

Article

An Optimal Feature Selection Method for Human Activity Recognition Using Multimodal Sensory Data

Tazeem Haider , Muhammad Hassan Khan *  and Muhammad Shahid Farid 

Department of Computer Science, University of the Punjab, Lahore 54590, Pakistan

* Correspondence: hassankhan@pucit.edu.pk

Abstract: Recently, the research community has taken great interest in human activity recognition (HAR) due to its wide range of applications in different fields of life, including medicine, security, and gaming. The use of sensory data for HAR systems is most common because the sensory data are collected from a person's wearable device sensors, thus overcoming the privacy issues being faced in data collection through video cameras. Numerous systems have been proposed to recognize some common activities of daily living (ADLs) using different machine learning, image processing, and deep learning techniques. However, the existing techniques are computationally expensive, limited to recognizing short-term activities, or require large datasets for training purposes. Since an ADL is made up of a sequence of smaller actions, recognizing them directly from raw sensory data is challenging. In this paper, we present a computationally efficient two-level hierarchical framework for recognizing long-term (composite) activities, which does not require a very large dataset for training purposes. First, the short-term (atomic) activities are recognized from raw sensory data, and the probabilistic atomic score of each atomic activity is calculated relative to the composite activities. In the second step, the optimal features are selected based on atomic scores for each composite activity and passed to the two classification algorithms: random forest (RF) and support vector machine (SVM) due to their well-documented effectiveness for human activity recognition. The proposed method was evaluated on the publicly available CogAge dataset that contains 890 instances of 7 composite and 9700 instances of 61 atomic activities. The data were collected from eight sensors of three wearable devices: a smartphone, a smartwatch, and smart glasses. The proposed method achieved the accuracy of 96.61% and 94.1% by random forest and SVM classifiers, respectively, which shows a remarkable increase in the classification accuracy of existing HAR systems for this dataset.

Keywords: human activity recognition; activities of daily living; composite activities; atomic activities; signal processing; sensory data analysis



Citation: Haider, T.; Khan, M.H.; Farid, M.S. An Optimal Feature Selection Method for Human Activity Recognition Using Multimodal Sensory Data. *Information* **2024**, *15*, 593. <https://doi.org/10.3390/info15100593>

Academic Editor: Francesco Fontanella

Received: 5 August 2024

Revised: 12 September 2024

Accepted: 26 September 2024

Published: 29 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human activity recognition (HAR) is extensively used to monitor the daily life activities of persons in homes, especially elderly people, to support healthy and active aging. Such systems are being used for other medical or security purposes. Previously, video cameras were being used for data collection of these systems, but most people are not comfortable with the cameras, as everyone is concerned about their privacy. Moreover, the installation and maintenance cost of video cameras is also non-negligible [1]. With the advancement of technology, cheaper and highly accurate motion sensors in mobile phones and wearable devices, e.g., smartwatches and smart glasses, are now being used to monitor the activities of daily living (ADL) [2,3]. These wearable devices use different sensors, including accelerometer, gyroscope, heart rate, barometer, and magnetometer. The HAR systems are mainly designed to detect two types of activities: atomic and composite. Atomic activities are short-term smaller activities like sitting, standing, bending, walking, etc. These activities can be recognized from raw sensory data or smaller video frames directly. Composite activities are long-term activities, usually composed of two or more

atomic activities. For example, the activity of washing hands is a composite activity that is composed of different atomic activities like opening tap water, applying soap, rubbing hands, closing tap water, etc. [4].

It is challenging to use raw sensory data for the recognition of composite activities directly [5] because, if multiple examples of the same composite activity are performed by the same person but with a change in the order of underlying atomic activities, the sensory data stream will be different. In general, the meaningful features from raw sensory data are first extracted, and then using those features, the composite activities are recognized. As an atomic activity is of short duration of time, it can be easily recognized from sensory data. We can say that in such cases, we can use atomic activities as meaningful features for composite activity recognition because the diversity of sensory data is minimal in terms of atomic activities. Thus, each example of the same composite activity produces different raw data but, in terms of atomic activities, it is considered as the same [6].

HAR systems are usually designed to follow a standard sequence of pattern recognition to recognize the composite activities: data collection, pre-processing, feature extraction, selection of meaningful features, and classification in the respective classes [7,8], as shown in Figure 1. These systems use different machine and deep learning techniques for feature extraction and selection of related features. In addition to these techniques, ontology-based learning is also used for activity recognition and reasoning systems.

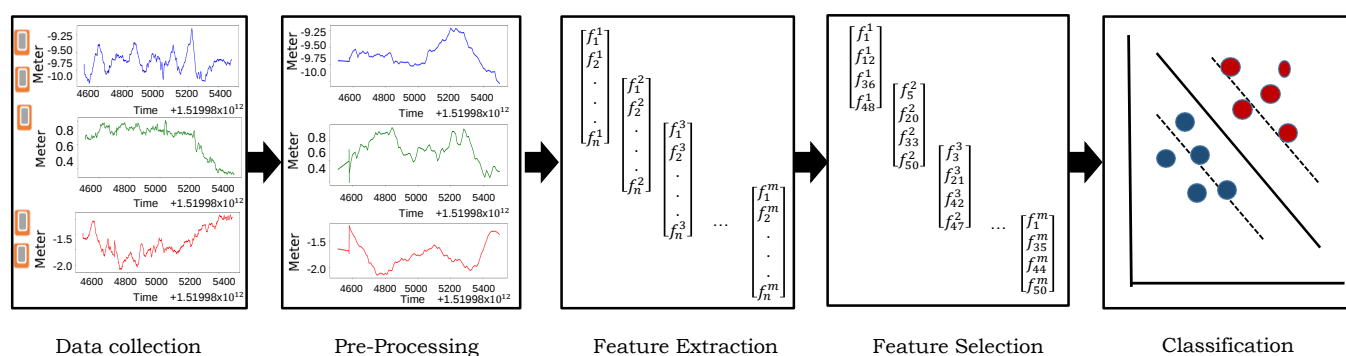


Figure 1. Standard sequence of steps for human activity recognition from raw sensory data: The first two boxes in the figure depict the sensory signals, where the x-axis represents the time and the y-axis represents the information provided by the respective sensor, e.g., accelerometers measure a changing acceleration (in meters) on the sensor, gyroscopes measure changing angular motion, etc.

Feature encoding and optimal feature selection are important parts of HAR systems. The techniques used for feature encoding of HAR can be classified into three categories. Some studies have used the handcrafted features [9,10] that are manually calculated as statistical measures (average, variance, skewness) from the raw data and passed to the machine learning classifiers as input [11–14]. Codebook-based approaches are also used for feature encoding in various HAR systems. In these approaches, the input data are divided into different clusters using clustering algorithms like k-means. These clusters act as a dictionary, with the center of each cluster as a codeword. These codewords are then used to create a histogram-based feature vector for the classification of activities [15,16]. Deep learning-based approaches are also utilized for the classification of human activities. These techniques use automatic feature extraction for the classification model. Convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM), and many hybrid deep learning models are commonly used [17,18]. The existing feature encoding techniques have some limitations. For example, the handcrafted-based techniques are time-consuming and require more manual effort, which leads to the possibility of human error. However, these techniques can give better results on smaller datasets. The deep learning-based techniques have proven to give better results, but they require extensive data for efficient performance and are computationally more expensive. Codebook-based techniques provide a compact representation of data, which reduces the dimensionality.

Moreover, these are less error-prone than handcrafted techniques, and they are also less complex and computationally intensive than deep learning-based techniques. However, if we add more classes in the dataset, we need to recompute codebooks. Keeping in mind these limitations, we have proposed a method using a codebook based-approach for feature encoding, which can perform efficiently in less time and on a limited dataset.

In this paper, we present a generic framework for the recognition of human ADLs. Since human ADLs are long-term activities consisting of several repetitive short series of actions, the recognition of the composite activities directly from raw sensory data is challenging because of the diversity of data in two or more sequences of the same activity. However, there can be some similarities in the temporal sequence of atomic activities. Therefore, to address this problem, we propose a two-level hierarchical method to recognize the composite activities from the sensory data. In the first step, the atomic activities are recognized from raw sensory data, then the composite activities are recognized in the second step. Initially, the low-level features are calculated from raw sensory data using the basic bag-of-features (BOF) codebook approach. These features are passed to the SVM for atomic activities classification; the SVM is designed to produce results ranging from 0 to 1 for each atomic activity. These values are named as *atomic scores* for each atomic activity relative to the specific composite activity. Based on the percentage of these atomic scores, the optimal features are extracted for the classification of composite activities and are passed to the classifiers: SVM and RF. We used the CogAge dataset for the evaluation of the proposed method. The dataset contains seven composite activities collected from six individuals wearing smart glasses, smartwatches, and a smartphone placed in their pockets. Each composite activity is represented by the combination of 61 atomic activities (i.e., 61-dimensional feature vector). The major contributions of this study include:

- A generic framework is presented to recognize complex, long-term human activities of daily living (i.e., composite activities) using a two-level hierarchical approach to address the challenge of recognizing composite activities from raw sensory data;
- An optimal feature learning technique is presented that uses atomic scores of the underlying atomic activities in a composite activity;
- The proposed method was evaluated on the CogAge dataset, demonstrating a significant improvement in accuracy compared to existing methods, which highlights the effectiveness of the proposed framework.

2. Related Work

Activity recognition has gained the attention of researchers in the last few decades for its applications in security, medicine, rehabilitation, and gaming. Previously, the study was limited to recognizing simple actions (atomic activities) like sitting, standing, walking, etc. However, human daily life activities are long-term activities usually performed hierarchically. Therefore, the recognition of human long-term (composite) activities is a hierarchical task. Recently, different studies have been conducted to recognize both shorter and long-term activities of daily living using different techniques. The existing HAR methods can be divided into three groups based on their implementation technique: handcrafted feature-based techniques, codebook-based techniques, and deep learning-based techniques, as shown in Figure 2. A brief discussion of each technique is given in the following subsections.

2.1. Handcrafted Feature-Based Techniques

In this domain, the researchers compute the desired features and fuse them to make a feature vector, which is then fed to any classification algorithm. Handcrafted features are usually the simplest statistical measures, like mean, median, standard deviations, etc., or some complex features, like power spectral density, Fourier transform, etc. These features, also known as manual features, are calculated directly from the raw sensory data after some pre-processing and can be used directly or as an input to the machine learning models for HAR systems [19–22]. Different studies have used handcrafted features for HAR with deep

and shallow models [23–27]. Some of the commonly used handcrafted features for HAR systems are shown in Table 1.

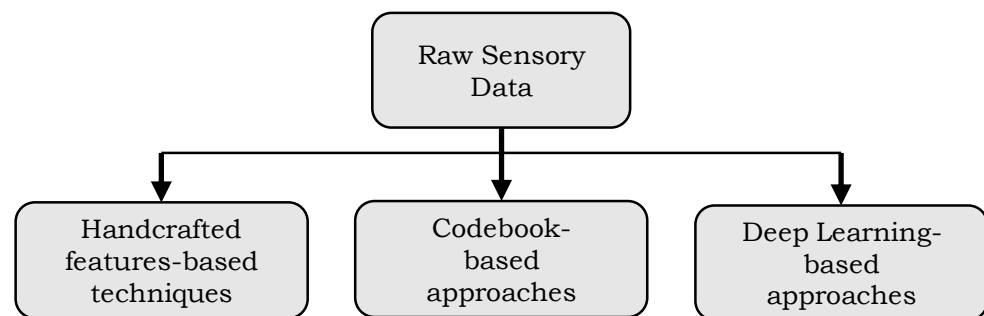


Figure 2. Classification of existing features encoding techniques for HAR systems.

Table 1. Commonly used handcrafted features in different HAR studies. These features are mostly the simplest features that are calculated directly from raw sensory data and used as an input of any classification algorithm.

Feature	Formula
Minimum	$\min = \min_{j=1,\dots,n}(x_j)$
Maximum	$\max = \max_{j=1,\dots,n}(x_j)$
Mean	$\bar{x} = \frac{1}{n} \sum_{x=1}^n x_i$
Median	$\text{Me} = x_{0.5}$
Standard Deviation	$s = \sqrt{\frac{1}{n-1} \sum_{x=1}^n (x_i - \bar{x})^2}$
Variance	$s^2 = \frac{1}{n-1} \sum_{x=1}^n (x_i - \bar{x})^2$
Skewness	$S = \frac{\frac{1}{n} \sum_{x=1}^n (x_i - \bar{x})^3}{s^3}$
Kurtosis	$K = \frac{\frac{1}{n} \sum_{x=1}^n (x_i - \bar{x})^4}{s^4}$
Root Mean Square	$\text{RMS} = \sqrt{\frac{1}{n} \sum_{x=1}^n (x_i)^2}$

The authors in [28] calculated the statistical handcrafted features in both frequency and time domains (including minimum, maximum, standard deviation, skewness, kurtosis, weighted average, etc.) from the sensory data of mobile phone sensors. These features represent the properties of different activities in two domains. They passed these handcrafted features and raw data to the shallow and deep models. The handcrafted features with shallow models gave the best results for their data. The authors in [7] used a two-level hierarchical method for the recognition of human composite activities. In the first step, the atomic activities were recognized using seventeen handcrafted features, including mean, median, first-order norm, skewness, etc. In the second step, these handcrafted features were used to recognize the composite activities.

In [21], the authors investigated a combination of handcrafted features with traditional classifiers and compared their results with deep learning-based models on four datasets. A total of 63 handcrafted features (including mean, standard deviation, interquartile difference, skewness, etc.) were computed, listing 21 features in each of the x , y , and z axes. There was not a remarkable difference in the results of handcrafted features with machine learning algorithms and deep learning-based models. However, the results of the use of raw data were very low as compared to both. In [29], the authors proposed the orthodox aspects of handcrafted features for HAR. They computed vector point (VP) and absolute distance (AD), which reduced the data size three to six times, thus reducing the time complexity and increasing the accuracy. Handcrafted-based features are efficient in computations, can easily be calculated, and are also effective when we have limited labeled data. However, computing handcrafted features may be time-consuming and require manual effort where human error is possible [16,30]. Moreover, we require a re-computation of features if more classes are added to the training.

2.2. Codebook-Based Feature Extraction

These techniques convert the raw sensory data into compact high-level features by following the channel of the bag-of-features (BOF) approach [31]. In such techniques initially, the data are clustered using any clustering algorithm (like k-means). These clusters form a codebook, and the center of each codebook is a codeword. Using these codewords, a histogram is constructed that represents the data as compact high-level feature vectors. These feature vectors are used as an input of any classification algorithm [32].

The authors in [33] used a codebook-based approach in combination with template matching to increase the computational and space efficiency in human locomotion activity recognition (HLAR). In another study [34], the authors proposed a codebook-based approach using k-means clustering for the assessment of human movements of different physical activities. A hierarchical method for HAR systems was proposed in [6]. First, the atomic activities were detected using a codebook-based approach and their recognition score for each composite activity was calculated. In the second step, they encoded the atomic scores for atomic activities using the temporal pooling method (rank pooling) for composite activity recognition. Other studies using codebook-based approaches for feature encoding in the HAR systems include [31,35–37].

Codebook-based techniques represent the data as compact feature sets by reducing the dimensionality of the feature space. These techniques specifically retain the relevant features and remove noise by ignoring the irrelevant features, thus highlighting the important ones. However, important data may be lost by dimensionality reduction in such techniques. The classification is mainly dependent on the quality of the codebook, so a poor choice of codewords may lead to lower accuracy. Moreover, such techniques can be computationally more expensive when we have to manage large codebooks and require additional computation if we add new classes to the dataset [38].

2.3. Deep Learning-Based Techniques

Deep learning models have had remarkable performance in wearable human activity recognition. Such models are end-to-end networks that can learn complicated patterns from raw data directly using a fully automated feature extraction method. Deep learning models, including multilayer perceptron and artificial neural networks, are also being used in HAR systems [39–43]. Other deep learning models used for HAR include convolutional neural networks (CNN) [44], recurrent neural networks RNN [45], hybrid models [46,47], and gated recurrent units (GRU) [48]. Auto-associative learning modules, generally called auto-encoders, are commonly used for feature learning as unsupervised algorithms. Different studies used the features extracted from auto-encoders as an input of classifiers in activity recognition systems to improve the performance [49–53]. As auto-encoders do not require annotations, so many activity recognition systems have used autoencoders for dimensionality reduction of the very large HAR datasets [50,54–61].

The authors in [62] proposed a temporal convolutional hourglass auto-encoder (TCHA) in which they learned both spatial and temporal features from videos. They evaluated the proposed method on their own compiled dataset and also on some state-of-the-art datasets and claimed a good average accuracy for all the datasets. The authors in [63] used a deep convolutional neural network with a hierarchical-split (HS) model to increase the performance of activity recognition. The idea of the HS module gives better results as compared to the basic modules with the same model complexity.

Yang Li et al. [64] proposed a method for sensor data contribution significance analysis (CSA) for HARs based on sensor status. In this method, they measured the contribution of different sensors to the recognition of a particular behavior. Then, the spatial distance of the matrix was created to enhance context awareness and minimize data noise, thus proposing a HAR algorithm based on a wide time-domain convolutional neural network and multi-environment sensor data (HAR-WCNN) for daily behavior recognition. The authors in [47] used five different hybrid deep learning models and proposed a multi-layer hybrid LSTM-GRU model for human activity recognition using sensory data. The model

was evaluated extensively on two different datasets. Many studies have been made for activity recognition using RNN and LSTM models [65–70] in HAR systems. As a composite activity is a long-term activity, these networks outperform for HAR systems.

A recent study in [71] proposed a one-dimensional (1D) CNN using a deep learning approach for human activity recognition. In addition to activity recognition, it detected activity-related anomalies in the daily lives of individuals. The authors in [72] proposed a two-step hierarchical system to detect suspicious activities. In the first step, a two-dimensional (2D) CNN model is used to extract important characteristics from input data, while an LSTM is trained in the second step to predict whether the activity is suspicious. The authors in [73] proposed a WiFi-based HAR using a deep learning approach. They used a recurrent neural network (RNN) with an LSTM model to classify indoor activity sequences.

In addition to simple deep learning models, several recent studies have proposed ensemble deep learning models for activity recognition using sensory data. For example, the authors in [74] addressed the inadequacy problem by evaluating fifteen data augmentation techniques. Additionally, they recognized human ADLs by proposing a multi-branch hybrid Conv-LSTM network for automatic feature learning. They assessed their ensemble model on two datasets, achieving better results. Another study [75] proposed deep neural networks with the combination of a set of convolutional layers and capsule network (DCapsNet) to automatically extract the activity features for HAR or gait recognition. The authors in [76] combined three deep learning models, including a bidirectional long-short-term memory, a convolutional neural network, and a bidirectional gated recurrent unit, for activity recognition directly from raw sensory data. The method can efficiently recognize both long-term and short-term activities from unprocessed data. In another study [77], the authors proposed a time convolution network recognition model with attention mechanism (TCN-Attention-HAR) for human activity recognition. They used the TCN for extracting temporal features and attention model to assign higher weights to important features, claiming to get excellent recognition performance compared to the other advanced models.

Although the deep learning-based models perform best on various datasets, it is challenging to rigorously assess their performance for feature learning methods. In deep learning-based models, the whole process becomes highly complex due to many factors, including the lack of knowledge [8,78], the selection of appropriate network architecture [79,80], and determining the appropriate parameter values [7,81,82] to extract the meaningful features for accurately representing the activities. Despite the fact that deep learning techniques can autonomously extract important features from raw sensory data, standalone deep learning models are not sufficient to capture the in-depth association between the patterns [47,83]. To this end, the ensemble deep learning models have proven to be effective to obtain better results as compared to the traditional deep learning models; however, they are computationally complex and increase the model's complexity, requiring extensive data. Moreover, the hyper-parameter tuning is a challenging task in such models [84]. Although the deep learning-based models require a lot of data for training purpose to get better results [85,86], this limitation can be overcome by using the pre-trained models or data augmentation techniques.

The existing HAR systems are either limited to short-term activities or their performance is not very effective. A system is required for the effective recognition of human activities of daily living (ADLs) for handling real-world uncertainty. Optimal feature extraction is very important in such systems. However, existing feature extraction techniques have some limitations: handcrafted-based techniques are not very reliable, as they are manually extracted and the possibility of human error is higher in such techniques; codebook-based and deep learning-based techniques are computationally expensive. Therefore, a computationally effective method for feature extraction is also required for long-term activity recognition. All these factors encourage us to propose a technique for the recognition of ADLs with an effective feature extraction method.

3. Proposed Method

Because composite activities are long-term activities consisting of short sequences of small actions, traditional methods struggle to recognize them directly from raw sensory data. For example, if we take multiple examples of the same composite activity performed by the same individual but with a change in the order of underlying atomic activities, the stream of sensory data will be different. However, we can observe a similarity in terms of the data stream of atomic activities. Although deep learning-based techniques have shown the capability to recognize these composite activities directly from raw data, they require extensive data for training purposes to get better results. Considering these limitations, we propose a two-level hierarchical method. In the first level, the short-term (atomic) activities are recognized from raw sensory data, and the atomic score of each atomic activity is calculated. Second, the optimal features are selected based on the percentage of atomic score. These features are then passed to the classifiers for recognition of composite activities. The overview of the proposed method is described in Figure 3. The implementation details of the proposed methodology are provided in the following subsections.

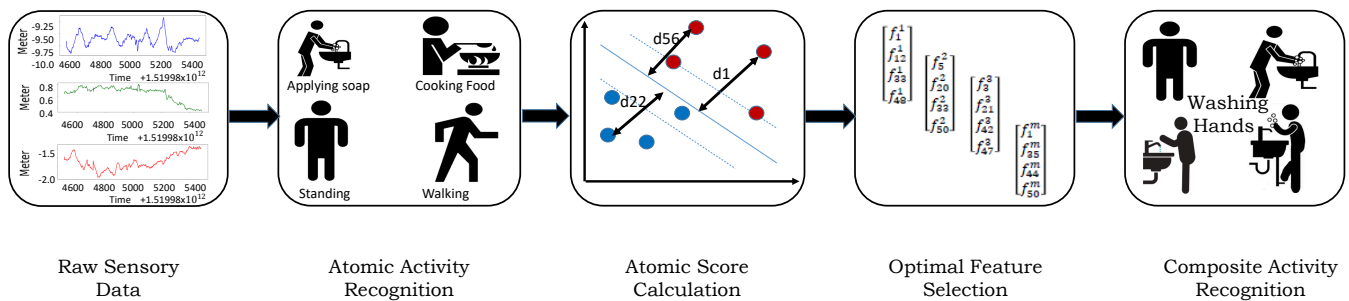


Figure 3. The proposed method works in a two-level hierarchical manner: First, the atomic activities are recognized directly from raw sensory data and the atomic score of each atomic activity is calculated. In the second step, the optimal features are selected on the basis of atomic score percentage and fed to the classifiers for composite activities recognition. The first box in the figure depicts the sensory signals, where the x -axis represents the time and the y -axis represents the information provided by the respective sensor, e.g., accelerometers measure a changing acceleration (in meters) on the sensor, gyroscopes measure changing angular motion, etc.

3.1. Atomic Activities Recognition

Human ADLs are long-term activities that consist of a sequence of small actions (atomic activities). Accurate recognition of the atomic activities is very important, as these results serve as a baseline for our next steps. There are a total of 61 atomic activities in the CogAge dataset, including state and behavioral activities like sitting, standing, cutting, opening a tap, etc. Therefore, each composite activity is represented by a combination of the 61-dimensional feature vector in our dataset. The raw sensory data are processed to get feature vectors of atomic activities. For the sensory data, let us first clarify some terms typically used. When we say *time-series data*, we refer to the whole data used for training and testing, but when we say a *sequence*, it may be a segment of the data. For example, the data recorded from different individuals in 24 h can be referred to as time series data and a small window of data is the sequence. The *subsequence* refers to the smaller window within a sequence (see Figure 4). In this step, our objective is to predict the labels of a sequence. For this purpose, we need a feature vector for each sequence as a robust representation of the data. Using the distribution of each subsequence, the feature vectors of sequences are constructed by the codebook approach.

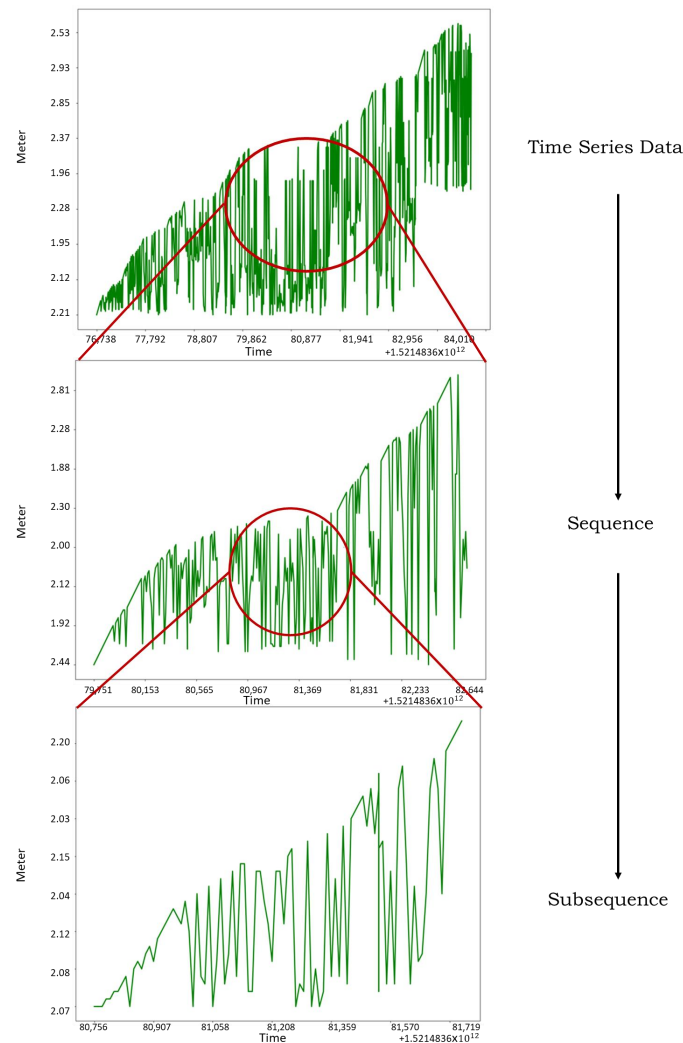


Figure 4. An illustration of decomposition of time series data into sequences and subsequences. The first two boxes in the figure depict the sensory signals, where the x -axis represents the time and the y -axis represents the information provided by the respective sensor, e.g., accelerometers measure a changing acceleration (in meters) on the sensor, gyroscopes measure changing angular motion, etc.

The atomic activity recognition method comprises two main phases: feature extraction and model training. In the feature extraction phase, a feature vector representing the characteristics of each sensor's corresponding sequence is extracted. Subsequently, the features of all the sensors are fused to make a high-dimensional feature vector representing each composite activity. This phase uses a codebook approach, in which subsequences are selected using a sliding window then, using k-mean clustering, a *codebook* is constructed. The center of each codebook is a *codeword*. After the construction of the codebook, codeword assignment takes place, in which each subsequence is assigned to the most similar codeword, thus obtaining the features of each subsequence. The codebook-based approach is explained in Figure 5. Finally, the features extracted from sequences for each sensor are fused using an early fusion approach (as discussed in [87]), where features for S sensors are simply concatenated into a high-dimensional feature. The construction of the codebook-based feature vector is explained below.

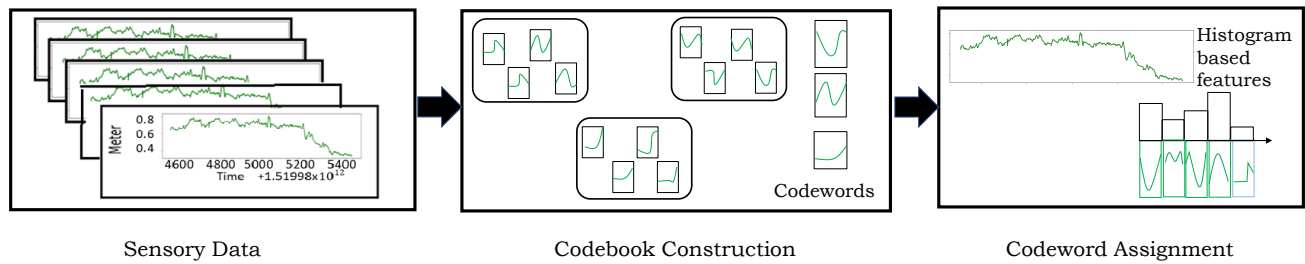


Figure 5. Codebook-based feature computation process. The codebook is constructed by grouping similar subsequences using a k-means algorithm. The center of each codebook is a codeword. The features are computed by assigning each subsequence to the most similar codeword. The first two boxes in the figure depict the sensory signals, where the x-axis represents the time and the y-axis represents the information provided by the respective sensor, e.g., accelerometers measure a changing acceleration (in meters) on the sensor, gyroscopes measure changing angular motion, etc.

3.1.1. Codebook Construction

The first step of codebook construction is to identify suitable codewords. From the training data, subsequences of the same length are extracted. These subsequences can be defined as the sliding windows of size w that are shifted from start to end of the dataset. The window is positioned at every l th time point, where a smaller l results in a large overlap between neighboring subsequences. Note that the window size w and sliding size l are the hyper-parameters, so they must be selected carefully. However, we can choose w independently from the expected duration of activities, as the final classification will be based on the histogram of codewords within a given time interval.

The next step involves the grouping of subsequences into N different clusters. We have used k-mean clustering for identifying the N clusters of similar subsequences. We measured the distance between two subsequences using the Euclidean distance. We can define it mathematically as

$$J = \sum_{i=1}^N \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (1)$$

where N is the number of clusters, S_i is the set of subsequences assigned to the i th cluster, x represents a subsequence, and μ_i is the centroid of the i th cluster. The Euclidean distance $\|x - \mu_i\|^2$ measures the distance between a subsequence x and the cluster centroid μ_i . In this way, a codebook consisting of N clusters is generated, where the center of each cluster is the codeword.

3.1.2. Codeword Assignment

After the construction of the codebook, a histogram is constructed representing the distribution of N codewords for conversion of a sequence into a feature representation using the simplest bag-of-features (BOF) approach due to its effectiveness in the action recognition and image classification [31]. Each subsequence is assigned to a similar codeword and its frequency is incremented. Subsequently, the histogram of codewords is normalized so that the sum of the frequency of N codewords is equal to one. This assignment approach is known as hard assignment.

3.1.3. Model Training

We trained a one-vs.-all classifier to distinguish between training sequences with labels of target activity and of other activities. As the dimensions of the features are higher, we used SVM for its effectiveness in handling the high dimensional feature vectors [88]. The SVM positions the boundary at an equal distance from positive and negative sequences to draw a classification boundary using the margin maximization principle. For recognition of N activities, N SVMs are trained as binary classifiers, with one SVM for each activity. Among all these classifiers, we used SVM with the highest score for the recognition of

test sequences. For all these SVMs, a radial basis function (RBF) kernel is used that has only a single parameter γ to control the complexity of the classification boundary. It is observed from previous studies [89] that if we set γ as mean of squared Euclidean distances among all training sequences, it gives a stable and computationally cheaper performance. Additionally, the value of hyper-parameter C to controls the penalty of misclassification is set to 2 because of the effectiveness of this setting for activity recognition tasks as mentioned in [90]. The decision function used in SVM with an RBF kernel is the same as in Equation (6) described in the next subsection.

In addition to classification, the SVM is trained to output a scoring value between 0 and 1, indicating the likelihood of the example containing the atomic activity. These values are actually the recognition scores (known as probabilistic atomic scores in our study), which play a major role in the selection of optimal features.

3.2. Feature Selection for Composite Activities

A composite activity is made up of a sequence of atomic activities, and each atomic activity is associated with a probabilistic atomic score. For example, in the “Brushing Teeth” composite activity, the atomic score of “Standing” is higher, but the scores of “Lying”, “Clean Floor”, or “Clean Surface” are much lower. Conversely, in the “Cleaning Room” composite activity, The atomic score of “Clean Floor” and “Clean Surface” will be higher than all the other atomic activities. To handle this uncertainty, the atomic activities are selected on the percentage of their participation (atomic score) in particular composite activities as best features. For example, in Brushing Teeth, out of 61 atomic activities, 24 activities have a greater than 90% atomic score, 14 activities have a greater than 80% score, and only 8 activities have a score greater than 70% of all atomic activities score. Therefore, instead of taking all 61 atomic activities as features, we can set any threshold on the percentage participation of the atomic activities and use only those atomic activities as best features to train the model for that specific composite activity as they participate more in recognition of that composite activity. The activities with their percentage participation are described in Table 2.

Table 2. The count of atomic activities in composite activities across different percentages of atomic scores.

Composite Activity Name	Atomic Activities Count in		
	90%	80%	70%
Brushing Teeth	24	14	8
Cleaning Room	20	11	8
Handling Medication	25	15	9
Preparing Food	26	15	9
Styling Hair	17	9	6
Using Phone	18	10	7
Washing Hands	16	9	6

If we use all 61 atomic activities as features, this will negatively impact the training, as some activities have very low atomic scores, i.e., nearly equal to 0, and it also slows down the training of the classifier. It means that these atomic activities do not participate in the performance of that composite activity. We extracted those atomic activities and eliminated them from the training of that composite activity. Similarly, using 80% or fewer activities as features can increase the chance to lose important data, as the number of activities is too low in these percentages. Here, we have set a threshold of 90% participation of atomic activities because it contains a sufficient number of atomic activities for each composite activity. We can define it as:

$$A_{\text{selected}} = \{a_i \in A \mid \sum_{i=1}^{n_a} s_i \geq 0.90\} \quad (2)$$

where A_{selected} is the set of selected atomic activities, a_i is an atomic activity, A is the set of all 61 atomic activities, and s_i is the atomic score of the i th atomic activity. Then, the feature vector F_c is created for each composite activity c_i .

$$F_c = [F_1, F_2, \dots, F_n] \quad (3)$$

where n is the total count of atomic activities. Moreover, each feature F_i is defined as

$$F_i = \begin{cases} 1 & \text{if } a_i \in A_{\text{selected}} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In this way, the features were selected for each composite activity and their union was used as combined features for multi-class classification.

$$F_{\text{combined}} = \bigcup_{c \in C} F_c \quad (5)$$

where F_{combined} is the selected combined feature vector for composite activities classification, and F_c is the feature vector of a composite activity.

3.3. Composite Activity Recognition

The classification of composite activities was performed using the selected features. We used two classifiers for this purpose—support vector machine and random forest—due to their well-documented effectiveness for human activity recognition. Moreover, these classifiers are more robust to the high dimensional features and can effectively handle large datasets with complex decision boundaries by avoiding overfitting. A detailed discussion of both classifiers is given below.

3.3.1. Support Vector Machine

SVM is a powerful machine-learning algorithm for classification purposes. It is a commonly used algorithm, especially for human activity recognition, due to the following advantages: it performs efficiently on high dimensional feature vectors, it is less prone to overfitting, it performs well on less or medium-sized data, and is flexible for different kernel functions. SVM is commonly used for binary classification, but we can also perform multi-class classification with one-vs.-one or one-vs.-all strategies. The classification function used for the SVM classifier can be explained as:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(x, x_i) + b \right) \quad (6)$$

where $f(x)$ is the label assigned by SVM for input x , α_i is dual coefficient, y_i is the class label, $K(x, x_i)$ is the kernel function that computes the correlation between x and support vectors x_i , and b is the intercept term.

For our experiments, we used simple linear SVM. After training, the testing was performed for composite activities classification.

3.3.2. Random Forest

Random forest is also a very powerful classifier that works by making multiple decision trees. It gives higher accuracy than other machine learning algorithms that use a single decision tree. RF can also handle complex non-linear relationships between features and targets [91,92]. The algorithm partitions the feature vector recursively for training purposes and makes different decision trees. The classification function can be described as:

$$f(x) = \text{mode}(T_1(x), T_2(x), \dots, T_N(x)) \quad (7)$$

where $f(x)$ is the prediction made by RF for input x and $T_i(x)$ is the prediction of the i th decision tree for input x .

In addition to SVM, we also used RF for classification purposes and obtained a remarkable increase in testing accuracy. The results of both classifiers are discussed in detail in Section 4.2.

3.4. Experimental Setup

The proposed method is a two-level hierarchy. In the first step, we recognize the atomic activities from raw data by using the simplest BOF approach. In the second step, the composite activities are recognized on the basis of these atomic activities using two classifiers: SVM and RF. The method was evaluated on the publicly available CogAge dataset containing 61 atomic and 7 composite activities. The dataset contains a total of 890 instances of composite activities performed by different individuals using either left, right, or both hands. The experiments were performed on activity data that was done using both hands (i.e., the smartwatch was placed on the subject's left or right arm). The experimental settings of both classifiers are described in Table 3. All the experiments were performed on a machine with an Intel i5-4310U CPU with 16 GB RAM (Dell, Siegen, Germany).

Table 3. Experimental settings for support vector machine and random forest classifiers used in this study.

Support Vector Machine	Random Forest
Cross-validation: 5-folds	Cross-validation: 5-folds
Random State: 42	Random State: 42
Kernel: Linear	Number of Trees: 200
	Maximum Depth: 10

4. Results and Discussion

4.1. Dataset

We used the cognitive village (CogAge) dataset introduced in [6]. The dataset was collected for 7 composite and 61 atomic activities from wearable devices including the LG G5 smartphone, Huawei smartwatch, and JINS MEME glasses.

From these devices, data were obtained from the following eight sensors:

1. Smartphone's accelerometer (sp-acc): Provides measurements of acceleration (including gravity) in meters per second squared (m/s^2) across three dimensions x , y , and z ;
2. Smartphone's gyroscope (sp-gyro): Provides angular velocity measurement in radians per second (rad/s) across three dimensions x , y , and z ;
3. Smartphone's gravity (sp-grav): Provides the sequence of gravity force in meters per second squared (m/s^2) across three dimensions x , y , and z ;
4. Smart-phone's linear accelerometer (sp-linacc): Provides measurements of acceleration (excluding gravity) in meters per second squared (m/s^2) across three dimensions x , y , and z ;
5. Smartphone's magnetometer (sp-mag): Provides intensities of earth's magnetic field in tesla (t) across three dimensions x , y , and z , which help in determining the orientation of the mobile phone;
6. Smartwatch's accelerometer (sw-acc): Provides measurements of acceleration (including gravity) in meters per second squared (m/s^2) across three dimensions x , y , and z ;
7. Smartwatch's gyroscope (sw-gyro): Provides angular velocity measurement in radians per second (rad/s) across three dimensions x , y , and z ;
8. Smart glasses' accelerometer (sg-acc): Provides sequence of acceleration forces in meters per second squared (m/s^2) across three dimensions x , y , and z .

The sample data of a few sensors from the dataset are shown in Table 4. The data are collected at different timestamps across three axes: x , y , and z . A complete description of the dataset can be found in [6].

Table 4. Sample data of a few sensors collected from an individual for Brushing Teeth activity. The data are collected at different timestamps across the x, y, and z coordinates. The smartphone's accelerometer and gravity sensors provide measurements in meters per second squared (m/s^2), the gyroscope provides angular velocity measurement in radians per second (rad/s), and magnetometer provides intensities of earth's magnetic field in tesla across three dimensions x, y, and z.

Sensor	Timestamp	x-Axis	y-Axis	z-Axis
Smartphone Accelerometer (m/s^2)	1522061483421	0.5344391	3.7631073	9.24202
	1522061483425	0.54881287	3.7607117	9.191742
	1522061483430	0.57754517	3.7559204	9.160614
Smartphone Gyroscope (rad/s)	1521205075060	0.89419556	−0.5268707	−0.1633606
	1521205075065	1.170105	0.0015106201	−0.18891907
	1521205075070	1.074234	−0.015533447	−0.24751282
Smartphone Magnetometer (tesla)	1521205075001	−78.93925	−21.0886	−95.36343
	1521205075011	−79.0144	−21.200943	−95.100975
	1521205075021	−78.97682	−21.575928	−95.0634
Smartphone Gravity (m/s^2)	1521205075001	5.426985	7.1766686	3.9004674
	1521205075006	5.4208546	7.198336	3.8689373
	1521205075011	5.416293	7.219007	3.836677

The data of 61 atomic activities (including state and behavioral activities) were collected from 8 individuals, creating a total of 9700 instances. The data for 7 composite activities were collected from 6 individuals, creating 890 instances, which are further split into 447 training and 443 testing instances. The total duration of composite activities varies from 30 s to 5 min because some activities, like cleaning the room or preparing food, take more time, but handling medication takes a shorter time to complete. The dataset details are given in Table 5.

Table 5. The summary of the cognitive village (CogAge) dataset. The dataset contains two types of activities: atomic and composite. Instances of both activities were performed using either one hand or both hands.

Dataset	CogAge
No. of Atomic Activities	61
No. of Composite Activities	7
No. of Participants for Atomic Activities	8
No. of Participants for Composite Activities	6
Instances of Atomic Activities	9700
Instances of Composite Activities	890
Devices and Their Placement	Smart glasses (eyes) Smartwatch (wrist) Smartphone (waist)
Sensors (unit)	8 Smartphone accelerometer (m/s^2) Smartphone gyroscope (radian/s) Smartphone gravity (m/s^2) Smartphone linear accelerometer (m/s^2) Smartphone magnetometer (tesla) Smartwatch accelerometer (m/s^2) Smartwatch gyroscope (radian/s) Smart glasses accelerometer (m/s^2)
Composite Activity Feature Representation	61

Training and testing data were collected separately on different days by each individual. Table 6 shows the composite activities and the number of instances for each composite activity in the dataset. Atomic activities are listed in Table 7.

Table 6. List of composite activities and number of instances of each activity in the CogAge dataset. A total of 890 instances were collected from 6 individuals using either only one hand or both hands.

Sr. No	Activity Name	Example Points
1.	Brushing Teeth	128
2.	Cleaning Room	129
3.	Handling Medication	137
4.	Preparing Food	113
5.	Styling Hair	127
6.	Using Phone	128
7.	Washing Hands	128

Table 7. List of atomic activities in the CogAge dataset. Atomic activities include 6 state activities and 51 behavioral activities. A total of 9700 instances of atomic activities were collected from 8 participants in this dataset.

Bending	Lying	Sitting	Squatting
Standing	Walking	Bring	CleanFloor
CleanSurface	CloseBigBox	CloseDoor	CloseDrawer
CloseLidByRotate	CloseOtherLid	CloseSmallBox	Drink
DryOffHand	DryOffHandByShake	EatSmall	Gargle
GettingUp	Hang	LyingDown	OpenBag
OpenBigBox	OpenDoor	OpenDrawer	OpenLidByRotate
OpenOtherLid	OpenSmallBox	PlugIn	PressByGrasp
PressFromTop	PressSwitch	PutFromBottle	PutFromTapWater
PutOnFloor	PutHighPosition	Read	Rotate
RubHands	ScoopPut	SittingDown	SquattingDown
StandingUp	StandUpFromSquatting	TakeFromFloor	TakeFromHighPosition
TakeOffJacket	TakeOut	TalkByTelephone	ThrowOut
ThrowOutWater	TouchSmartPhoneScreen	Type	Unhang
Unplug	WearJacket	Write	CloseTapWater
OpenTapWater			

4.2. Results

We used the features computed by our algorithm for the composite activities classification. We performed two types of experiments. In the first experiment, seven one-vs.-all SVM classifiers were trained, one for each composite activity. For this purpose, a feature vector of each composite activity c_i was calculated based on the probabilistic atomic score of underlying atomic activities. These probabilistic atomic scores can also be referred to as the percentage participation of each atomic activity in the performance of a specific composite activity.

Here, the threshold setting on the percentage of atomic scores is a crucial task because if we set a low threshold, we can lose the important features. Conversely, if we use all the features, it will negatively impact our classifier, as some atomic activities do not participate in the performance of a specific composite activity, like “Lying” in “Preparing Food” composite activity, as discussed earlier in Section 3.2. Therefore, we have set a threshold of 90% participation in the performance of composite activities, as it retains the important features and also seems a good choice in terms of the size of the feature vector. The selected features for each composite activity are shown in Table 8. These selected features were then passed to the SVM classifier for the one-vs.-all classification of each of the seven composite activities.

Table 8. Features used for one-vs.-all classification of composite activities. These features are selected based on 90% participation of atomic activities in the performance of a composite activity.

Sr. No	Composite Activity	No. of Selected Features
1.	Brushing Teeth (BT)	24
2.	Cleaning Room (CR)	20
3.	Handling Medication (HM)	25
4.	Preparing Food (PF)	26
5.	Styling Hair (SH)	17
6.	Using Phone (UP)	18
7.	Washing Hands (WH)	16

The results of one-vs.-all classification for all seven composite activities are shown in Figure 6.

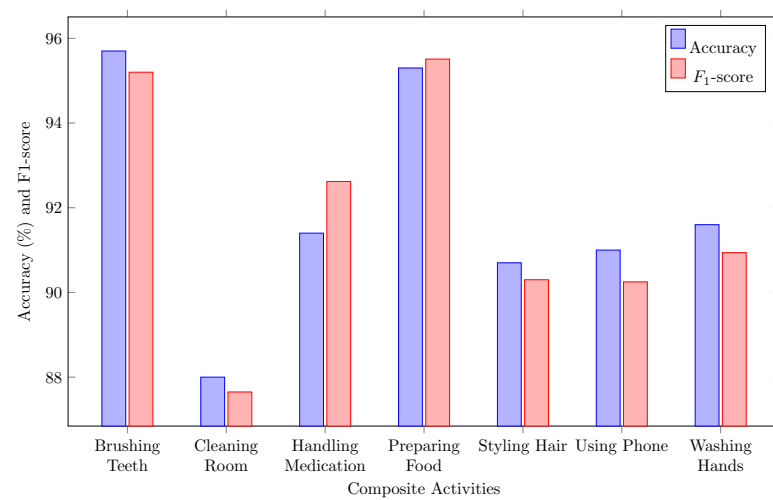


Figure 6. Results of one-vs.-all classification of composite activities. The performance of the proposed method is measured using two matrices, namely, accuracy and F_1 score.

These results were compared with the results computed in [6] for one-vs.-all classification of composite activities on the same dataset. There was a remarkable increase in the accuracy and F1-score of composite activities recognition in our system as shown, in Figure 7 below.

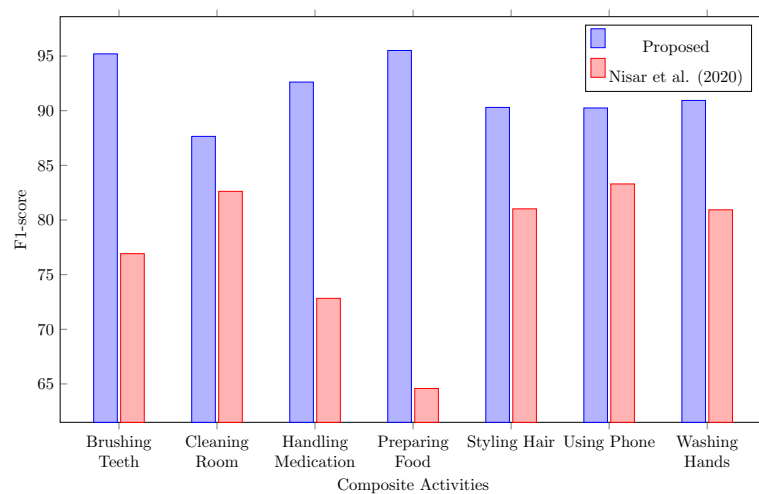


Figure 7. The performance comparison of proposed one-vs.-all classification with the results computed in [6]. The proposed model consistently performed better for each composite activity recognition.

The second experiment was performed for multiclass classification of composite activities. The most effective features, based on atomic score, were extracted for each composite activity c_i resulting in seven feature vectors $F_1, \dots, F_7$ for “Brushing Teeth” to “Washing Hands”, respectively. The combined feature vector $F_{combined}$ (as described in Equation (5)) for multiclass classification was obtained by taking the union of all these individual feature vectors. This combined feature vector was then passed to the two classifiers SVM and RF as input and the classification was performed.

4.3. Performance Evaluation

The performance of both classifiers is measured using two matrices: accuracy and F_1 -score. Accuracy is measured by calculating the ratio of correctly identified instances to the total number of instances, as described in the following equation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

where TP , TN , FP , and FN are true positive, true negative, false positive, and false negative, respectively. The F_1 score is the harmonic mean of precision and recall.

$$F_1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where the precision can be obtained when all the positive predictions are divided by the actual positive predictions. Recall is the ratio of successfully predicted positive instances to all positive instances.

The results of both classifiers are shown in Table 9. The selected features performed best for the RF classifier for all the experiments because of its property of handling large datasets with complex decision boundaries by avoiding overfitting. However, there is a small difference in the results of SVM for the same features. The features were selected based on atomic score, and experiments were performed for different values of atomic score percentage. It is seen that the accuracy of both classifiers also decreases as the number of selected features falls, thus removing some important features. However, if we pass all the features to the classifier, the training gets slow and there is no increase in the accuracy. Both classifiers performed best for features containing greater than or equal to 90% atomic score because the feature vector obtained is of optimal size and there is less loss of important features in this score. The highest accuracy obtained for the recognition system is 94.1% for SVM and 96.61% for RF. The line graph for these results is shown in Figure 8. The loss curves of both classifiers are shown in Figure 9.

Table 9. Based on atomic scores percentage, features were selected for composite activities classification. The first column shows the atomic score percentage. In the second column, the selected features are obtained by taking common features from all the composite activities in that percentage. These selected features were passed to the SVM and RF classifiers. Both classifiers gave best results on 90% atomic scores, which shows that the features selected by the proposed model are the best features. The best scores are highlighted in bold.

Atomic Score (%)	Selected Features	SVM Accuracy (%)	RF Accuracy (%)
65	19	84.4	95.48
70	22	87.3	95.25
75	28	86.0	95.93
80	35	90.5	96.38
85	39	92.55	95.03
90	48	94.1	96.61
100	61	93.9	96.16

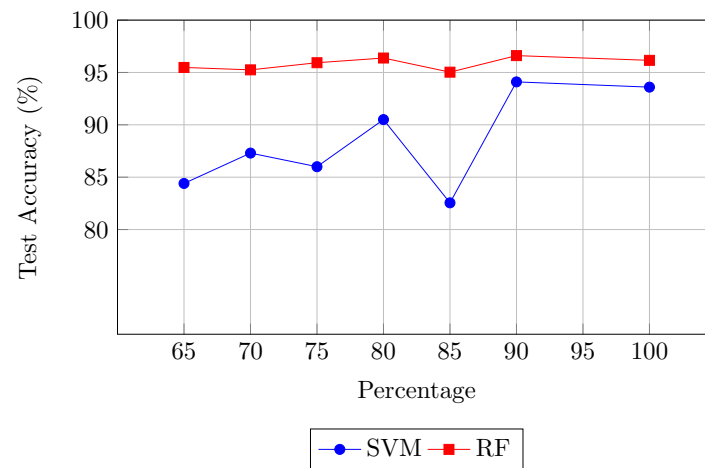


Figure 8. Accuracy of SVM and RF classifiers for composite activities recognition on different percentages of atomic score. The graph shows that both the classifiers performed best for 90% atomic score as their input features.

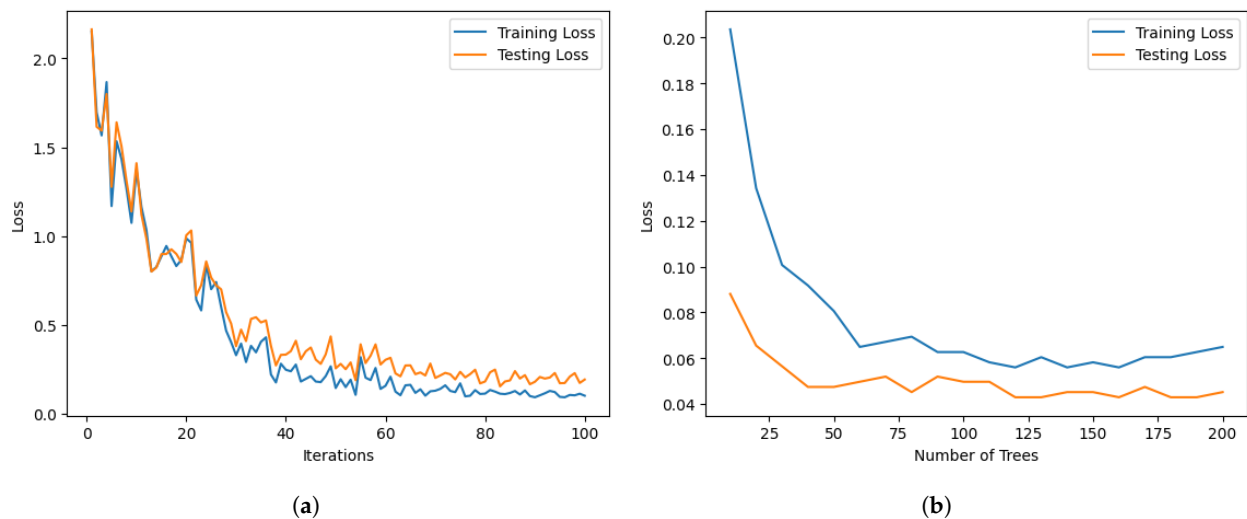


Figure 9. An illustration of training and testing loss of both classifiers used in this study: (a) SVM loss curve, (b) random forest loss curve.

4.4. Performance Comparison with Existing State-of-the-Art Techniques

The results of the proposed algorithm were compared with the existing techniques on the CogAge dataset (see Table 10). The table shows the results of different studies performed for composite activity recognition on this dataset. Nisar et al. in [6] used a rank pooling approach along with max and average pooling for dimensionality reduction and the classification was performed. Amjad et al. in [7] used handcrafted features for classifier input and obtained 79% accuracy, which is higher than that of the previous study on this dataset. The study in [93] proposed a novel self-attention Transformer model for the classification of ADLs and was able to achieve the better performance than the LSTM networks. Another study was performed [94] using the deep learning technique of hierarchical multitask learning and it obtained better accuracy than the existing methods at that time. A recent study [74] proposed a multi-branch hybrid Conv-LSTM network to classify human activities of daily living using multimodal data of different wearable smart devices. A significant increase of 12–14% in the accuracy achieved by our both classifiers over the existing techniques on this dataset can be seen, which shows that the features computed by our method performed best as classifier input, proving the effectiveness of our proposed method.

Table 10. Comparison of existing techniques on the CogAge HAR dataset. The proposed method performed better for both classifiers than existing techniques for this dataset, which shows the effectiveness of our feature selection method.

Methods	Year	Accuracy (%)
Rank pooling fused with max and average pooling [6]	2020	68.65
Handcrafted features [7]	2021	79
Transformers [93]	2022	73.36
Hierarchical multi-task learning [94]	2023	82.31
Multi-branch hybrid Conv-LSTM network [74]	2024	82.16
Proposed method (using SVM)	2024	94.1
Proposed method (using RF)	2024	96.61

We also computed the computational complexity of the proposed algorithm in terms of time required to recognize an activity sequence, and the results are compared with existing state-of-the-art techniques. Table 11 demonstrates the comparison of the computational complexity of the study in [74] with the proposed method. Unlike the method in [74], which requires 5.2 ms to recognize an unknown activity sequence, the proposed technique is significantly faster, requiring only 1.1 ms to classify an activity sequence.

Table 11. Analysis of the computational time to recognize an activity sequence in the CogAge dataset. The classification time explains the average time to recognize an activity sequence. The terms “ms” refers to milliseconds.

Method	Classification Time (ms)
Asfaq et al. [74]	5.5
Proposed method (using SVM)	3.3
Proposed method (using RF)	1.1

5. Conclusions

This paper presents a generic framework for the recognition of composite activities from sensory data. Since human composite activities are long-term activities consisting of several atomic activities, we cannot recognize the composite activities directly from raw sensory data. To overcome this limitation, we have proposed a method that works in a two-level hierarchical manner. First, the raw sensory data are represented as a compact feature vector using a codebook-based approach. The atomic activities are recognized using these features, and the atomic score of each atomic activity is calculated relative to the composite activities. In the second step, the optimal features are selected based on the percentage participation of atomic activities in a composite activity by using the atomic scores. Finally, the selected features are fed to the two classifiers: random forest and SVM for the recognition of composite activities. The proposed method was evaluated on the CogAge dataset and it achieved the highest recognition rate of 96.61%. Since the proposed method heavily relies on the scores of atomic activities, incorrect recognition of atomic scores may negatively impact the recognition accuracies of composite activities. Additionally, the proposed two-step hierarchical method may increase computation time and cost in some cases. In the future, we plan to propose a simple method to recognize composite activities using raw sensory data directly.

Author Contributions: Conceptualization, T.H., M.H.K. and M.S.F.; methodology, T.H. and M.H.K.; software, T.H. and M.S.F.; validation, T.H.; investigation, T.H., M.H.K. and M.S.F.; writing—original draft preparation, T.H.; writing—review and editing, M.H.K. and M.S.F.; supervision, M.H.K. and M.S.F.; project administration, M.H.K. and M.S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the Higher Education Commission, Pakistan within the project “HumCare: Human Activity Analysis in Healthcare” (Grant Number: 15041).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The CogAge dataset used in this study is publicly available.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Koping, L.; Shirahama, K.; Grzegorzec, M. A general framework for sensor-based human activity recognition. *Comput. Biol. Med.* **2018**, *95*, 248–260. [[CrossRef](#)] [[PubMed](#)]
2. Civitarese, G.; Sztyler, T.; Riboni, D.; Bettini, C.; Stuckenschmidt, H. POLARIS: Probabilistic and ontological activity recognition in smart-homes. *IEEE Trans. Knowl. Data Eng.* **2019**, *33*, 209–223. [[CrossRef](#)]
3. Yang, R.; Wang, B. PACP: A position-independent activity recognition method using smartphone sensors. *Information* **2016**, *7*, 72. [[CrossRef](#)]
4. Khan, M.H. *Human Activity Analysis in Visual Surveillance and Healthcare*; Logos Verlag GmbH: Berlin, Germany, 2018; Volume 45.
5. Fan, S.; Jia, Y.; Jia, C. A feature selection and classification method for activity recognition based on an inertial sensing unit. *Information* **2019**, *10*, 290. [[CrossRef](#)]
6. Nisar, M.A.; Shirahama, K.; Li, F.; Huang, X.; Grzegorzec, M. Rank pooling approach for wearable sensor-based ADLs recognition. *Sensors* **2020**, *20*, 3463. [[CrossRef](#)]
7. Amjad, F.; Khan, M.H.; Nisar, M.A.; Farid, M.S.; Grzegorzec, M. A Comparative Study of Feature Selection Approaches for Human Activity Recognition Using Multimodal Sensory Data. *Sensors* **2021**, *21*, 2368. [[CrossRef](#)] [[PubMed](#)]
8. Li, F.; Shirahama, K.; Nisar, M.A.; Köping, L.; Grzegorzec, M. Comparison of feature learning methods for human activity recognition using wearable sensors. *Sensors* **2018**, *18*, 679. [[CrossRef](#)]
9. Lara, O.D.; Labrador, M.A. A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1192–1209. [[CrossRef](#)]
10. Ke, S.R.; Thuc, H.L.U.; Lee, Y.J.; Hwang, J.N.; Yoo, J.H.; Choi, K.H. A Review on Video-Based Human Activity Recognition. *Computers* **2013**, *2*, 88–131. [[CrossRef](#)]
11. Rani, V.; Kumar, M.; Singh, B. Handcrafted features for human gait recognition: CASIA-A dataset. In Proceedings of the International Conference on Artificial Intelligence and Data Science, Hyderabad, India, 17–18 December 2021; pp. 77–88.
12. Schonberger, J.L.; Hardmeier, H.; Sattler, T.; Pollefeys, M. Comparative evaluation of hand-crafted and learned local features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1482–1491.
13. Saba, T.; Mohamed, A.S.; El-Affendi, M.; Amin, J.; Sharif, M. Brain tumor detection using fusion of hand crafted and deep learning features. *Cogn. Syst. Res.* **2020**, *59*, 221–230. [[CrossRef](#)]
14. Patel, B.; Srikanthan, S.; Asani, F.; Agu, E. Machine learning prediction of tbi from mobility, gait and balance patterns. In Proceedings of the 2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), Washington, DC, USA, 16–18 December 2021; pp. 11–22.
15. Khan, M.H.; Farid, M.S.; Grzegorzec, M. Spatiotemporal features of human motion for gait recognition. *Signal Image Video Process.* **2019**, *13*, 369–377. [[CrossRef](#)]
16. Fatima, R.; Khan, M.H.; Nisar, M.A.; Doniec, R.; Farid, M.S.; Grzegorzec, M. A Systematic Evaluation of Feature Encoding Techniques for Gait Analysis Using Multimodal Sensory Data. *Sensors* **2024**, *24*, 75. [[CrossRef](#)] [[PubMed](#)]
17. Zhao, B.; Lu, H.; Chen, S.; Liu, J.; Wu, D. Convolutional neural networks for time series classification. *J. Syst. Eng. Electron.* **2017**, *28*, 162–169. [[CrossRef](#)]
18. Tran, L.; Hoang, T.; Nguyen, T.; Kim, H.; Choi, D. Multi-model long short-term memory network for gait recognition using window-based data segment. *IEEE Access* **2021**, *9*, 23826–23839. [[CrossRef](#)]
19. Sargano, A.B.; Angelov, P.; Habib, Z. A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition. *Appl. Sci.* **2017**, *7*, 110. [[CrossRef](#)]
20. Dong, M.; Han, J.; He, Y.; Jing, X. HAR-Net: Fusing deep representation and hand-crafted features for human activity recognition. In Proceedings of the International Conference on Signal and Information Processing, Networking and Computers, Ji’nan, China, 23–25 May 2018; pp. 32–40.
21. Ferrari, A.; Micucci, D.; Mobilio, M.; Napolitano, P. Hand-crafted features vs residual networks for human activities recognition using accelerometer. In Proceedings of the 2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT), Ancona, Italy, 19–21 June 2019; pp. 153–156.
22. Khan, M.A.; Sharif, M.; Akram, T.; Raza, M.; Saba, T.; Rehman, A. Hand-crafted and deep convolutional neural network features fusion and selection strategy: An application to intelligent human action recognition. *Appl. Soft Comput.* **2020**, *87*, 105986. [[CrossRef](#)]
23. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In Proceedings of the Ambient Assisted Living and Home Care: 4th International Workshop, IWAAL 2012, Vitoria-Gasteiz, Spain, 3–5 December 2012; pp. 216–223.

24. Ronao, C.A.; Cho, S.B. Human activity recognition using smartphone sensors with two-stage continuous hidden Markov models. In Proceedings of the 2014 10th International Conference on Natural Computation (ICNC), Xiamen, China, 19–21 August 2014; pp. 681–686.
25. Rana, R.; Kusy, B.; Wall, J.; Hu, W. Novel activity classification and occupancy estimation methods for intelligent HVAC (heating, ventilation and air conditioning) systems. *Energy* **2015**, *93*, 245–255. [\[CrossRef\]](#)
26. Seera, M.; Loo, C.K.; Lim, C.P. A hybrid FMM-CART model for human activity recognition. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 5–8 October 2014; pp. 182–187.
27. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. In Proceedings of the Esann, Bruges, Belgium, 24–26 April 2013; Volume 3, p. 3.
28. Chen, Z.; Zhang, L.; Cao, Z.; Guo, J. Distilling the knowledge from handcrafted features for human activity recognition. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4334–4342. [\[CrossRef\]](#)
29. Abid, M.H.; Nahid, A.A. Two Unorthodox Aspects in Handcrafted-feature Extraction for Human Activity Recognition Datasets. In Proceedings of the 2021 International Conference on Electronics, Communications and Information Technology (ICECIT), Khulna, Bangladesh, 14–16 September 2021; pp. 1–4.
30. Liang, H.; Sun, X.; Sun, Y.; Gao, Y. Text feature extraction based on deep learning: A review. *EURASIP J. Wirel. Commun. Netw.* **2017**, *2017*, 211. [\[CrossRef\]](#)
31. Khan, M.H.; Farid, M.S.; Grzegorzec, M. A comprehensive study on codebook-based feature fusion for gait recognition. *Inf. Fusion* **2023**, *92*, 216–230. [\[CrossRef\]](#)
32. Khan, M.H.; Farid, M.S.; Grzegorzec, M. A generic codebook based approach for gait recognition. *Multimed. Tools Appl.* **2019**, *78*, 35689–35712. [\[CrossRef\]](#)
33. Azmat, U.; Jalal, A. Smartphone Inertial Sensors for Human Locomotion Activity Recognition based on Template Matching and Codebook Generation. In Proceedings of the 2021 International Conference on Communication Technologies (ComTech), Rawalpindi, Pakistan, 21–22 September 2021; pp. 109–114. [\[CrossRef\]](#)
34. Ryu, J.; McFarland, T.; Haas, C.T.; Abdel-Rahman, E. Automatic clustering of proper working postures for phases of movement. *Autom. Constr.* **2022**, *138*, 104223. [\[CrossRef\]](#)
35. Shirahama, K.; Grzegorzec, M. On the generality of codebook approach for sensor-based human activity recognition. *Electronics* **2017**, *6*, 44. [\[CrossRef\]](#)
36. Uddin, M.Z.; Thang, N.D.; Kim, J.T.; Kim, T.S. Human activity recognition using body joint-angle features and hidden Markov model. *ETRI J.* **2011**, *33*, 569–579. [\[CrossRef\]](#)
37. Siddiqui, S.; Khan, M.A.; Bashir, K.; Sharif, M.; Azam, F.; Javed, M.Y. Human action recognition: A construction of codebook by discriminative features selection approach. *Int. J. Appl. Pattern Recognit.* **2018**, *5*, 206–228. [\[CrossRef\]](#)
38. Khan, M.H.; Farid, M.S.; Grzegorzec, M. Person identification using spatiotemporal motion characteristics. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 166–170.
39. Gaikwad, N.B.; Tiwari, V.; Keskar, A.; Shivaprakash, N. Efficient FPGA implementation of multilayer perceptron for real-time human activity classification. *IEEE Access* **2019**, *7*, 26696–26706. [\[CrossRef\]](#)
40. Mesquita, C.M.; Valle, C.A.; Pereira, A.C. Dynamic Portfolio Optimization Using a Hybrid MLP-HAR Approach. In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, Australia, 1–4 December 2020; pp. 1075–1082.
41. Rustam, F.; Reshi, A.A.; Ashraf, I.; Mehmood, A.; Ullah, S.; Khan, D.M.; Choi, G.S. Sensor-based human activity recognition using deep stacked multilayered perceptron model. *IEEE Access* **2020**, *8*, 218898–218910. [\[CrossRef\]](#)
42. Azmat, U.; Ghadi, Y.Y.; Shloul, T.A.; Alsuhibany, S.A.; Jalal, A.; Park, J. Smartphone Sensor-Based Human Locomotion Surveillance System Using Multilayer Perceptron. *Appl. Sci.* **2022**, *12*, 2550. [\[CrossRef\]](#)
43. Hassan, M.M.; Uddin, M.Z.; Mohamed, A.; Almogren, A. A robust human activity recognition system using smartphone sensors and deep learning. *Future Gener. Comput. Syst.* **2018**, *81*, 307–313. [\[CrossRef\]](#)
44. Mekruksavanich, S.; Jitpattanakul, A. Deep convolutional neural network with rnns for complex activity recognition using wrist-worn wearable sensor data. *Electronics* **2021**, *10*, 1685. [\[CrossRef\]](#)
45. Chung, S.; Lim, J.; Noh, K.J.; Kim, G.; Jeong, H. Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning. *Sensors* **2019**, *19*, 1716. [\[CrossRef\]](#)
46. Tong, L.; Ma, H.; Lin, Q.; He, J.; Peng, L. A novel deep learning Bi-GRU-I model for real-time human activity recognition using inertial sensors. *IEEE Sens. J.* **2022**, *22*, 6164–6174. [\[CrossRef\]](#)
47. Batool, S.; Khan, M.H.; Farid, M.S. An ensemble deep learning model for human activity analysis using wearable sensory data. *Appl. Soft Comput.* **2024**, *159*, 111599. [\[CrossRef\]](#)
48. Khodabandelou, G.; Moon, H.; Amirat, Y.; Mohammed, S. A fuzzy convolutional attention-based GRU network for human activity recognition. *Eng. Appl. Artif. Intell.* **2023**, *118*, 105702. [\[CrossRef\]](#)
49. Varamin, A.A.; Abbasnejad, E.; Shi, Q.; Ranasinghe, D.C.; Rezatofighi, H. 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, New York, NY, USA, 5–7 November 2018; pp. 246–253.
50. Malekzadeh, M.; Clegg, R.G.; Haddadi, H. Replacement autoencoder: A privacy-preserving algorithm for sensory data analysis. In Proceedings of the 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), Orlando, FL, USA, 17–20 April 2018; pp. 165–176.

51. Jia, G.; Lam, H.K.; Liao, J.; Wang, R. Classification of electromyographic hand gesture signals using machine learning techniques. *Neurocomputing* **2020**, *401*, 236–248. [\[CrossRef\]](#)
52. Rubio-Solis, A.; Panoutsos, G.; Beltran-Perez, C.; Martinez-Hernandez, U. A multilayer interval type-2 fuzzy extreme learning machine for the recognition of walking activities and gait events using wearable sensors. *Neurocomputing* **2020**, *389*, 42–55. [\[CrossRef\]](#)
53. Gavrilin, Y.; Khan, A. Across-sensor feature learning for energy-efficient activity recognition on mobile devices. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–7.
54. Mohammadian Rad, N.; Van Laarhoven, T.; Furlanello, C.; Marchiori, E. Novelty detection using deep normative modeling for imu-based abnormal movement monitoring in parkinson's disease and autism spectrum disorders. *Sensors* **2018**, *18*, 3533. [\[CrossRef\]](#)
55. Prasath, R.; O'Reilly, P.; Kathirvalavakumar, T. *Mining Intelligence and Knowledge Exploration*; Springer: Berlin/Heidelberg, Germany, 2014.
56. Almaslukh, B.; AlMuhtadi, J.; Artoli, A. An effective deep autoencoder approach for online smartphone-based human activity recognition. *Int. J. Comput. Sci. Netw. Secur.* **2017**, *17*, 160–165.
57. Mohammed, S.; Tashev, I. Unsupervised deep representation learning to remove motion artifacts in free-mode body sensor networks. In Proceedings of the 2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Eindhoven, The Netherlands, 9–12 May 2017; pp. 183–188.
58. Malekzadeh, M.; Clegg, R.G.; Cavallaro, A.; Haddadi, H. Protecting sensory data against sensitive inferences. In Proceedings of the 1st Workshop on Privacy by Design in Distributed Systems, Porto, Portugal, 23–26 April 2018; pp. 1–6.
59. Malekzadeh, M.; Clegg, R.G.; Cavallaro, A.; Haddadi, H. Mobile sensor data anonymization. In Proceedings of the International Conference on Internet of Things Design and Implementation, Montreal, QC, Canada, 15–18 April 2019; pp. 49–58.
60. Gao, X.; Luo, H.; Wang, Q.; Zhao, F.; Ye, L.; Zhang, Y. A human activity recognition algorithm based on stacking denoising autoencoder and lightGBM. *Sensors* **2019**, *19*, 947. [\[CrossRef\]](#)
61. Bai, L.; Yeung, C.; Efstratiou, C.; Chikomo, M. Motion2Vector: Unsupervised learning in human activity recognition using wrist-sensing data. In Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers, London, UK, 9–13 September 2019; pp. 537–542.
62. Wahla, S.Q.; Ghani, M.U. Visual Fall Detection from Activities of Daily Living for Assistive Living. *IEEE Access* **2023**, *11*, 108876–108890. [\[CrossRef\]](#)
63. Tang, Y.; Zhang, L.; Min, F.; He, J. Multi-scale deep feature learning for human activity recognition using wearable sensors. *IEEE Trans. Ind. Electron.* **2022**, *70*, 2106–2116. [\[CrossRef\]](#)
64. Li, Y.; Yang, G.; Su, Z.; Li, S.; Wang, Y. Human activity recognition based on multienvironment sensor data. *Inf. Fusion* **2023**, *91*, 47–63. [\[CrossRef\]](#)
65. Ordóñez, F.J.; Roggen, D. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors* **2016**, *16*, 115. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Rivera, P.; Valarezo, E.; Choi, M.T.; Kim, T.S. Recognition of human hand activities based on a single wrist imu using recurrent neural networks. *Int. J. Pharma Med. Biol. Sci* **2017**, *6*, 114–118. [\[CrossRef\]](#)
67. Aljarrah, A.A.; Ali, A.H. Human activity recognition using PCA and BiLSTM recurrent neural networks. In Proceedings of the 2019 2nd International Conference on Engineering Technology and its Applications (IICETA), Al-Najef, Iraq, 27–28 August 2019; pp. 156–160.
68. Hu, Y.; Zhang, X.Q.; Xu, L.; He, F.X.; Tian, Z.; She, W.; Liu, W. Harmonic loss function for sensor-based human activity recognition based on LSTM recurrent neural networks. *IEEE Access* **2020**, *8*, 135617–135627. [\[CrossRef\]](#)
69. Martindale, C.F.; Christlein, V.; Klumpp, P.; Eskofier, B.M. Wearables-based multi-task gait and activity segmentation using recurrent neural networks. *Neurocomputing* **2021**, *432*, 250–261. [\[CrossRef\]](#)
70. Singh, N.K.; Suprabhath, K.S. HAR Using Bi-directional LSTM with RNN. In Proceedings of the 2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), Hyderabad, Indi, 25–27 August 2021; pp. 153–158.
71. Kaya, Y.; Topuz, E.K. Human activity recognition from multiple sensors data using deep CNNs. *Multimed. Tools Appl.* **2024**, *83*, 10815–10838. [\[CrossRef\]](#)
72. Reddy, C.R.; Sreenivasulu, A. Deep Learning Approach for Suspicious Activity Detection from Surveillance Video. In Proceedings of the 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 5–7 March 2020.
73. Abuhoureyah, F.S.; Wong, Y.C.; Isira, A.S.B.M. WiFi-based human activity recognition through wall using deep learning. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107171. [\[CrossRef\]](#)
74. Ashfaq, N.; Khan, M.H.; Nisar, M.A. Identification of Optimal Data Augmentation Techniques for Multimodal Time-Series Sensory Data: A Framework. *Information* **2024**, *15*, 343. [\[CrossRef\]](#)
75. Sezavar, A.; Atta, R.; Ghanbari, M. DCapsNet: Deep capsule network for human activity and gait recognition with smartphone sensors. *Pattern Recognit.* **2024**, *147*, 110054. [\[CrossRef\]](#)
76. Lalwani, P.; Ramasamy, G. Human activity recognition using a multi-branched CNN-BiLSTM-BiGRU model. *Appl. Soft Comput.* **2024**, *154*, 111344. [\[CrossRef\]](#)

77. Wei, X.; Wang, Z. TCN-attention-HAR: Human activity recognition based on attention mechanism time convolutional network. *Sci. Rep.* **2024**, *14*, 7414. [[CrossRef](#)] [[PubMed](#)]
78. Koutroumbas, K.; Theodoridis, S. *Pattern Recognition*; Academic Press: Cambridge, MA, USA, 2008.
79. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature selection in machine learning: A new perspective. *Neurocomputing* **2018**, *300*, 70–79. [[CrossRef](#)]
80. Wang, H.; Kläser, A.; Schmid, C.; Liu, C.L. Dense trajectories and motion boundary descriptors for action recognition. *Int. J. Comput. Vis.* **2013**, *103*, 60–79. [[CrossRef](#)]
81. Durgesh, K.S.; Lekha, B. Data classification using support vector machine. *J. Theor. Appl. Inf. Technol.* **2010**, *12*, 1–7.
82. Nurhanim, K.; Elamvazuthi, I.; Izhar, L.; Ganesan, T. Classification of human activity based on smartphone inertial sensor using support vector machine. In Proceedings of the 2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation (ROMA), Kuala Lumpur, Malaysia, 19–21 September 2017; pp. 1–5.
83. Khatun, M.A.; Yousuf, M.A.; Ahmed, S.; Uddin, M.Z.; Alyami, S.A.; Al-Ashhab, S.; Akhdar, H.F.; Khan, A.; Azad, A.; Moni, M.A. Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor. *IEEE J. Transl. Eng. Health Med.* **2022**, *10*, 1–16. [[CrossRef](#)] [[PubMed](#)]
84. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ.-Comput. Inf. Sci.* **2023**, *35*, 757–774. [[CrossRef](#)]
85. Khan, M.H.; Farid, M.S.; Grzegorzec, M. Vision-based approaches towards person identification using gait. *Comput. Sci. Rev.* **2021**, *42*, 100432. [[CrossRef](#)]
86. Khan, M.H.; Farid, M.S.; Grzegorzec, M. A non-linear view transformations model for cross-view gait recognition. *Neurocomputing* **2020**, *402*, 100–111. [[CrossRef](#)]
87. Snoek, C.G.; Worring, M.; Smeulders, A.W. Early versus late fusion in semantic video analysis. In Proceedings of the 13th Annual ACM International Conference on Multimedia, Singapore, 6–11 November 2005; pp. 399–402.
88. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* **1999**, *10*, 988–999. [[CrossRef](#)]
89. Zhang, J.; Marszałek, M.; Lazebnik, S.; Schmid, C. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Comput. Vis.* **2007**, *73*, 213–238. [[CrossRef](#)]
90. Shirahama, K.; Köping, L.; Grzegorzec, M. Codebook approach for sensor-based human activity recognition. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, Heidelberg, Germany, 12–16 September 2016; pp. 197–200.
91. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
92. Cutler, A. *Random Forests for Regression and Classification*; Utah State University: Ovronnaz, Switzerland, 2010.
93. Augustinov, G.; Nisar, M.A.; Li, F.; Tabatabaei, A.; Grzegorzec, M.; Sohrabi, K.; Fudickar, S. Transformer-based recognition of activities of daily living from wearable sensor data. In Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence, Rostock, Germany, 19–20 September 2022; pp. 1–8.
94. Nisar, M.A.; Shirahama, K.; Irshad, M.T.; Huang, X.; Grzegorzec, M. A Hierarchical Multitask Learning Approach for the Recognition of Activities of Daily Living Using Data from Wearable Sensors. *Sensors* **2023**, *23*, 8234. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.