Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-021-02931-5

ORIGINAL RESEARCH

S-THAD: a framework for sensor-based temporal human activity detection from continuous data streams

Muhammad Ehatisham-ul-Haq[1,3] · Muhammad Awais Azam[2] · Fiza Murtaza[3] · Yasar Amin[1] · Qiang Fu[4]

Received: 25 July 2020 / Accepted: 28 January 2021

Abstract
With recent evolvement in smart sensing systems, sensor-based activity recognition has endured numerous research studies that mainly emphasize classifying pre-segmented data chunks having a fixed duration. Each data chunk generally involves a single human activity for classification into one of the predefined activity classes. Such activity recognition models, trained with pre-segmented and fixed-size data chunks, cannot adapt well to natural human activities in real-time settings, where the duration of activities varies. Also, the real-time data available from smart devices is in a continuous form and not discrete chunks. As a result, the real-time implementation of activity-aware applications based on the existing models becomes impractical. Therefore, in this paper, a novel framework, i.e., "S-THAD", is proposed for sensor-based temporal human activity detection from the continuous and untrimmed 3D motion sensor data. The proposed method is capable of detecting when and which activity of interest appears in a continuous data stream. The publicly available PAMAP2 dataset is used to test the proposed scheme, which entails long and untrimmed data streams from the wearable inertial sensors placed at three different body positions, including hand, chest, and ankle. The experimental results indicate that the proposed scheme achieves significant detection performance on this dataset.

Keywords Activity recognition · Continuous data stream · Machine learning · Proposal classification · Temporal activity detection · Wearable sensor

1 Introduction

The advancement in smart sensing devices had led to an immense increase in the research work related to sensor-based human activity recognition (HAR) studies and its applications (Diete and Stuckenschmidt 2019; Wu et al. 2019; Ye et al. 2019; Asghari et al. 2020). Many researchers have proposed different methods to classify human activities based on the data available from smart devices (Igwe et al. 2020; Noor et al. 2017; Mohammed Hashim and Amutha 2020). These methods are generally based on the classification of pre-segmented short chunks of data, containing only a single human activity, into one of the predefined activity classes (Hassan et al. 2017; Morales and Akopian 2017). In this regard, the data from smart devices is generally recorded and processed in the form of small discrete chunks to train and test the HAR systems. As follows, most of the earlier studies in the field of sensor-based HAR focused on a *fixed-size* sliding window for data segmentation and feature extraction (Lara and Labrador 2013; Antos et al. 2014; Xing et al. 2014; Ehatisham-ul-Haq et al. 2017). The window size is set based on the types of activities for recognition and the sampling rate of the recording device. Generally, using a large window size for feature extraction provides rich information for classification. However, it also increases the misclassification rate in the case when some transitions occur within a single window. In contrast, the small window size is capable of recognizing human activities accurately,

✉ Muhammad Ehatisham-ul-Haq
ehatishamuet@gmail.com

[1] Faculty of Telecom and Information Engineering, University of Engineering and Technology (UET), Taxila, Punjab, Pakistan

[2] Technology and Innovation Research Group, School of Information Technology, Whitecliffe , Wellington, New Zealand

[3] Sino-Pak Center for Artificial Intelligence, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur, Khyber Pakthunkhwa, Pakistan

[4] School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

Published online: 26 February 2021

but the computational overhead of the system increases. A *fixed-size* window segmentation approach can be useful for experimenting with different types of case-studies relating to HAR. However, for activity segmentation *in-the-wild*, using a fixed window size is not a practical choice since the duration of natural human activities varies in real-time. There are many expedient applications of sensor-based data obtained from the smart devices in real-time, such as health and fitness monitoring (Mekruksavanich and Jitpattanakul 2020; Subasi et al. 2020), child-care (Li et al. 2018), driving assistance (Ma et al. 2017), user identification (Ehatisham-ul-haq et al. 2018), activity tracking (Alemdar and Ersoy 2017), and context awareness and recognition (Rault et al. 2017; Kim and Yoon 2018). However, the manual segmentation of sensor data into fixed-size chunks is not advantageous for these real-time applications owing to varying lengths of human activity patterns.

In recent years, many researchers have proposed novel approaches for sensor-based data segmentation and time-series classification of human activities to alleviate the issues concerning fixed-size activity segmentation (San-Segundo et al. 2016; Zebin et al. 2018; Chambers and Yoder 2020). Ma et al. (2020) also presented a novel approach for generating an adaptive sliding window, based on Multivariate Gaussian Distribution (MGD), to recognize human activities for assisted living. Liono et al. (2016) investigated the optimal window size for activity segmentation by maximizing the class separability between different temporal segments. Zhuang and Xue (2019) proposed an interval-based method for detecting and recognizing sports-related activities using a smartwatch, where they emphasized the detection of motion states for activity segmentation. The experimental results validated that their proposed scheme performs better than the conventional sliding window-based HAR methods. Akbari et al. (2018) presented a hierarchical window-based approach for signal segmentation to effectively recognize human activities. In this aspect, they first extracted the features over large chunks of data, using a large window size, to recognize the activity. Afterward, they further decomposed the data segments (appearing to entail more than one activity) into smaller chunks for fine-grained labeling and recognition. Li et al. (2019) recognized the basic and transitional physical activities using a sliding window-based segmentation method that entails k-means clustering for aggregating the activity fragments. They tested their proposed scheme on a publicly available smartphone-based HAR dataset and endorsed its effectiveness. Kozina et al. (2011) presented a threshold-based method for dynamic signal segmentation, which analyzes the consecutive data samples using a windowing approach to calculate the threshold value. The authors computed the difference between the positive and negative signal peaks in the consecutive windows to segment the signal. Malhotra et al. (2018) proposed a correlation-based method for time-series classification of human activities. A few research studies (Triboan et al. 2017, 2019) also proposed semantic data segmentation approaches for HAR. Minh Dang et al. (2020) presented a comprehensive survey of the latest research work carried out in the field of HAR. They categorized the existing HAR methods into different groups and subgroups based on the underlying data, feature extraction techniques, and training/classification methods. They also analyzed the pros and cons of different HAR approaches and discussed the future research challenges in HAR domain.

In recent years, with the advancement in deep learning algorithms/approaches, various researchers have put their efforts into the utilization of these algorithms for developing sensor-based HAR methods (Dawar and Kehtarnavaz 2018; Nweke et al. 2018; Ahad et al. 2021; Uddin et al. 2020). Wang et al. (2020) utilized the hybrid deep learning techniques for HAR based on the wearable sensor data. Chen et al. (2019) proposed a multi-agent HAR, which is based on the spatial–temporal attention model. Wan et al. (2020) presented a smartphone accelerometer-based architecture design for HAR, which is based on convolutional neural network (CNN). Norgaard et al. (2020) utilized multiple sensors for time-series classification of variable length human activities based on deep neural network (DNN). Varamin et al. (2018) proposed deep auto-set, a multi-label architecture for HAR based on the wearable sensors. The authors explored HAR as a set prediction problem and adopted the Maximum Posteriori (MAP) estimation to predict activity sets using DNN. They also predicted the number of activities involved in a data chunk (i.e., window) and the probability of occurrence for each alternative activity within a data chunk. Yao et al. (2018) performed dense labeling of human activity sequences and applied fully convolutional network (FCN) design for activity prediction based on each time step (i.e., sample). In this regard, the authors utilized human activity data from the wearable sensors and assigned an activity label to each data sample for sample-based prediction. Zhang et al. (2018) also performed sample-based activity prediction based on U-Net. Wang et al. (2019) and Chen et al. (2020) presented the detailed surveys on the use of deep learning approaches for sensor-based HAR and reviewed their advantages and shortcomings. Although, deep learning-based HAR schemes tend to be successful in improving the classification results, however, these approaches are computationally costly and thus not practicable for real-time operations on battery-constrained devices, i.e., smartphone, smartwatch, or portable/wearable sensing system. Moreover, training and updating a deep learning-based activity detection model in real-time requires a large amount of data, which is a very time-consuming task. Thus, there is a need to develop a robust yet computationally efficient solution for activity detection based on the sensor data from smart devices.

This research work aims to address the existing limitations associated with sensor-based activity detection/recognition methods, particularly temporal human activity detection. Therefore, it is worth mentioning that the current HAR techniques in this regard are mostly concerned with the task of *activity segmentation* only, which involves dividing the sensor data into small fragments for better feature extraction and thus classification. Thus, these schemes do not emphasize finding out the starting and ending duration of an activity in a long and continuous data stream, which is crucial for enabling real-time HAR applications. Likewise, most of these techniques entail experimentation as a case-study for HAR, thus overlooking the notion of temporal activity detection in real-time. Generally, the data available from smartphone-embedded sensors or wearable inertial sensors is continuous and thus untrimmed. This results in long streams of data comprising multiple activities of interest as well as the background activities with varying durations. Consequently, determining any transition from one activity to another becomes essential for enabling context-aware applications in-the-wild. Therefore, the automatic detection of human activities from continuous sensor data is indispensable. In this regard, an automatic method is required to simultaneously detect the starting and ending time of each activity and its corresponding label from the continuous data stream. This task is comprehended as *temporal human activity detection* (THAD), which aims to find out when and which activity of interest appears in the continuous and untrimmed data stream. As follows, in this paper, a novel framework, i.e., "S-THAD", is presented for *sensor-based temporal human activity detection*, based on 3D motion sensor data. Thus, given a continuous sensor data stream, the proposed S-THAD method aims to generate a small number of temporal locations as output. These locations are termed as *temporal activity proposals* and are expected to encompass a human activity of interest. An efficient THAD method should be capable of retrieving tight activity locations with high precision while discarding the background activity regions. Also, the retrieved locations or activity proposals should take into account a high temporal overlap with the actual locations or ground truths for their accurate classification into one of the predefined activity classes. The proposed S-THAD method is inspired by the observation that different human activities entail varying patterns with different durations. Therefore, finding the discriminating features from the successive non-overlapping data chunks can help to detect human activities of daily living (ADLs) from continuous data streams. Moreover, it can also differentiate between the activities of interest and background activities, thus providing an opportunity to cope with the *open-set* activities in real-time HAR applications. The proposed scheme can be effectively applied for THAD using any type of sensor data, including ambient data, smartphone-embedded inertial sensor data, or wearable sensor data. Furthermore, it is capable of processing other types of signal data as well, including electrocardiogram (ECG) and electroencephalograph (EEG) signals, for segmentation and event detection purposes.

The main contributions of this paper are as follows:

- A novel method is proposed for temporal detection of human activities from the continuous sensor data based on supervised learning.
- Multiple wearable sensor positions are incorporated in the proposed framework, where *twelve* ADLs (including a background activity class) are chosen for experimentation.
- Automated model selection is carried out to choose the best-case classifier (i.e., Random Forest) for investigating the proposed scheme results in detail. The best-case detection results are also compared with those obtained with other standard classifiers.
- A detailed experimental analysis is conducted to investigate the effect of different sensor positions on THAD performance.
- A *decision-level* fusion method is applied to combine the wearable sensors data, available from three body positions, to improve the detection performance.

The remaining paper is organized as follows. Section 2 explains the proposed methodology for S-THAD in detail. Section 3 comprehensively analyzes the performance of the proposed method and discusses the obtained results in detail. Finally, Sect. 4 concludes this research work and provides recommendations for future work.

2 Proposed method for S-THAD

Figure 1 presents the block diagram of the proposed S-THAD method that entails four key steps: (1) data segmentation and snippet generation, (2) feature extraction from snippet, (3) snippets accumulation into candidate activity proposal, and (4) temporal activity detection. The details related to each of these steps are provided in the subsequent sections.

2.1 Data acquisition, segmentation, and snippet generation

There are many publicly available datasets for testing the performance of a HAR model, e.g., PAMS (Esfahani and Malazi 2018), SisFall (Sucerquia et al. 2017), ExtraSensory (Vaizman et al. 2018), and MobiAct (Vavoulas et al. 2016). These datasets entail pre-segmented and discrete chunks

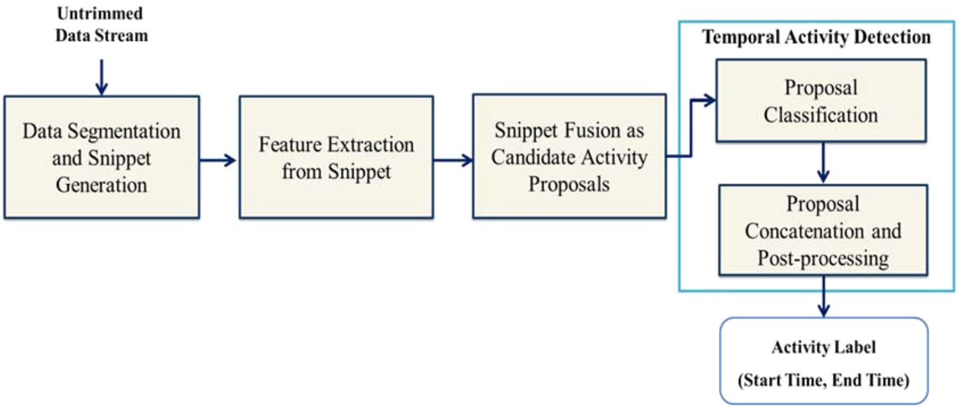
Untrimmed Data Stream
Data Segmentation and Snippet Generation
Feature Extraction from Snippet
Snippet Fusion as Candidate Activity Proposals
Temporal Activity Detection
Proposal Classification
Proposal Concatenation and Post-processing
Activity Label (Start Time, End Time)

Fig. 1 Block diagram of the proposed S-THAD method

of data, where each data segment comprises only a single human activity labeled as the ground truth. Thus, it is not possible to utilize these datasets for modeling and testing the continuous sensor data. For evaluating a THAD model, an untrimmed or continuous data stream is required, which involves multiple activities having a random time duration. Thus, for implementation of the proposed S-THAD framework, a publicly available human activity dataset, i.e., PAMAP2 dataset (Reiss and Stricker 2012), is utilized as it satisfies the requirements for testing a THAD model. This dataset entails nine (09) participants' data for overall eighteen (18) activities. Besides, the dataset also contains a *background* activity class covering the supplementary user activity (such as *transient activities* or *waiting states*). Each participant's data is provided in the form of a continuous data stream with labeled timestamps and activities. Based on these labels, it is possible to find out the starting and ending time of each activity in the data stream to form the ground truths. Hence, this dataset fits well into the pipeline of our proposed framework. The duration, length, and order of the performed activities are not similar for each participant. Furthermore, each activity is not necessarily performed by all the participants. Hence, for a few activities, there exist only a small number of data instances across all the participants. To implement and test the proposed scheme, the detection of eleven (11) most frequent activities is emphasized, which entail a sufficient number of samples for the system training and testing. These activities include *background (a0)*, *lying (a1)*, *sitting (a2)*, *standing (a3)*, *walking (a4)*, *running (a5)*, *cycling (a6)*, *Nordic walking (a7)*, *upstairs (a8)*, *downstairs (a9)*, *vacuuming (a10)*, and *ironing (a11)*. To keep the continuity of the data stream, the remaining activities are set as *background* activities. The data from the accelerometer and gyroscope sensors is utilized for implementation, which is collected at a sampling rate of nearly 100 Hz by placing wearable inertial measurement units (IMUs) at three body positions (i.e., *ankle*, *chest*, and *hand* position) of a subject. The duration of average data labeled per subject in the PAMAP2 dataset is equal to one hour approximately. Hence, this dataset entails an adequate amount of continuous data to effectively train and test any THAD scheme.

In the first step of the proposed method, the untrimmed sensor data streams are divided and labeled in the form of small non-overlapping data chunks having a duration of 1-s. These data segments are termed as *snippets*. The primary reason for using a 1-s *snippet* is to reduce the computational complexity of processing each data sample separately while retrieving the activity locations as tight as possible based on the windowing approach.

2.2 Feature extraction from snippets

After data segmentation and snippet generation, a set of twenty (20) time-domain statistical signal attributes are extracted as features from the non-overlapping snippets to get useful information about the data. These features are as follows: maximum and minimum signal amplitudes, peak-to-peak signal value and time, mean, variance, standard deviation, kurtosis, skewness, third and fourth moments of the signal, minimum and maximum latencies, signal percentiles (i.e., 25th, 50th, and 75th), energy, mean of the first and second difference of the signal, and signal entropy. By extracting these features from both 3D sensors (including accelerometer and gyroscope), a final feature vector of size 1×120 is obtained per snippet. The main reason for adopting these hand-crafted features is their efficient performance in state-of-the-art HAR methods (Shoaib et al. 2014; Hassan et al. 2017; Bharti et al. 2019; Ehatisham-ul-Haq and Azam 2020; Ehatisham-Ul-Haq et al. 2020) on account of low computational cost. Unlike high-level deep features extracted automatically from the data, the *hand-crafted* features do not entail any noise-driven features. Thus, the analysis and visualization of these features and their individual impact on the classification problem is easier than deep learning-based features. Moreover, deep features require a very large amount of data for efficient training and their performance cannot be generalized if the underlying training dataset is small or imbalanced. Besides, the computational complexity of deep learning-based feature extraction models is generally very high as compared to simple time-domain features. As a result, using deep learning-based models for real-time activity detection/recognition on the battery-constrained devices is not a viable solution.

2.3 Snippet accumulation into candidate activity proposals

After feature extraction from each snippet, these features are accumulated over multiple sized windows to generate the candidate proposals containing the targeted activities or background regions. The window size is calculated in terms of the second (sec). As natural human activities have varying durations in time, therefore using a single snippet (of size 1-s) to detect the ADLs is impractical. Hence, the accumulation of multiple snippets is performed using different sized windows to accurately detect human activities with varying lengths. The impact of using different window sizes for candidate proposal generation/classification is evaluated in Sect. 3.2. To accumulate the snippet features into a proposal representation p, the sum pooling is used as given in Eq. (1):

$$p_w = \sum_{i=s}^{w+s-1} f_i \tag{1}$$

where, s and w represent start and size of the window, respectively, as shown in Fig. 2. It is essential to mention here that non-overlapping windows of different sizes are used for accumulation, which means that if the first window starts at $s = 1$ and ends at w, then the next window will start and end at $s = w + 1$ and $2w$, respectively. This process keeps on repeating until the data is available from the sensing device or data stream.

2.4 Temporal activity detection

The final step of the proposed method is activity detection, which consists of two building blocks, i.e., *proposal classification* and *proposal concatenation*. The following sections provide details relevant to these blocks.

2.4.1 Proposal classification

This step classifiers each candidate proposal (generated after accumulating the snippets) into one out of $M + 1$ classes using supervised machine learning. In this aspect, Random Forest (RF) classifier is used for proposal classification, which is an ensemble classifier based on the collection of decision trees. RF classifier obtains class predictions from all individual trees, which are combined using a majority voting principle to generate the final output prediction. Each tree generates the output prediction based on a random subset of features or samples to improve the overall group prediction (Breiman 2001). In this way, RF provides improved classification performance as compared to other classifiers, which is one of the key reasons for adopting this classifier. For training the proposed S-THAD system, RF classifier is fed with the final feature vector concerning each candidate proposal and its respective class label as inputs. During the testing stage, we classify each candidate activity proposal based on the pre-trained model to find the respective class label.

2.4.2 Proposal concatenation

As the candidate activity proposals are non-overlapping, hence, it is required to merge multiple candidate proposals if they have the same class label. It means that the consecutive proposals with the same labels can be merged into a single region, as shown in Fig. 3. However, there is a possibility that one (or a small sequence of the proposals) may get misclassified within a long stream of candidate proposals relating to the same class. Thus, the temporal locations of the classified activity proposals will change, which degrades the system performance.

To further optimize the proposal concatenation, some post-processing is performed to filter out and correct the

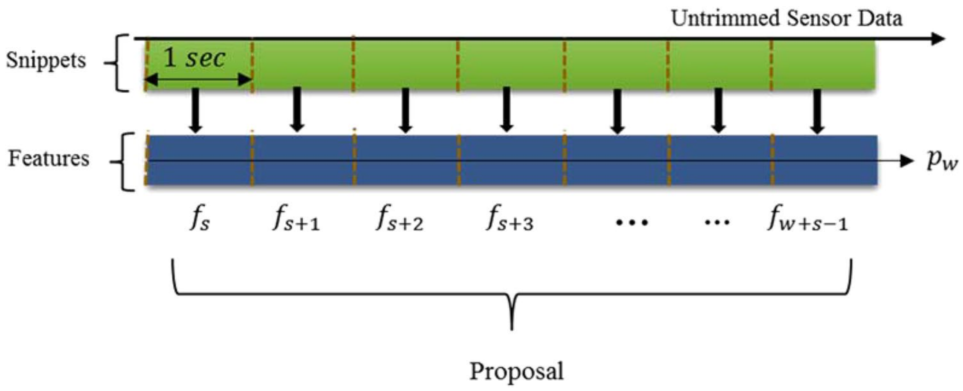

Fig. 2 Accumulating *snippets* into *candidate proposals* using a window size equal to *w* snippets

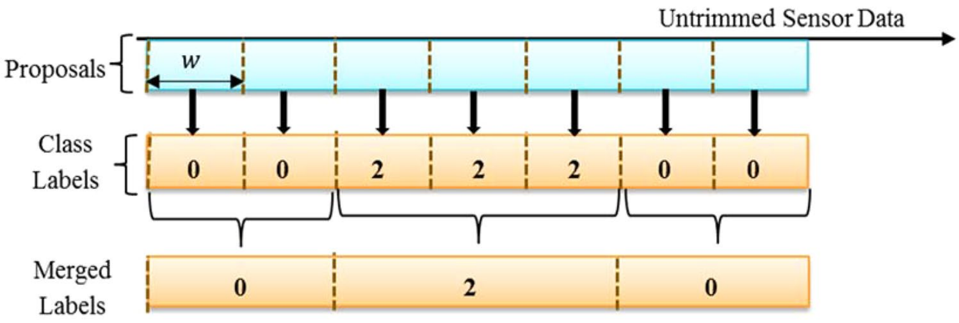

Fig. 3 Concatenation of candidate activity proposals into temporal activity regions

wrongly classified candidate proposals. The process of post-processing is explained hereby with the help of an example. Consider, a sequence of u successive candidate proposals is classified, where q proposals are recognized as belonging to the class having label L. However, among these u candidate proposals, r number of consecutive proposals are found to be classified to some other classes, where $r = u - q$. If $1 \leq r < l_s/2$, these r activity proposals are detected as misclassified and assigned the same class label (i.e., L) as the remaining q proposals. The symbol l_s denotes the length of the smallest activity in terms of the number of candidate proposals, which is computed from the training data streams. In this way, the performance of the proposed S-THAD framewk is further enhanced by optimizing the temporal detection of different activities in the continuous datstream.

3 Experimental results and analysis

This section presents the performance analyses of the proposed S-THAD method in detail. To evaluate the proposed scheme, an automated model selection technique is applied at first to pick the best classification algorithm (with optimized hyperparameter values) for the underlying data. In this regard, the "Auto-WEKA" method (Thornton et al. 2013) is employed for model selection using the WEKA tool (Holmes et al. 2002). This method takes a few parameters (e.g., time limit, memory limit, optimization metric, and number of best configurations) as input and returns the best classifier configuration as output. The default input parameter values are used in this study when applying the "Auto-WEKA" method. As a result, RF classifier (with the number of iterations equal to 100) is obtained as an optimal choice of the classifier, which is further utilized for training and testing the proposed scheme based on the *leave-one-subject-out* (LOSO) validation scheme. LOSO is a *subject-specific* cross-validation scheme that takes into account an unseen subject in the testing stage to generalize the system performance for real-time settings.

3.1 Performance evaluation metrics

The existing literature on sensor-based human activity analysis mainly focuses on evaluating the activity classification performance in terms of accuracy, precision, recall, and F1-score. However, there are no standard parameters for assessing THAD performance based on the sensor data. It is crucial to mention here that THAD does not only entail recognizing an activity but also finding out its location (i.e., starting and ending time). A few researchers (Minnen et al. 2006; Ward et al. 2006, 2011) have suggested parameters like the number of *events*, *insertions*, *deletions*, *fragmentations*, and *substitutions* for assessing the detection performance; however, these measures are never adopted worldwide. In the past few years, computer vision-based research work has produced several studies on THAD in long untrimmed videos with state-of-the-art parameters for evaluating their performance (Mettes et al. 2015; Heilbron et al. 2016; Murtaza et al. 2020). These studies depict that we can assess the performance of any THAD scheme based on its *activity localization* and *recognition* quality. In this regard, "temporal Intersection over Union (tIoU)" finds out how well our predicted location matches with the corresponding ground truth location for a specific activity in a continuous data stream. The tIoU parameter is based on Jaccard's similarity coefficient (Hancock et al. 2004). It measures the ratio of the area of intersection and area of union pertaining to the predicted and ground truth activity locations (Mettes et al. 2015; Heilbron et al. 2016). Equation (2) characterizes the tIoU of two boundaries, i.e., time intervals, where, B_{actual} and $B_{\text{predicted}}$ represent the ground truth and predicted locations, respectively.

$$tIoU = \frac{B_{\text{predicted}} \cap B_{\text{actucal}}}{B_{\text{predicted}} \cup B_{\text{actucal}}} \tag{2}$$

The tIoU metric gives a value ranging from 0 to 1, representing the worst and perfect matching between the ground truth and predicted locations, respectively. Besides, it also determines if a predicted location is *true* *positive*, *false positive*, or *false negative* based on a predefined threshold value (δ) using the criteria given below.

- If tIoU $> \delta$ with the same output activity label as ground truth, the predicted location is considered as *true positive*.
- If tIoU $< \delta$ and the output activity label is the same as ground truth, the predicted location is considered as *false positive*.
- If tIoU $> \delta$ and the output activity label is different from ground truth, the predicted location is considered as *false negative*.

Generally, a median value of tIoU is used for thresholding, i.e., $\delta = 0.5$, which signifies an overlap of 50% between the predicted boundary and actual ground truth. However, this threshold criterion is not fixed and may vary for different case-studies. After evaluating each predicted location as *true positive*, *false positive*, or *false negative*, the THAD performance can be measured in terms of *average precision* (AP) that signifies the area under the *precision-recall* curve (Heilbron et al. 2016). It is important to note that the AP metric is different from conventional *precision* and is represented as given in Eq. (3).

$$AP(c) = \frac{\sum_{i=1}^{m} \text{precision}(i) \times \text{rel}(i)}{\sum_{i=1}^{m} \text{rel}(i)} \tag{3}$$

where, m represents total instances of the testing class c, *precision*(i) denotes the precision at the i-th instance, and *rel(i)* function indicates a value 1 or 0 if the i-th instance is detected as a *true positive* or *false positive/false negative*, respectively. The AP values obtained for all individual classes are then averaged to calculate the *mean average precision* (mAP). Furthermore, the mAP value against the entire range of tIoU values (i.e., 0–1) to generalize the THAD performance.

3.2 S-THAD results and discussions

As mentioned earlier, human activities are not always fixed and have different lengths over time. In the case of natural human activity detection in a continuous data stream, the duration of an activity cannot be predetermined. The same activity can be performed multiple times with varying time intervals. Hence, it is necessary to investigate different window lengths when concatenating the snippets for activity detection. In this aspect, multiple sized windows (with $w = 5$ –30) are utilized in this scheme for accumulating the snippets into candidate activity proposals, where w represents the window size in terms of the number of snippets. In addition, a detailed investigation is carried out to find out how different window sizes affect the output results for proposal classification and thus activity detection. Table 1 presents the average results obtained when classifying the candidate activity proposals using RF classifier. It can be observed from the table that for each sensing position, the best-case activity proposal classification results (in terms of F1-score) are achieved using the window size $w = 10$. The overall best performance (i.e., F1-score = 82.4%) is attained when the sensors are placed at the *hand* position with $w = 10$, whereas the worst results are obtained in the case of *chest* position. It can also be analyzed from Table 1 that increasing the window size from $w = 10$ leads to a decrease in the overall F1-score obtained for activity proposal classification.

Table 1 Activity proposal classification results obtained using RF classifier with different window sizes

Sensors position	Window size in *"w"* snippets	Precision	Recall	F1-Score	Accuracy
Hand	5	0.867	0.740	0.772	0.964
	10	**0.929**	**0.797**	**0.824**	**0.970**
	15	0.902	0.787	0.819	0.969
	20	0.887	0.781	0.808	0.969
	25	0.872	0.761	0.781	0.962
	30	0.766	0.748	0.748	0.956
Chest	5	0.790	0.677	0.695	0.949
	10	**0.812**	**0.738**	**0.749**	**0.956**
	15	0.747	0.714	0.678	0.954
	20	0.709	0.711	0.677	0.950
	25	0.700	0.710	0.667	0.948
	30	0.647	0.517	0.537	0.934
Ankle	5	0.816	0.683	0.723	0.954
	10	**0.863**	**0.757**	**0.790**	**0.964**
	15	0.853	0.725	0.777	0.961
	20	0.841	0.721	0.764	0.960
	25	0.828	0.717	0.750	0.958
	30	0.826	0.704	0.741	0.958

The numerical values in bold represent the best average results obtained for activity proposal classification corresponding to each sensor position

Figures 4, 5, 6 present a comparison of the mAP values obtained for the proposed THAD method using wearable sensors placed at the *hand*, *chest*, and *ankle* positions, respectively. It can be observed from these figures that better mAP values are obtained when the window size is usually smaller, where the best-case results are obtained with $w = 10$. If the window size is further increased, the output detection results become insignificant, showing no particular trend to be reported as reliable results. It is owing to the reason that using a large window size may result in significant overlap with the background, which poorly affects the detection performance. In contrast, merging snippets based on small window size often produces more

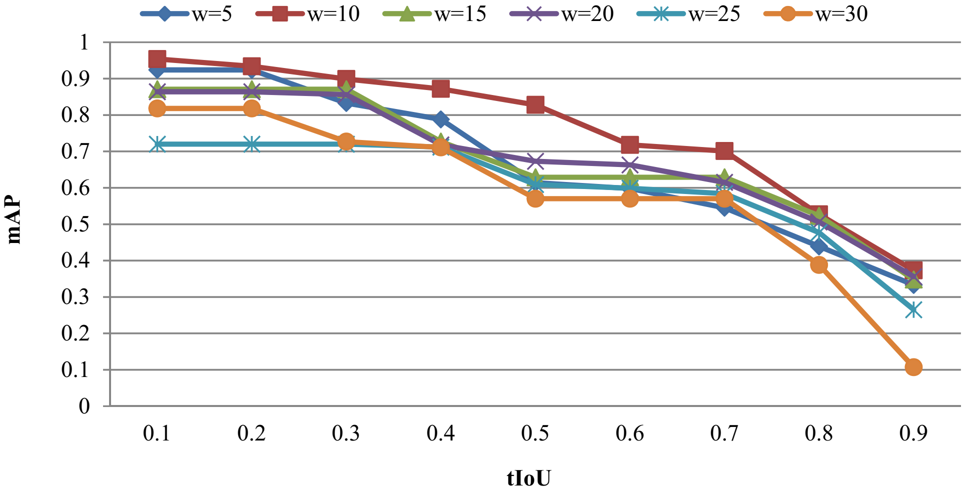

Fig. 4 Comparison of the mAP obtained for the proposed S-THAD method at multiple tIoU values and window sizes with sensors placed at the "hand" position

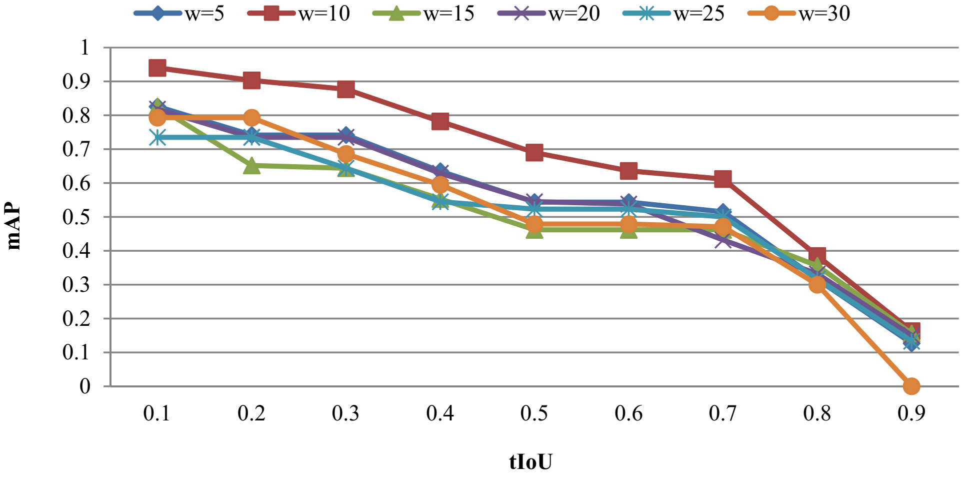

Fig. 5 Comparison of the mAP obtained for the proposed S-THAD method at multiple tIoU values and window sizes with sensors placed at the "chest" position

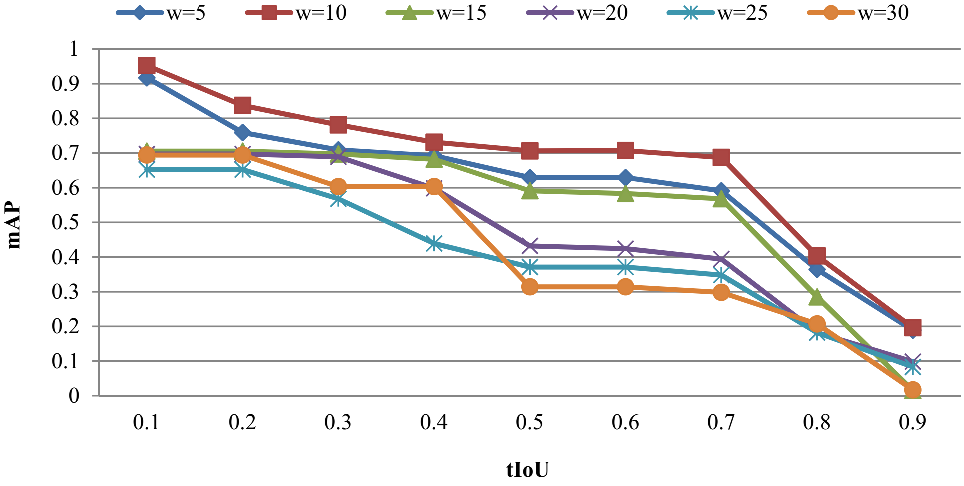

Fig. 6 Comparison of the mAP obtained for the proposed S-THAD method at multiple tIoU values and window sizes with sensors placed at the "ankle" position

tight detection with the ground truth, thus improving the overall detection results. However, a very small window is not a practical choice to be adopted as it cannot fully represent the action and also results in more interclass similarity. Hence, it can be concluded that a 10-s window length is well-suited for the proposed scheme based on the discussed results. Therefore, the remaining parts of the results in this paper are presented and investigated for w = 10 only.

By analyzing the results reported in Figs. 4, 5, 6, it can be analyzed that the values of mAP mostly decrease as the tIoU parameter increases. It is owing to the reason that a high value of the tIoU means a more fitting temporal overlap of the predicted and actual boundaries, which is harder to achieve above a particular threshold value (i.e., 0.5). However, the results produced at higher threshold values are more significant and have higher confidence as there is more overlap between the actual and predicted locations. The proposed scheme achieves the best-case mAP value of 0.718, 0.636., and 0.707 corresponding to the *hand*, *chest*, and *ankle* position, respectively, at tIoU = 0.6 and w = 10. For the same positions, the obtained results are decreased by a value of 1.7%, 1.4%, and 2.0%, respectively, in the case of tIoU = 0.7. Nevertheless, this reduction in the mAP values provides more confidence in the obtained results. By comparing the mAP obtained for different sensing positions, it can be said that the best overall results are achieved when placing the sensors at the *hand* position. Table 2 compares the mAP values obtained for the proposed scheme based on different classifiers (including RF, Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbors (K-NN), and Naïve Bayes (NB)). These results signify the best-case performance is achieved for each classifier at tIoU = 0.6 and tIoU = 0.7. Furthermore, RF provides the overall best performance for S-THAD, whereas NB classifier achieves poor results among all the classifiers. These results signify the efficacy of using RF classifier for S-THAD.

Table 3 the individual activity detection results for all sensor positions based on RF classifier, which are computed using a window size w = 10. It can be observed from the table that many activities are detected with varying results at three different positions based on the tIoU parameter. It is due to the fact that wearable inertial sensors entail local body movements only. Furthermore, some activities are only well-suited to be detected/recognized when the sensors are placed at a particular position. For example, the *upstairs* activity is not well detected with the sensors placed at the *hand* or *chest* position. However, it attains promising detection results in the case of *ankle* position. Similarly, the *upstairs* activity also achieves better results for the *ankle* position. The activities of *running* and *walking* are best detected with the sensors placed at the *chest/ankle* and *hand* position, respectively. The static activities, i.e., *sitting*, *lying*, and *standing*, are better detected based on the hand movements. Likewise, the *ironing* activity and other *background* activities are also detected better using the *hand* position. Table 3 also depicts that the individual activity detection performance mostly decreases as the tIoU metric is increased.

Furthermore, it can be observed from Table 3 that only a few activities are detected in a better way for all body positions. However, most of the activities are *position-dependent* and only achieve good detection results for some specific sensor position. For addressing this challenge, the use of *decision-level fusion*, based on the *logarithmic opinion pool* (LOGP) (Li et al. 2015), is envisaged for integrating the individual activity detection results at multiple positions. LOGP entails a soft fusion of the individual posterior probability $\rho_f(\tau|\gamma)$

Table 2 Comparison of mAP values obtained for S-THAD using different classifiers (with w = 10)

Sensing position	Classifier	tIoU								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Hand	SVM	0.985	0.977	0.970	0.970	0.795	0.674	0.674	0.568	0.432
	K-NN	0.992	0.970	0.917	0.894	0.727	0.629	0.629	0.333	0.333
	DT	0.955	0.955	0.871	0.697	0.424	0.364	0.364	0.182	0.167
	NB	0.795	0.750	0.667	0.583	0.417	0.333	0.333	0.333	0.167
	RF	0.954	0.934	0.899	0.872	0.828	**0.718**	**0.701**	0.527	0.373
Chest	SVM	0.909	0.652	0.652	0.644	0.636	0.583	0.477	0.348	0.161
	K-NN	0.902	0.735	0.727	0.682	0.568	0.561	0.477	0.348	0.113
	DT	0.911	0.867	0.809	0.739	0.627	0.567	0.485	0.317	0.150
	NB	0.598	0.598	0.598	0.432	0.265	0.265	0.265	0.227	0.144
	RF	0.940	0.903	0.877	0.782	0.690	**0.636**	**0.612**	0.385	0.163
Ankle	SVM	0.909	0.826	0.818	0.712	0.545	0.538	0.538	0.432	0.265
	K-NN	0.833	0.826	0.720	0.689	0.515	0.462	0.462	0.333	0.212
	DT	0.826	0.818	0.720	0.629	0.523	0.485	0.394	0.295	0.167
	NB	0.750	0.750	0.750	0.583	0.500	0.462	0.462	0.379	0.167
	RF	0.952	0.837	0.781	0.731	0.706	**0.707**	**0.687**	0.403	0.196

The numerical values in bold represent the overall best average results obtained for the proposed S-THAD method

Table 3 Individual activity detection results at multiple positions based on RF classifier (using $w = 10$)

Activities		tIoU								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
		Hand position								
$a0$	Background	0.91	0.91	0.82	0.82	0.73	0.69	0.64	0.27	0.00
$a1$	Lying	1.00	1.00	1.00	1.00	0.98	0.94	0.94	0.68	0.00
$a2$	Sitting	1.00	1.00	1.00	1.00	1.00	0.96	0.90	0.92	0.71
$a3$	Standing	1.00	1.00	1.00	1.00	0.98	0.96	0.96	0.91	0.88
$a4$	Walking	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.88	0.82
$a5$	Running	1.00	1.00	0.91	0.62	0.33	0.10	0.10	0.06	0.00
$a6$	Cycling	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.90	0.82
$a7$	Nordic Walking	1.00	1.00	1.00	1.00	1.00	0.91	0.90	0.88	0.61
$a8$	Upstairs	0.54	0.30	0.08	0.08	0.00	0.00	0.00	0.00	0.00
$a9$	Downstairs	1.00	1.00	0.98	0.96	0.96	0.22	0.20	0.00	0.00
$a10$	Vacuuming	1.00	1.00	1.00	1.00	0.98	0.88	0.88	0.00	0.00
$a11$	Ironing	1.00	1.00	1.00	0.98	0.98	0.96	0.96	0.82	0.64
mAP→		**0.954**	**0.934**	**0.899**	**0.872**	**0.828**	**0.718**	**0.701**	**0.527**	**0.373**
		Chest position								
$a0$	Background	1.00	1.00	1.00	0.91	0.73	0.55	0.55	0.27	0.00
$a1$	Lying	1.00	1.00	1.00	1.00	0.98	0.98	0.96	0.81	0.64
$a2$	Sitting	0.96	0.87	0.79	0.66	0.20	0.10	0.08	0.00	0.00
$a3$	Standing	1.00	1.00	1.00	1.00	1.00	0.94	0.94	0.62	0.00
$a4$	Walking	1.00	1.00	1.00	0.40	0.33	0.21	0.18	0.00	0.00
$a5$	Running	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.74	0.62
$a6$	Cycling	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.82	0.00
$a7$	Nordic Walking	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.78	0.69
$a8$	Upstairs	0.32	0.20	0.16	0.10	0.00	0.00	0.00	0.00	0.00
$a9$	Downstairs	1.00	1.00	1.00	1.00	0.98	0.98	0.95	0.58	0.00
$a10$	Vacuuming	1.00	1.00	1.00	0.98	0.86	0.76	0.76	0.00	0.00
$a11$	Ironing	1.00	0.76	0.57	0.34	0.20	0.12	0.12	0.00	0.00
mAP→		**0.940**	**0.903**	**0.877**	**0.782**	**0.690**	**0.636**	**0.612**	**0.385**	**0.163**
		Ankle position								
$a0$	Background	0.91	0.82	0.82	0.73	0.64	0.55	0.55	0.27	0.18
$a1$	Lying	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.68	0.00
$a2$	Sitting	1.00	0.48	0.21	0.00	0.00	0.12	0.10	0.00	0.00
$a3$	Standing	0.52	0.31	0.12	0.00	0.00	0.00	0.00	0.00	0.00
$a4$	Walking	1.00	0.44	0.28	0.11	0.00	0.00	0.00	0.00	0.00
$a5$	Running	1.00	1.00	0.98	0.98	0.96	0.96	0.94	0.77	0.54
$a6$	Cycling	1.00	1.00	1.00	1.00	1.00	1.00	0.95	0.81	0.22
$a7$	Nordic walking	1.00	1.00	0.96	0.96	0.91	0.91	0.88	0.77	0.64
$a8$	Upstairs	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.88	0.55
$a9$	Downstairs	1.00	1.00	1.00	1.00	1.00	0.98	0.92	0.12	0.00
$a10$	Vacuuming	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.22	0.10
$a11$	Ironing	1.00	1.00	1.00	1.00	0.98	0.98	0.98	0.31	0.12
mAP→		**0.952**	**0.837**	**0.781**	**0.731**	**0.706**	**0.707**	**0.687**	**0.403**	**0.196**

corresponding to each classifier for estimating the global probability function $P(\tau|\gamma)$, as given in Eq. (4).

$$P(\tau|\gamma) = \prod_{f=1}^{N_c} \rho_f(\tau|\gamma)^{\alpha_f} \tag{4}$$

where, τ represents a class label, N_c is the number of classifiers (i.e., $N_c = 3$ in this case), and $\alpha_f = \frac{1}{N_c}$. The final class label for τ is then assigned to the class with the largest probability $P(\tau|\gamma)$.

Table 4 summarizes the results obtained for THAD using the *decision-level* fusion of all sensor positions (i.e., *hand*, *chest*, and *ankle*). It can be observed from these results that the individual detection performance for most of the activities is significantly improved using the fusion approach, particularly at higher values of the tIoU parameter (i.e., as 0.6 and 0.7), as highlighted in the table. This analysis depicts the fact that if we can combine the resulting information available simultaneously from the sensors placed at the *hand*, *chest*, and *ankle* position, we can improve the overall detection performance of the proposed scheme as well as the individual activities. It is owing to the reason that some activities get better recognized using a combination of local movements from different body parts.

Figure 7 compares the overall mAP obtained corresponding to individual positions and their *decision-level* fusion based on RF classifier. It can be analyzed that the mAP values for the proposed scheme are significantly improved to 88.9% and 87.4% at tIoU = 0.6 and tIoU = 0.7, respectively. These results demonstrate the efficacy of using *decision-level* fusion for improving the S-THAD results based on multiple sensor positions. In general, the proposed scheme achieves satisfactory performance for temporal detection of ADLs and can be successfully utilized for real-time HAR applications.

3.3 Comparison with state-of-the-arts

As discussed earlier, the proposed scheme entails the temporal detection of human activities based on the continuous and untrimmed sensor data. Hence, it is not realistic to provide a one-to-one comparison of the proposed scheme results with the existing HAR methods that mainly involve activity classification based on the discrete chunks of data. However, to demonstrate the efficacy of the proposed S-THAD method over the existing studies, Table 5 compares the primary attributes of some well-known sensor-based HAR schemes with the proposed method. It can be examined from the table that the existing schemes only emphasize the recognition of human activities, such as ADLs, home tasks, or elderly activities. In this regard, different types of sensing modalities (including smartphones, wearable sensing systems, and ambient sensing devices) and their combination are used for data recording purposes. The acquired sensor data is first divided into small fragments such that each data segment entails a single activity. After that, feature extraction is carried out for model training and thus classification. Both deep learning models and standard machine learning algorithms are employed in the existing studies for HAR, which achieved successful recognition results. However, the existing schemes do not entail finding out the starting and ending duration of an activity in a long and continuous data stream, which is indispensable for enabling real-time HAR applications. As the raw data available from the sensing devices is always continuous and thus untrimmed, hence, it becomes crucial to detect the transitions between different activities in real-time scenarios. In this regard, the proposed S-THAD scheme enables temporal detection of human activities in the continuous data streams, thus enabling real-time activity-aware applications. The existing HAR methods (as depicted in Table 5) fail to address these issues associated with the real-time applications of HAR. Henceforth, the efficacy of the proposed S-THAD scheme is justified over the existing HAR schemes.

The proposed framework is designed such that it can easily be adapted and applied to real-time sensor data for temporal activity detection. In this aspect, the data can be recorded and processed continuously in the form of 1-s snippets to extract the time-domain features (as mentioned in Sect. 2.2). A non-overlapping window of a certain duration can then be applied to accumulate the snippet features and generate the final feature vector for activity proposal classification. In this way, the transitions between different activities can be detected based on the current and previous candidate activity labels to find out the starting and ending duration of an activity in the continuous data. The sensing device does not need to store the entire data stream, but only a few snippets of data and the output activity labels to accurately detect the activities of interest.

4 Conclusions

This paper proposes a novel framework for temporal activity detection from long and continuous data streams available from wearable sensors, such as accelerometer and gyroscope. For this purpose, twelve (12) ADLs (including the background activities) are selected from the publicly available PAMAP2 dataset to validate the proposed scheme. Furthermore, three different sensing positions (i.e., *hand*, *chest*, and *ankle*) are chosen for activity detection. At first, the statistical signal attributes are computed as features from the inertial sensors data in the form of snippets. Then, sum pooling is applied to accumulate the feature vectors over different sized windows to form candidate activity proposals. Finally, RF classifier is applied for evaluating the proposed scheme performance in terms of mAP at multiple tIoU values. Overall, the best mAP value (i.e., 71.8% at tIoU = 0.6)

Table 4 Activity detection results achieved for the *decision-level* fusion of all sensor positions based on RF classifier (using $w = 10$)

Activities		tIoU								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$a0$	Background	1.00	1.00	1.00	0.91	0.82	0.77	0.71	0.27	0.20
$a1$	Lying	1.00	1.00	1.00	1.00	0.98	0.96	0.96	0.96	1.00
$a2$	Sitting	1.00	1.00	1.00	1.00	0.96	0.92	0.92	0.92	0.16
$a3$	Standing	1.00	1.00	1.00	1.00	0.98	0.96	0.96	0.96	1.00
$a4$	Walking	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00
$a5$	Running	1.00	1.00	1.00	1.00	0.46	0.28	0.16	0.33	0.10
$a6$	Cycling	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	1.00
$a7$	Nordic walking	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00
$a8$	Upstairs	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00	0.33
$a9$	Downstairs	1.00	1.00	1.00	1.00	0.88	0.88	0.88	0.24	0.21
$a10$	Vacuuming	1.00	1.00	1.00	1.00	0.91	0.90	0.90	0.42	0.12
$a11$	Ironing	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.91	0.82
mAP→		**1.000**	**1.000**	**1.000**	**0.992**	**0.912**	**0.889**	**0.874**	**0.731**	**0.578**

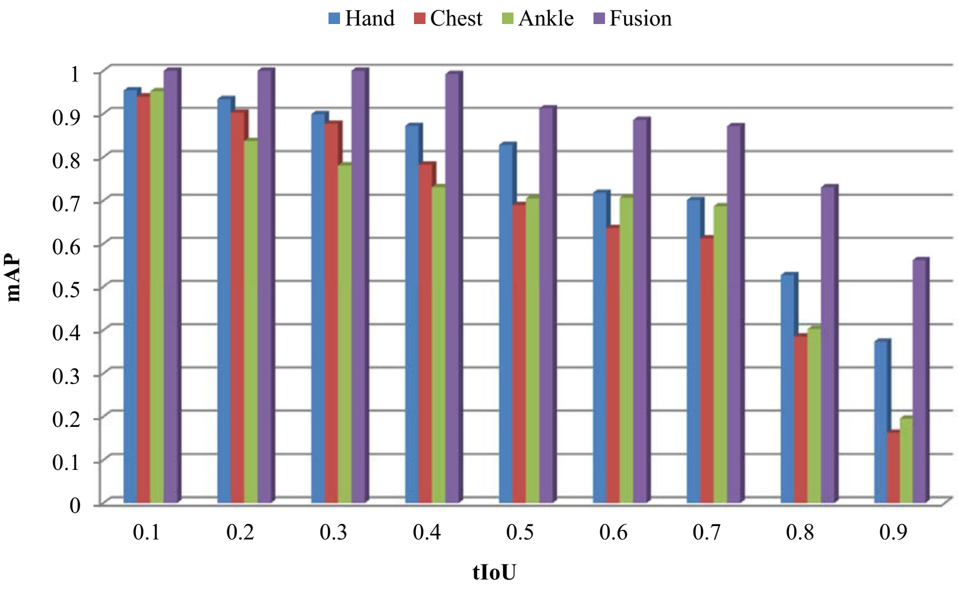

Fig. 7 Comparison of the mAP values obtained for the proposed scheme using individual sensing positions and their *decision-level* fusion based on RF classifier

is obtained for the *hand* position. However, by applying the *decision-level fusion* at multiple positions, the mAP value is significantly improved to 88.9% at tIoU = 0.6.

In future studies, the accuracy of the proposed framework can be improved using the novel feature extraction and activity segmentation approaches. Besides, this research work can be expanded to detect and assimilate new activities into the classification model for improving system performance. The proposed method can also be incorporated into the real-time implementation of wide-ranging HAR applications and context-aware systems for event detection.

Funding This research work is sponsored by the Directorate of Advanced Studies, Research, and Technological Development (ASR&TD), University of Engineering and Technology (UET), Taxila, Punjab, Pakistan.

Availability of data and material A public domain dataset, i.e., “PAMAP2”, is used for evaluating the proposed scheme, which is adequately cited in the paper.

Code availability Code for data processing and analysis will be shared/uploaded to the journal repository once the paper has been conditionally accepted.

Table 5 Comparison of the proposed scheme with the existing state-of-the-arts

Study/Task	Activity type (# of Activities)	Sensing device	Model/Classifier	Results
(Hassan et al. 2018)/ Activity recognition	ADLs (12)	W (A+G)	DBN	Accuracy = 97.5
(Ignatov 2018)/ Activity recognition	ADLs (06)	SP (A)	Customized CNN	Accuracy = 97.6%
(Zhao et al. 2018)/ Activity recognition	ADLs (06)	SP (A+G+M.)	Residual LSTM	Accuracy = 93.6%
(Rivera et al. 2017)/ Activity recognition	Gestures (06)	SP (A+G+M.)	LSTM	Accuracy = 80%
(Catal et al. 2015)/ Activity recognition	ADLs (06)	SP (A)	MLP, LR, DT (Decision-level Fusion)	F1-Score = 91.8%
(Ravi et al. 2017)/ Activity recognition	ADLs (06)	SP (A)	CNN	F1-Score = 97.4%
	ADLs (07)	SP (A+G)		F1-Score = 93.1%
(Hassan et al. 2017)/ Activity recognition	ADLs (12)	SP (A+G)	NN, SVM, DBN	Accuracy = 89.61% (DBN)
(Garcia-Ceja et al. 2018)/ Activity recognition	Home tasks (07)	SP (A+Mic.); W (A.)	RF	Accuracy = 94.1%
(Guiry et al. 2014)/ Activity recognition	ADLs (09)	SP (A+G+M); Pressure Sensor;	DT, NB, SVM, MLP	Accuracy = 92.8% (MLP)
(Fatima et al. 2013)/ Activity recognition	Home tasks (10)	Motion Sensor; Ambient (T)	HMM, NN, CRF, SVM, CE (using Genetic Algorithm)	F1-Score = 90.1% (CE)
	Home tasks (11)	Motion Sensor; Item; EU; Ambient (D, T, L);		F1-Score = 81.9% (CE)
	Home tasks (15)	Motion Sensor; Ambient (D);		F1-Score = 85.7% (CE)
(Wang et al. 2018)/ Activity recognition	Elderly activities (17)	W (B+T+A+G+M); Ambient (PIR);	SVM	Accuracy = 98.3%
Proposed/activity classification	ADLs (12)	W (A+G);	SVM, K-NN, DT, NB, RF	Accuracy = 96.3% (RF)

Proposed/temporal activity detection	ADLs (12)	W (A+G);	SVM, K-NN, DT, NB, RF	mAP = 88.9% @ tIou = 0.6 (RF)
				mAp = 87.4% @ tIoU = 0.7(RF)

A accelerometer, *B* barometer, *CE* classifier ensemble, *CRF* conditional random fields; *D* door sensor, *DBN* deep belief network, *EU* electricity usage, *G* gyroscope, *HMM* hidden markov model, *L* light sensor, *LR* logistic regression, *M* magnetometer, *Mic.* microphone, *MLP* multilayer perceptron, *NN* neural network, *PIR* passive infrared sensor, *T* temperature sensor, *W* wearable

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Ahad MAR, Antar A Das, Ahmed M (2021) Deep learning for sensor-based activity recognition: recent trends. In: IoT Sensor-Based Activity Recognition: Human Activity Recognition, vol 173. Springer, pp 149–173. https://doi.org/10.1007/978-3-030-51379-5_9

Akbari A, Wu J, Grimsley R, Jafari R (2018) Hierarchical signal segmentation and classification for accurate activity recognition. In: UbiComp/ISWC 2018 Adjunct Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers, pp 1596–1605

Alemdar H, Ersoy C (2017) Multi-resident activity tracking and recognition in smart environments. J Ambient Intell Humaniz Comput 8:513–529. https://doi.org/10.1007/s12652-016-0440-x

Antos SA, Albert MV, Kording KP (2014) Hand, belt, pocket or bag: practical activity tracking with mobile phones. J Neurosci Methods 231:22–30. https://doi.org/10.1016/j.jneumeth.2013.09.015

Asghari P, Soleimani E, Nazerfard E (2020) Online human activity recognition employing hierarchical hidden Markov models. J Ambient Intell Humaniz Comput 11:1141–1152. https://doi.org/10.1007/s12652-019-01380-5

Bharti P, De D, Chellappan S, Das SK (2019) HuMAn: complex activity recognition with multi-modal multi-positional body sensing. IEEE Trans Mob Comput 18:857–870. https://doi.org/10.1109/TMC.2018.2841905

Breiman L (2001) Random Forrest. Mach Learn. https://doi.org/10.1023/A:1010933404324

Catal C, Tufekci S, Pirmit E, Kocabag G (2015) On the use of ensemble of classifiers for accelerometer-based activity recognition. Appl Soft Comput J 37:1018–1022. https://doi.org/10.1016/j.asoc.2015.01.025

Chambers RD, Yoder NC (2020) Filternet: a many-to-many deep learning architecture for time series classification. Sensors (Switzerland). https://doi.org/10.3390/s20092498

Chen K, Yao L, Zhang D, et al (2019) Multi-agent attentional activity recognition. In: IJCAI International Joint Conference on Artificial Intelligence, pp 1344–1350

Chen K, Zhang D, Yao L, et al (2020) Deep learning for sensor-based human activity recognition: overview, challenges and opportunities.https://arxiv.org/abs/2001.07416

Dawar N, Kehtarnavaz N (2018) Action detection and recognition in continuous action streams by deep learning-based sensing fusion. IEEE Sens J 18:9660–9668. https://doi.org/10.1109/JSEN.2018.2872862

Diete A, Stuckenschmidt H (2019) Fusing object information and inertial data for activity recognition. Sensors (Switzerland). https://doi.org/10.3390/s19194119

Ehatisham-ul-Haq M, Azam MA (2020) Opportunistic sensing for inferring in-the-wild human contexts based on activity pattern recognition using smart computing. Futur Gener Comput Syst 106:374–392. https://doi.org/10.1016/j.future.2020.01.003

Ehatisham-ul-Haq M, Azam MA, Loo J et al (2017) Authentication of smartphone users based on activity recognition and mobile sensing. Sensors (Switzerland). https://doi.org/10.3390/s17092043

Ehatisham-ul-haq M, Awais M, Naeem U et al (2018) Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing. J Netw Comput Appl 109:24–35. https://doi.org/10.1016/j.jnca.2018.02.020

Ehatisham-Ul-Haq M, Azam MA, Amin Y, Naeem U (2020) C2FHAR: coarse-to-fine human activity recognition with behavioral context modeling using smart inertial sensors. IEEE Access 8:7731–7747. https://doi.org/10.1109/ACCESS.2020.2964237

Esfahani P, Malazi HT (2018) PAMS: a new position-aware multi-sensor dataset for human activity recognition using smartphones. In: 2017 19th International Symposium on Computer Architecture and Digital Systems, CADS 2017, pp 1–7

Fatima I, Fahim M, Lee YK, Lee S (2013) A genetic algorithm-based classifier ensemble optimization for activity recognition in smart homes. KSII Trans Internet Inf Syst 7:2853–2873. https://doi.org/10.3837/tiis.2013.11.018

Garcia-Ceja E, Galván-Tejada CE, Brena R (2018) Multi-view stacking for activity recognition with sound and accelerometer data. Inf Fusion 40:45–56. https://doi.org/10.1016/j.inffus.2017.06.004

Guiry JJ, van de Ven P, Nelson J (2014) Multi-sensor fusion for enhanced contextual awareness of everyday activities with ubiquitous devices. Sensors (Switzerland) 14:5687–5701. https://doi.org/10.3390/s140305687

Hancock JM, Zvelebil MJ, Hancock JM (2004) Jaccard Distance (Jaccard Index, Jaccard Similarity Coefficient). Dictionary of bioinformatics and computational biology. John Wiley and Sons Ltd, New Jersey

Hassan MM, Uddin MZ, Mohamed A, Almogren A (2017) A robust human activity recognition system using smartphone sensors and deep learning. Futur Gener Comput Syst. https://doi.org/10.1016/j.future.2017.11.029

Hassan MM, Huda S, Uddin MZ et al (2018) Human activity recognition from body sensor data using deep learning. J Med Syst. https://doi.org/10.1007/s10916-018-0948-z

Heilbron FC, Niebles JC, Ghanem B (2016) Fast temporal activity proposals for efficient detection of human actions in untrimmed videos. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 1914–1923

Holmes G, Donkin A, Witten IH (2002) WEKA: a machine learning workbench. Springer, Berlin, pp 357–361

Ignatov A (2018) Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. Appl Soft Comput J 62:915–922. https://doi.org/10.1016/j.asoc.2017.09.027

Igwe OM, Wang Y, Giakos GC, Fu J (2020) Human activity recognition in smart environments employing margin setting algorithm. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-02229-y

Kim S, Yoon YI (2018) Ambient intelligence middleware architecture based on awareness-cognition framework. J Ambient Intell Humaniz Comput 9:1131–1139. https://doi.org/10.1007/s12652-017-0647-5

Kozina S, Lustrek M, Gams M (2011) Dynamic signal segmentation for activity recognition. Proc Int Jt Conf Artif Intell 5:1–12

Lara OD, Labrador M (2013) A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor 15:1192–1209. https://doi.org/10.1109/SURV.2012.110112.00192

Li W, Chen C, Su H, Du Q (2015) Local binary patterns and extreme learning machine for hyperspectral imagery classification. IEEE Trans Geosci Remote Sens 53:3681–3693. https://doi.org/10.1109/TGRS.2014.2381602

Li X, Malebary S, Qu X, et al (2018) ICare: Automatic and user-friendly child identification on smartphones. In: HotMobile 2018 Proceedings of the 19th International Workshop on Mobile Computing Systems and Applications, pp 43–48

Li JH, Tian L, Wang H et al (2019) Segmentation and recognition of basic and transitional activities for continuous physical human activity. IEEE Access 7:42565–42576. https://doi.org/10.1109/ACCESS.2019.2905575

Liono J, Qin AK, Salim FD (2016) Optimal time window for temporal segmentation of sensor streams in multi-activity recognition. In: ACM International Conference Proceeding Series, pp 10–19

Ma C, Dai X, Zhu J et al (2017) DrivingSense: dangerous driving behavior identification based on smartphone autocalibration. Mob Inf Syst. https://doi.org/10.1155/2017/9075653

Ma C, Li W, Cao J et al (2020) Adaptive sliding window based activity recognition for assisted livings. Inf Fusion 53:55–65. https://doi.org/10.1016/j.inffus.2019.06.013

Malhotra A, Schizas ID, Metsis V (2018) Correlation analysis-based classification of human activity time series. IEEE Sens J 18:8085–8095. https://doi.org/10.1109/JSEN.2018.2864207

Mekruksavanich S, Jitpattanakul A (2020) Exercise activity recognition with surface electromyography sensor using machine learning approach. In: 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, ECTI DAMT and NCON 2020, pp 75–78

Mettes P, Van Gemert JC, Cappallo S, et al (2015) Bag-of-fragments: Selecting and encoding video fragments for event detection and recounting. In: ICMR 2015 Proceedings of the 2015 ACM International Conference on Multimedia Retrieval, pp 427–434

Minh Dang L, Min K, Wang H et al (2020) Sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Recognit. https://doi.org/10.1016/j.patcog.2020.107561

Minnen D, Westeyn TL, Starner T et al (2006) Performance metrics and evaluation issues for continuous activity recognition. Proc Int Work Perform Metrics Intell Syst. https://doi.org/10.1145/1889681.1889687

Mohammed Hashim BA, Amutha R (2020) Human activity recognition based on smartphone using fast feature dimensionality reduction technique. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-020-02351-x

Morales J, Akopian D (2017) Physical activity recognition by smartphones, a survey. Biocybern Biomed Eng 37:388–400. https://doi.org/10.1016/j.bbe.2017.04.004

Murtaza F, Yousaf MH, Velastin SA, Qian Y (2020) Vectors of temporally correlated snippets for temporal action detection. Comput Electr Eng. https://doi.org/10.1016/j.compeleceng.2020.106654

Noor MHM, Salcic Z, Wang KIK (2017) Adaptive sliding window segmentation for physical activity recognition using a single triaxial accelerometer. Pervasive Mob Comput 38:41–59. https://doi.org/10.1016/j.pmcj.2016.09.009

Norgaard S, Saeedi R, Gebremedhin AH (2020) Multi-sensor time-series classification for activity tracking under variable length. IEEE Sens J 20:2701–2709. https://doi.org/10.1109/JSEN.2019.2953938

Nweke HF, Teh YW, Al-garadi MA, Alo UR (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst Appl 105:233–261

Rault T, Bouabdallah A, Challal Y, Marin F (2017) A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications. Pervasive Mob Comput 37:23–44. https://doi.org/10.1016/j.pmcj.2016.08.003

Ravi D, Wong C, Lo B, Yang GZ (2017) A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE J Biomed Heal Inform 21:56–64. https://doi.org/10.1109/JBHI.2016.2633287

Reiss A, Stricker D (2012) Introducing a new benchmarked dataset for activity monitoring. In: Proceedings International Symposium on Wearable Computers, ISWC, pp 108–109

Rivera P, Valarezo E, Choi M-T, Kim T-S (2017) Recognition of human hand activities based on a single wrist IMU using recurrent neural networks. Int J Pharma Med Biol Sci 6:114–118. https://doi.org/10.18178/ijpmbs.6.4.114-118

San-Segundo R, Montero JM, Barra-Chicote R et al (2016) Feature extraction from smartphone inertial signals for human activity segmentation. Signal Process 120:359–372. https://doi.org/10.1016/j.sigpro.2015.09.029

Shoaib M, Bosch S, Durmaz Incel O et al (2014) Fusion of smartphone motion sensors for physical activity recognition. Sensors (Switzerland) 14:10146–10176. https://doi.org/10.3390/s140610146

Subasi A, Khateeb K, Brahimi T, Sarirete A (2020) Human activity recognition using machine learning methods in a smart healthcare environment. Innovation in health informatics. Springer, Berlin, pp 123–144

Sucerquia A, López JD, Vargas-Bonilla JF (2017) SisFall: a fall and movement dataset. Sensors (Switzerland). https://doi.org/10.3390/s17010198

Thornton C, Hutter F, Hoos HH, Leyton-Brown K (2013) Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 847–855

Triboan D, Chen L, Chen F, Wang Z (2017) Semantic segmentation of real-time sensor data stream for complex activity recognition. Pers Ubiquitous Comput 21:411–425. https://doi.org/10.1007/s00779-017-1005-5

Triboan D, Chen L, Chen F, Wang Z (2019) A semantics-based approach to sensor data segmentation in real-time ACTIVITY RECOGNition. Future Gener Comput Syst 93:224–236. https://doi.org/10.1016/j.future.2018.09.055

Uddin MZ, Hassan MM, Alsanad A, Savaglio C (2020) A body sensor data fusion and deep recurrent neural network-based behavior recognition approach for robust healthcare. Inf Fusion 55:105–115. https://doi.org/10.1016/j.inffus.2019.08.004

Vaizman Y, Ellis K, Lanckriet G, Weibel N (2018) ExtraSensory app: data collection in-the-wild with rich user interface to self-report behavior. Proc CHI. https://doi.org/10.1145/3173574.3174128

Varamin AA, Abbasnejad E, Shi Q, et al (2018) Deep auto-set: a deep auto-encoder-set network for activity recognition using wearables. In: Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, pp 246–253

Vavoulas G, Chatzaki C, Malliotakis T, et al (2016) The MobiAct dataset: recognition of activities of daily living using smartphones. In: Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health, pp 143–151

Wan S, Qi L, Xu X et al (2020) Deep Learning Models for Real-time Human Activity Recognition with Smartphones. Mob Netw Appl 25:743–755. https://doi.org/10.1007/s11036-019-01445-x

Wang Y, Cang S, Yu H, Member S (2018) A data fusion-based hybrid sensory system for older people ' s daily activity and daily routine recognition. IEEE Sens J 18:6874–6888. https://doi.org/10.1109/JSEN.2018.2833745

Wang J, Chen Y, Hao S et al (2019) Deep learning for sensor-based activity recognition: a survey. Pattern Recognit Lett 119:3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang H, Zhao J, Li J et al (2020) Wearable sensor-based human activity recognition using hybrid deep learning techniques. Secur Commun Netw. https://doi.org/10.1155/2020/2132138

Ward JA, Lukowicz P, Tröster G (2006) Evaluating performance in continuous context recognition using event-driven error characterisation. Lecture notes in computer science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, pp 239–255

Ward JA, Lukowicz P, Gellersen HW (2011) Performance metrics for activity recognition. ACM Trans Intell Syst Technol. https://doi.org/10.1145/1889681.1889687

Wu D, Zhang H, Niu C et al (2019) Inertial sensor based human activity recognition via reduced kernel PCA. Internet of things. Springer, Cham, pp 447–456

Xing R, Tong H, Ji P (2014) Activity recognition with smartphone sensors. Tsinghua Sci Technol 19:235–249. https://doi.org/10.1109/TST.2014.6838194

Yao R, Lin G, Shi Q, Ranasinghe DC (2018) Efficient dense labelling of human activity sequences from wearables using fully convolutional networks. Pattern Recognit 78:252–266. https://doi.org/10.1016/j.patcog.2017.12.024

Ye J, Dobson S, Zambonelli F (2019) Lifelong learning in sensor-based human activity recognition. IEEE Pervasive Comput 18:49–58. https://doi.org/10.1109/MPRV.2019.2913933

Zebin T, Sperrin M, Peek N, Casson AJ (2018) Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks. Conf Proc Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf 2018:1–4. https://doi.org/10.1109/EMBC.2018.8513115

Zhang Y, Zhang Y, Zhang Z et al (2018) Human activity recognition based on time series analysis using U-Net. J Eng. https://doi.org/10.1155/2018/4752191

Zhao Y, Yang R, Chevalier G et al (2018) Deep residual Bidir-LSTM for human activity recognition using wearable sensors. Math Probl Eng. https://doi.org/10.1155/2018/7316954

Zhuang Z, Xue Y (2019) Sport-related human activity detection and recognition using a smartwatch. Sensors (Switzerland). https://doi.org/10.3390/s19225001