properties | Details |
|---|---|
| Name of classes | Sitting, standing, lying, walking, falling |
| No. of classes | 5 |
| Length of feature vector | 1*39 |
| Number of images | 20,000 |
| Number of feature values | 780,000 |

vector.

From Eq. (5), the total number of feature values is 780,000.

The dataset properties are summarized in Table 2 as below.

### 2.2. Human activity recognition

In an implemented system as shown in Fig. 4, the BlazePose framework is utilized, from real-time video capture to landmark detection. At the first stage, real-time video gets converted into frames. Then each frame is processed through a person detector for human identification. Once the human has been identified, the k-NN algorithm serves to recognize human skeleton landmarks.

With identified landmarks, key joint angles, Euclidean distance between the key joints, and key joint slopes are calculated using Eqs. (2)–(4), respectively. These three features together form the feature vector. A newly formed set of data is used to train and verify a compiled feature vector. Further, four different algorithms are used to classify the postures and activities of the elderly.

The SVM primarily supports binary classification. Also in multiclass classification, it applies binary classification. It breaks the data of multiclass classification into multiple binary classification problems [22]. Paper presents a support vector machine algorithm to implement the HAR system. Here, the radial basis function (RBF) kernel SVM's parameter Gamma is analyzed and described in the result analysis section.

A non-parametric supervised learning technique called decision trees. It is utilized for both regression as well as classification applications. Its goal is to learn simple decision rules derived from data features in order to build a testable hypothesis for the value of the desired component. In this algorithm, the maximum tree depth is a boundary to prevent further splitting of nodes [23]. The authors have discussed the maximum tree depth analysis over the performance of the decision tree algorithm.

**Table 3**
Subjective analysis of a system in a real time environment.

| Algorithm | n | TP + TN | FP | FN | Accuracy |
|---|---|---|---|---|---|
| SVM | 200 | 190 | 05 | 05 | 0.956 |
| DT | 200 | 191 | 04 | 05 | 0.958 |
| RF | 200 | 196 | 03 | 01 | 0.986 |
| Ensemble | 200 | 197 | 02 | 01 | 0.988 |

A technique for machine learning called random forest uses several decision trees to address regression as well as classification issues. The estimator is an important parameter in this algorithm. It determines the quantity of trees within the forest [24]. In the result section, the authors have discussed 'the maximum tree depth' and the number of estimators' analyses over the performance of the random forest algorithm.

The goal of ensemble techniques is to enhance overall generalization and robustness compared to an individual estimator by amalgamating forecasts from multiple base estimators developed using a specific learning process [25]. Here, the authors have combined random forest, SVM, and decision tree algorithms.

To assess a machine learning model's performance on untested data authors have used the cross-validation approach. The given data is divided into several folds or subsets, one of which is used as a validation set, the rest of the folds serve as tools to educate the model. With ten folds, k-fold cross-validation is employed here. The implemented cross-validation technique effectively overcomes the overfitting problem.

The implemented system is validated using subjective and objective analysis. Subjective analysis involves interpreting information based on personal opinions, feelings, interpretations, and perspectives. while Objective analysis involves examining information or data based solely on facts, evidence, and mathematical criteria.

In the Evaluation section, subjective analyses of accuracy, cost, power, and elderly compatibility with the implemented system in a real-time environment are discussed.

## 3. Evaluation

The implemented systems underwent rigorous real-time scenario testing, comprising 200 trials for each of the four techniques, distributed randomly among all participants. Throughout the testing of each technique, 150 instances of activities such as sitting, standing, walking, and

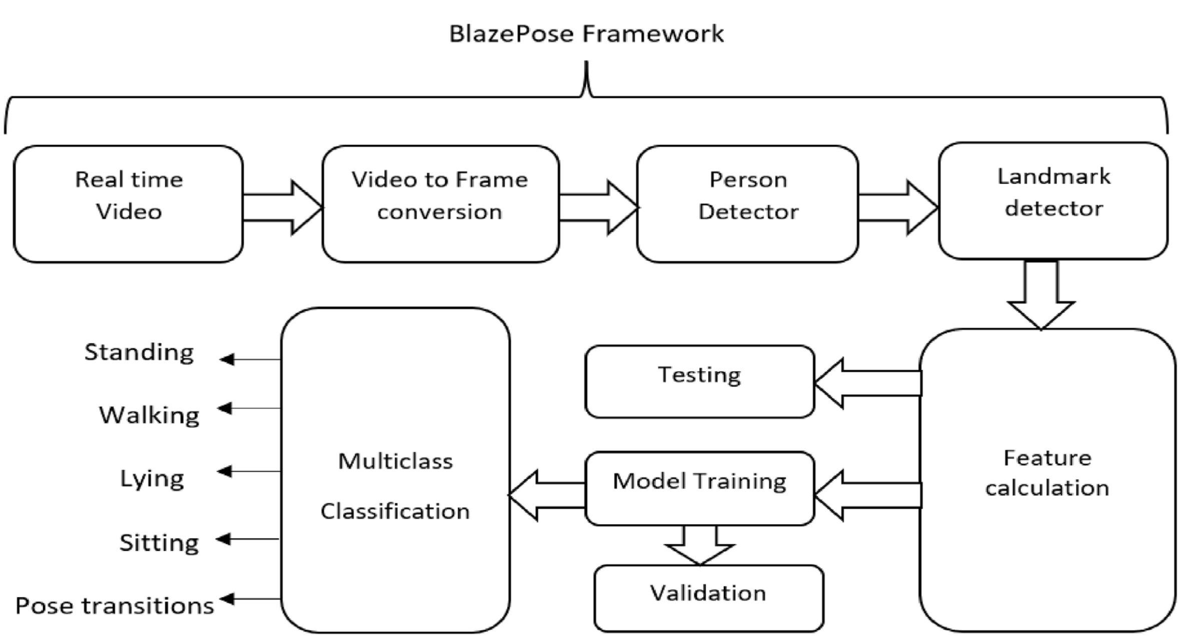


**Fig. 4.** Human activity recognition.

**Table 4**
Feature compilation effect on accuracy of different algorithms.

| Algorithm | Key joint Angle (KJ) | Key joint Angle + Euclidean distance (ED) | Key joint Angle + Euclidean distance + Key joint slopes (SL) |
|---|---|---|---|
| SVM | 0.78 ± 2 | 0.91 ± 2 | 0.94 ± 1 |
| Decision Tree | 0.78 ± 2 | 0.93 ± 2 | 0.96 ± 1 |
| Random Forest | 0.84 ± 2 | 0.95 ± 2 | 0.98 ± 1 |
| Ensemble method | 0.87 ± 2 | 0.94 ± 3 | 0.98 ± 2 |

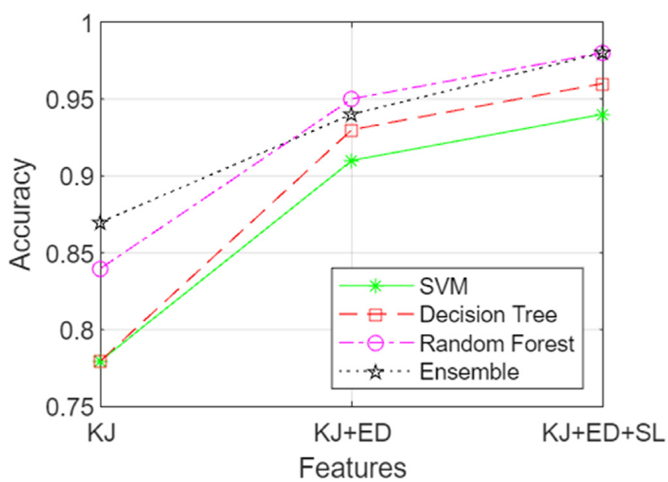


**Fig. 5.** Feature compilation effect on accuracy.

lying were recorded, while 50 fall events were captured. Importantly, all techniques demonstrated exceptional real-time performance, proficiently distinguishing between various activity events. The accuracy of each technique is given in Table 3. In a real-time scenario, SVM achieved an accuracy of 95.6 %, DT achieved 95.8 % accuracy, RF achieved 98.6 % accuracy, and the ensemble approach achieved an impressive accuracy of 98.8 %. These figures underscore the system's efficacy in real-time scenarios, portraying consistent and reliable results across diverse participants.

A supplementary random test is also performed. It involved two male and two female participants who were not originally part of the study. This random test also yielded consistent accuracy, affirming the system's stability. During random examinations, it was noticed that although the original dataset put together by the authors consisted exclusively of male participants, later assessments have revealed that there is no substantial influence on the identification of skeletal landmarks. This is due to the intentional choice of clothing that was selected to enhance the clarity of essential anatomical landmarks. Additionally, the BlazePose framework has been trained on a wide array of datasets that cover both male and female subjects. Consequently, the framework showcases its capability to precisely detect and recognize landmarks for individuals of any gender.

In a real-time scenario, Subjective analysis of cost, power, and elderly people's compatibility with the implemented system is also done. The wall-mounted system, consisting of a Jetson nano computation board, a 12-megapixel camera with a 5 V/2 A rating each, and a sound alarm with a 5 V/30 mA rating, has a power consumption of 6 W. This energy-efficient setup ensures extended operational runtime and reduced electricity costs. Its lightweight design of 1.2 kg makes installation hassle-free, while the absence of wearables eliminates any discomfort or compatibility issues for elderly users. With a cost-effective estimate, the system offers a non-intrusive solution for human activity recognition in

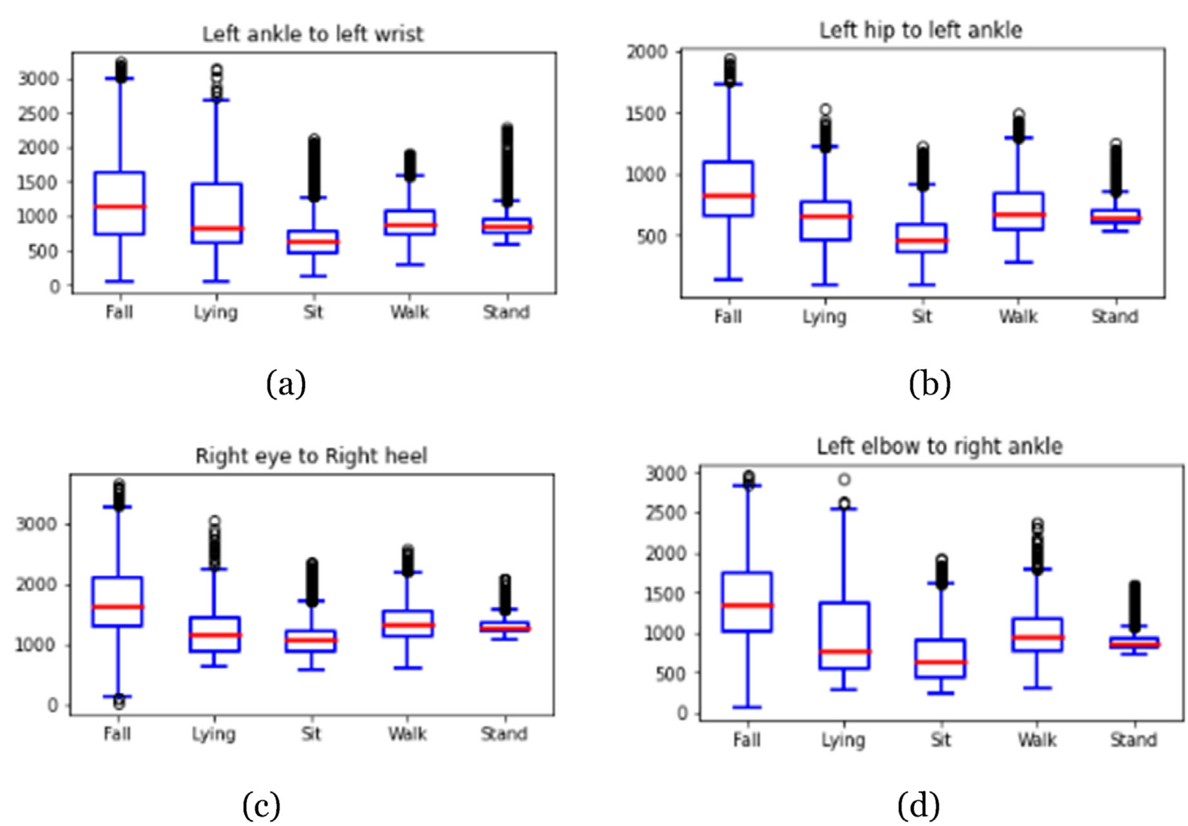


**Fig. 6.** Box plot sample analysis.

**Table 5**
Performance analysis of a SVM algorithm.

| Activity | Precision | Recall | F-1 |
|---|---|---|---|
| Fall | 0.97 | 0.94 | 0.96 |
| Lying | 0.95 | 0.94 | 0.94 |
| Sit | 1 | 1 | 1 |
| Stand | 0.9 | 0.98 | 0.94 |
| Walk | 0.97 | 0.91 | 0.94 |

**Table 6**
Performance analysis of a SVM algorithm with varying Gamma function.

| Gamma | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| 0.01 | 0.72 ± 0.2 | 0.67 | 0.74 | 0.69 |
| 0.001 | 0.96 ± 0.2 | 0.94 | 0.98 | 0.97 |
| 0.0001 | 0.8 ± 0.1 | 0.78 | 0.81 | 0.75 |
| 0.00001 | 0.86 ± 0.3 | 0.83 | 0.86 | 0.82 |

assisted living environments. Its compatibility with smart home setups further enhances monitoring and safety, making it a valuable tool for elderly care.

## 4. Results and discussion

In the results and discussion, the authors have presented an objective analysis of the implemented system. This section provides a comprehensive discussion of the performance metrics and performance analysis of all four implemented systems.

### *4.1. Performance metrics and methods*

The implemented system has been validated using the confusion matrix. It is an efficient method to assess the classification algorithm. A confusion matrix is a plot of correct predictions versus incorrect predictions. In the implemented system, the authors have used four different methods for posture and activity classification. The dataset is segregated into two subsets. The initial subset serves as the training dataset, whereas the subsequent one functions as the testing dataset. To assess the efficacy of implemented models, a testing and training dataset is used. Based on these validations, "Accuracy, Recall, Precision and F1 scores" are calculated to represent the performance of an implemented system.

### *4.2. Feature vector compilation analysis*

As given in Table 4, there are three types of features: Key joint angle, Euclidean distance between the key joints, and slope between the key joints. Authors have observed the effects of these features in different combinations. It is observed that with all three combined feature types, results are improved.

As shown in Fig. 5, graphical analysis clearly shows that with the use of three combined features, the accuracy of all implemented algorithms is increased.

Based on the initial results of different algorithms, three combined feature types are used to create a feature vector of 1*39 length. Fig. 6 shows some box plot samples. These box plots indicate how well features are differentiating the five postures.

### *4.3. Performance analysis of the SVM*

As outlined within the suggested framework segment, an RBF kernel-supported support vector machine is adopted to categorize participants' activities and postures. For each activity listed in Table 5, the gamma function is optimally set at 1 divided by the number of features. Across all indicated everyday activities, the executed system achieved an average accuracy of 95 %, a recall rate of 95.8 %, an F1 score of 95.4 %, and a precision level of 95.6 %. The SVM computation yielded results within a 3-s timeframe.

As shown in Table 6, further SVM algorithm performance analysis can be observed by varying the gamma function. At gamma = 0.001, SVM had the highest accuracy of 96 %.

As shown in Fig. 7, graphical analysis shows that at Gamma = 0.001, SVM outperforms in all quality parameters.

### *4.4. Performance analysis of the decision tree algorithm*

As discussed in the proposed architecture section, a DT algorithm is used to classify the posture and activities of the participants. With the maximum depth function at 10. For all activities, performance analysis is given in Table 7.

Across all the specified day-to-day living tasks, the operational system achieved an average accuracy rate of 96 %, a recall rate of 97.4 %, an F1 score of 96.4 %, and a precision rate of 96 %. The outcome was generated through the utilization of the decision tree algorithm, completed within a 2-s time frame. Further Decision tree algorithm performance analysis is done by varying the maximum depth function. At maximum depth = 50,

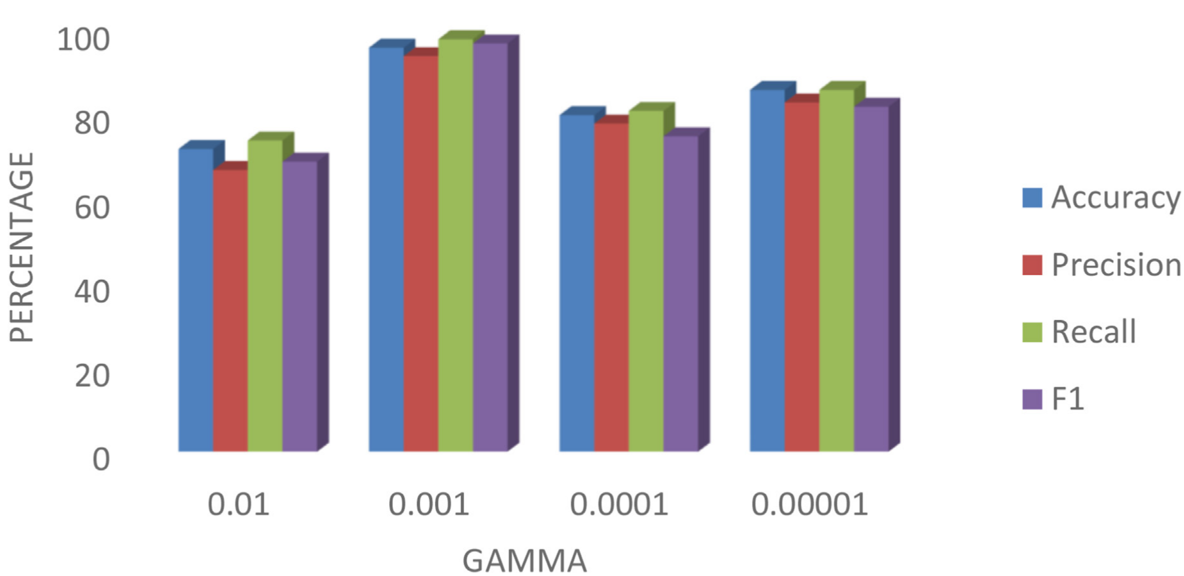


**Fig. 7.** SVM performance analysis with varying Gamma function.

**Table 7**
Performance analysis of a Decision tree algorithm.

| Activity | Precision | Recall | F-1 |
|---|---|---|---|
| Fall | 0.97 | 0.96 | 0.96 |
| Lying | 0.96 | 0.97 | 0.96 |
| Sit | 0.98 | 1.0 | 0.98 |
| Stand | 0.95 | 0.98 | 0.97 |
| Walk | 0.94 | 0.96 | 0.95 |

**Table 8**
Performance analysis of a DT algorithm with varying maximum depth function.

| Max. Depth | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| 10 | $0.96 \pm 0.3$ | 0.95 | 0.96 | 0.96 |
| 30 | $0.96 \pm 0.3$ | 0.96 | 0.95 | 0.96 |
| 50 | $0.97 \pm 0.1$ | 0.98 | 0.97 | 0.97 |
| 80 | $0.95 \pm 0.2$ | 0.95 | 0.95 | 0.96 |
| 100 | $0.94 \pm 0.1$ | 0.93 | 0.92 | 0.94 |

an algorithm got highest accuracy 97 %. Other parameters values are given in the Table 8.

As shown in Fig. 8, graphical analysis shows that at maximum depth = 50, DT algorithm outperforms in all quality parameters.

### 4.5. Performance analysis of the random forest algorithm

As discussed in the proposed architecture section, an RF algorithm is used to classify the activities and posture of participants. For all activities given in Table 9, the maximum depth function is kept at 10. Across all the specified day-to-day activities, the executed system achieved an average accuracy rating of 98 %, a recall rate of 98.4 %, an F1 score of 98.4 %, and a precision level of 98.4 %. Decision tree algorithm took 2 s execution time for result.

Further Random Forest algorithm performance analysis is done by varying the maximum depth function. At maximum depth = 50, an algorithm got highest accuracy of 99 %. Other parameters values are given in the Table 10.

Also, the Random Forest algorithm performance analysis is done by varying the number of estimators. At 10 number of estimators, an algorithm got highest accuracy of 98 %. Other parameters values are given in the Table 11.

**Table 9**
Performance analysis of a Random Forest algorithm.

| Activity | Precision | Recall | F-1 |
|---|---|---|---|
| Fall | 0.99 | 0.97 | 0.98 |
| Lying | 0.98 | 0.98 | 0.98 |
| Sit | 0.99 | 1.0 | 0.99 |
| Stand | 0.99 | 0.99 | 0.99 |
| Walk | 0.97 | 0.98 | 0.98 |

**Table 10**
Performance analysis of a RF algorithm with varying maximum depth function.

| Max. Depth | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| 10 | $0.98 \pm 0.3$ | 0.98 | 0.95 | 0.98 |
| 30 | $0.98 \pm 0.3$ | 0.99 | 0.96 | 0.99 |
| 50 | $0.99 \pm 0.1$ | 1 | 0.97 | 0.99 |
| 80 | $0.99 \pm 0.2$ | 0.99 | 0.97 | 0.99 |
| 100 | $0.98 \pm 0.1$ | 0.99 | 0.96 | 0.98 |

**Table 11**
Performance analysis of a RF algorithm with varying number of estimators.

| Estimator nos. | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| 10 | $0.98 \pm 0.1$ | 0.99 | 0.94 | 0.96 |
| 30 | $0.97 \pm 0.2$ | 0.97 | 0.94 | 0.98 |
| 50 | $0.98 \pm 0.3$ | 0.99 | 0.95 | 0.99 |
| 80 | $0.97 \pm 0.2$ | 1 | 0.95 | 0.98 |
| 100 | $0.98 \pm 0.2$ | 1 | 0.96 | 0.97 |

**Table 12**
Performance analysis of the Ensemble method.

| Algorithm | Accuracy | Precision | Recall | F-1 |
|---|---|---|---|---|
| SVM | 0.95 | 0.94 | 0.94 | 0.95 |
| DT | 0.96 | 0.95 | 0.96 | 0.96 |
| RF | 0.99 | 0.98 | 0.96 | 0.98 |
| Ensemble | 0.99 | 0.98 | 0.97 | 0.99 |

### 4.6. Ensemble methods performance analysis

The ensemble method has shown an accuracy of 99 %. It took 3 s for the result to appear. In Table 12, it is clearly visible that as compared to the other three methods, the ensemble method has an edge over others. Other parameters are mentioned in Table 12.

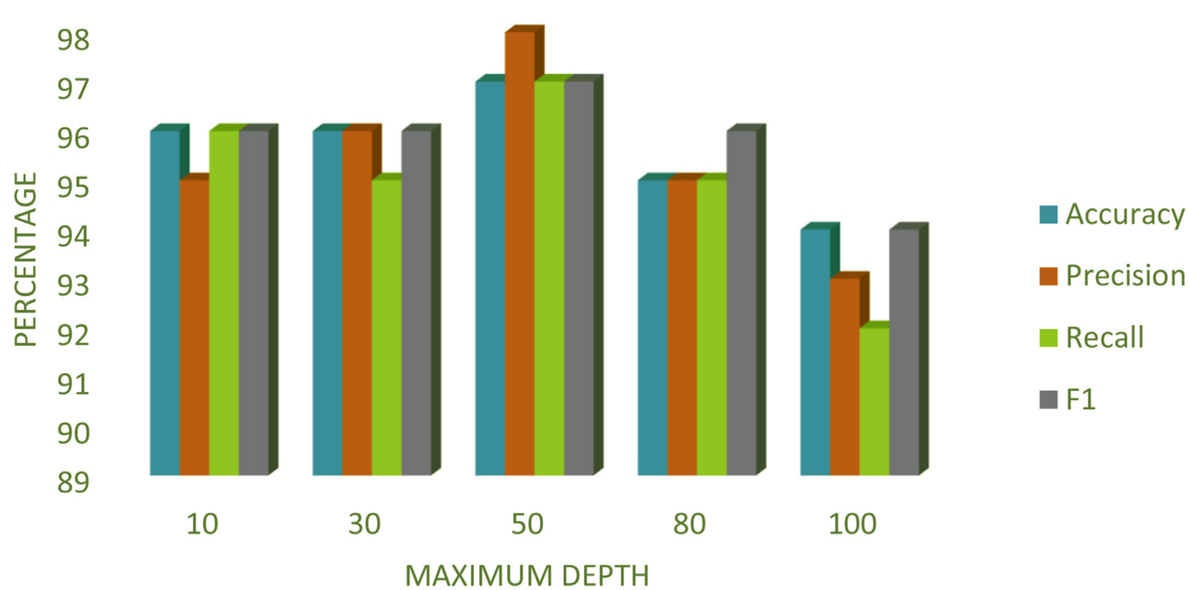


**Fig. 8.** DT algorithm performance analysis with varying maximum depth.

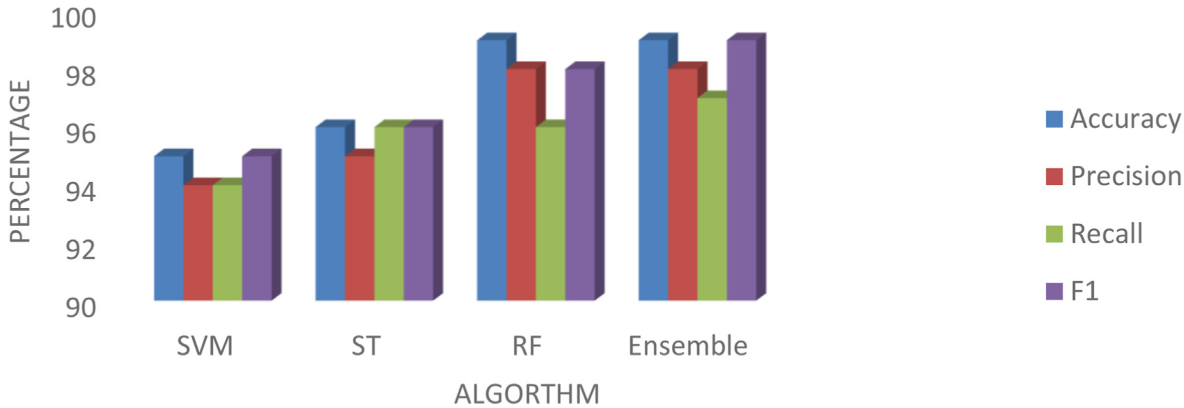


**Fig. 9.** A Comparative analysis of an implemented algorithm.

As shown in Fig. 9, graphical analysis shows that the ensemble approach outperforms other implemented algorithms.

Based on discussions and findings in different sections of this paper, in the next section authors have given the conclusion.

## 5. Conclusion

The paper introduces a novel human activity recognition (HAR) system utilizing computer vision for assisted living systems (ALS) with a focus on supporting elderly individuals. The existing HAR systems using wired or wireless sensor networks have several limitations, including high costs, power issues, weight, and discomfort for the elderly when wearing and carrying such devices. In contrast, the proposed computer vision-based approach presents a breakthrough by overcoming these shortcomings.

The uniqueness of the proposed system lies in its creation of a custom dataset for training. Instead of relying on memory-intensive image datasets, the authors employ a numerical-based, low memory-consuming dataset generated through real-time event simulation on diverse subjects. This approach not only reduces training delays but also significantly lowers memory consumption, making the system more efficient and practical for real-world implementation. An important focal point of this paper lies in the efficient utilization of features within the operational system. The feature compilation section demonstrates the careful selection and incorporation of key-joint angles, distances between key joints, and slopes between them, enabling accurate human posture recognition. The utilization of these features facilitates the training of machine learning algorithms and enables the system to detect five different human postures with impressive precision and accuracy. Four machine learning algorithms, comprising decision tree (DT), random forest (RF), support vector machine (SVM), and an ensemble approach, were put into practice and subjected to testing. The results demonstrate that the ensemble method outperforms the other algorithms, achieving outstanding metrics, including a remarkable accuracy of 99 %, precision of 98 %, recall of 97 %, and an F1 score of 99 %. These exceptional performance metrics serve as strong validation of the effectiveness and reliability of the implemented system.

The implemented computer vision based HAR system serves as a promising solution for assisted living systems, particularly for the elderly population. Its ability to alleviate the limitations of traditional sensor networks, its efficient memory consumption, and its superior recognition accuracy make it a highly practical and valuable tool for enhancing the quality of life and safety of elderly individuals in assisted living environments.

The paper's contributions and findings emphasize the potential of the proposed HAR system to revolutionize assisted living for the elderly. By harnessing the power of computer vision, the system addresses existing challenges and achieves exceptional results. Further advancements and real-world applications of this technology hold great promise for positively impacting the lives of elderly individuals and promoting their independence and well-being.

## Funding sources

No financial assistance was obtained by the authors from any funding organizations.

## Credit author statement

The study was designed by Sudhir Gaikwad and Shripad Bhatlawande. Sudhir Gaikwad undertook the study, contributing to the drafting of the manuscript. Swati Shilaskar and Anjali Solanke provided support in terms of analysis, literature review, and manuscript revisions. All authors participated in the review and approval process of the final manuscript version.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] Government of India. Indian ministry report, elderly in India. https://mospi.gov.in/sites/default/files/publication_reports/Elderly%20in%20India%202021.pdf; 2021.

[2] www.ncbi.nlm.nih.gov, Accessed on 9 April 2023.

[3] United Nations. Report by economic and social affairs population division, world population ageing. 2020. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd-2020_world_population_ageing_highlights.pdf.

[4] Huynh-The Thien, Hua Cam-Hao, Nguyen Anh Tu, Kim Dong-Seong. Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health. IEEE Internet Things J 2021;8(3):1533–43. https://doi.org/10.1109/JIOT.2020.3013272.

[5] Roy W, Hopper Andy, Veronica F, Gibbons Jonathan. the active badge location system. ACM transaction on information system; 1992. p. 91–102. https://dl.acm.org/doi/pdf/10.1145/128756.128759.

[6] S. Gaikwad, S. Bhatlawande, A. Solanke, S. Shilaskatr, Assisted Living System: A Review of 25 Years of Research, Int J Eng Trends Technol (IJETT-2022), 317-330. https://ijettjournal.org/Volume-70/Issue-6/IJETT-V70I6P233.pdf.

[7] Zhou F, Zhou Feng, Roger Jiao Jianxin, Chen Songlin, Zhang Daqing. A case-driven ambient intelligence system for elderly in-home assistance applications. IEEE Trans Syst Man Cybern C Appl Rev 2011:179–89. https://doi.org/10.1109/TSMCC.2010.2052456.

[8] Wang Simin, Zabir Salim, Leibe Bastian. Lying pose recognition for elderly fall detection. In: Robotics:Science and systems VII. MIT Press; 2012. p. 345–53. https://doi.org/10.7551/mitpress/9481.003.0049.

[9] Tsinganos P, Skodras A. A smartphone-based fall detection system for the elderly". In: IEEE,10th international symposium on image and signal processing and analysis. ISPA; 2017. p. 53–8. https://doi.org/10.1109/ISPA.2017.8073568.

[10] Ali Chelli, Patzold Matthias. A machine learning approach for fall detection and daily living activity recognition. IEEE Access 2019;7:1–17. https://doi.org/10.1109/ACCESS.2019.2906693.

[11] Henry Rimminen, Lindström Juha, Linnavuo Matti, Sepponen Raimo. Detection of falls among the elderly by a floor sensor using the electric near field. IEEE Trans Inf Technol Biomed 2010;14(6):1474–6. https://doi.org/10.1109/TITB.2010.2051956.

[12] Sehrawat Deepti, Gill Nasib Singh. Smart sensors: analysis of different types of IoT sensors. In: IEEE 3rd international conference on trends in electronics and informatics; 2019. p. 523–8. https://doi.org/10.1109/ICOEI.2019.8862778.

[13] Khurana Rajat, Singh Kushwaha Alok Kumar. Deep learning approaches for human activity recognition in video surveillance - a survey. In: IEEE first international conference on secure cyber computing and communication (ICSCCC); 2018. https://doi.org/10.1109/ICSCCC.2018.8703295.

[14] Nagpal Diana. Shikha gupta "human activity recognition and prediction: overview and research gaps". In: IEEE 8th international conference for convergence in technology; 2023. p. 1–5. https://doi.org/10.1109/I2CT57861.2023.10126458.

[15] Raeis Hossein, Kazemi Mohammad, Shirmohammadi Shervin. InARMS: individual activity recognition of multiple subjects with FMCW radar. In: IEEE international instrumentation and measurement technology conference (I2MTC); 2022. https://doi.org/10.1109/I2MTC48687.2022.9806694.

[16] Tahvilian Ehsan, Partovi Ehsan, Ejtehadi Mehdi, Bakhshayesh Parsa Riazi, Behz Saeed. Accuracy improvement in simple and complex human activity recognition using a CNN-BiLSTM multi-task deep neural network. In: IEEE, 8th Iranian conference on signal processing and intelligent systems (ICSPIS); 2022. https://doi.org/10.1109/ICSPIS56952.2022.10043933.

[17] Guiping Lin, Weiwei Jiang, Sicong Xu, Xiaobo Zhou, Xing Guo, Yujun Zhu et al., Human Activity Recognition Using Smartphones with WiFi Signals, IEEE Trans. Hum.-Mach. Syst. 202353(1):142-153. https://doi.org/10.1109/ICSPIS56952.2022.10043933.

[18] Shahbazian Reza, Trubitsyna Irina. Human sensing by using radio frequency signals: a survey on occupancy and activity detection. IEEE Access 2023;11: 40878–904. https://doi.org/10.1109/ACCESS.2023.3269843.

[19] Han Junwei, Xiang Ji, Hu Xintao, Guo Lei, Liu Tianming. Arousal recognition using audio-visual features and FMRI-based brain response. IEEE Trans. 2015;6:337–47. https://doi.org/10.1109/TAFFC.2015.2411280.

[20] Abdel-Basset Mohamed, Hawash Hossam, Chang Victor I, Chakrabortty R, Ryan M. Deep learning for heterogeneous human activity recognition in complex IoT applications. IEEE Internet Things J 2022;9(8):5653–65. https://doi.org/10.1109/jiot.2020.3038416.

[21] Babangida Lawal, Perumal Thinagaran, Mustapha Norwati, Yaakob Razali. Internet of things (IoT) based activity recognition strategies in smart homes: a review. IEEE Sensor J 2022;22(9):8327–36. https://doi.org/10.1109/JSEN.2022.3161797.

[22] Shuvo Md MH, Ahmed N, Nouduri Koundinya, Palaniappan Kannappan. A hybrid approach for human activity recognition with support vector machine and 1D convolutional neural network. In: IEEE applied imagery pattern recognition workshop (AIPR); 2020. https://doi.org/10.1109/AIPR50011.2020.9425332.

[23] Charan S, Saravanan M, Surendran R. Prediction of sufficient accuracy for human activity recognition using novel long short-term memory in compared with decision tree. In: IEEE international conference on sustainable computing and data communication systems (ICSCDS); 2023. https://doi.org/10.1109/ICSCDS56580.2023.10104587.

[24] Radhika V, Rajendra Prasad Ch, Chakradhar A. Smartphone-based human activities recognition system using random forest algorithm. In: IEEE international conference for advancement in technology (ICONAT); 2022. https://doi.org/10.1109/ICONAT53423.2022.9726006.

[25] Rojanavasu Pornthep, Jantawong Ponnipa, Jitpattanakul Anuchit, Mekruksavanich Sakor. In: IEEE joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunications engineering (ECTI DAMT & NCON); 2023. https://doi.org/10.1109/ECTIDAMTNCON57770.2023.10139689.