Human Activity Recognition with a Time Distributed Deep Neural Network

Gunjan Pareek[1], Swati Nigam[1,2](✉), Anshuman Shastri[2], and Rajiv Singh[1,2]

[1] Department of Computer Science, Banasthali Vidyapith, Rajasthan 304022, India
swatinigam.au@gmail.com

[2] Centre for Artificial Intelligence, Banasthali Vidyapith, Rajasthan 304022, India
anshumanshastri@banasthali.in

Abstract. Human activity recognition (HAR) is necessary in numerous domains, including medicine, sports, and security. This research offers a method to improve HAR performance by using a temporally distributed integration of convolutional neural networks (CNN) and long short-term memory (LSTM). The proposed model combines the advantages of CNN and LSTM networks to obtain temporal and spatial details from sensor data. The model efficiently learns and captures the sequential dependencies in the data by scattering the LSTM layers over time. The proposed method outperforms baseline CNN, LSTM, and existing models, as shown by experimental results on benchmark datasets UCI-Sensor and Opportunity-Sensor dataset and achieved an accuracy of 97% and 96%, respectively. The results open up new paths for real-time applications and research development by demonstrating the promise of the temporally distributed CNN-LSTM model for improving the robustness and accuracy of human activity recognition from sensor data.

Keywords: Sensor Data · Action Identification · Convolution Neural Network · Time Distributed Feature Extraction · Long Short-Term Memory

1 Introduction

Healthcare monitoring, sports analysis, and human-computer interaction (HCI) are just a few of the areas where human activity recognition (HAR) is finding increasing use [1–3]. Recognizing human behavior accurately from sensor data in real-time is crucial for delivering individualized and contextualized support. Recent years have seen encouraging results from deep learning models in HAR, since they can automatically generate discriminative characteristics from raw sensor data.

Owing to their different strengths in capturing temporal and spatial correlations, deep learning architectures such as convolutional neural networks (CNN) [6] and long short-term memory (LSTM) [4] have been extensively used. However, there are limits to using either CNN or LSTM models. While CNN models are more inclined towards spatial data than temporal dynamics, LSTM models have difficulty capturing long-term relationships.

B. J. Choi et al. (Eds.): IHCI 2023, LNCS 14532, pp. 127–136, 2024.
https://doi.org/10.1007/978-3-031-53830-8_13

In order to alleviate the existing limitations, we offer a technique to improve HAR performance by applying a time-distributed CNN-LSTM model to sensor data. To extract both temporal and spatial characteristics from sensor input, the temporally distributed CNN-LSTM network associates the improvements of CNN and LSTM architectures. To better recognize activity patterns across time, the proposed model uses a time-distributed LSTM to capture the sequential dependencies in the data. However, the model can gather important information across several sensor channels since the CNN layers focus on extracting spatial characteristics from sensor input.

The aim of this study is to assess the effectiveness of the proposed time-distributed CNN-LSTM model in enhancing HAR, relative to the conventional CNN and LSTM models. We test the model's efficiency using publically available datasets. We expect the suggested technique to greatly enhance the accuracy and reliability of human activity detection from sensor data by harnessing the combined strength of CNN and LSTM architectures.

The remainder of the paper is structured as follows: In Sect. 2, we examine relevant work in the area of sensor data and the limits of existing methods. The proposed temporally distributed CNN-LSTM model is described in depth, including its architecture and training method, in Sect. 3. The experimental design, including datasets, measures for success, and implementation specifics, is presented in Sect. 4. The research finishes with Sect. 5, in which the contributions are summarized.

2 Related Works

There has been a lot of work done on HAR. Researchers have investigated a wide range of approaches to improve the robustness and precision of action recognition systems. Here, we summarize recent research that has improved HAR using sensor data.

Representational analysis of neural networks for HAR using transfer learning is described by An et al. [1]. To compare and contrast the neural network representations learned for various activity identification tasks, they proposed a transfer learning strategy. The results show that the suggested strategy is useful for increasing recognition accuracy with little additional training time.

Ismail et al. [2] offer AUTO-HAR, an automated CNN architecture design-based HAR system. They present a system that mechanically generates an activity-recognition-optimized CNN structure. The recognition performance is enhanced due to the framework's excellent accuracy and flexibility across datasets.

A storyline evaluation of HAR in an AI frame is provided by Gupta et al. [4]. This study compiles and assesses many techniques, equipment, and datasets that have been requested for the problem of human activity identification. It gives an overview of the state-of-the-art techniques and talks about the difficulties and potential future developments in this area.

Gupta et al. [6] offer a method for HAR based on deep learning and the information gathered from wearable sensors. In particular, convolutional and recurrent neural networks, two types of deep learning models, are investigated for their potential. Findings show that the suggested method is efficient in obtaining high accuracy for activity recognition.

A transfer learning strategy for human behaviors employing a cascade neural network architecture is proposed by Du et al. [11]. The approach takes the lessons acquired from one activity recognition job and applies them to another similar one. This research shows that the cascade neural network design is superior at identifying commonalities across different types of motion.

Wang et al. [13] provide a comprehensive overview of deep learning for HAR based on sensor data. Their work summarizes many deep learning models and approaches that have been applied to the problem of activity recognition. It reviews the previous developments and talks about the difficulties and potential future paths.

For HAR using wearable sensors, CNN is proposed by Rueda et al. [14]. The research probes several CNN designs and delves into the merging of sensor data from various parts of the body. Findings prove that CNN can reliably identify actions from data collected by sensors worn on the body.

A multi-layer parallel LSTM network for HAR using smartphone sensors is presented by Yu et al. [15]. In order to extract both three-dimensional and sequential characteristics from sensor input, the network design makes use of parallel LSTM layers. The experimental findings demonstrate the effectiveness of the proposed network in performing activity recognition tasks. A few other methods are described in [15–17].

3 The Proposed Method

The block diagram of the proposed method for improving human activity identification using a temporally distributed CNN-LSTM model using sensor data is shown in Fig. 1. Each component of the block diagram is described here.

3.1 Input Dataset

The study uses two datasets, namely UCI-Sensor [2] and Opportunity-Sensor [5], as input data. These datasets contain sensor readings captured during various human activities.

3.2 Data Pre-processing

The input data undergoes pre-processing steps, including null removal and normalization. Null removal involves handling missing or incomplete data, while normalization ensures that the data is scaled and standardized for better model performance.

3.3 Time Distributed Frame Conversion

The pre-processed data is then converted into time-distributed frames. This step involves splitting the data into smaller frames based on a specific time step and the total number of sensor channels. This enables the model to capture temporal dynamics and extract features from the data.

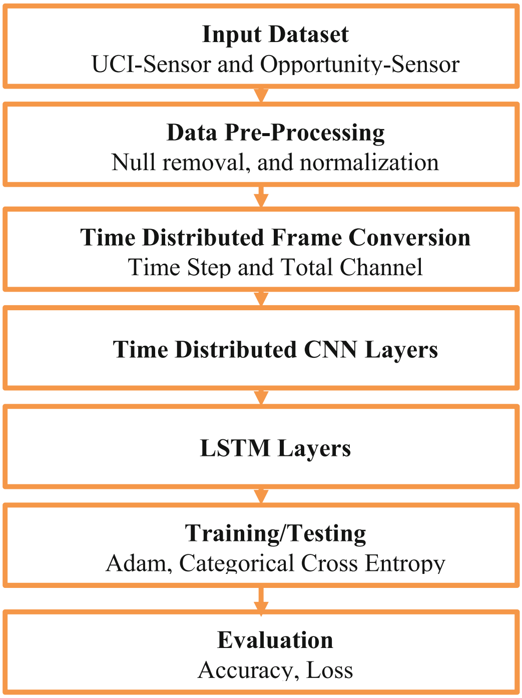

Input Dataset
UCI-Sensor and Opportunity-Sensor
Data Pre-Processing
Null removal, and normalization
Time Distributed Frame Conversion
Time Step and Total Channel
Time Distributed CNN Layers
LSTM Layers
Training/Testing
Adam, Categorical Cross Entropy
Evaluation
Accuracy, Loss

Fig. 1. Block diagram of the proposed time distributed CNN-LSTM model.

3.4 Time Distributed CNN Layers

Convolutional neural network (CNN) layers play a crucial role in handling the time-distributed frames. These CNN layers are designed to enable the model to identify significant patterns and structures by extracting spatial attributes from the input sensor data. A typical convolutional layer consists of numerous convolution kernels or filters.

Let us designate the number of convolution kernels as K. Each individual kernel is tasked with capturing distinct features, thereby generating a corresponding feature matrix. When employing K convolution kernels, the convolutional operation's output would consist of K feature matrices, which can be illustrated as:

$$\mathrm{Zk} = \mathrm{f}(\mathrm{WK}^{*}\mathrm{X} + \mathrm{b}) \tag{1}$$

In this given context, let X denote the input data with dimensions $m \times n$. The Kth convolution kernel with dimensions $k_1 \times k_2$ is represented by W_K, and the bias is denoted by 'b'. The convolution operation is depicted by ' $*$ '. The dimension of the Kth feature matrix Z_k depends on the chosen stride and padding method during the convolution operation. For instance, when using a stride of (1,1) and no padding, the size of Z_k becomes $(m - k_1 + 1) \times (n - k_2 + 1)$. The function f signifies the selected nonlinear activation function, applied to the output of the convolutional layer. Common activation functions include sigmoid, tanh, and ReLU.

3.5 LSTM Layers

The layers get the results from the CNN layers. Temporal dependencies in the data may be captured and learned by the LSTM layers. The network's ability to learn and anticipate future activity sequences is greatly enhanced by the addition of LSTM layers. LSTM utilizes three gates to manage the information flow within the network. The forget gate (ft) regulates the extent to which the previous state ($ct - 1$) is preserved. The input gate (it) decides whether the current input should be employed to update the LSTM's information. The output gate (ot) dictates the specific segments of the current cell state that should be conveyed to the subsequent layer for further iteration.

$$ft = \sigma(W(f)xt + V(f)ht - 1 + bf) \tag{2}$$

$$it = \sigma(W(i)xt + V(i)ht - 1 + bi) \tag{3}$$

$$ot = \sigma(W(o)xt + V(o)ht - 1 + bo) \tag{4}$$

$$ct = ft \otimes ct - 1 + it \otimes \tanh(W(c)xt + V(c)ht - 1 + bc) \tag{5}$$

$$ht = ot \otimes \tanh(ct) \tag{6}$$

Here, xt represents the input data fed into the memory cell during training, while ht signifies the output within each cell. Additionally, W, V, and b denote the weight matrix and biases correspondingly. The function σ refers to the sigmoid activation, which governs the significance of the message being propagated, and ⊗ indicates the dot product operation.

3.6 Training and Testing

Loss function "categorical cross-entropy" and "Adam" as an optimizer are used during training and testing. During training, the model uses the annotated data to fine-tune its settings and becomes better at identifying people at work.

3.7 Evaluation

Metrics like accuracy and loss are used to assess the trained model's performance. The accuracy and loss metrics gauge the model's effectiveness in categorizing human behaviors by measuring its precision and accuracy, respectively. The model's overall performance and its capacity to reliably distinguish various actions may be depicted from these assessment indicators.

4 Experimental Results and Discussion

4.1 UCI Sensor Dataset [2] Results

Six basic human activities—walking, sitting, standing, laying down, walking upstairs and downstairs are represented in the UCI-HAR [2] machine learning repository dataset. The information was collected from 30 people (aged 19 to 48) using an Android mobile device (Galaxy S2) equipped with inertial sensors. This dataset also includes transitions between other types of stationary postures, such as standing to sit, sitting to stand, lying to sit, laying to stand, and standing to laying.

The accuracy and loss calculated for each epoch for the proposed CNN-LSTM model are shown in Fig. 2. The confusion matrix for the proposed method is shown in Fig. 3 for six activities, and classification report is shown in Fig. 4 for the UCI-Sensor dataset. A comparison with the state of the art [1, 2, 4, 6], and baseline CNN and LSTM models is shown in Table 1. From this comparative analysis, one can conclude that the proposed model performs better.

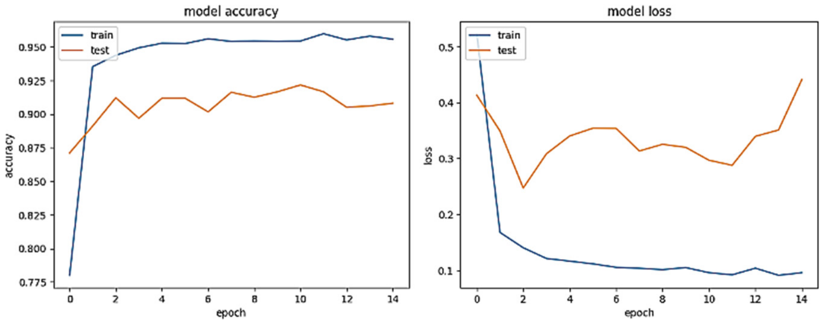
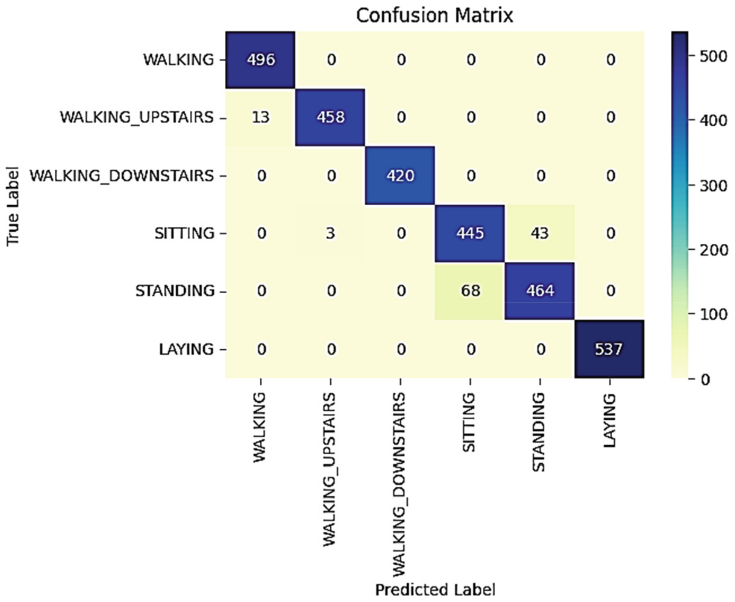
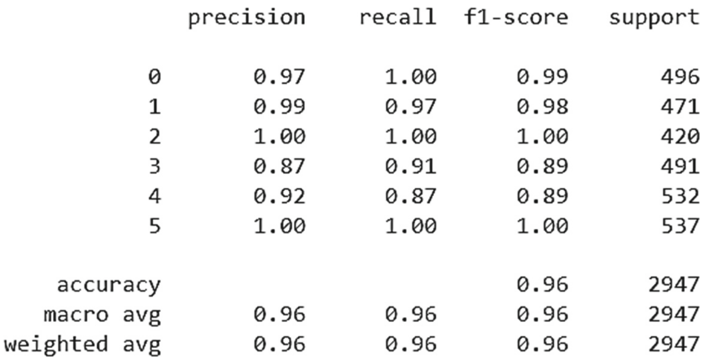


Fig. 2. Accuracy-loss plot for the proposed CNN-LSTM model.



	WALKING	WALKING_UPSTAIRS	WALKING_DOWNSTAIRS	SITTING	STANDING	LAYING
WALKING	496	0	0	0	0	0
WALKING_UPSTAIRS	13	458	0	0	0	0
WALKING_DOWNSTAIRS	0	0	420	0	0	0
SITTING	0	3	0	445	43	0
STANDING	0	0	0	68	464	0
LAYING	0	0	0	0	0	537

Fig. 3. Confusion matrix for the proposed CNN-LSTM model.



```
              precision    recall  f1-score   support

           0       0.97      1.00      0.99       496
           1       0.99      0.97      0.98       471
           2       1.00      1.00      1.00       420
           3       0.87      0.91      0.89       491
           4       0.92      0.87      0.89       532
           5       1.00      1.00      1.00       537

    accuracy                           0.96      2947
   macro avg       0.96      0.96      0.96      2947
weighted avg       0.96      0.96      0.96      2947
```

Fig. 4. Classification report for the proposed CNN-LSTM model.

Table 1. UCI-sensor dataset comparative analysis.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Time-Series CNN [1]	93.09	91.10	92.10	92.10
Parallel LSTM [2]	94.34	93.86	93.34	93.80
Feature Learning CNN [4]	67.51	66.80	66.78	67.35
Auto-Har [6]	94.80	94.65	94.70	95
Baseline CNN	74	75	73	73
Baseline LSTM	43	43	42	38
Time Distributed CNN-LSTM	**96**	**96**	**96**	**96**

4.2 OPPORTUNITY Sensor Dataset Results

Standing, laying down, walking, and navigating the stairwell are only some of the six basic human actions included in the Opportunity [5] machine learning repository dataset. Thirty people, ranging in age from 19 to 48, were surveyed using Android smartphones (Samsung Galaxy S II) equipped with inertial sensors. This dataset also includes transitions between other static postures, such as sitting, standing, lying, laying, sitting, lying, and standing.

The accuracy and loss calculated for each epoch for the proposed CNN-LSTM model are shown in Fig. 5. The confusion matrix for the proposed method is shown in Fig. 6 for six activities and classification report is shown in Fig. 7 for OPPORTUNITY-Sensor dataset. A comparison with the state of the art [11, 13–15], and baseline CNN and LSTM models is shown in Table 2. From this comparative analysis, one can conclude that the proposed model performs better.

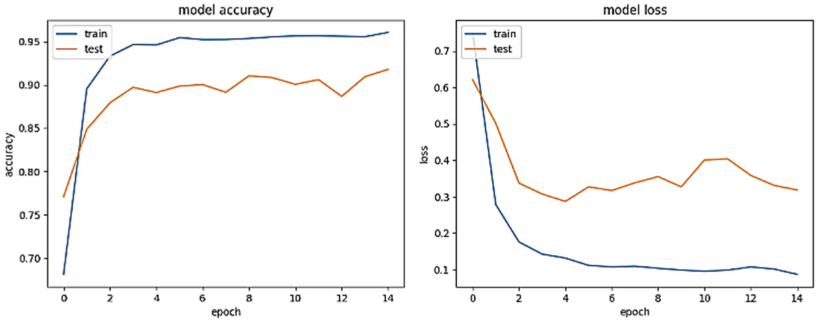
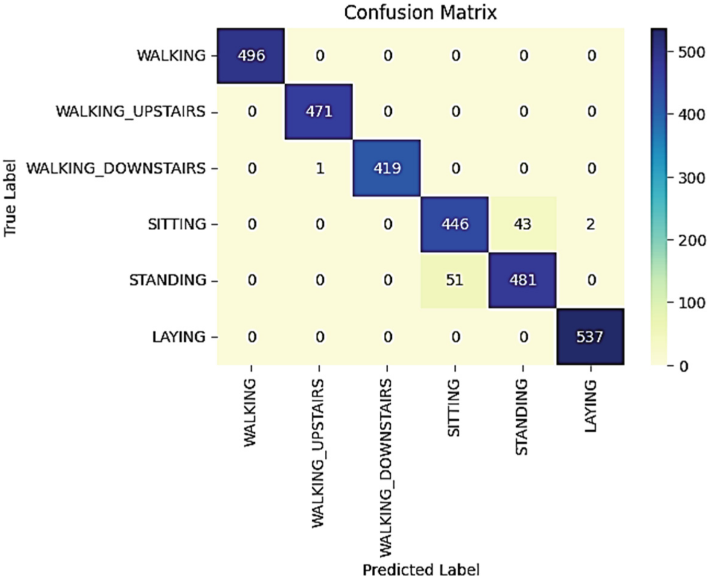
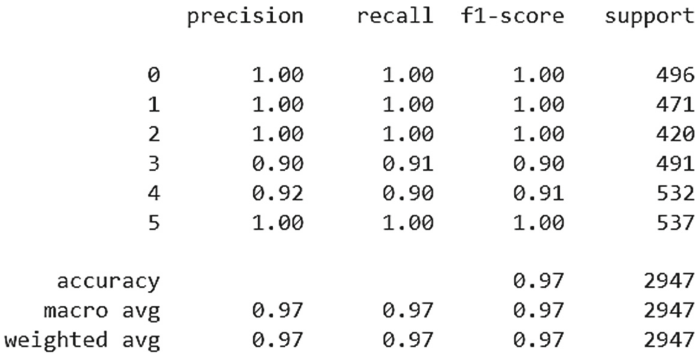


Fig. 5. Accuracy-loss plot for the proposed CNN-LSTM model.



	WALKING	WALKING_UPSTAIRS	WALKING_DOWNSTAIRS	SITTING	STANDING	LAYING
WALKING	496	0	0	0	0	0
WALKING_UPSTAIRS	0	471	0	0	0	0
WALKING_DOWNSTAIRS	0	1	419	0	0	0
SITTING	0	0	0	446	43	2
STANDING	0	0	0	51	481	0
LAYING	0	0	0	0	0	537

Fig. 6. Confusion matrix for the proposed CNN-LSTM model.



```
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       496
           1       1.00      1.00      1.00       471
           2       1.00      1.00      1.00       420
           3       0.90      0.91      0.90       491
           4       0.92      0.90      0.91       532
           5       1.00      1.00      1.00       537

    accuracy                           0.97      2947
   macro avg       0.97      0.97      0.97      2947
weighted avg       0.97      0.97      0.97      2947
```

Fig. 7. Classification report for the proposed CNN-LSTM model.

Table 2. Opportunity dataset comparative analysis.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Hybrid M [11]	46.68	46.75	46.70	47
b-LSTM-S [13]	92.70	92.45	92.10	92.90
InnoHAR [14]	94.60	94.20	94.20	94.80
CNN [15]	93.70	93.70	93.70	93.70
CNN	73	74	72	72
LSTM	38	34	35	27
Time Distributed CNN-LSTM	**97**	**97**	**97**	**97**

5 Conclusions

This research shows that a time-distributed CNN-LSTM model using sensor data significantly improves the performance of human activity recognition. The proposed model outperforms baseline CNN and LSTM, and other existing models, as shown by experimental results on the UCI-Sensor dataset and the Opportunity-Sensor dataset. The temporally distributed CNN-LSTM model achieved 97% accuracy for the Opportunity-Sensor dataset and 96% accuracy for the UCI-Sensor dataset across the board. These results demonstrate the value of integrating CNN and LSTM architectures to better capture temporal and spatial characteristics, which in turn enhances the accuracy and reliability of human activity classification from sensor data. Improving the effectiveness and scalability of the proposed model may require more investigation into broadening the assessment to other datasets and investigating optimization strategies.

References

1. An, S., Bhat, G., Gumussoy, S., Ogras, U.: Transfer learning for human activity recognition using representational analysis of neural networks. ACM Transactions on Computing for Healthcare **4**(1), 1–21 (2023)
2. Ismail, W.N., Alsalamah, H.A., Hassan, M.M., Mohamed, E.: AUTO-HAR: An adaptive human activity recognition framework using an automated CNN architecture design. Heliyon **9**(2), e13636 (2023). https://doi.org/10.1016/j.heliyon.2023.e13636
3. Nigam, S., Singh, R., Misra, A.K.: A review of computational approaches for human behavior detection. Archives of Computational Methods in Engineering **26**, 831–863 (2019)
4. Gupta, N., Gupta, S.K., Pathak, R.K., Jain, V., Rashidi, P., Suri, J.S.: Human activity recognition in artificial intelligence framework: a narrative review. Artif. Intell. Rev. **55**(6), 4755–4808 (2022)
5. Ciliberto, M., Fortes Rey, V., Calatroni, A., Lukowicz, P., Roggen, D.: Opportunity++: A multimodal dataset for video- and wearable, object, and ambient sensors-based human activity recognition. Frontiers in Computer Science **3**, 1–7 (2021). https://doi.org/10.3389/fcomp.2021.792065
6. Gupta, S.: Deep learning based human activity recognition (HAR) using wearable sensor data. Int. J. Info. Manage. Data Insights **1**(2), 100046 (2021). https://doi.org/10.1016/j.jjimei.2021.100046

7. Lv, T., Wang, X., Jin, L., Xiao, Y., Song, M.: Margin-based deep learning networks for human activity recognition. Sensors **20**(7), 1871 (2020)
8. Cruciani, F., et al.: Feature learning for human activity recognition using convolutional neural networks: a case study for inertial measurement unit and audio data. CCF Trans. Pervasive Comp. Interac. **2**(1), 18–32 (2020)
9. Shuvo, M.M.H., Ahmed, N., Nouduri, K., Palaniappan, K.: A hybrid approach for human activity recognition with support vector machine and 1d convolutional neural network. A hybrid approach for human activity recognition with support vector machine and 1D convolutional neural network. In: 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), pp. 1–5. IEEE, Washington DC, USA (2020)
10. Nematallah, H., Rajan, S.: Comparative study of time series-based human activity recognition using convolutional neural networks. In: 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6. IEEE, Dubrovnik, Croatia (2020)
11. Du, X., Farrahi, K., Niranjan, M.: Transfer learning across human activities using a cascade neural network architecture. In: 2019 ACM International Symposium on Wearable Computers, pp. 35–44. London United Kingdom (2019)
12. Xu, C., Chai, D., He, J., Zhang, X., Duan, S.: InnoHAR: A deep neural network for complex human activity recognition. IEEE Access **7**, 9893–9902 (2019)
13. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: a survey. Pattern Recogn. Lett. **119**, 3–11 (2019)
14. Rueda, F.M., Grzeszick, R., Fink, G.A., Feldhorst, S., Ten Hompel, M.: Convolutional neural networks for human activity recognition using body-worn sensors. Informatics **5**(2), 1–17 (2018)
15. Yu, T., Chen, J., Yan, N., Liu, X.: A multi-layer parallel LSTM network for human activity recognition with smartphone sensors. In: 2018 10th International conference on wireless communications and signal processing (WCSP), pp. 1–6. IEEE, Hangzhou, Zhejiang, China (2018)
16. Hammerla, N.Y., Halloran, S., Plötz, T.: Deep, convolutional, and recurrent models for human activity recognition using wearables. In: 25th International Joint Conference on Artificial Intelligence (IJCAI), pp. 1533–1540. New York, USA (2016)
17. Nigam, S., Singh, R., Singh, M.K., Singh, V.K.: Multiview human activity recognition using uniform rotation invariant local binary patterns. J. Ambient. Intell. Humaniz. Comput. **14**(5), 4707–4725 (2023)