



ARTICLE

Cascade Human Activity Recognition Based on Simple Computations Incorporating Appropriate Prior Knowledge

Jianguo Wang¹, Kuan Zhang^{1,*}, Yuesheng Zhao^{2,*}, Xiaoling Wang² and Muhammad Shamrooz Aslam²

¹School of Biomedical Engineering, Capital Medical University, Beijing, 100054, China

²School of Automation, Guangxi University of Science and Technology, Liuzhou, 545006, China

*Corresponding Authors: Kuan Zhang. Email: kzhang@ccmu.edu.cn; Yuesheng Zhao. Email: zhyuesheng@126.com

Received: 21 March 2023 Accepted: 06 September 2023 Published: 31 October 2023

ABSTRACT

The purpose of Human Activities Recognition (HAR) is to recognize human activities with sensors like accelerometers and gyroscopes. The normal research strategy is to obtain better HAR results by finding more efficient eigenvalues and classification algorithms. In this paper, we experimentally validate the HAR process and its various algorithms independently. On the base of which, it is further proposed that, in addition to the necessary eigenvalues and intelligent algorithms, correct prior knowledge is even more critical. The prior knowledge mentioned here mainly refers to the physical understanding of the analyzed object, the sampling process, the sampling data, the HAR algorithm, etc. Thus, a solution is presented under the guidance of right prior knowledge, using Back-Propagation neural networks (BP networks) and simple Convolutional Neural Networks (CNN). The results show that HAR can be achieved with 90%–100% accuracy. Further analysis shows that intelligent algorithms for pattern recognition and classification problems, typically represented by HAR, require correct prior knowledge to work effectively.

KEYWORDS

Human activities recognition; prior knowledge; physical understanding; sensors; HAR algorithms

1 Introduction

In recent years, Human Activities Recognition (HAR) has received tremendous attention from scholars due to its widespread application, such as rehabilitation training and competitive sports. The purpose of HAR is to unambiguously identify human poses and behaviors using strap-down sampled data from triaxial accelerometers, triaxial gyroscopes, and other sensors.

Currently, HAR as discussed in the academic community mainly refers to the recognition of human activities including sitting, standing, lying, walking, running, going upstairs and downstairs. HAR is to identify each of these types of human activity and give a conclusion as to what type of activity the human body is currently doing. These body activities are usually measured by sensors such as triaxial accelerometers and triaxial gyroscopes, based on which the recognition results are given by an AI algorithm.



The processes of human activity involve a wide range of complex movements such as triaxial acceleration, deceleration, bending, lifting, lowering and twisting of the upper limbs, chest, waist, hips, lower limbs and other parts of the human body. Correspondingly, measuring devices may be strap-down and fixed to different parts of the human body, or may be placed randomly in a pocket on the surface of the body. A considerable number of methods have been employed by scholars to analyze and utilize measured samples of human activity.

Li et al. [1] proposed and implemented a detection algorithm for the user's behavior based on the mobile device. After improving the directional independence and stride, the proposed algorithm enhances the adaptability of the HAR algorithm based on the frequency-domain and time-domain features of the accelerometer samples. A recognition accuracy of 95.13% was achieved when analyzing several sets of motion samples.

Based on accelerometers and neural networks, Zhang et al. [2] proposed a HAR method. A total of 100% of human activities are classified identically, including walking, sitting, lying, standing and falling.

The study conducted by Liu et al. [3] was based on the data analysis of the triaxial accelerometer and gyroscope of smartphones in order to extract the eigenvalue vector of human activities. It also selects four typical statistical methods to create HAR models separately. Model decisions are used to find the optimal HAR model, which achieves an average recognition rate of 92% for six activities: standing, sitting, going upstairs and going downstairs.

Zhou et al. proposed a multi-sensor-based HAR system in [4]. An algorithm is designed to identify eight common human activities using two levels of classification based on a decision tree. The identification rate averaged 93.12 percent.

Based on Coordinate Transformation and Principal Component Analysis (CT-PCA), Chen et al. [5] developed a robust HAR system. The Online Support Vector Machine (OSVM) was able to recognize 88.74% in terms of the variations of orientation, placement, and subject.

Using smartphones in different positions, Yang et al. [6] studied HAR with smartphones and proposed Parameter Adjustment Corresponding to Smartphone Position (PACP), a novel position-independent method that improves HAR performance. PACP achieves significantly higher accuracy than previous methods with over 91 percent accuracy.

Bandar et al. [7] developed a CNN model, which aimed at an effective smartphone-based HAR system. Two public datasets collected by smartphones, University of California, Irvine (UCI) and Wireless Sensor Data Mining (WISDM), were used to assess the performance of the proposed method. Its performance is evaluated by the F1-score, which is achieved at 97.73% and 94.47% on the UCI and WISDM datasets, respectively.

Andrey [8] developed an approach based on user-independent deep learning algorithm for online HAR. It is shown that the proposed algorithm exhibits high performance at low computational cost and does not require manual selection of feature values. The recognition rate of the proposed algorithm is 97.62% and 93.32% on the UCI and WISDM datasets, respectively.

Based on the wearable HAR system (w-HAR) and HAR framework, Bhat et al. [9] presented the w-HAR containing marked data from seven activities performed by 22 volunteers. This framework is capable of 95 percent accuracy, and the online system can improve HAR accuracy by up to 40 percent.

Ronao et al. [10] used smartphone sensors to exploit the inherent features of human activities. A deep CNN is proposed to provide an approach to extract robust eigenvalues automatically adaptively from one-dimensional (1D) time-series raw data. Based on a benchmark dataset collected from 30 volunteers, the proposed CNN outperforms other HAR techniques, exhibiting a combined performance of 94.79% on the raw dataset. Due to the fact that HAR processes 1D time series data, the CNNs used in HAR are mainly 1D-CNNs rather than the 2D-CNNs commonly used in image recognition.

Based on the above studies, it can be seen that the accelerometer used in HAR is the usual triaxial accelerometer in smartphones. The algorithms used in HAR include Decision Trees (DT), Naive Bayesian Network, Ada-Boost, Principal Component Analysis (PCA), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Convolutional Neural Networks (CNN) and so on. More and more scholars have adopted deep learning as the main HAR algorithm, and increasing innovative ideas are being introduced, as in [11–25].

Therefore, many HAR researchers are trying to find better eigenvalues and more efficient algorithms to implement HAR. Currently, the dominant HAR research strategy is how to achieve a complex multi-class classification task of human activities in a single step.

In this paper, we experimentally validate the process of HAR and the various algorithms independently and discuss the key role of prior knowledge step by step. Theoretical analysis and experimental validation show that the single-step multi-class classification task for complex human activities is difficult to implement and difficult to understand, while the solution of the proposed cascade structure model is clear to understand and easy to implement. The essential difference between the two HAR strategies is whether the HAR is based on the correct prior knowledge for the intelligent algorithm. A priori knowledge, which involves a physical understanding of human activity and the HAR algorithm, plays an important role in HAR. Our study shows that the recognition of nine daily activities from each other can be made clear and simple. The main reason for this effect is that the correct prior knowledge is incorporated into the proposed solution, which simplifies and improves HAR.

Section 2 of this paper presents the background of our confirmatory experiments. The results of our confirmation experiments are presented in Section 3 and discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2 Confirmatory Experiments

In this work, the HAR process and various HAR algorithms are tested to refresh our understanding of current developments in HAR. First, some measurement devices are used to obtain HAR raw data in the HAR experiment.

2.1 Measurement Devices, HAR Experiments, and Raw Sampling Data

With respect to measuring equipment, specialized equipment is designed and produced. Inside the device are triaxial accelerometers, gyroscopes and magnetometers. As shown in Fig. 1, the volunteer's ankles were fitted with two sets of such devices on his lateral sides.



Figure 1: HAR devices and fixing positions

The activities performed by the volunteers included sitting, standing, lying right and left, supine and prone positions, going upstairs and downstairs. The raw sampling data from these HAR devices was computed using a variety of so-called intelligent algorithms.

The experiment was also repeated using three smartphones. A third smartphone was attached to the forehead of the volunteers, while two smartphones were attached to their ankles. Furthermore, we used additional public datasets including UCI in [26], and WISDM in [27].

The UCI dataset contains inertial measurements generated from 30 volunteers aged between 19 and 48 while performing daily activities: standing, sitting, lying, walking, going upstairs and downstairs. Smartphones (Samsung Galaxy SII) fixed on the volunteers' waists were used to collect triaxial acceleration and angular velocity signals, which were produced by the triaxial accelerometer and the triaxial gyroscope built in the smartphones. The sampling frequency of these signals is 50 Hz.

The WISDM dataset contains triaxial acceleration data from 29 volunteers performing six daily activities: sitting, standing, walking, jogging, walking upstairs and walking downstairs. While collecting the data, the volunteers kept their smartphones in their foreleg pockets. The signal of the triaxial acceleration of the smartphone is collected at a sampling frequency of 20 Hz.

2.2 Data Pre-Processing: Bad Data Identification and Data Marking

Raw sampling data can be applied directly when a product is in its final form, but not during the research phase. Identifying bad data and labeling the data with human activity is crucial in research.

Constantly, measurement experiments are always imperfect, and thus bad data is frequently encountered. Causes of bad data include device power failure, device sliding, person sliding, etc. For example, volunteers did not rigorously complete all activities in a standard manner and equipment did not always work smoothly and properly. In addition, various noises and disturbances are constantly present during the experiments.

In our practice, about 10% of the raw sampled data can be visually identified as bad data and can be manually removed, as shown in Fig. 2. Further, it is reasonable to assume that the fraction of unseen and unrecognized bad data is not less than 10%, since the fraction of visible and identifiable bad data is approximately 10%.

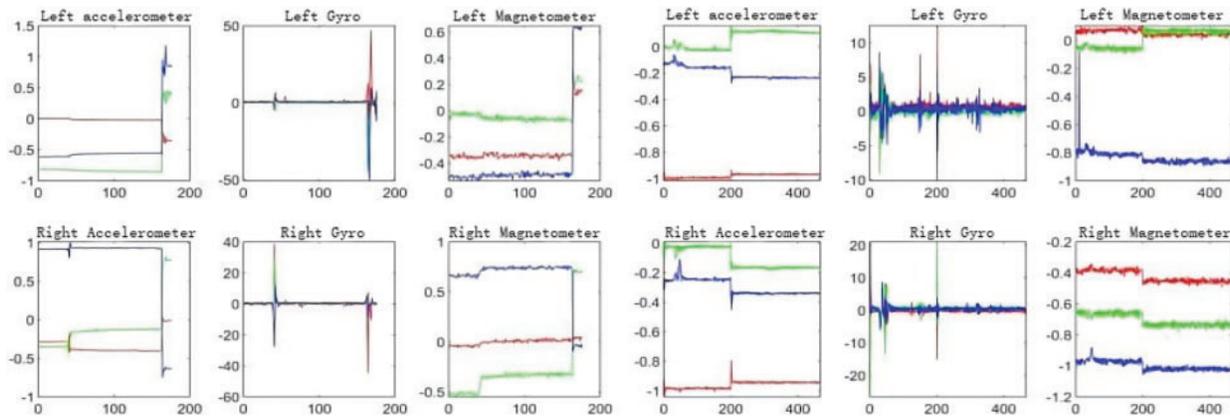


Figure 2: Examples of visible bad data

After removing the visible bad data, it is necessary to label the remaining 90% of the raw sampled data. This is to annotate which data are samples of walking and which are samples of sitting, standing, running, going upstairs and downstairs, etc.

During the data annotation phase, any error decreases the HAR accuracy. It is reasonable to conjecture that the maximum accuracy of HAR is close to 80 percent or even lower for raw data, since the fraction of valid data in the total raw data is close to 80 percent.

The subsequent calculations in this paper use the remaining data after the bad data has been manually removed.

2.3 Validation of the HAR Algorithm

A number of HAR algorithms were analyzed and tested in our study. As described in the references, these algorithms include Naive Bayesian Networks (NBN), Decision Trees (DT), Convolutional Neural Networks (CNN), Principal Component Analysis (PCA), Support Vector Machines (SVM), K-Nearest Neighbor Algorithms (KNA) and Ada-Boost.

After a comprehensive comparison, BP neural networks and 1D-CNNs are used for HAR in this paper. The pseudo-code of Cascade HAR for the 1D-CNNs and BP neural networks used in this paper is shown in Fig. 3, and the algorithm parameters for the 1D-CNNs or BP neural networks used in this paper are shown in Tables 1 and 2. Compared to other intelligent algorithms such as Decision Tree, we can achieve HAR accuracy of 90.5%–100% using simple BP neural networks or convolutional neural networks. The main difference between our solution and other HAR solutions is that it adopts a cascade identification structure, which is the result of a deep understanding of human activity.

3 Validation of HAR Algorithms

Derived from the motion analysis of HAR process, “the vector sum of triaxial acceleration” or “the X-axis acceleration” should be used according to various HAR needs. That is, the acceleration samples in the vertical direction of the human torso and the forward direction of the human torso are the key data for HAR. This is an important starting point of the prior knowledge in this paper.

```

Begin
  INPUT: Sampling for 2-3 seconds
  Make a binary classification between dynamic and static human activities
  If dynamic then
    {
      Make a binary classification between running and walking/upstairs/downstairs
      OUTPUT: If running then output "running"
      If walking/upstairs/downstairs then
        {
          Make a binary classification between walking and upstairs/downstairs
          OUTPUT: If walking then output "walking"
          If upstairs/downstairs then
            {
              Make a binary classification between upstairs and downstairs
              OUTPUT: If upstairs then output "walking"
              OUTPUT: If downstairs then output "downstairs"
            }
          }
        }
      }
  If static then
    {
      Make a binary classification between lying and standing/sitting
      OUTPUT: If lying then output "lying"
      If standing/sitting then
        {
          Make a binary classification between standing and sitting
          OUTPUT: If standing then output "standing"
          OUTPUT: If sitting then output "standing"
        }
      }
    }
End

```

Figure 3: Pseudo-code of cascade HAR with CNNs or BP neural networks in this paper

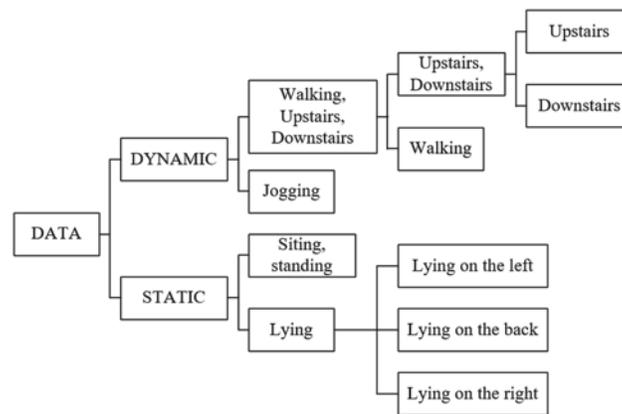
Table 1: Recognition algorithms and their parameters used in this paper

Algorithms	Parameters
BP neural networks	BP neural network with a two-layer network structure with about 10 nodes in each layer and varying number of nodes in each layer.
CNN	Same structure at all levels. Four convolution layers and two max-pooling layers. There are 128 convolution kernels in the first two layers and 512 in the next two layers. The size of the convolution kernel is 3, the epoch is 100 and the batch size is 8, the activation function in the convolution layer is relu and the activation functions in the density layer are relu and softmax, respectively.

Table 2: Structure and parameters of our CNN

Layer	Output shape	Parameters
Conv1d_1 (Conv1D)	(None, 20, 128)	512
Conv1d_2 (Conv1D)	(None, 18, 128)	49280
Max_pooling1d_1 (Maxpooling1d)	(None, 9, 128)	0
Conv1d_3 (Conv1D)	(None, 7, 512)	197120
Conv1d_4 (Conv1D)	(None, 7, 512)	786944
Max_pooling1d_2 (Maxpooling1d)	(None, 2, 512)	0
Flatten_1 (Flatten)	(None, 1024)	0
Dense_1 (Dense)	(None, 100)	102500
Dense_2 (Dense)	(None, 2)	202

Based on the correct prior knowledge, we transform the complex single-step multi-class classification task into a set of hierarchical binary classification recognition tasks as shown in Fig. 4, and adopt the following Steps 1–5 to validate the HAR of neural networks and Step 6 to validate the HAR of CNNs, respectively.

**Figure 4:** Binary classification cascade structure model for HAR

In order to verify the HAR effect of the Back-Propagation (BP) neural networks through Steps 1–5, the samples from the left accelerometer are separated into 466 groups, 30 samples in each group, which are the number of samples for 3 s at the sampling frequency of 10 Hz. The eigenvalues of each group are then calculated, including their maximum, minimum, mean, standard deviation, variance, skewness, kurtosis, and range of the data groups. These eigenvalues are separated into training sets, validation sets, and test sets for the two-layer BP neural network, which are then trained, validated and tested based on these eigenvalues. As for the convolution neural networks in Step 6, the sampled data need to be labeled with the corresponding activities and the eigenvalue calculation for 466 groups data does not need to be performed as in Steps 1–5.

Step 1 distinguishes between walking/running and sitting/lying, which means identifying the dynamic activities of a human body from its static attitudes. The above sets of eigenvalues are sent to the BP neural networks for network training and model validation. In these datasets, dynamic data are placed in target class 1, and static data are placed in target class 2. A comparison of the HAR accuracy of training dataset, validation dataset, test dataset, and all data are shown in Fig. 5. All accuracy values are 100%. This suggests that the dynamic activities of the subjects can be clearly identified from their static attitudes.



Figure 5: Step 1, binary classification task of static body attitudes and dynamic activities using a BP neural network

Step 2 distinguishes between walking/going upstairs/going downstairs and running. The above eigenvalues sets are sent to the BP neural networks for network training and model validation. In these datasets, the data of walking/going upstairs/going downstairs are placed in target class 1, and the running data are placed in target class 2. A comparison of HAR accuracy for training dataset, validation dataset, test dataset, and all data is shown in Fig. 6. All accuracy values are 100%. This shows that the activities of walking/going upstairs/going downstairs can be accurately distinguished from the activity of running.



Figure 6: Step 2, the binary classification task of running and walking/going upstairs/going downstairs using a BP neural network

In Step 2, the total numbers of data for target class 1 and class 2 are significantly different. The data for the three activities, walking/going upstairs/going downstairs, are placed in target class 1, while only the running data are placed in target class 2. But it is not difficult to understand that this does not affect the identification rate.

Step 3 distinguishes between walking and going upstairs/downstairs. The above eigenvalues sets are sent to the BP neural networks for network training and model validation. In these datasets, the data of going upstairs/going downstairs are placed in target class 1, and the walking data are placed in target class 2. A comparison of HAR accuracy of the training dataset, validation dataset, test dataset, and all data is shown in Fig. 7. Accuracy values of 91.9%, 90.5%, 90.5%, and 91.5% can be found. Due to the high probability, this means that the going upstairs and downstairs activities of the subjects can be clearly distinguished from their walking activities.

Step 4 distinguishes between going upstairs and going downstairs. The above eigenvalues sets are sent to the BP neural networks for network training and model validation. Among these datasets, the going upstairs data are placed in target class 1, and the going downstairs data are placed in target class 2. A comparison of the HAR accuracy of the training dataset, validation dataset, test dataset, and all data is shown in Fig. 8. It can be found that since all results are 100% correct, there is no error in distinguishing between the upstairs and downstairs activities of the subjects.



Figure 7: Step 3, the binary classification task of walking and going upstairs/going downstairs using a BP neural network



Figure 8: Step 4, the binary classification task of going upstairs and downstairs using a BP neural network

As shown in Fig. 9, Step 5 distinguishes between sitting/standing and lying. As physically understood, accelerometers measure only the acceleration of gravity when the body is at rest, not the acceleration of other body movements. Since the gravity directions of sitting/standing and lying are respectively in the longitudinal and transverse directions of the human body, these activities can be identified by using “the X-axis accelerometer samples” in the vertical direction of the human body.



Figure 9: Step 5, the binary classification task of lying and sitting/standing using a BP neural network

In addition, since the measuring devices used here was fixed separately on both ankles of the subjects, it is likely that the position and orientation of the lower leg will be the same for the two human poses of sitting and standing. Therefore, for such mounting positions, the sitting and standing poses cannot be identified by any recognition algorithm.

In the same way as Steps 1–4 above, the eigenvalues of the left accelerometer (sitting/standing data are placed in target class 1 and lying data in target class 2) are sent to the BP neural networks for network training and model validation. Fig. 9 shows the HAR accuracy for the training, validation, test, and all data. It can be found that there is no error in distinguishing between lying and sitting/standing, since all results are 100%.

The total recognition rate of the cascade model for HAR is calculated from the recognition rates of BP neural network models at each layer, as shown in Table 3. The recognition rate for walking, going upstairs and downstairs is as low as 91.5%, which was caused by the recognition in the third step and needs to be improved.

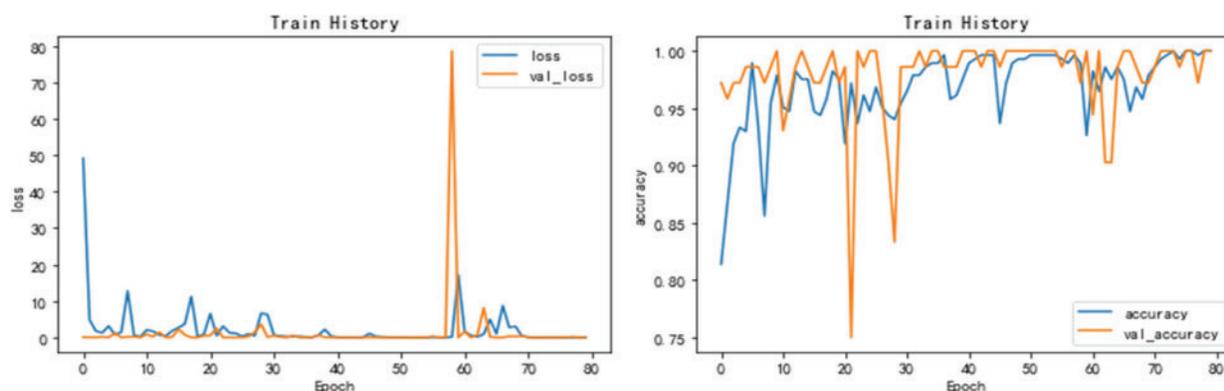
Table 3: Comprehensive recognition rate of human activities using BP neural networks

Activities	Recognition rate
Lying	100%
Sitting	100%
Standing	100%
Running	100%
Walking	91.5%
Going upstairs	91.5%
Going downstairs	91.5%

As shown in Table 4 and Figs. 10–13, Step 6 is to validate the intelligent algorithms of the CNNs for HAR. Compared with the two-layer BP neural networks in previous Steps 1–5, CNNs have no need for “Feature Engineering”, that is, the measured data can be directly fed into the CNN model, while there is no need to calculate the eigenvalues of these data in advance.

Table 4: Step 6, comprehensive recognition accuracy with 1D-CNN

Human activities	Recognition accuracy
Sitting	98.36%
Standing	98.36%
Lying	98.36%
Walking	93.83%
Running	93.83%
Going upstairs and downstairs	93.83%

**Figure 10:** Identifying dynamic and static human activities with a CNN. The HAR accuracy is 99.35% and 100% for the test and training datasets, respectively

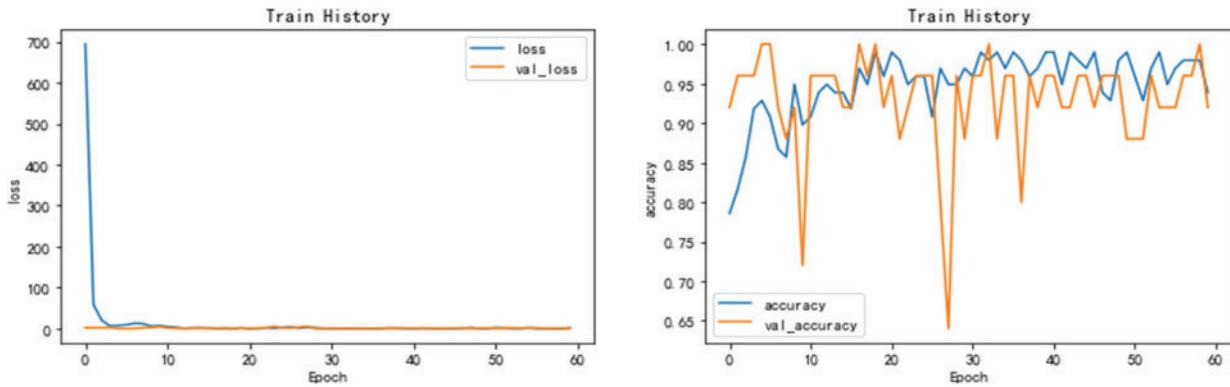


Figure 11: Identifying running and walking/going upstairs/going downstairs. The HAR accuracy is 94.44% and 97.56% for the test and training datasets, respectively

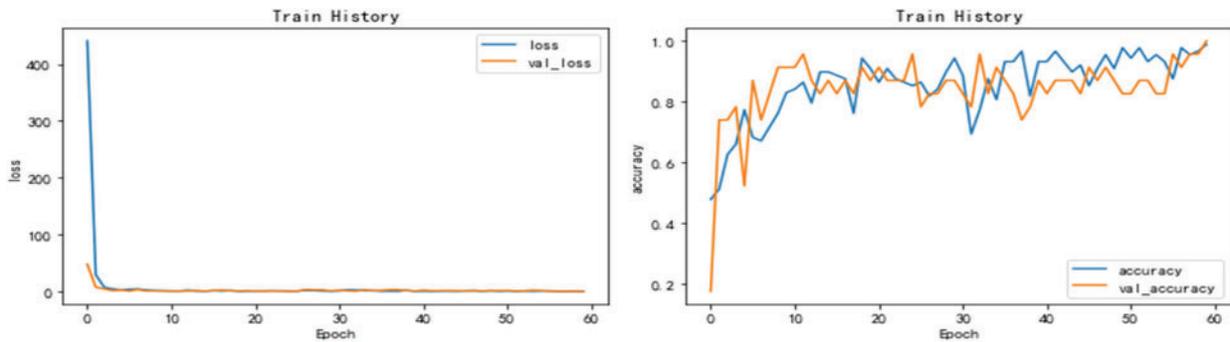


Figure 12: Identifying walking and going upstairs/downstairs. HAR accuracy is 100% on both test and training datasets

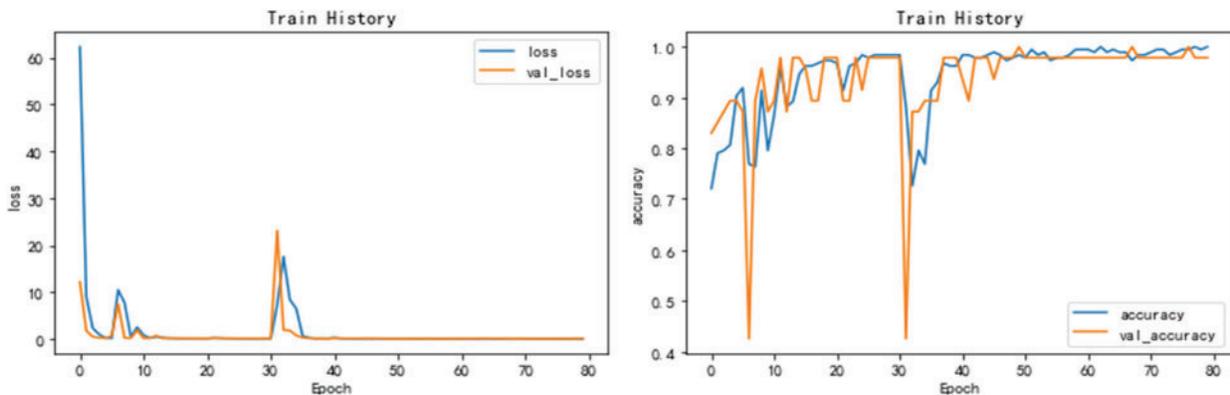


Figure 13: Identifying lying and sitting/standing. The HAR accuracy is 99.0% and 99.14% for the test and training datasets, respectively

A CNN is tested to distinguish between dynamic and static human activities as shown in Fig. 10. The HAR accuracy is 99.35% and 100% for the test and training datasets, respectively. Another CNN is used to distinguish between running and walking/going upstairs and downstairs as shown in Fig. 11.

The HAR accuracy is 94.44% and 97.56% for the test and training datasets, respectively. A third CNN is used to distinguish between walking and going upstairs and downstairs, as shown in Fig. 12. The HAR accuracy is 100% on both test and training datasets. A fourth CNN is used to distinguish between lying and sitting/standing as shown in Fig. 13. The HAR accuracy is 99.0% and 99.14% for the test and training datasets, respectively.

HAR with 1D-CNNs would fail on the task of single-step multi-class classification, distinguishing between sitting/lying/standing/walking/running/going upstairs and going downstairs simultaneously. The solution proposed in this paper is the hierarchical recognition model of “cascade layered”. In this hierarchical recognition model, a two-layer BP neural network with the same structure is used for each layer in Steps 1–5, and a slightly different 1D-CNN is used for each layer in Step 6.

In Step 6, the training dataset accounts for 70% of all data, and the test dataset accounts for the remaining 30%. Of the training dataset, 20% is used as a validation dataset. Table 4 shows the overall recognition rate with 1D-CNNs. That is, the human static behaviors of sitting, standing and lying were recognized with 98.36% accuracy and the dynamic human activities including walking, running, going upstairs and downstairs were recognized with 93.83% accuracy.

The accuracy rate of dynamic recognition in Table 4 is as low as 93.83%, while the accuracy rate of static recognition is 98.36%. Why? The immediate reason is that the HAR accuracy between running and walking/going upstairs/going downstairs is only 94.44%, and the main reason for the 94.44% accuracy is that there are too few sampled data for running compared to other activities. The theoretical basis for this conjecture is that the classification effect of CNNs is related to the total amount of data. The larger the amount of data, the higher the classification accuracy can be achieved.

4 Discussion about Confirmation Experiments

In this paper, some typical human activities are measured and identified using some specially customized measuring devices and three smartphones. Our confirmatory experiments found that standing/sitting/lying/walking/running/going upstairs and downstairs can be recognized with 90.5%–100% accuracy with only a simple BP neural network and basic CNNs. This result follows from the physical understanding of the sampling device, the sampling process, the sampling data, and the HAR algorithm. This is the prior knowledge that this paper focuses on, which leads to sufficient data preparation and algorithm design, especially the cascade model of “hierarchical recognition”.

In the first step of our HAR, it is necessary to distinguish between static poses and dynamic activities of the human body. It is easy to identify static poses and dynamic activities, as one property is static and the other is dynamic. It is true that a variety of simple algorithms can perform this binary classification task accurately.

In the second step of our HAR, running and walking (including going upstairs and downstairs) need to be distinguished. Since walking and running are movements of the same nature and similar scales, both BP neural networks and CNNs have the classification ability to distinguish them.

In an additional recognition step, we also perform a set of binary classifications to distinguish one human activity from another. The two types of human activity are either of a different nature or of the same nature and on a similar scale.

Evidently, BP neural networks and 1D CNNs are unable to accurately distinguish human activities of the same nature and similar scale while recognizing human activities of different nature.

The point we want to make here is that correct prior knowledge, including an understanding of the capabilities and characteristics of the algorithm, is essential. We believe that it is difficult to accurately identify human static postures (sitting, standing and lying), going upstairs and downstairs, walking, and running at the same time by using a single-step multi-classification model, because these classification objects are of different natures and different scales.

Essentially, single-step multi-class recognition is a research strategy adopted by numerous HAR scholars, but this research strategy ignores the understanding of algorithmic and human activity.

Different human activities have different physical properties and observational scales. For example, walking, running, going upstairs and going downstairs are all periodic behaviors. The main difference lies in the frequency and amplitude of these periodic steps, which have a timescale of seconds.

Static behavior and transient events are perfectly distinct from walking, running, going upstairs and downstairs. The main difference between static poses, such as standing and lying, is that gravity is applied in a different direction to the human body, with a timescale that can exceed 10 s; Some transient events, such as falls and collisions, differ mainly in the direction and value of the acceleration generated by the human event, which may have a timescale of milliseconds.

The three types of typical human activities, poses, and events mentioned above have perfectly different physical properties and observational scales, and therefore their key characteristics are different. Compared to a single-step multi-class recognition, a chain of binary classification recognition is significantly simpler and more accurate.

The prior knowledge contained in this point is analogous to the relation between a telescope and a microscope. By analogy, although the principles of telescopes and microscopes are the same, it is unlikely that they can be integrated into an instrument called a micro + telescope, which can resolve not only stars but also accurately observe different microscopic particles of matter. Similarly, even though various classification algorithms have different degrees of classification power, while increasing their classification accuracy, their generalization power inevitably decreases. Second, recognition tools or recognition algorithms can identify objects with the same properties at different scales, or objects with different properties at the same scale, but they typically cannot handle both scales and properties.

Therefore, simple algorithms are used to achieve the discriminative effect possessed by complex algorithms. This phenomenon includes appealing inspirations for generalized classification problems. We will discuss these inspirations further in a future study.

5 Conclusions

The HAR procedure and various HAR algorithms have been independently validated experimentally in this paper, so the key role of prior knowledge can be discussed step by step. Theoretical analysis and experimental validation show that the single-step multi-class classification task for common human activities is difficult and complex, whereas the solution of the proposed cascade structure model is relatively simple and easy. The essential difference between the two HAR strategies lies in the different understanding of intelligent algorithms.

Prior knowledge plays an essential role in HAR. The study presented in this paper demonstrates that the identification of nine common human activities from each other can be relatively clear and simple. The main reason for this effect is that proper prior knowledge is incorporated into this solution, which can simplify and improve HAR.

The task of HAR is to identify certain common human activities from each other. These human activities include standing, sitting, walking, running, lying right and left, prone, supine, going up and downstairs, etc. Activity can be measured using sensors such as accelerometers and gyroscopes. We implement HAR on the measured data using BP neural networks and deep learning networks in a cascaded structure.

Numerous existing algorithms are capable of recognizing common human activities, poses, and events. These common algorithms include Principal Component Analysis (PCA), BP neural network, Ada Boost algorithm, Random Forest (RF), Decision Trees (DT), Support Vector Machine (SVM), K-nearest Neighbor algorithm, Convolutional Neural Network (CNN), etc. However, correct prior knowledge is crucial for proper use of general algorithms. For example, BP neural networks and basic CNNs can perform accurate HAR in the cascade models, but cannot in non-cascade models because a series of cascaded binary classification tasks can sufficiently utilize the recognition capabilities of BP neural networks and basic CNNs, while other HAR solutions require considerably more complex models and algorithms to achieve similar accuracy.

This can be likened to a general prior knowledge: at the level of contemporary science and technology, it is difficult to integrate microscopes and astronomical telescopes into a single instrument. Similarly, existing general-purpose algorithms find it difficult to identify human activities with different natures, different scales, and different poses in a single step. While such a piece of prior knowledge is easy to understand, it is difficult to be applied flexibly for HAR improvement.

In fact, a thorough knowledge of the subject under study is always required. Additional comparative experiments will be presented in our future research to explore the importance of prior knowledge for HAR and pattern recognition.

Acknowledgement: The authors are deeply grateful to Dr. Jiao Wenhua for his great help.

Funding Statement: This work was supported by the Guangxi University of Science and Technology, Liuzhou, China, sponsored by the Researchers Supporting Project (No. XiaoKeBo21Z27, The Construction of Electronic Information Team Supported by Artificial Intelligence Theory and Three-Dimensional Visual Technology, Yuesheng Zhao). This work was supported by the Key Laboratory for Space-based Integrated Information Systems 2022 Laboratory Funding Program (No. Space-InfoNet20221120, Research on the Key Technologies of Intelligent Spatio-Temporal Data Engine Based on Space-Based Information Network, Yuesheng Zhao). This work was supported by the 2023 Guangxi University Young and Middle-Aged Teachers' Basic Scientific Research Ability Improvement Project (No. 2023KY0352, Research on the Recognition of Psychological Abnormalities in College Students Based on the Fusion of Pulse and EEG Techniques, Yutong Lu).

Author Contributions: Study conception and design: Jianguo Wang, Kuan Zhang, Yuesheng Zhao; Data collection: Jianguo Wang, Xiaoling Wang; Analysis and interpretation of results: Yuesheng Zhao, Xiaoling Wang, Kuan Zhang; Draft manuscript preparation: Yuesheng Zhao, Xiaoling Wang, Muhammad Shamrooz Aslam. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The public datasets used in this study, UCI and WISDM, are accessible as described in references [26] and [27]. Our self-built dataset can be accessed at the website: https://github.com/NBcaixukun/Human_pose_data.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] F. Li and J. Pan, "Human motion recognition based on triaxial accelerometer," *Journal of Computer Research and Development*, vol. 53, no. 3, pp. 621–631, 2016.
- [2] L. Zhang, Z. Kuang, K. Li, K. Wei, Z. Wang *et al.*, "Human activity behavior recognition based on acceleration sensors and neural networks," *Modern Electronics Technique*, vol. 42, no. 16, pp. 71–74, 2019.
- [3] B. Liu, H. Liu, X. Jin and D. Guo, "Human activity recognition based on smart phone sensor," *Computer Engineering and Applications*, vol. 52, no. 4, pp. 188–193, 2016.
- [4] L. Zhou, L. Lei and L. Yang, "Human behavior recognition system based on multi-sensor," *Transducer and Microsystem Technologies*, vol. 35, no. 3, pp. 89–91, 2016.
- [5] Z. Chen and Q. Zhu, "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3070–3080, 2017.
- [6] R. Yang and B. Wang, "PACP: A position-independent activity recognition method using smartphone sensors," *Information*, vol. 7, no. 4, pp. 72, 2016.
- [7] A. Bandar and J. Al Muhtadi, "A robust convolutional neural network for online smartphone-based human activity recognition," *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, vol. 35, no. 2, pp. 1609–1620, 2018.
- [8] I. Andrey, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.
- [9] G. Bhat, N. Tran, H. Shill and U. Y. Ogras, "w-HAR: An activity recognition dataset and framework using low-power wearable devices," *Sensors*, vol. 20, no. 18, pp. 5356, 2020.
- [10] C. A. Ronao and S. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, 2016.
- [11] T. Liu, S. Wang and Y. Liu, "A lightweight neural network framework using linear grouped convolution for human activity recognition on mobile devices," *The Journal of Supercomputing*, vol. 78, no. 5, pp. 6696–6716, 2022.
- [12] Y. Li and L. Wang, "Human activity recognition based on residual networks and BiLSTM," *Sensors*, vol. 22, no. 2, pp. 635, 2022.
- [13] J. He, Z. L. Guo and L. Y. Liu, "Human activity recognition technology based on sliding window and convolutional neural network," *Journal of Electronics & Information Technology*, vol. 44, no. 1, pp. 168–177, 2022.
- [14] N. T. H. Thu and D. S. Han, "HiHAR: A hierarchical hybrid deep learning architecture for wearable sensor-based human activity recognition," *IEEE Access*, vol. 9, pp. 145271–145281, 2021.
- [15] A. S. Syed, D. Sierra-Sosa and A. Kumar, "A deep convolutional neural network-XGB for direction and severity aware fall detection and activity recognition," *Sensors*, vol. 22, no. 7, pp. 2547, 2022.
- [16] Z. Xiao, X. Xu and H. Xing, "A federated learning system with enhanced feature extraction for human activity recognition," *Knowledge-Based Systems*, vol. 229, no. 5, pp. 107338, 2021.
- [17] Y. Li, S. Zhang and B. Zhu, "Accurate human activity recognition with multi-task learning," *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 4, pp. 288–298, 2020.
- [18] N. Dua, S. N. Singh and V. B. Semwal, "Multi-input CNN-GRU based human activity recognition using wearable sensors," *Computing*, vol. 103, no. 7, pp. 1461–1478, 2021.
- [19] G. Zheng, "A novel attention-based convolution neural network for human activity recognition," *IEEE Sensors Journal*, vol. 21, no. 23, pp. 27015–27025, 2021.
- [20] C. T. Yen, J. X. Liao and Y. K. Huang, "Feature fusion of a deep-learning algorithm into wearable sensor devices for human activity recognition," *Sensors*, vol. 21, no. 24, pp. 8294, 2021.
- [21] J. Wang, K. Zhang and Y. Zhao, "Human activities recognition by intelligent calculation under appropriate prior knowledge," in *Proc. of 2021 5th ACAIT*, Haikou, China, pp. 690–696, 2021.

- [22] W. S. Lima, E. Souto, K. El-Khatib, R. Jalali and J. Gama, "Human activity recognition using inertial sensors in a smartphone: An overview," *Sensors*, vol. 19, no. 14, pp. 3213, 2019.
- [23] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami *et al.*, "Deep CNN-LSTM with self-attention model for human activity recognition using wearable sensor," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 2700316, 2022.
- [24] S. B. U. D. Tahir, A. B. Dogar, R. Fatima, A. Yasin, M. Shafiq *et al.*, "Stochastic recognition of human physical activities via augmented feature descriptors and random forest model," *Sensors*, vol. 22, no. 17, pp. 6632, 2022.
- [25] N. Halim, "Stochastic recognition of human daily activities via hybrid descriptors and random forest using wearable sensors," *Array*, vol. 15, pp. 100190, 2022.
- [26] D. Anguita, A. Ghio, L. Oneto, P. P. Xavier, O. Reyes *et al.*, "A public domain dataset for human activity recognition using smartphones," in *Proc. of the 21th Int. European Symp. on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, pp. 437–442, 2013.
- [27] J. R. Kwapisz, G. M. Weiss and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.