Int. j. inf. tecnol. (August 2021) 13(4):1615–1624
https://doi.org/10.1007/s41870-021-00719-6

ORIGINAL RESEARCH

Feature fusion using deep learning for smartphone based human activity recognition

Dipanwita Thakur[1] · Suparna Biswas[2]

Received: 12 February 2021 / Accepted: 23 May 2021 / Published online: 12 June 2021
© Bharati Vidyapeeth's Institute of Computer Applications and Management 2021

Abstract Identification of human physical activities is an active research area since long due to its application in personalized health and fitness monitoring. The performance accuracy of human activity recognition (HAR) models mainly depend on the features which are extracted from domain knowledge. The features are the input of the classification algorithm to efficiently identify human physical activities. Manually extracted features (handcrafted) need expert domain knowledge. Thus these features have significant importance to identify different human activities. Recently deep learning methods are utilized to extract the features automatically from raw sensory data for HAR models. However, state-of-the-art HAR literature established that the importance of handcrafted features can't be ignored as it is extracted from expert domain knowledge. Thus, in this paper we use the fusion of both the handcrafted features and automatically extracted features using deep learning (DL) for HAR model to enhance the performance of HAR. Extensive experimental results demonstrate that our proposed feature fusion based HAR model gives higher accuracy compared with state-of-the-art HAR literature for both the self collected and public dataset.

Keywords HAR · Feature fusion · Smartphone sensors · Deep learning

✉ Dipanwita Thakur
dipanwita.thakur@gmail.com

Suparna Biswas
mailtosuparna@gmail.com

[1] Banasthali Vidyapith, Vanasthali, Rajasthan, India

[2] Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India

1 Introduction

HAR systems are gaining popularity in this digital age as they are used to identify various context-aware activities to get accurate medical facilities in the early stage of the disease and various ambient intelligence applications [16, 41]. HAR systems can provide medical facilities using remote health monitoring. The collection of physical human activity signals plays a crucial role. Several sensors are used to accumulate the physical signals. Wearable sensors are commonly used to accumulate physical signals. However, it is problematic for the users to wear various body sensors. In addition, hardware cost is also associated with wearable sensors and sophisticated signal processing techniques are required for it [12]. Ambient sensors are also used to collect data for physical human activities. Ambient sensors can only be used in some specific area. Moreover, it suffers from privacy policy. Recent advancement of the smartphone, with many in-built sensors, has become a powerful tool to collect the physical human activity signals. Due to the noninvasive property and diversity of in-built sensors the smartphone is widely adopted for HAR. Continuous physical activity data can be collected using a smartphone as every person carries a smartphone with themselves. So, a massive amount of data can be collected using smartphone based in-built sensors. Accelerometer and gyroscope are two in-built sensors of smartphones popularly used in HAR [4, 11, 36].

1.1 Motivation

Smartphone based HAR models can be implemented using shallow Machine Learning (ML) or DL algorithms. In shallow ML algorithms, feature engineering is an important phase to extract the relevant features. Then the features are fed to any classification algorithm to identify various human physical activities efficiently and correctly [12, 22]. As the raw smartphone sensory signals are not in the appropriate form, as they are highly fluctuating and oscillatory. Thus it is very difficult to identify the rudimentary patterns using these raw signals. So without feature engineering the classifiers are inadequate to give correct results. Moreover, without extracting the proper features the classifier fails to identify similar human physical activities such as walking and moving upstairs. Moreover, the manually extracted statistical features are used to overcome the problem regarding the position of the smartphone, while collecting the data. In HAR literature, most of the HAR solutions are position specific. That means, while collecting the data the position of the smartphone is fixed and the accuracy of the classifiers vary according to the position of the smartphone. The manually extracted features are not influenced by the position of the smartphone while collecting the data [2]. As a result, some standard statistical manually extracted features are gathered from raw smartphone sensory data. After extracting the handcrafted features, the shallow ML classifiers are used to recognize various human physical activities. Hence, shallow ML algorithms rely on handcrafted features [1, 3, 20]. The DL algorithms are one step ahead from shallow ML algorithms as DL algorithms automatically learn relevant features from raw sensor data without any human interference and also recognize human physical activities at the same time [18, 40].

Both the shallow ML algorithms with handcrafted feature extraction and DL algorithms with automatically learned features have attained great success to implement smartphone based HAR models. Hence, it is so obvious to believe that the combination of manually extracted features with automatically learned features from any DL method will improve the potentiality of smartphone based HAR model [13]. Chen et al. [13], proposed feature fusion of handcrafted and automatic learning features using deep “Long Short-Term Memory” (LSTM) to enhance the accuracy of their proposed smartphone based HAR model. However, the authors did not consider the training time and testing time for the proposed model. The authors used two stacked LSTM which increase the training time and test time. In the HAR literature, enhancing the accuracy is not the only concern. We need an efficient and accurate HAR solution with admissible accuracy as well as the training and test time should be less. Moreover, the HAR solution should be simple enough so that it consumes minimum storage and energy. Convolutional Neural Network (CNN) is one of the popular simple DL algorithms in various application domains of smart health [14, 23, 31]. The aforementioned issues motivate us to use feature fusion of handcrafted and automatic learning features using CNN to escalate the performance of HAR models using smartphone sensors in terms of accuracy, training time and testing time. According to the best of our knowledge, the feature fusion using handcrafted features and automatically extracted features using CNN never used before in the HAR domain with acceptable accuracy, training time and testing time.

We summarize our contribution is as follows:

- We propose an experimental feature fusion model which efficiently integrate both the handcrafted features and DL algorithm to improve the efficiency of HAR model using smartphone sensors.
- We use both the self-collected and public standard dataset of smartphone based HAR for the evaluation of our proposed method. Our exhaustive experimental results reveal the proposed approach notably outperforms state-of-the-art approaches.

2 Related work

In this section, some relevant works for smartphone based HAR using various ML and DL algorithms are explored. In the literature of the last ten years, the majority of smartphone based HAR solutions using machine learning algorithms relied on feature extraction.

2.1 Shallow machine learning algorithms

Feature engineering and activity identification are two major phases of shallow ML algorithms. Since raw smartphone sensory data are noisy and ineffective to efficiently identify various human physical activities, it is necessary to extract relevant features with expert domain knowledge. Then the extracted features are fed to shallow machine learning algorithms to implement efficient smartphone based HAR models. The researchers in Ref. [5] investigated different human activities using four different classifiers. The authors used different locations to collect the data using the accelerometer sensor of smartphones and also extracted different features to feed to the classifiers. Decision tree (C4.5) classifier has given the highest accuracy of 95.2% among all the four classifiers. Antos et al. [6] investigated the identification of 3 different human activities using two different classifiers. The output of the SVM classifier is given as an input to the hidden Markov model to increase the accuracy. The data is collected using the smartphone’s accelerometer sensor. Thirteen different features are used to do the experiment. The pant pocket location of accelerometer sensory data gets the highest accuracy of 95.2%. Bayat et al. [9], proposed a recognition system for human activities using accelerometer sensors built in smartphones. The fusion of five classifiers are used to classify the activities using self collected data and leading to an accuracy of 91.15%. In Ref. [37], the authors used K-nearest neighbor to identify six different human physical activities with 88.1% accuracy. Whereas, the authors in Ref. [35] achieved the classification accuracy of 96% using SVM classifier for seven different activities while keeping the mobile in the pant pocket. Unsupervised learning algorithm used in Ref. [26], to classify six human physical activities. In Ref. [20], SVM and k-nearest neighbor algorithms were used to classify five different human activities with the accuracy of 97.12%. An extreme learning machine was proposed in Ref. [12] to classify six different human activities with 98.88% accuracy. Similarly, in Ref. [8] and [36], random forest and multi-layer perceptron algorithms were used to classify human activities with 99.86% and 93% accuracy, respectively. The authors in Ref. [20], proposed smartphone based HAR system. In this work, the authors used time and frequency domain features. Moreover, the authors used two descriptors to extract feature sets from signals. SVM and KNN classifiers were used to recognize human physical activities. Recently, a hybrid method of filter and wrapper feature selection was proposed in Ref. [1] in smartphone based HAR systems. SFFS was used to extract various features with SVM. The SVM was used as a classifier in this. In a recent review article, [34], the authors compared various HAR systems with different perspectives such as position of smart phone, classification algorithms, human activities and accuracy. Almost, all the proposed HAR models using traditional ML are using handcrafted feature extraction to efficiently identify several human physical activities.

2.2 Deep learning algorithms

DL algorithms are able to learn the relevant features automatically. As a result, the performance of DL algorithms is remarkably high in case of smartphone based HAR models. CNN is gaining popularity in the HAR domain due to its hierarchical feature extraction capability. For example, Zeng et al. [42], proposed a CNN based HAR using mobile sensors. Three different publicly available datasets including “Antitraker” are used in this research work for experiment. The authors have taken the advantages of “local dependency” and “scale invariance” of CNN to achieve the accuracy of 88.19%, 76.83% and 96.88% using three different datasets respectively. In Ref. [24], the authors presented a smartphone based HAR model which can automatically learn features using “Sparse Auto-Encoder (SAE)”. The tri-axial accelerometer, gyroscope and the magnitude of both are used as different channels. In this work, using statistical metrics the classification accuracy is achieved by 97.55%. Ronao et al. [29], proposed another HAR solution using CNN. In this work, the authors used a HAR smartphone dataset from “UCI” repository. Also, the authors used manual hyperparameter tuning to achieve 95.75% accuracy to identify human physical activities. In their other work [30], the authors presented a CNN which is used to learn effective features from raw smartphone sensor data. In this work, they used temporal “fast Fourier transformation” on the raw data with CNN to implement the HAR model. Bevilacqua et al. [10], proposed CNN based HAR to identify sixteen different lower limb activities using five different sensors including accelerometer and gyroscope. In another work, Jiang et al. [21], proposed a CNN based HAR solution using “UCI” public dataset with 97.5% accuracy. The authors used “Adam” for hyperparameter optimization. Ignatov et al. [19], proposed another HAR solution using both the handcrafted features and CNN. Two different public datasets (WISDM and UCI) are used in this work. Using WISDM public dataset the authors has achieved the average accuracy of 90.42% and using UCI public dataset the authors has achieved the average accuracy of 94.35%. Also, the authors achieved 95.32% accuracy of HAR solution without taking the handcrafted features using UCI dataset. Zhou et al. [43] proposed a CNN based HAR solution for nine different activities using accelerometer, magnetometer, gyroscope and smartphone barometer. The authors achieved 98% accuracy. Dhanraj et al. [14], proposed a CNN based HAR solution using UCI public dataset with accuracy of 93.926%. In this work, the authors have mentioned the training and testing time as 3.4274 seconds and 372.6 ms respectively.

2.3 Handcrafted features with DL algorithms

In Ref. [27], the authors used time and frequency domain features with CNN to implement a smartphone based HAR system and achieved accuracy of 96.41%. The authors in Ref. [32], used two directional features with bidirectional long short-term memory (BLSTM) for incremental learning in HAR and achieved approximately 93% of average accuracy. Almaslukh et al. [2] proposed a robust position independent HAR system using CNN with 88% accuracy. In this work, the data was collected using smartphones and smartwatches located in seven different positions of the body. Also, the authors used some handcrafted features to locate the position of the smartphone. However, the authors did not integrate the handcrafted features and automatically extracted features to enhance the performance of the proposed HAR model. In Ref. [13], the authors used fusion of handcrafted features with automatic learned features using deep LSTM, to enhance the performance of the proposed HAR model. In this work, the authors achieved the average accuracy of 96.44% and 98.67% using public dataset from UCI and their own collected dataset respectively.

According to the smartphone based HAR literature, proper feature fusion in between handcrafted features and automatic learning features is performed only in Ref. [13] using a deep LSTM algorithm. In real application, both the manually extracted features with expert domain knowledge and automatic extracted features using DL are used to implement efficient smartphone based HAR models. In this paper, our aim is to implement a smartphone based HAR model using both the features to escalate the performance of the proposed HAR model using CNN.

3 The proposed model

This research work consists of different phases to identify human activities with higher accuracy and lower time cost.

3.1 Data collection

In this work, the data is gathered using Samsung Galaxy On-Max android smartphone. We create one android application to collect sensor data of six different human physical activities such as sitting, standing, walking, lying, walking upstairs and walking downstairs. The android smartphone application uses tri-axial accelerometer and gyroscope sensors with frequency of 50 Hz to accumulate the data, keeping the device in the front pant pocket or in hand. The data is collected from 25 subjects including 15 females and 10 males aged about 15–45 years, height about 163–172 cm and weight about 52–65 kg. The subjects are totally healthy without any medical complications. All subjects are asked to perform six normal human physical activities. All the activities are performed for three minutes with the repetition of five times by each of the subjects. All the activities are performed in both the indoor and outdoor conditions. Figure 1 shows the experimental setup of data collection in our experiment. This dataset consists of 15562 instances and 146 features.

3.2 Data pre-processing

The raw data collected from the smartphone in-built sensors consists of noise. As the data is collected using mobile applications there may be possibilities of missing and redundant data as there is a difference in individual data collection. To remove the high frequency noise and the gravitational acceleration from the signal, a low-pass elliptic filter with 20 Hz cutoff frequency followed by a high-pass elliptic filter with 0.5 Hz cutoff frequency are applied respectively [15].

Each signal is divided in 5 s sliding window with an overlap of 2 s between two consecutive windows as of state-of-the-art literatures [12, 20] show that 2–5 s sliding window with 20–50 Hz frequency is the ideal situation for the segmentation of the collected data.

3.3 Handcrafted features

Feature extraction is the key major step in conventional machine learning to build a good classifier. Smartphone sensors that collect raw data are noisy. Hence, using this data we are not able to recognize physical human activities efficiently. Various time and frequency domain statistical features are useful for HAR. Time domain features are useful to separate various static features such as standing and laying. Frequency domain features are useful to differentiate in between static and dynamic features. We have taken the extracted features similar as [4, 7, 12, 28], which are the standard features used in various HAR literature shown in Table 1. To improve the training performance much more features have been extracted to describe each activity window.

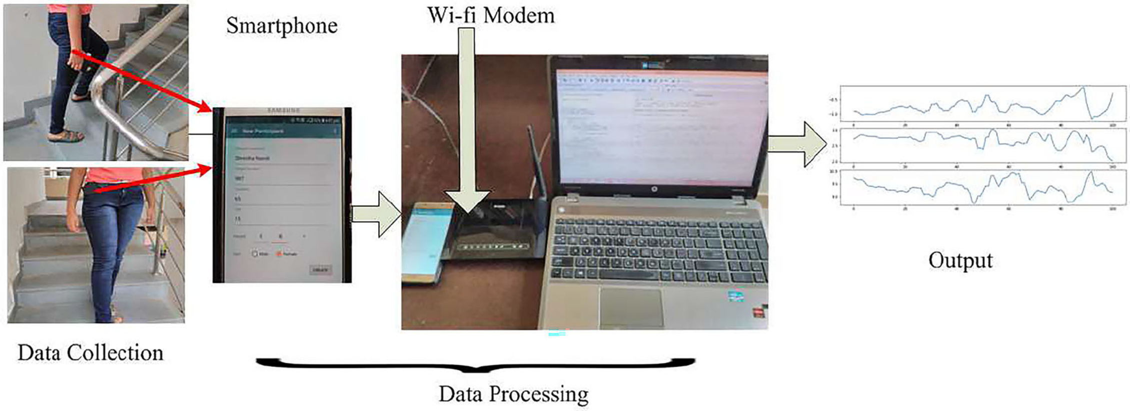

Fig. 1 Experimental setup

Table 1 Extracted features

Domain	Features
Time	Minimum
	Standard Deviation
	Correlation Coefficient
	Median Absolute Value
	Signal Entropy
	Interquartile Range
	Average Sum of Squares
	Mean Value
	Auto Regression Coefficients
	Signal Magnitude Area
	Maximum
Frequency	Weighted Average
	Kurtosis
	Largest Frequency Component
	Angle between two Vectors
	Skewness
	Energy of a Frequency Interval

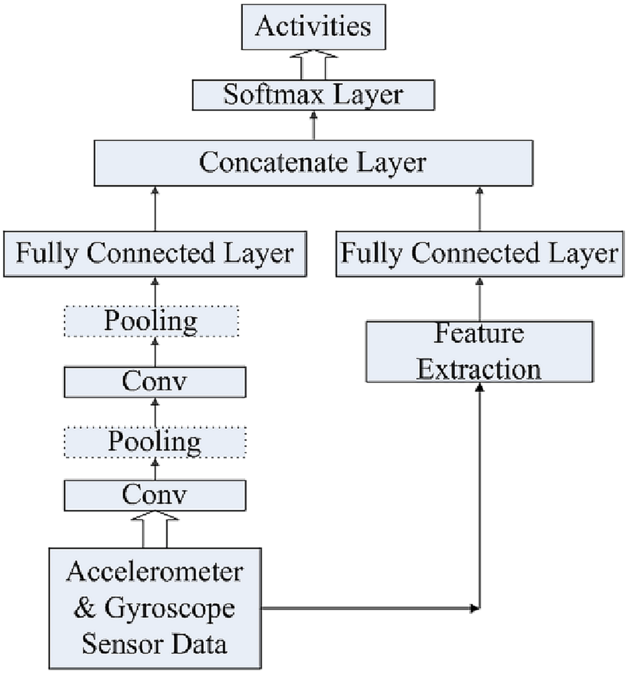

Fig. 3 The proposed feature fusion framework

3.4 Automatic feature learning

The great advantage of DL algorithms is the ability to learn relevant features automatically from raw sensory data. The smartphone sensor data is time-series data [25]. Research has shown that there are many major advantages over other strategies in the use of CNNs for time series classification [39]. CNN is a highly noise-resistant model. Using CNN we can extract highly relevant and deep features.

A typical architecture of CNN is shown in Fig. 2. The network architecture of our CNN consists of an input layer, convolution layer which extracts the features from raw sensory data, a pooling layer to reduce the size of the extracted features, fully connected layer to integrate all the extracted features and a softmax layer to make differentiation between activities.

3.5 CNN based HAR

This section demonstrates the feature extraction method used in CNN. Our proposed method is shown in Fig. 3. Triaxial accelerometer time-series data is represented with overlapping window sizes of length w. This data is given as input to CNN. There are three different types of layers in our CNN-based model.

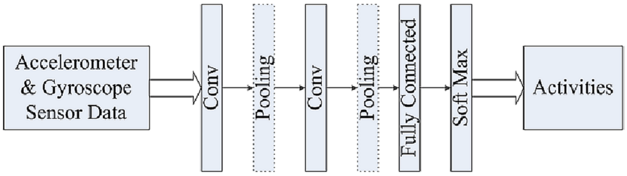

Fig. 2 Traditional CNN

- Input layer consists of h_i^0 units. The values of the input layer are decided by the input data.
- Hidden layers consists of h_i^l units. The values of the next hidden layer, l is dependent on the previous layer $l-1$.
- Output layer consists of h_i^L units. The value of the output layer is dependent on the last hidden layer.

A weight matrix $w_{i,j}^l$ is adjusted to learn the network. It is the input of h_i^l's unit and output of the h_j^{l+1} unit. ith unit of layer l is denoted by u_i^l. x_i^l and y_i^l are the total input and output of u_i^l and h_i^l units respectively.

3.5.1 *Convolutional input layer*

Convolutional layer is the first layer of the CNN, extracts the important features and patterns from data. In order to form a generous representation of data, multiple feature maps are included in this layer. Multiple feature maps of accelerometer data are represented as $x_i^{:,j}, j = 1, \cdots, J$. Suppose there is an N unit layer as input, followed by a convolutional layer. If we use kernel or filter size as M, the convolutional layer consists of $N - M + 1$ units. The output of the convolutional layer is represented as :

$$x_i^{l,j} = \sigma\left(b_j + \sum_{m=1}^{M} w_m^j x_{i+m-1}^{l-1,j} \right) \tag{1}$$

Here, the activation function is σ, the j-th feature map is b_j, the weight for the j-th feature map and m-th filter index is w_m^j.

3.5.2 *Max-pooling Layer*

The extracted features by convolutional layer are partitioned in different sets. In each set, max operation is applied to get the output values, given by

$$P_i^{l,j} = \max_{r\in R}\left(x_{(i-1)\times S+r}^{l-1,j} \right) \tag{2}$$

Here, polling size and stride are R and S respectively. The convolutional and polling layer stacked one after another to extract discriminant features.

3.5.3 *Training procedure*

Initially, the forward propagation is performed in a convolutional layer with N number of nodes using Eq. 1. The size of the convolutional layer is $N - M + 1$, where M is the filter size. The max-pooling layer takes the input from the output of the convolutional layer. The max-pooling is done using Eq. 2.

If the input layer consists of N nodes then the max-pooling layer consists of $\frac{N}{r}$ nodes, where r is converted to a single value window width using the max function. After that a fully connected layer is followed by the max-pooling layer. The forward propagation is represented as:

$$x_i^l = \sum_j w_{j,i}^{l-1} \sigma\left(x_i^{l-1}\right) + b_i^{l-1} \tag{3}$$

The topmost layer is the softmax classifier.

The feature vector extracted from the convolutional and pooling layers is given as an input to the softmax classifier. The feature vector $f^k = [f_1, f_2, \cdots, f_I]$. The topmost pooling layer consists of I number of units. The softmax classifier is denoted as

$$P(c|f) = \underset{c\in C}{\operatorname{argmin}} \frac{exp(f^{M-1}w^M + b^M)}{\sum_{j=1}^{N_c} exp\left(f^{M-1}w_j\right)} \tag{4}$$

where

- c = activity class
- M = index of the last layer
- N_c = total number of activity classes

After the first iteration of forward propagation, error value is generated using loss function L. We update w, using gradient descent. In our case, we use "cross entropy cost function" as a loss function. The gradient is measured as follows for the fully connected layer:

$$\frac{\partial L}{\partial w_{i,j}^l} = y_i^l \frac{\partial L}{\partial x_j^{l+1}} \tag{5}$$

where $y_i^l = \sigma(x_i^l) + b_i^l$ is the non-linear mapping function. The node j of $(l+1)$-th layer is x_j^{l+1}, $x_i^l = \sum_j w_{j,i}^{l-1} y_j^{l-1}$. The gradient in convolutional layer is calculated as:

$$\frac{\partial L}{\partial w_{a,b}} = \sum_{i=1}^{N-M-1} y_{i+m}^{l-1} \frac{\partial L}{\partial y_i^L} \sigma^l x_i^l \tag{6}$$

We stop the CNN training phase if it does not boost its output for five successive epochs on the validation sample.

3.6 Proposed feature fusion

Extracted features with expert domain knowledge and the features learned by DL algorithms both have significant impact in HAR model. To take the advantage of both the features, we propose a feature fusion concept to improve the identification accuracy of various human physical activities using smartphone in-built sensors. The proposed feature fusion model is outlined in Fig. 3. In this work, we have explored CNN for feature learning. The raw sensory data is given as input in 2D CNN for feature learning. At the end, the learned features are taken as input by the fully connected layer. Simultaneously, the manually extracted features as shown in Table 1, are taken as input by another fully connected layer to get more conceptual features. Finally, both the features in combination are given as input to the softmax layer for human physical activity classification.

4 Performance evaluation

In this section, we explain the overall performance of our proposed fusion method after exhaustive experiment.

4.1 Experimental setup

The description of our own dataset used in our experiment is illustrated in Sect. 3.1 in detail. To establish the veracity of the proposed feature fusion method, in our experiment, we use another public dataset which is taken from open source "UCI Machine Learning Repository" [3] for smartphone based HAR. This smartphone sensor based dataset consists of 30 subjects aged about 19 to 48 years, with 6 different human physical activities such as sitting, standing, walking, lying, walking upstairs and walking downstairs. Here, a waist-mounted smartphone (*SamsungGalaxySII*) with in-built sensors is used. Both the accelerometer and gyroscope sensors are used to collect the data. In this dataset total number of instances are 10299, number of features are 561. The public UCI dataset is already pre-processed using noise filters. Here, Butterworth low pass filter was used to separate the gravitational components from the body motion components. In this, a low pass filter with 0.3 HZ cutoff frequency was used, assuming the gravitational force only low frequency components. Then fixed-width sliding windows of 2.56 s with 50% overlapping was used.

In both the datasets, 70% of the total data are used for training the model and the rest 30% of the dataset are used for testing purposes. We use both the time domain and frequency domain features in our experiment as mentioned in Table 1. To validate the performance of the recommended model, we compare with two conventional ML methods such as SVM [3], Ensemble Learning Machine [33], and one DL of CNN [17]. Default parameters are used for all the benchmark schemes. In our proposed feature fusion, the CNN algorithm is used for feature learning. There are so many hyperparameters used by CNN algorithms. Using the trial and error method we have fixed the hyperparameter values such as batch size is of 1024, number of epochs varies from 100 to 200, Initial learning rate is 0.005, filter-size is set to 20, max-pooling size is set to 3, number of filter map 180, 60, and 30 respectively. The top two fully connected hidden layers have 1024 and 30 nodes respectively.

4.2 Experimental results

4.2.1 Results on our own collected dataset

The experimental results using self-collected dataset are tabulated in Table 2. Using our own dataset, automatic feature learning using CNN outperforms the conventional ML methods such as SVM and ELM with manually extracted features. This signifies that the automatic extracted features are more relevant to identify human physical activities using our own collected dataset. However, the proposed feature fusion method outperforms all the 3 benchmark schemes. Figure 4, shows the classification accuracy in terms of percentage using our own collected dataset.

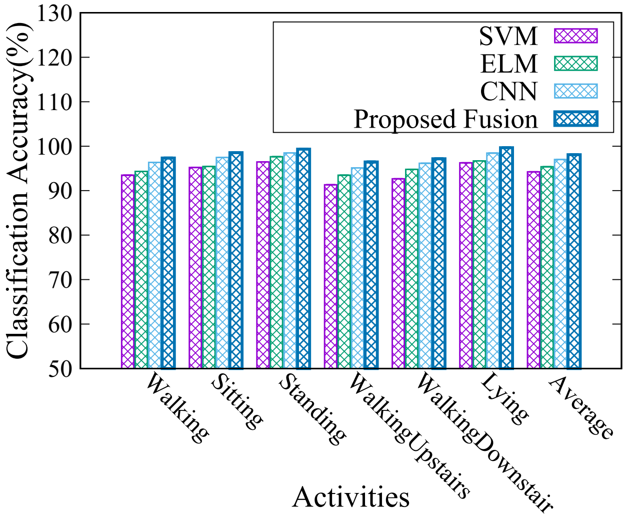

Fig. 4 Classification accuracy of all the approaches using our collected dataset

4.2.2 Results on UCI public dataset

The experimental results using UCI public dataset are tabulated in Table 3. According to our experimental result, it is clear that using our feature fusion approach we are able to get higher accuracy compared to all the 3 benchmark schemes we have considered. Our proposed method outperforms the other benchmark schemes. Using UCI public dataset, the performances of conventional ML methods such as SVM and ELM outperforms the DL method, CNN. Both SVM and ELM have used handcrafted features. Whereas, CNN utilizes the automatic feature learning concept. Therefore, there is a significant impact of manually extracted features to recognize human physical activities. In our proposed feature fusion, we have fused both the features to enhance the performance of classifiers. Figure 5, shows the classification accuracy in terms of percentage using UCI public dataset.

The proposed feature fusion method outperforms using both the datasets. However, the UCI public dataset gives higher classification accuracy compared to our own collected dataset. The higher number of features can be a

Table 2 Accuracy of the classifiers for each activities using different approaches using self collected dataset

Method	Walk	Sit	Stand	Upstair	Downstair	Lie	Average
SVM	0.9348	0.9523	0.9645	0.9134	0.9266	0.9627	0.9424
ELM	0.9432	0.9544	0.9762	0.9347	0.9478	0.9669	0.9539
CNN	0.9637	0.9748	0.9849	0.9512	0.9615	0.9846	0.9701
Proposed fusion	0.9734	0.9856	0.9932	0.9645	0.9717	0.9967	0.9809

Table 3 Accuracy of the classifiers for each activities using different approaches using UCI public dataset

Method	Walk	Sit	Stand	Upstair	Downstair	Lie	Average
SVM	0.9678	0.9739	0.9869	0.9557	0.9587	1.0000	0.9739
ELM	0.9776	0.9856	0.9890	0.9622	0.9748	1.0000	0.9815
CNN	0.9556	0.9689	0.9754	0.9368	0.9458	0.9935	0.9627
Proposed fusion	0.9874	0.9936	0.9989	0.9785	0.9878	1.0000	0.9910

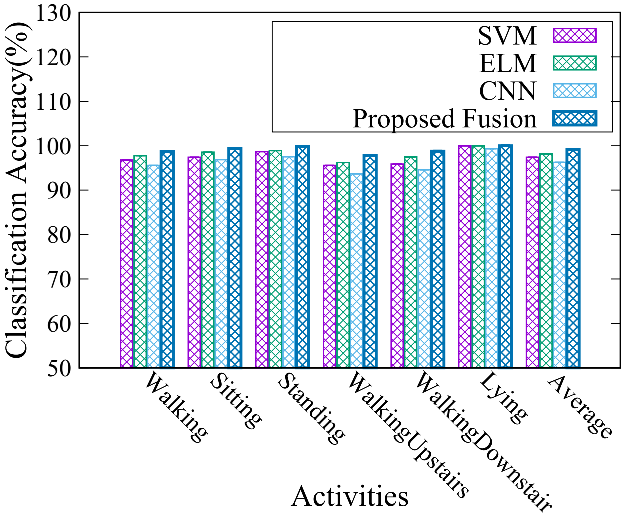

Fig. 5 Classification accuracy of all the approaches using UCI public dataset

reason to get higher accuracy using UCI dataset. There can be other reasons as well such as the position and orientation of the smartphone while collecting the data. So, using different datasets sometimes the conventional ML methods with manually extracted features give higher classification accuracies and sometimes DL method with automatic feature extraction gives higher classification accuracy. Therefore, we can conclude that both the manually and automatically extracted features are relevant to identify different human physical activities. As a result, the fusion of these features outperforms all other aforementioned methods using both the datasets which utilizes the advantages of both the features.

4.3 Compared with state-of-the-art smartphone based HAR literature

We compare our proposed feature fusion method with some state-of-the-art smartphone based HAR literature as mentioned in Table 4. Most of them have not mentioned the training and testing time of their proposed model. The most relevant research work [13], has not mentioned the training and testing time. However, the performance accuracy of our proposed fusion model is higher than the proposed model of Chen et al. [13] using two different datasets. Compared to the research work of Dhanraj et al. [14], the accuracy of our proposed method is much higher. However, the training and testing time are higher using our proposed method. Due to the feature fusion, it may be possible to take higher training and testing time but it is admissible. Otherwise, in terms of accuracy our proposed method using two different datasets out performs the state-of-the- art smartphone based HAR literature.

5 Application

The significant uses of HAR are in dynamic and helped living frameworks for smart homes, medical care observing applications, checking and reconnaissance frameworks for indoor and outside exercises, and tele-submersion applications. Moreover, studies of HAR are helpful to identify various lifestyle diseases such as obesity, diabetes. Moreover, it can also be used for rehabilitation purposes such as cardiac rehabilitation or any neurological rehabilitation.Therefore, the research community is always trying to enhance the HAR systems using various ML or DL

Table 4 Comparison with other approaches using Smartphone based HAR literature

Method	Dataset	Accuracy (%)	Training time	Testing time
CNN [29]	UCI	95.75	–	–
CNN [19]	UCI	94.35	–	–
CNN [21]	UCI	97.50	–	–
CNN [14]	UCI	93.93	3.4274 s	372.6 ms
Feature fusion [13]	UCI	96.44	–	–
Feature fusion [13]	Collected	98.67	–	–
Proposed fusion model	Our own	98.09	4.5674 s	448.8 ms
Proposed fusion model	UCI	99.10	5.2341 s	538.6 ms

methods based on sensory data. Recently, in Ref. [38], the authors proposed a “SpatioTemporal Human Activity Model (STHAM), for simulating SARS-CoV-2 transmission dynamics”. The movement of the people and their interaction with individuals in the society is the primary cause of the flow of “SARS-COV-2” [38].

6 Conclusion and future scope

In this work, we propose a feature fusion model using manually extracted features with expert domain knowledge and automatic learned features by CNN for smartphone based HAR. In this, we have used both the self collected dataset and UCI public dataset to assess the performance of the recommended feature fusion method. The experimental results show that the performance of the classifiers is not always the same. For instance, the performance of the shallow ML algorithms (SVM and ELM) with handcrafted features are better than CNN. Therefore, there is a significant impact of the handcrafted features in case of HAR. However, the performance of CNN is better than shallow ML algorithms (SVM and ELM) with handcrafted features using our own collected dataset. So, it is obvious to conclude that the feature fusion of handcrafted feature and automatic learned feature enhance the performance of the HAR model. In our experiment, we use CNN for automatic feature selection. As a result, the training time and testing time is highly admissible in our proposed fusion based HAR model. Finally, from the experimental results we conclude that our proposed approach outperforms the state-of-the-art smartphone based HAR literature for both the datasets in terms of accuracy, training time and testing time.

In future work one can consider different dynamic activities to identify with several other deep learning approaches. The position and orientation of the smartphone to escalate the performance of the smartphone based HAR system can be considered in future.

Declarations

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

References

1. Ahmed N, Rafiq J, Islam M (2020) Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model. Sensors 20(1):317
2. Almaslukh B, Artoli AM, Al-Muhtadi J (2018) A robust deep learning approach for position-independent smartphone-based human activity recognition. Sensors 18(11):3726
3. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2012) Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In: International workshop on ambient assisted living, pp 216–223
4. Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL (2013) A public domain dataset for human activity recognition using smartphones. In: European symposium on artificial neural networks, computational intelligence and machine learning
5. Anjum A, Ilyas MU (2013) Activity recognition using smartphone sensors. In: IEEE 10th consumer communications and networking conference (CCNC), pp 914–919
6. Antos SA, Albert MV, Kording KP (2014) Hand, belt, pocket or bag: practical activity tracking with mobile phones. J Neurosci Methods 231:22–30
7. Attal F, Dedabrishvili M, Oukhellou FC, Amirat Y (2015) Physical human activity recognition using wearable sensors. Sensors 15(12):31314–31338
8. Barua A, Masum AKM, Hossain ME, Bahadur EH, Alam MS (2019) A study on human activity recognition using gyroscope, accelerometer, temperature and humidity data. In: 2019 International conference on electrical, computer and communication engineering (ECCE), pp 1–6
9. Bayat A, Pomplun M, Tran DA (2014) A study on human activity recognition using accelerometer data from smartphones. Proc Comput Sci 34:450–457
10. Bevilacqua A, MacDonald K, Rangarej A, Widjaya V, Caulfield B, Kechadi T (2018) Human activity recognition with convolutional neural networks. In: Machine learning and knowledge discovery in databases
11. Bulbul E, Cetin A, Dogru IA (2018) Human activity recognition using smartphones. In: 2018 2nd international symposium on multidisciplinary studies and innovative technologies (ISMSIT), pp 1–6
12. Chen Z, Jiang C, Xie L (2019) A novel ensemble Elm for human activity recognition using smartphone sensors. IEEE Trans Ind Inf 15(5):2691–2699
13. Chen Z, Jiang C, Xiang S, Ding J, Wu M, Li X (2020) Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori. IEEE Trans Instrum Meas 69(7):3992–4001
14. Dhanraj S, De S, Dash D (2019) Efficient smartphone-based human activity recognition using convolutional neural network. In: 2019 International conference on information technology (ICIT), pp 307–312
15. Fahrenberg J, Foerster F, Smeja M, Müller W (1997) Assessment of posture and motion by multichannel piezoresistive accelerometer recordings. Psychophysiology 34(5):607–612
16. Franco A, Magnani A, Maio D (2020) A multimodal approach for human activity recognition based on skeleton and rgb data. Pattern Recogn Lett 131:293–299
17. Fukushima K (1988) Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw 1(2):119–130
18. Hammerla NY, Halloran S, Ploetz T (2016) Deep, convolutional, and recurrent models for human activity recognition using wearables. CoRR abs/1604.08880
19. Ignatov A (2018) Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl Soft Comput 62:915–922
20. Jain A, Kanhangad V (2018) Human activity classification in smartphones using accelerometer and gyroscope sensors. IEEE Sens J 18(3):1169–1177
21. Jiang X, Lu Y, Lu Z, Zhou H (2018) Smartphone-based human activity recognition using cnn in frequency domain. In: APWeb/WAIM Workshops

22. Lara OD, Labrador MA (2013) A survey on human activity recognition using wearable sensors. IEEE Commun Surv Tutor 15(3):1192–1209
23. Lei F, Liu X, Dai Q, Ling BWK (2019) Shallow convolutional neural network for image classification. SN Appl Sci 2(1):97
24. Li Y, Shi D, Ding B, Liu D (2014) Unsupervised feature learning for human activity recognition using smartphone sensors. Mining Intelligence and Knowledge Exploration, Lecture Notes In Computer Science 8891:99–107
25. Liu Y, Nie L, Liu L, Rosenblum DS (2016) From action to activity: sensor-based activity recognition. Neurocomputing 181:108–115
26. Mejia-Ricart LF, Helling P, Olmsted A (2017) Evaluate action primitives for human activity recognition using unsupervised learning approach. In: 2017 12th International conference for internet technology and secured transactions (ICITST), pp 186–188
27. Qin Z, Hu L, Zhang N, Chen D, Zhang K, Qin Z, Choo KR (2019) Learning-aided user identification using smartphone sensors for smart homes. IEEE Internet Things J 6(5):7760–7772
28. Quiroz JC, Banerjee A, Dascalu SM, Lau SL (2017) Feature selection for activity recognition from smartphone accelerometer data. Intelligent Automation and Soft Computing
29. Ronao C, Cho SB (2015) Deep convolutional neural networks for human activity recognition with smartphone sensors. In: Neural information processing, pp 46–53
30. Ronao C, Cho SB (2016) Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst Appl 59:235–244
31. Taheri S, Ezoji M, Sakhaei SM (2020) Convolutional neural network based features for motor imagery eeg signals classification in brain-computer interface system. SN Appl Sci 2(4):555
32. Tao D, Wen Y, Hong R (2016) Multicolumn bidirectional long short-term memory for mobile devices-based human activity recognition. IEEE Internet Things J 3(6):1124–1134
33. Thakur D, Biswas S (2020a) A novel human activity recognition strategy using extreme learning machine algorithm for smart health. In: Emerging technologies in data mining and information security
34. Thakur D, Biswas S (2020b) Smartphone based human activity monitoring and recognition using ML and DL: a comprehensive survey. J Ambient Intell Humaniz Comput 11:5433–5444
35. Tian Y, Chen W (2016) Mems-based human activity recognition using smartphone. In: 2016 35th Chinese control conference (CCC), pp 3984–3989
36. Voicu RA, Dobre C, Bajenaru L, Ciobanu RI (2019) Human physical activity recognition using smartphone sensors. Sensors 19(3):458
37. Wang A, Chen G, Yang J, Zhao S, Chang C (2016) A comparative study on human activity recognition using inertial sensors in a smartphone. IEEE Sens J 16(11):4566–4578
38. Wang Y, Li B, Gouripeddi R, Facelli JC (2021) Human activity pattern implications for modeling sars-cov-2 transmission. Comput Methods Programs Biomed 199:105896
39. Wu W, Zhang Y (2019) Activity recognition from mobile phone using deep cnn. In: 2019 Chinese Control Conference (CCC), pp 7786–7790
40. Yang JB, Nhut N, San P, li X, Shonali P (2015) Deep convolutional neural networks on multichannel time series for human activity recognition. IJCAI
41. Yao L, Sheng QZ, Benatallah B, Dustdar S, Wang X, Shemshadi A, Kanhere SS (2018) Wits: an iot-endowed computational framework for activity recognition in personalized smart homes. Computing 100(4):369–385
42. Zeng M, Nguyen LT, Yu B, Mengshoel OJ, Zhu J, Wu P, Zhang J (2014) Convolutional neural networks for human activity recognition using mobile sensors. In: 6th International conference on mobile computing, applications and services, pp 197–205
43. Zhou B, Yang J, Li Q (2019) Smartphone-based activity recognition for indoor localization using a convolutional neural network. Sensors (Basel, Switzerland) 19(3):621