DOI: 10.19884/j.1672-5220.202309005

A Novel Deep Learning Framework for Location Information Assisted Complex Human Activity Recognition

YU Jingwei[1], ZHANG Lei[1,2*], GAO Zhenyu[1], NI Qin[3]
1. College of Information Science and Technology, Donghua University, Shanghai 201620, China
2. Engineering Research Center of Digitalized Textile & Fashion Technology, Ministry of Education, Donghua University, Shanghai 201620, China
3. Key Laboratory of Multilingual Education with AI, Shanghai International Studies University, Shanghai 201620, China

Abstract: With the popularization of smart living and the rapid development of wearable terminal technology in recent years, sensor-based human activity recognition (HAR) has attracted widespread attention and has significant academic research and commercial application value. This paper focuses on enhancing the HAR model's recognition of users' daily simple activities (SAs) and complex activities (CAs), and proposes a deep learning (DL) model. Firstly, two publicly available datasets, UCI HAR and Shoaib CHA, are normalized. Then the characteristics of distinct activities are retrieved by the proposed model for HAR. Given the high association between users' activities and locations, location information is integrated into the dataset by the one-hot encoding technique to boost the model's classification performance. In addition, the proposed DL model is evaluated against eight traditional machine learning (ML) algorithms and six DL algorithms. Finally, the effect of various types of activities on the HAR model's recognition ability is studied. The experimental findings reveal that the proposed model achieves the highest classification accuracy on UCI HAR and Shoaib CHA datasets, with 96.77% and 99.13%, respectively. The classification accuracy of the HAR model is also greatly enhanced for both SAs and CAs by adding location information to the datasets.

Key words: human activity recognition (HAR); machine learning (ML); deep learning (DL); wearable sensor; convolutional neural network; long short-term memory (LSTM) neural network
CLC number: TP212 **Document code**: A
Article ID: 1672-5220(2024)03-0231-10

Open Science Identity (OSID)

0 Introduction

Wearable computing emphasizes a human-centered approach to creating a pervasive and ever-present computing environment. It aims to integrate devices into people's living and working environments in order to provide the best possible service. In mobile and wearable computing, human activity as a key component of the environment is a major area of current study. Simultaneously, human activity recognition (HAR) plays a significant part in human-computer interaction. Various HAR systems have been created based on diverse methods, and may be divided into three categories: activity recognition based on videos[1], WiFi signals[2] and sensors[3]. Non-invasive sensor-based activity recognition techniques have been widely employed in a variety of applications, such as smart homes[4], physical exercise[5-6] and remote health monitoring[7], and have demonstrated acceptable performance with consideration of constraints like light, obstructions, privacy and cost.

The majority of current HAR research focuses on finding solutions to the recognition of simple activities (SAs) like walking, jogging, sitting and lying down. Additionally, the issue of falls in the elderly is a hot study area[8]. However, researchers have not focused much on the issue of identifying complex activities (CAs) like working, eating, and performing chores. In reality, individuals spend lengthy amounts of time executing CAs (basic activities and various behaviors) in addition to conducting SAs (repetitive motions or single postures)[9]. CAs are usually long in duration and have high-level semantics, and they are more reflective of a user's daily life. Identifying CAs is a challenge and requires the building of sophisticated and capable models. A significant difference between CAs and SAs is the strong association between CAs and locations (typically one-dimensional (1D) information)[10]. For example, the place for biking is normally outdoors and the place for

Received date: 2023-09-15
Foundation items: National Natural Science Foundation of China (Nos. 62371118, 6210020445 and 61901104); Natural Science Foundation of Shanghai, China (Nos. 21ZR1446900 and 21511100102); Science and Technology Research Project of Shanghai Songjiang District, China (No. 20SJKJGG4C)
* Correspondence should be addressed to ZHANG Lei, email: lei.zhang@dhu.edu.cn
Citation: YU J W, ZHANG L, GAO Z Y, et al. A novel deep learning framework for location information assisted complex human activity recognition[J]. *Journal of Donghua University (English Edition)*, 2024, 41(3): 231-240.

eating is typically a restaurant. Based on the strong association between users' activities and locations in real life, obtaining information about locations not only assists in recognizing CAs, but also distinguishes activities with similar features. To determine if a user is biking or riding an exercise bike, for instance, it may be necessary to know whether he/she is outdoors or at the gym. Thus the HAR model's classification performance may be significantly enhanced by fully using the location information that corresponds to the place.

Most wearable gadgets, such as smartphones and smartwatches, are now packed with a plethora of integrated sensor devices[11], allowing researchers to gather human physiological data to monitor daily life activities[12-13]. The machine learning (ML)-based traditional HAR model may achieve great classification outcomes for SAs using sensor data such as acceleration, angular velocity, and magnetic intensity data. However, the model's recognition performance for CAs is frequently disappointing and has several limitations. On the one hand, the traditional HAR approach is based on manual methods to extract and select prominent characteristics. On the other hand, due to the fact that CAs have more complicated structures, the final selected prominent aspects frequently fail to characterize the CAs. As a result, shallow ML models are challenging to adapt to new complicated HAR sites. Deep learning (DL) significantly simplifies massive feature engineering by replacing ML's laborious feature extraction with automatic feature extraction by end-to-end neural networks. In addition to overcoming the constraints of ML, DL-based HAR models may learn more sophisticated and advanced features to improve classification performance. Gupta[14] proposed a novel hybrid DL model, convolutional neural network-gate recurrent unit (CNN-GRU), for classifying CAs. The results showed that the hybrid DL model could efficiently extract spatio-temporal features from raw sensor data and provided a higher accuracy than other DL models. Mekruksavanich et al.[15] proposed a DL model ResNet-squeeze-and-excitation (ResNet-SE) comprised of convolutional layers and residual networks to improve recognition performance in complex HAR tasks. The findings revealed that the deep residual network outperformed previous models in terms of durability and activity recognition.

Several studies have shown that, in addition to the regularly utilized acceleration, angular velocity and magnetic intensity data, the use of location data can also aid in the identification of human activities. Given that CAs are strongly related to the user's location, Peng et al.[16] built base classifiers for acceleration, vital signs and location data, respectively, and then fused the three base classifiers to ensure the independence of different contextual data. The experiments demonstrated that location data could effectively improve the performance of recognizing CAs.

The main contributions of this work are summarized as follows.

1) A novel multi-channel 1D-ResNet-BiLSTM is proposed to significantly improve the classification performance of the HAR model for SAs as well as CAs.

2) The location information is added to the experimental dataset via one-hot encoding. The model's ability of HAR is significantly improved by the addition of location characteristics.

3) The proposed multi-channel 1D-ResNet-BiLSTM model is compared with eight classical ML algorithms and six DL algorithms. The classification performance of the proposed model is also compared to that in previous studies.

4) The association between SAs/CAs and sensor position data is further investigated. By subdividing the behavior types, we explore the model's classification performance for different position sensor data under different behavior types.

1 Experiments

1.1 Experimental materials and methods

Figure 1 depicts the flowchart and the four basic phases of the research: data collection based on wearable sensors; data processing as well as data segmentation using a sliding window; training classification models; classification evaluation of the models for various behaviors in the sensor data.

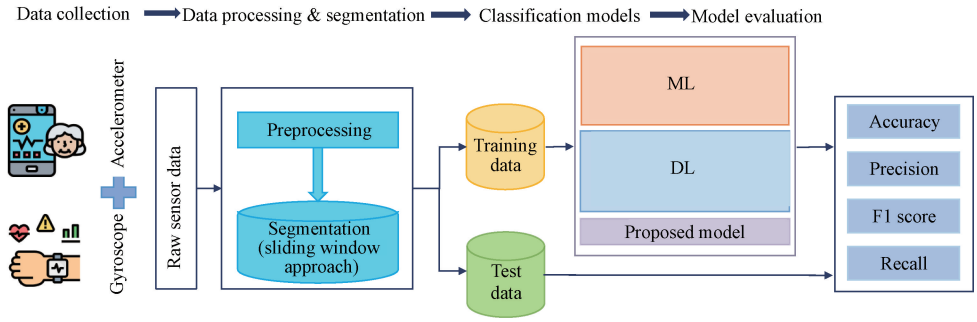


Fig. 1 HAR flowchart

1.1.1 Experimental dataset

In this study, we use two datasets: the UCI HAR dataset[17] and the Shoaib CHA dataset[18]. The UCI HAR dataset and the Shoaib CHA dataset involve SAs and CAs, respectively. Table 1 provides a summary of the experimental datasets.

Table 1 Experimental dataset

Dataset	Subject count	Age/year	Device	Sensor
UCI HAR	30	19–48	Smartphone (Samsung Galaxy SII)	Tri-axial accelerometer and gyroscope
Shoaib CHA	10	23–35	Two smartphones (Samsung Galaxy SII)	Tri-axial accelerometer, gyroscope, magnetometer and linear accelerometer

Dataset	Sensor position	Sampling frequency/Hz	Activity
UCI HAR	Waist	50	Walking, lying, sitting, standing and walking upstairs and downstairs
Shoaib CHA	Right pocket and right wrist	50	Walking, lying, sitting, standing, walking upstairs and downstairs, biking, typing, writing, drinking, eating, talking and smoking

Given the high association between activities and locations, location information was employed to improve the model's recognition of CAs. While the Shoaib CHA dataset does not record location information, location characteristics are assigned to each activity based on prior knowledge. This study defines five locations for the Shoaib CHA dataset: bedrooms, restaurants, offices, corridors and outdoors. The same activity might occur in more than one location. For instance, sitting can take place in offices, restaurants and bedrooms. In the Shoaib CHA dataset, there are 90 000 examples for each activity. It is assumed that each activity occurs at the same frequency in each feasible location, i. e., the likelihood of sitting occurring at each location is 1/3. As a result, the activity is evenly dispersed with 30 000 samples at each location. The samples for the remaining activities are similarly allocated to different locations as shown in Table 2.

Table 2 Activities and corresponding locations in Shoaib CHA dataset

Location	Activity
Bedroom	Walking, sitting and standing
Restaurant	Walking, sitting, standing, drinking, talking and eating
Office	Walking, sitting, standing, typing, writing, drinking and talking
Corridor	Standing and walking upstairs and downstairs
Outdoor	Walking, jogging, standing, biking and smoking

Location is typically 1D information. These 1D traits are also challenging to employ in detecting complicated activities. As a result, one-hot encoding is used on the location information in this study, and the one-hot encoded location information is utilized as location characteristics. Table 3 displays the location code.

Table 3 One-hot encoding of different locations

Location	Code
Bedroom	10000
Restaurant	01000
Office	00100
Corridor	00010
Outdoor	00001

1.1.2 Data processing

The raw data acquired by the sensor are pre-processed. The data are adjusted using Z-score normalization so that the mean is 0 and the standard deviation is 1. Data cleaning is done by removing duplicates and null values. Before training the classification model, the raw time series data must be partitioned into multiple windows by selecting the appropriate window size. These windows are either fed directly to the DL network or used for feature extraction in classical ML algorithms. Longer windows can contain more data and are expected to achieve a higher classification accuracy. Shorter windows can detect activity changes faster. In this study, a sliding window with a window size of 90 and an overlap of 50% is selected to segment the preprocessed sensor data, as shown in Fig. 2. For training, the dataset is further randomly divided into a training set (70%) and a test set (30%).

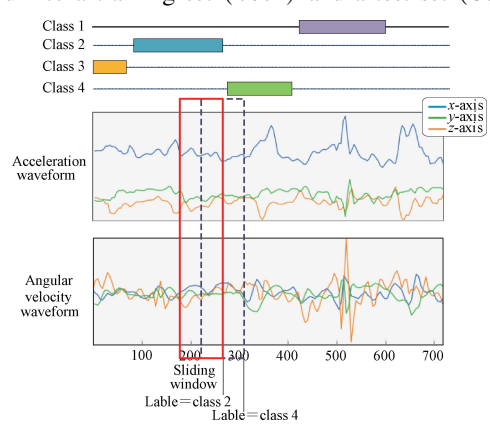


Fig. 2 Data segmentation using overlapping sliding windows

1.2 Classifiers

1.2.1 Classical ML models

Eight ML models are employed: support vector machine (SVM)[19], random forest (RF)[20], Naive Bayes (NB)[21], *k*-nearest neighbors (KNN)[22], stochastic gradient descent (SGD)[23], logistic regression (LR)[24], bagging[25] and gradient Boosting decision tree (GBDT)[26].

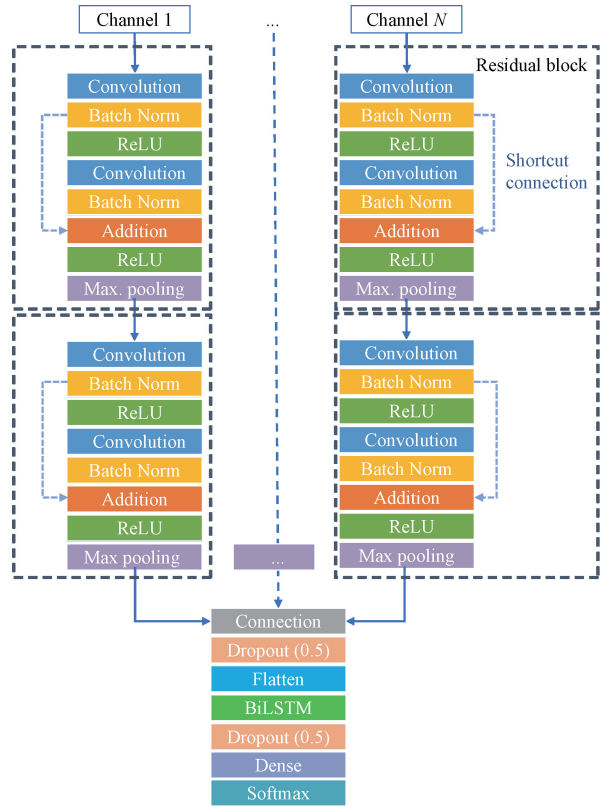
SVM, as a supervised learning algorithm, is a generalized linear classifier for two (multiple) classification problems, and its kernel function can be used to solve nonlinear problems. In this work, a radial basis kernel function (of type rbf) is used as the kernel function of SVM to map the samples to a higher dimensional space. In contrast to the verdict method used by the majority of other ML classification algorithms, NB is a typical generative learning method. In this work, a Gaussian plain Bayesian classifier is chosen for continuous data from sensors. KNN assigns the objects to be classified into the class of nearest neighbors based on their feature values, with the goal of representing the classification of the target data by the classification of the *k* closest sample data to the target. The Euclidean distance is chosen as the distance measure in this study. Furthermore, the choice of *k* value is critical to the final classification performance, and *k* is 7. SGD is mostly utilized for discriminative learning of linear classifiers with convex loss functions (support vector machines or logistic regression). In our experiments, we use a loss function of type hinge which is a (soft-margin) linear support vector machine. As a generalized linear regression analysis model, LR is a machine learning technique for resolving classification issues.

This research employs three classical ensemble models: RF, bagging and GBDT. RF, a more advanced method based on decision trees that can be readily trained in parallel for regression and classification applications with large-scale samples, is a sort of ensemble learning model. In our experiments, we set the *n*_estimators parameter which determines the number of numbers in the forest to be the default value of 100. The bagging algorithm creates training data by employing a sampling approach with put-back, which essentially introduces a sample perturbation that reduces variance by increasing sample unpredictability. Boosting is the process of combining numerous weak learners to generate a single strong learner. The idea of gradient Boosting, as a large class of algorithms in Boosting, is borrowed from the gradient descent method. Its basic principle is to train the newly added weak classifiers based on the negative gradient information of the current model's loss function, and then combine the trained weak classifiers into the existing model in the form of accumulation. The GBDT approach employs decision trees as weak classifiers. We used the log-likelihood loss function "deviance" in the experiments and set the *n*_estimators, i.e., the maximum number of iterations for weak learners, to 100.

1.2.2 DL models

DL models such as convolutional neural networks (CNNs) and long short-term memory (LSTM) neural networks can automatically extract local spatial and temporal features of sensor signals by learning a large number of samples, and have been widely used in sensor-based HAR tasks due to their powerful feature extraction capabilities[27-29]. ResNet[30], as an improved CNN model, can train very deep end-to-end networks with strong local spatial feature extraction by using residual connections. In HAR, hybrid models combining different DL architectures, such as 1DCNN with LSTM (1DCNN-LSTM) and 1DCNN with BiLSTM (1DCNN-BiLSTM), demonstrate significant performance advantages. These models, by integrating the features of their respective architectures, effectively extract features and process time-series data, thereby achieving high accuracy and efficiency in recognizing and analyzing complex behavior patterns.

Our proposed multi-channel 1D-ResNet-BiLSTM model is shown in Fig. 3.



Norm—normalization; Max—maximum.

Fig. 3 Proposed multi-channel 1D-ResNet-BiLSTM model

Before being input into the neural network, the signals from each sensor channel are pre-processed and data segmented. Following that, the processed signals are passed into the residual block for processing. Each residual block consists of two convolutional layers, two ReLU activation layers, two batch normalization layers, and a maximum pooling layer, where each convolutional

layer is followed by a batch normalization layer and a ReLU activation layer. To avoid the danger of model overfitting, each residual block eventually contains a maximum pooling layer for feature dimensionality reduction. Furthermore, the remaining blocks' shortcut connection eliminates the gradient vanishing problem. After the signals from each sensor channel have been extracted from the local spatial features by two residual blocks, these features are then fused and input to the dropout layer. Then, the features that have been flattened in the flatten layer are input to the BiLSTM to further extract features in the time dimension. A dropout layer is set after the BiLSTM, and by using dropout regularization, the dependency between neurons can be reduced and the generalization ability of the model can be improved. After the dense layer, the proposed model finally outputs the category probabilities for each of the different activity types through the Softmax layer.

1.3 Experimental setup

All classification algorithms in this work were implemented in Python 3.6. The detailed experimental configuration is shown in Table 4.

Table 4 Experimental configuration

Experimental tool	Python interpreter	ML	Deep neural network	System
Configuration	Python 3.6	Scikit-learn 0.24.2 machine learning toolkit	Tensorflow 1.15.0 framework, Pandas 1.1.5, and Numpy 1.19.5 environments	A laptop with 16.0 GB RAM, 3.20 GHz AMD Ryzen 7 6800H processor and 64-bit operating system

2 Results and Discussion

Accuracy A is used as a measure of classification quality to evaluate the performance of the proposed HAR model.

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + F_N},$$

where T_P refers to the number of samples that are actually positive and are classified as positive; T_N refers to the number of samples that are actually negative and are classified as negative; F_P refers to the number of samples that are actually negative but are classified as positive; F_N refers to the number of samples that are actually positive but are classified as negative. Additionally, in our experimental results, we also present performance evaluation metrics including precision P, sensitivity S (i.e., recall R) and the F1 score F:

$$P = \frac{T_P}{T_P + F_P},$$

$$S = R = \frac{T_P}{T_P + F_N},$$

$$F = \frac{2PR}{P + R}.$$

Based on the sensor data in the UCI HAR dataset, the classification performance of the multi-channel 1D-ResNet-BiLSTM model is compared with that of eight ML algorithms, and the results obtained are listed in Table 5. Table 5 shows that, among the eight classical ML models, SVM has the best classification results, with an accuracy of 88.72%. However, the other classical single models perform poorly in classifying sensor data, with a classification accuracy below 80%. Better classification outcomes may be obtained using ensemble models (RF, bagging and GBDT). In contrast, the proposed model exhibits the best classification performance on the UCI HAR dataset with a classification accuracy of 96.77%.

Table 5 Classification results of several algorithms on UCI HAR dataset

Classifier	Accuracy/%	Precision/%	Recall/%	F1 score/%
RF	84.22	84.22	84.21	84.22
NB	72.48	72.48	72.47	72.47
KNN	65.89	65.89	65.89	65.88
SGD	59.31	59.32	59.31	59.30
LR	63.56	63.57	63.57	63.56
SVM	88.72	88.73	88.73	88.71
Bagging	84.15	84.15	84.15	84.15
GBDT	87.21	87.21	87.21	87.20
Proposed model	96.77	96.78	96.77	96.77

The classification abilities of several DL models on the UCI HAR dataset are evaluated as shown in Fig. 4. It is observed that the proposed model has the highest accuracy (96.77%). Additionally, the hybrid model

also produces superior classification outcomes. The confusion matrix produced by the deep hybrid models on the UCI HAR dataset is shown in Fig. 5. It can be shown that sitting and standing are easily misclassified. The misclassification of walking upstairs and walking downstairs is also in a high possibility due to the similarities of the activities.

In order to evaluate the impact of user's location data on the classification performance of the model, experiments are further developed by various classical ML algorithms as well as DL algorithms based on sensor data from wrist and pocket locations on the Shoaib CHA dataset. The obtained results are listed in Table 6. Based on a comprehensive analysis of Table 6, we observe a significant phenomenon: when location information is integrated into the raw sensor data collected from the wrist and pocket (i. e., Wrist-lo and Pocket-lo), the classification accuracy of all models significantly improves compared to using only the raw data. This finding highlights the important role of location information in enhancing the interpretation of sensor data and improving the performance of model classification. Furthermore, all models can obtain superior classification accuracy for sensor data from the wrist as compared to sensor data from the pocket.

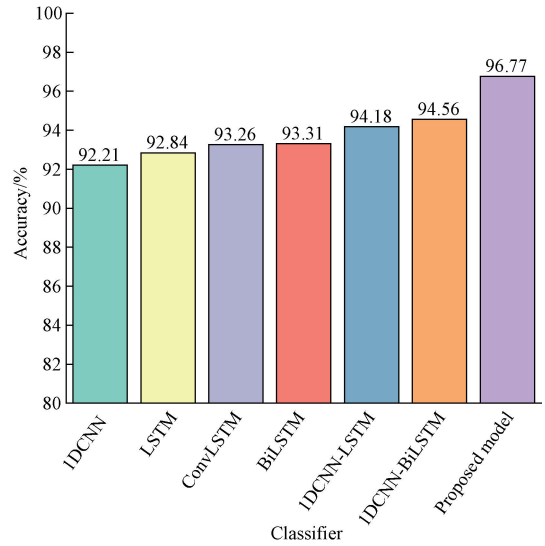


Fig. 4 Classification accuracy of DL models on UCI HAR dataset

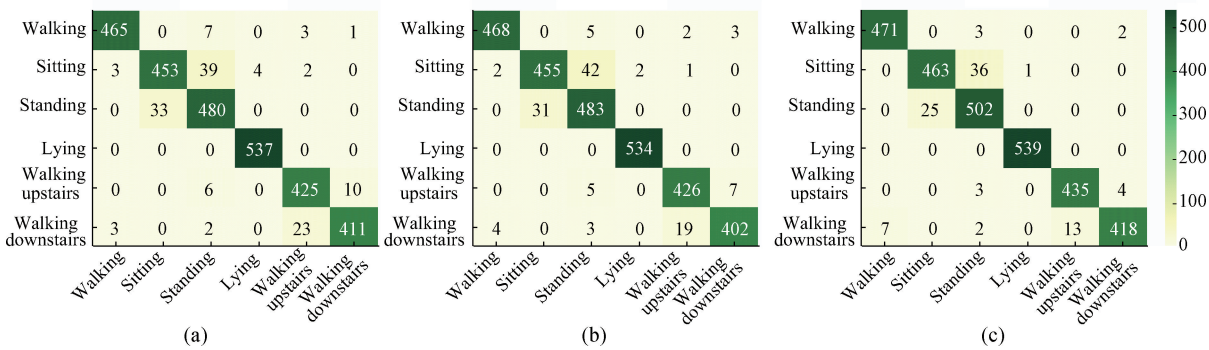


Fig. 5 Confusion matrix generated by hybrid models on UCI HAR dataset: (a) IDCNN-LSTM; (b) 1DCNN-BiLSTM; (c) proposed model

Table 6 Classification results of several algorithms on Shoaib CHA dataset

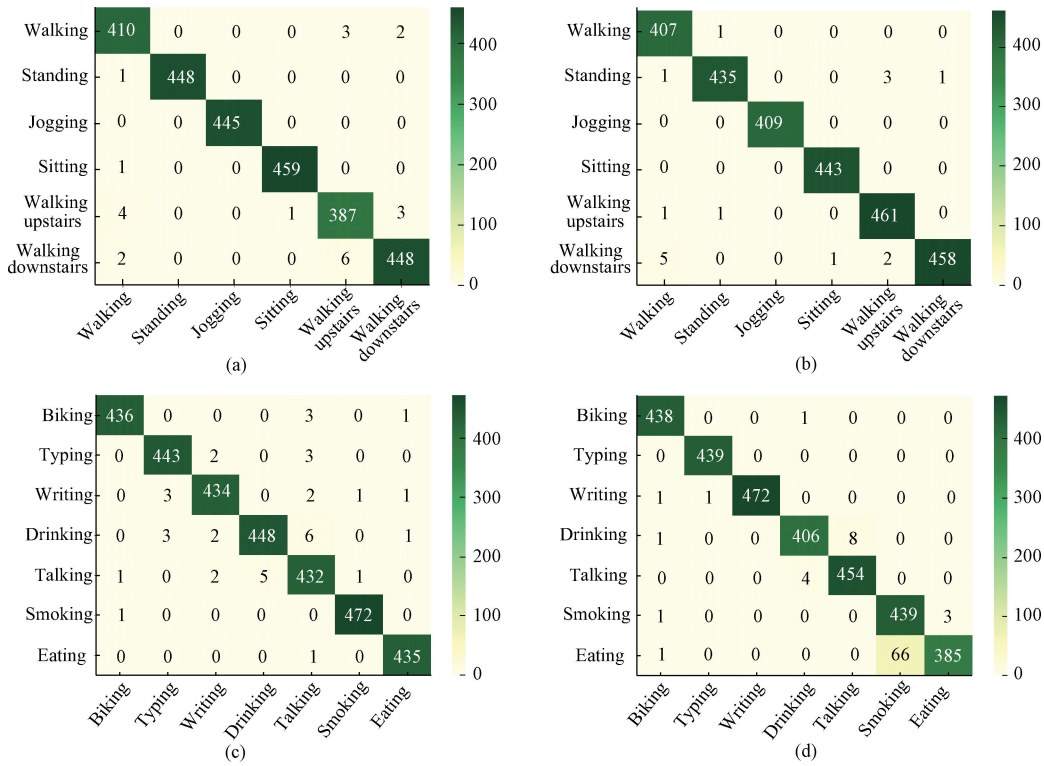
Classifier	Accuracy/%			
	Wrist-lo	Wrist	Pocket-lo	Pocket
RF	**97. 75**	97. 69	97. 68	97. 56
NB	**78. 41**	70. 07	72. 35	67. 48
KNN	**96. 08**	95. 89	95. 46	95. 13
SGD	**60. 01**	51. 17	56. 39	46. 18
LR	**71. 06**	62. 30	63. 41	53. 02
SVM	**94. 57**	94. 31	94. 25	94. 01
Bagging	**96. 36**	96. 25	96. 12	95. 89
GBDT	**97. 53**	97. 41	97. 42	96. 35
1DCNN	**97. 40**	97. 37	97. 33	96. 96
LSTM	**97. 67**	97. 45	97. 35	96. 15
BiLSTM	**97. 93**	97. 74	97. 71	97. 63
ConvLSTM	**98. 34**	98. 25	98. 23	98. 17
1DCNN-LSTM	**98. 73**	98. 69	98. 70	98. 61
1DCNN-BiLSTM	**98. 77**	98. 45	98. 35	98. 15
Proposed model	**99. 13**	98. 81	98. 52	98. 36

To explore the impact of various types of activities on recognition performance, the sensor data in the Shoaib CHA dataset are divided into three groups: SAs only, CAs only and all activities. Then, using 1DCNN-BiLSTM and the proposed model, trials are run on each group independently. The results are shown in Table 7. By using location data, the model's classification accuracy improves dramatically for both SAs and CAs. Furthermore, for SAs, the model performs better for sensing input from the pocket. The model works better for sensing input from the wrist for CAs and all activities.

Table 7 Classification model results for different activities

	Classifier	Accuracy/%			
		Wrist-lo	Wrist	Pocket-lo	Pocket
SA	1DCNN-BiLSTM	99.31	98.90	**99.40**	99.25
	Proposed model	99.45	99.17	**99.53**	99.38
CA	1DCNN-BiLSTM	**98.87**	98.53	98.45	98.26
	Proposed model	**99.18**	98.96	98.67	98.47
All	1DCNN-BiLSTM	**98.77**	98.45	98.35	98.15
	Proposed model	**99.13**	98.81	98.52	98.36

Figure 6 shows the proposed model's confusion matrix for SAs, CAs and all activities, where trials are performed for each behavior type by using sensor data from the wrist and pocket portions, respectively. For sensor data from the wrist, the model exhibits a higher misjudgment rate for SAs like walking upstairs and downstairs (Fig. 6(a)). By comparing Figs. 6(a) and 6(b), it is clear that mistake rates for SAs are higher when sensor data from the wrist is used than those when sensor data from the pocket is used. When sensor data from the pocket is used, the model consistently misidentifies smoking at a very high rate (Fig. 6(d)). As can be observed from Figs. 6(c) and 6(d), the model is better at detecting CAs when sensor data from the wrist is used. By comparing Figs. 6(e) and 6(f), the model provides superior classification performance for sensor data from the wrist to sensor data from the pocket for all activities. It can be deduced that when SA patterns are recognized, the model may obtain improved classification performance with sensor data from the lower limbs (pockets). When the identified activities include CAs, the model performs better with upper limb (wrist) sensor data.



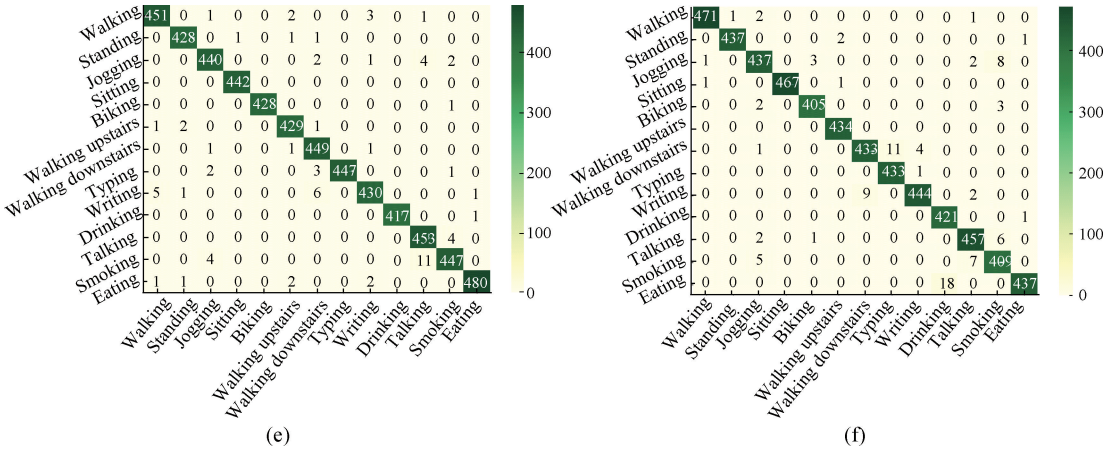


Fig. 6 Confusion matrix of proposed model for SAs, CAs and all activities: (a) classification of SA using wrist sensor data; (b) classification of SA using pocket sensor data; (c) classification of CA using wrist sensor data; (d) classification of CA using pocket sensor data; (e) classification of all using wrist sensor data; (f) classification of all using pocket sensor data

We compare the accuracy of the proposed model to the models in existing references using the two datasets, as shown in Table 8. According to the results, the hybrid model proposed in this study obtains the maximum accuracy on UCI HAR and Shoaib CHA datasets, with 96.77% and 99.13%, respectively.

Table 8 Accuracy comparison of proposed model with other classifiers in prior studies

Dataset	Model	Accuracy/%
UCI HAR	CNN-GRU[31]	96.67
	DMEFAM[32]	96.00
	Multibranch CNN-BiLSTM[33]	96.37
	Proposed model	**96.77**
Shoaib CHA	ResNet-SE[15]	98.67
	CNN-BiGRU[34]	98.78
	Deep Stacked Autoencoder[35]	97.13
	Proposed model	**99.13**

3 Conclusions

In this study, we propose a DL model (multi-channel 1D-ResNet-BiLSTM) with the goal of enhancing the HAR model's capability to identify users' SAs and CAs. Given the strong association between users' activities and locations, the one-hot encoding approach is used to assign location characteristics to each activity in the dataset based on prior information. The experimental findings suggest that the incorporation of location information improves the classification accuracy of the HAR model greatly when compared to the original data gathered from wrists and pockets. As for SAs, the model has a higher classification accuracy with lower limb (pocket) sensor data. When CAs are involved, the model performs better when sensor data from the upper limbs (wrist) is used. By comparing the proposed model with eight classical ML algorithms and six DL algorithms, it is found that the proposed model has the best classification performance on both UCI HAR and Shoaib CHA datasets, attaining 96.77% and 99.13% classification accuracy, respectively.

References

[1] ANDRADE-AMBRIZ Y A, LEDESMA S, IBARRA-MANZANO M A, et al. Human activity recognition using temporal convolutional neural network architecture[J]. *Expert Systems with Applications*, 2022, 191: 116287.

[2] LIN G P, JIANG W W, XU S C, et al. Human activity recognition using smartphones with WiFi signals[J]. *IEEE Transactions on Human-Machine Systems*, 2023, 53(1): 142-153.

[3] XU S G, ZHANG L, HUANG W B, et al. Deformable convolutional networks for multimodal human activity recognition using wearable sensors[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 1-14.

[4] LI Y, YANG G C, SU Z D, et al. Human activity recognition based on multienvironment sensor data[J]. *Information Fusion*, 2023, 91: 47-63.

[5] TANG B, GUAN W. CNN multi-position wearable sensor human activity recognition used in basketball training[J]. *Computational Intelligence and Neuroscience*, 2022, 2022: 9918143.

[6] MEKRUKSAVANICH S, JANTAWONG P, HNOOHOM N, et al. Badminton activity recognition and player assessment based on motion signals using deep residual network[C]// 2022 IEEE 13th International Conference on Software Engineering and Service Science

(ICSESS). New York: IEEE, 2022: 80-83.
[7] DASKALOS A C, THEODOROPOULOS P, SPANDONIDIS C, et al. Wearable device for observation of physical activity with the purpose of patient monitoring due to COVID-19 [J]. *Signals*, 2022, 3(1): 11-28.
[8] NADEEM A, MEHMOOD A, RIZWAN K. A dataset build using wearable inertial measurement and ECG sensors for activity recognition, fall detection and basic heart anomaly detection system[J]. *Data in Brief*, 2019, 27: 104717.
[9] PENG L Y, CHEN L, WU X J, et al. Hierarchical complex activity representation and recognition using topic model and classifier level fusion [J]. *IEEE Transactions on Biomedical Engineering*, 2017, 64(6): 1369-1379.
[10] WU S Y, FAN H H. Activity-based proactive data management in mobile environments [J]. *IEEE Transactions on Mobile Computing*, 2010, 9(3): 390-404.
[11] THAKUR D, BISWAS S. Smartphone based human activity monitoring and recognition using ML and DL: a comprehensive survey [J]. *Journal of Ambient Intelligence and Humanized Computing*, 2020, 11(11): 5433-5444.
[12] ABD RAHIM K N K, ELAMVAZUTHI I, IZHAR L, et al. Classification of human daily activities using ensemble methods based on smartphone inertial sensors[J]. *Sensors*, 2018, 18(12): 4132.
[13] JAIN A, KANHANGAD V. Human activity classification in smartphones using accelerometer and gyroscope sensors [J]. *IEEE Sensors Journal*, 2018, 18(3): 1169-1177.
[14] GUPTA S. Deep learning based human activity recognition (HAR) using wearable sensor data [J]. *International Journal of Information Management Data Insights*, 2021, 1(2): 100046.
[15] MEKRUKSAVANICH S, JITPATTANAKUL A, SITTHITHAKERNGKIET K, et al. ResNet-SE: channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors[J]. *IEEE Access*, 2022, 10: 51142-51154.
[16] PENG L Y, CHEN L, WU M H, et al. Complex activity recognition using acceleration, vital sign, and location data [J]. *IEEE Transactions on Mobile Computing*, 2019, 18(7): 1488-1498.
[17] ANGUITA D, GHIO A, ONETO L, et al. A public domain dataset for human activity recognition using smartphones [C]//The European Symposium on Artificial Neural Networks (ESANN). Bruges: [s. n.], 2013, 437-442.
[18] SHOAIB M, BOSCH S, INCEL O, et al. Complex human activity recognition using smartphone and wrist-worn motion sensors[J]. *Sensors*, 2016, 16(4): 426.
[19] CRISTIANINI N, SHAWE-TAYLOR J. An introduction to support vector machines: and other kernel-based learning methods [M]. Cambridge: Cambridge University Press, 2000.
[20] QUINLAN J R. Improved use of continuous attributes in C4.5 [J]. *Journal of Artificial Intelligence Research*, 1996, 4: 77-90.
[21] RUSSELL S J. Artificial intelligence a modern approach[M]. New York: Pearson Education, Inc., 2010.
[22] AHA D W, KIBLER D, ALBERT M K. Instance-based learning algorithms[J]. *Machine Learning*, 1991, 6(1): 37-66.
[23] ROBBINS H, MONRO S. A stochastic approximation method [J]. *The Annals of Mathematical Statistics*, 1951, 22(3): 400-407.
[24] KUTNER M H. Applied linear statistical models [M]. 5th ed. Boston: McGraw-Hill Irwin, 2005.
[25] BREIMAN L. Bagging predictors[J]. *Machine Learning*, 1996, 24(2): 123-140.
[26] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. *The Annals of Statistics*, 2001, 29(5): 1189-1232.
[27] BARUT O, ZHOU L, LUO Y. Multitask LSTM model for human activity recognition and intensity estimation using wearable sensor data[J]. *IEEE Internet of Things Journal*, 2020, 7(9): 8760-8768.
[28] BALOCH Z, SHAIKH F K, ALI UNAR M. CNN-LSTM-based late sensor fusion for human activity recognition in big data networks [J]. *Wireless Communications and Mobile Computing*, 2022, 2022: 3434100.
[29] AHMAD Z, KHAN N. CNN-based multistage gated average fusion (MGAF) for human action recognition using depth and inertial sensors[J]. *IEEE Sensors Journal*, 2021, 21(3): 3623-3634.
[30] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 770-778.
[31] LU L M, ZHANG C L, CAO K, et al. A multichannel CNN-GRU model for human activity recognition[J]. *IEEE Access*, 2022, 10: 66797-66810.
[32] WANG Y, XU H J, LIU Y X, et al. A novel deep multifeature extraction framework based on attention mechanism using wearable sensor data for human activity recognition[J]. *IEEE Sensors Journal*, 2023, 23(7): 7188-7198.

[33] CHALLA S K, KUMAR A, SEMWAL V B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data [J]. *The Visual Computer*, 2022, 38(12): 4095-4109.
[34] MEKRUKSAVANICH S, JITPATTANAKUL A. Deep convolutional neural network with RNNs for complex activity recognition using wrist-worn wearable sensor data[J]. *Electronics*, 2021, 10(14): 1685.
[35] ALO U R, NWEKE H F, TEH Y W, et al. Smartphone motion sensor-based complex human activity identification using deep stacked autoencoder algorithm for enhanced smart healthcare system[J]. *Sensors*, 2020, 20(21): 6300.

用于位置信息辅助复杂人体行为识别的新型深度学习框架

于静伟[1], 张 磊[1,2*], 高震宇[1], 倪 琴[3]
1. 东华大学 信息科学与技术学院, 上海 201620
2. 东华大学 数字化纺织服装技术教育部工程研究中心, 上海 201620
3. 上海外国语大学 多语种人工智能教育重点实验室, 上海 201620

摘 要: 随着近年来智能生活理念的普及和可穿戴终端技术的快速发展, 基于传感器数据的人体行为识别(human activity recognition, HAR)已引起广泛关注, 并且具有重要的学术研究和商业应用价值。该文研究了增强 HAR 模型对用户日常简单行为(simple activity, SA)和复杂行为(complex activity, CA)的识别, 并提出了一个深度学习(deep learning, DL)模型。首先, 使用两个可公开获取的数据集 UCI HAR 和 Shoaib CHA, 并对其进行标准化处理。其次, 使用所提出的模型提取各种动作的特征, 进行人体行为识别。鉴于用户行为和位置之间的高度关联, 通过独热编码技术将位置信息集成到数据集中, 以提高模型的分类性能。此外, 将所提出的模型与 8 种经典机器学习(machine learning, ML)算法和 6 种 DL 算法进行了对比。最后, 评估了不同行为类型对 HAR 模型识别性能的影响。实验结果表明, 所提出的模型在 UCI HAR 和 Shoaib CHA 数据集上的最高分类准确率分别达到了 96.77%和 99.13%。通过向数据集添加位置信息, HAR 模型对 SA 和 CA 的分类准确率得到了显著提高。

关键词: 人体行为识别(HAR); 机器学习(ML); 深度学习(DL); 可穿戴传感器; 卷积神经网络; 长短期记忆(LSTM)神经网络