

Static and Dynamic Human Activity Detection Using Multi CNN-ELM Approach



Shilpa Ankalaki and M. N. Thippeswamy

1 Introduction

Human Activity Recognition has its foremost impact on applications like smart health care and assisting elders, and it is necessary for humankind to assist human's daily life by recording the human movements from the sensors that are incorporated into the human body in the form of smart devices. Selecting proper methods for analyzing sensor data to make the correct decision is one of the challenging tasks. In this view, deep learning has immense benefits in HAR. HAR can be accomplished in two ways are Vision-Based Approach and Sensor-Based Approach. The former approach makes use of visual cameras to capture the image/video and data will be processed using image processing techniques and; analyzed by using machine learning and deep learning techniques whereas sensor-based HAR can be accomplished by using non-visual sensors. Vision-based HAR has its limitation in applications due to its consideration of secrecy concerns of mounting vision sensors in private space [1]. Non-visual sensors based HAR has the advantage of generality because sensors can be easily embedded in almost all devices including the human body. The primary difficulty of Sensor-based HAR is the representation of information captured by various sensors. Traditional classifiers exhibit limited performance for HAR because of the process of extraction of handcrafted features. This drawback is overcome by using deep learning techniques that provide the facility of automatic feature extraction from given data.

S. Ankalaki (✉) · M. N. Thippeswamy
Department of Computer Science and Engineering, Nitte Meenakshi Institute of Technology,
Bangalore 560064, India

M. N. Thippeswamy
e-mail: thippeswamy.mn@nmit.ac.in

There are a couple of approaches under sensor-based HAR, those are knowledge-driven approach and data-driven approach [2]. A knowledge-driven approach constructs activity models by taking advantage of complete prior knowledge in the area of interest. The data-driven approach employs publically available datasets to study the activity models by applying machine learning and deep learning techniques. The proposed work emphasizes data-driven solutions to HAR, also discussed the existing barriers of their application on UCI-HAR dataset. This work focuses on the recognition of human static and dynamic activities of the UCI-HAR dataset.

Followings are the objectives and key contributions of this proposed work:

- Proposed the hierarchical 1D-CNN approach for Classification of Human Activities into Static and Dynamic activities
- Proposed the novel hybrid 1D-CNN-ELM approach for classifying the static activities into sitting, standing, and lying; and dynamic activities into walking, walking_upstairs, and walking_downstairs.
- Performance evaluation of 1D CNN-ELM approach for classification of static and dynamic activities using precision, recall, confusion matrix, and accuracy measures.

The rest of the paper is organized as follows: Sect. 2 briefs about understanding the human activity and human activity recognition; a literature survey of deep learning approaches for sensor-based HAR. Section 3 provides the details of the proposed methodology employed to accomplish the above-stated objectives. Section 4 provides the experimentation results and analysis of the proposed work.

2 Related Work

This section provides the relevant literature about the understanding of human activity, human activity recognition, and deep learning methodologies for sensor-based HAR.

2.1 Understanding Human Activity

Human activities are the order of human movements operated by an individual over a period of time in a given ambient. In the view of sensor based HAR, the activity can be well-defined as the set of actions where each action consists of a sequence of events. Events are interpreted as a sequence of data generated by various sensors records, whereas usually sensors are incorporated in human bodies but in advanced HAR sensors are incorporated in the environment as well [3, 4]. The mathematical representation of activity definition is represented by the Eqs. (1)–(2).

$$A = (A_i)_{i=1}^m \quad (1)$$

In Eq. (1), A represents the existing activity set which is inclusive of ‘ m ’ number of various activities. The sequence of data that is captured by the sensor for a given period of time is represented by an Eq. (2).

$$S = \{r_1, r_2, \dots, r_t, \dots, r_n\} \quad (2)$$

where r_t represents the reading of the sensor at time t .

The objective here is to construct the model that predicts the series of activities that belong to set A depending on the sensor reading S .

2.2 Human Activity Recognition

In our everyday life, people perform 2 kinds of traditional physical activities those are namely, static and dynamic human activities. Sitting, standing and sleeping are some of the activities which are inclusive of static activities. Human physical activities can also be categorized as atomic, simple, and complex activities [5]. Atomic activities are static activities, Standing is one such example. Simple activities are a systematic sequence of static activities performed within a specified time interval, Walking is one such activity that belongs to the simple activity category. Complex activities are an assortment of more than one simple activity that takes place at a specified time. Dancing is an example of a complex activity. Researchers like Lara and Labrador [6] summarized seven types of activities recognized by the HAR system in their literature. Those activity groups are namely: Daily activities, Transportation, Fitness, Phone usage, Military, Ambulation, and Upper Body activities.

Sensor-based approaches for human activity recognition can be accomplished by incorporating the sensors in the human body and environment. An approach of incorporating the sensors in the human body provides sensor data by which the system can identify individual activities like walking, dancing, skipping, cooking, etc. without considering the location. Whereas sensors incorporated in an environment gathers information about the locations where activities are taken place [7].

Sensor-based approaches are created a successful path for recognition of human activities ranging from static to dynamic and simple to complex human activities. One such work is accomplished by AROMA which recognizes both simple and complex activities together [8]. Researchers [2] successfully created a smart environment in UJAmI Smart Lab by incorporating binary sensors in the kitchen area, bedroom area, Door, Laundry basket, sofa. The sensor-based HAR can be accomplished by using machine learning and deep learning approaches. Deep learning approaches for sensor-based HAR are discussed in the subsequent section.

2.3 *Deep Learning Methodologies for Sensor-Based HAR*

The recent literature provides evidence that the deep learning methods outperform machine learning algorithms for sensor-based HAR because of its unique characteristics of automatic feature extraction. One such survey has been conducted in [9] which provides an extensive study about deep learning approaches for sensor-based HAR and also focused on the numerous ways to address the challenges of HAR. The extensive study has been carried out by the researchers of [10] about feature learning using a convolutional neural network for HAR. The experiment was conducted on the datasets like DCASE 2017 dataset, Extrasensory dataset, UCI-HAR dataset, and real-world Extrasensory dataset; and analyzed the performance of various architectures of CNN and fine-tuning the CNN by changing the values of its hyperparameters. Apart from the recognition of Daily human activities, researchers also contributed to the recognition of athletic tasks using deep neural networks [11]. This task has been accomplished by using the hybrid approach by combining CNN and LSTM. The experiment has been carried out from the sensor data collected from 417 athletes where every athlete has 13 athletic movements. The work discussed in [12] proposed a novel approach for HAR which adopts pose reconstruction dataset AMASS along with virtual IMU data. This dataset is trained by using the collective framework of CNN along with an unsupervised penalty for HAR. The experiment was conducted by using the hybrid architecture of CNN-RNN on Opportunity dataset, PAMAP2 dataset, and Daphnet Gait dataset and concluded that bidirectional LSTM outstrips the other algorithms on the opportunity dataset [13].

One of the missing features in formerly described contributions about the deep learning architectures is the grouping of activities present in the dataset into static and dynamic activities. The deep learning approaches described in this section classifies the activities into the individual classes without considering whether it is a static and dynamic activity. The proposed work focuses on the hierarchical and hybrid 1D CNN-ELM approach for activity recognition and performance evaluation of the classification of each activity using various metrics.

3 **Data and Methodology**

The proposed work aims to assess the performance of the hybrid CNN-ELM approach and includes performance analysis by fine-tuning the hyperparameters. Figure 1 depicts the detailed pictorial representation of the proposed methodology.

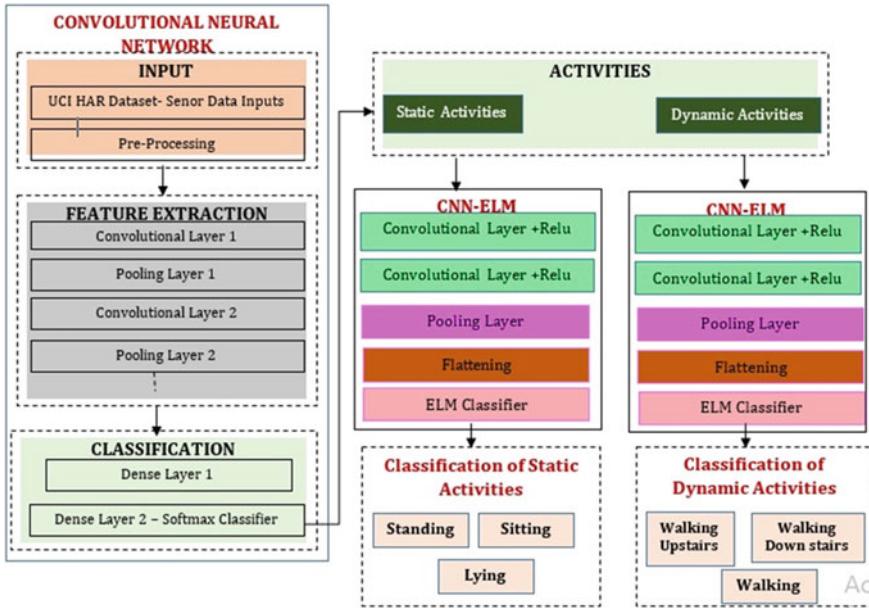


Fig. 1 Hierarchical multi CNN-ELM approach for classification of static and dynamic activities

3.1 Input Dataset

The proposed methodology has employed the UCI-HAR dataset to accomplish sensor-based HAR. The aforementioned dataset was constructed by [14], the dataset built by the activities performed by 30 individuals. Every individual has accomplished the following activities using wearable sensors, those are, lying, standing, sitting, walking_upstairs, walking, and walking_downstairs.

3.2 Pre-processing

In this stage, Data cleaning has been carried out on the UCI-HAR dataset by examining the dataset for missing values, null values, and balanced data distribution for each target class.

3.3 Feature Learning Using CNN

The proposed work employed, CNN for instinctive feature extraction from raw data of sensors which eliminates the step of handcrafted feature extraction. Use of CNN

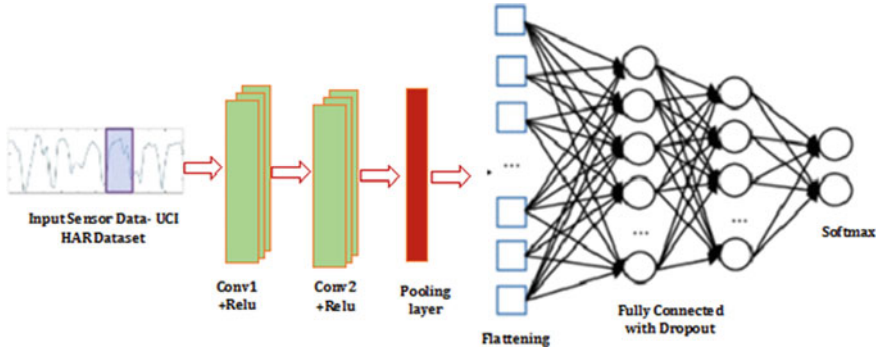


Fig. 2 Structure of 1-D CNN employed in proposed work

simplifies the activity recognition by removing the process of Handcrafted Feature extraction, which usually requires domain-specific experts to recognize preferable and distinct features [10]. The structure of CNN includes a set of layers namely, Convolutional pooling and dense layer. CNN accomplishes the task of feature extraction using a convolutional layer with varying the number and size of filters in each layer and non-linear feature extraction. Downsampling is accomplished by using the pooling layer and classification is accomplished by using a dense layer with a softmax activation function.

Usually, CNN's are categorized based on the dimension of kernels used in the convolutional layer those are, 1-D, 2-D, and 3-D CNN. The proposed work employed 1-D CNN as it uses the sensor data for feature extraction. Figure 2 describes the structure of CNN used in this work.

Figure 2 describes the structure of CNN employed in this work. Input is 1-dimensional sensor data, this raw data is passed through the two sets of the convolutional layer. In this stage, the features are extracted at the convolutional layer by performing the convolutional operations on raw sensor data using 1-dimensional kernels (filters). The nonlinear activation functions like Rectified Linear Unit (Relu) or Leaky Relu functions are applied to the resultant features obtained by the convolutional layer to make features highly nonlinear. The process of feature extraction using a one-dimensional kernel is depicted in Fig. 3.

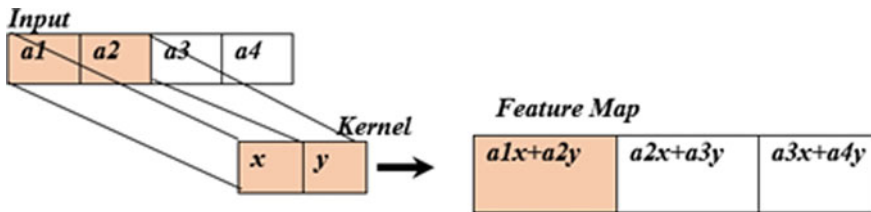


Fig. 3 Feature extraction using 1-D kernel in convolutional layer. Source [15]

Downsampling has been accomplished by using max or average pooling. The resultant features are flattened into a vector and it is considered as trainable learning which will be given as input to the fully connected network. The commonly used approach for the fully connected network is Multi-Layer Perceptron. The proposed work employed Hierarchical Multilevel CNN, in which Multi-Layer Perceptron (MLP) has been used at the root level to classify the activities into static and dynamic activities followed by two CNN–Extreme Learning Machine to classify sub-activities of static and dynamic activities.

3.4 Extreme Learning Machine (ELM)

The iterating nature which causes more computational effort is one of the foremost shortcomings of MLP [16]. In order to overcome this issue, the proposed work employed ELM instead of MLP at a fully connected neural network. The ELM was originally proposed by Huang et al. [17]. The procedure of ELM [16–18] is given in Table 1.

4 Results and Analysis

In this work, the hierarchical multi CNN-ELM approach has been used for classifying the activities of the UCI-HAR dataset. The first level involves the classification of

Table 1 Algorithm for extreme learning machine

| | |
|---|---|
| Input: Training Dataset (A, Y _i) (A _i stands for attributes, and Y _i is target class label), | |
| Output: Prediction of class labels | |
| Steps: | |
| Step 1: Initialization of arbitrarily engendered weights w _i and biases b _i from input to hidden layer | |
| Step 2: Compute the output matrix H for hidden layer, where H = {h _{ij} } (i = 1, ..., N and j = 1, ..., M) and | |
| $h_{ij} = g(w_j * a_j + b_j)$ | (3) |
| Step 3: Compute linear transfer function of output layer as showed in Eq. (4) | |
| $\sum_{i=1}^n \beta_i g(w_j * a_j + b_j) = T_j$ | (4) |
| Where, n – Number of hidden neurons | |
| a – Input Vector | |
| w and T – Weight and Output Vector respectively. | |
| β and $g(y)$ – Transfer function of Output and Hidden layer | |
| The matrix representation of Eq. (4) is showed in Eq. (5) | |
| $H\beta = T$ | (5) |
| Where, | |
| $H = [g(w_1 * a_1 + b_1) \dots g(w_n * a_1 + b_n) \dots g(w_1 * a_n + b_1) \dots g(w_n * a_n + b_n)]$ | |
| $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$ | $T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$ |
| Step 4: Eq. (6) represents the computation of the weights matrix for output layer | |
| $\beta = H \dagger T$ | (6) |
| where H [†] denotes the MP generalized inverse of matrix H. | |

human activities into static and dynamic activities using CNN with MLP classifier whereas the second level has the combination of 2 CNN-ELM classifiers to classify the static and dynamic activities into its respective sub-activities.

4.1 Classification of Static and Dynamic Human Activities

In the UCI-HAR dataset, the magnitudes are represented by features like tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, and tBodyGyroJerkMag.

Figure 4 represents the graph of static and dynamic activities with respect to the mean of body acceleration feature, i.e., tBodyAccMag_mean. From the analysis, we can conclude that static and dynamic activities are completely different and classification of these activities is accomplished by structural design of CNN depicted in Fig. 5. This structural design of CNN represents the first level binary classifier which classifies the activities into static and dynamic activities. The architecture consists of 2 successive convolutional layers with 32 filters in each layer; and size of the kernel 7 and 3 respectively followed by Max pooling (pool_size = 3) and flattening. The Relu activation function has been used in all convolution layers. The classification of activities has been accomplished by dense layer with a softmax activation function. The training has been performed for 20 epochs and with batch size 32. Categorical Cross entropy is employed for loss calculation and Adam optimizer with a learning rate of 0.004 is used for optimization. Figures 6 and 7 depicts the training and validation accuracy and training and validation loss of first level CNN architecture respectively.

In the second level, classification of static activities into sit, stand, and lying down; and moving activities into walking, walking upstairs, and walking downstairs is accomplished by using two 3-class CNN-ELM Classifier which is depicted in Fig. 8. The architecture consists of 2 successive convolutional layers with a number of filters 64 and 32; and the size of the kernel 7 and 3 respectively followed by Max

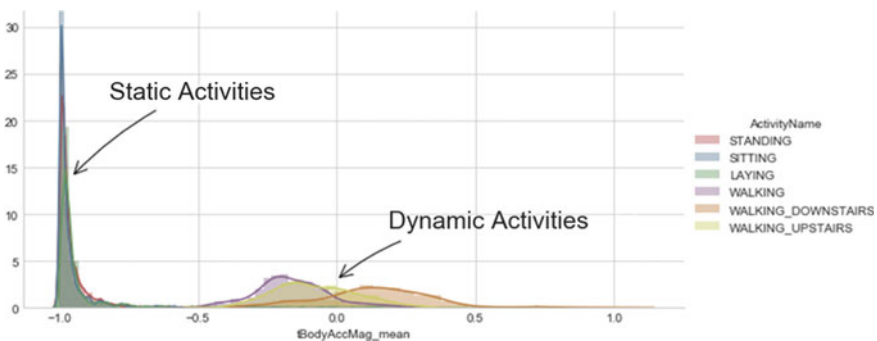


Fig. 4 Analysis of static and dynamic activities with respect to tBodyAccMag_mean feature

| Layer (type) | Output Shape | Param # |
|--------------------------------|-----------------|---------|
| conv1d_1 (Conv1D) | (None, 126, 32) | 896 |
| conv1d_2 (Conv1D) | (None, 124, 32) | 3104 |
| dropout_1 (Dropout) | (None, 124, 32) | 0 |
| max_pooling1d_1 (MaxPooling1D) | (None, 62, 32) | 0 |
| flatten_1 (Flatten) | (None, 1984) | 0 |
| dense_1 (Dense) | (None, 50) | 99250 |
| dense_2 (Dense) | (None, 2) | 102 |

Fig. 5 CNN architecture to classify activities into static and dynamic activities

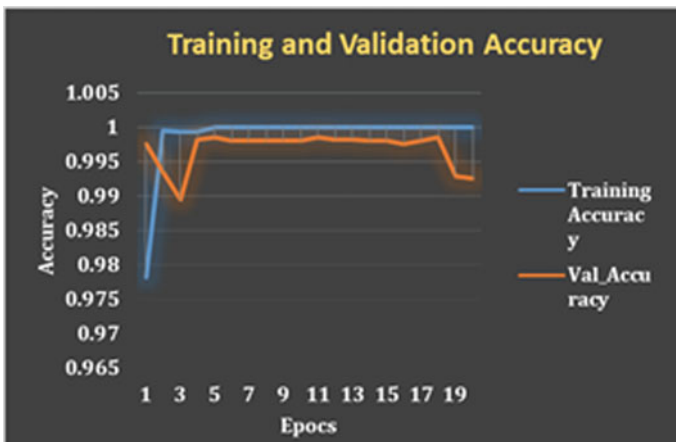


Fig. 6 Training and validation accuracy

pooling (pool_size = 3) and flattening. The Relu activation function has been used in all convolution layers. The classification of activities has been accomplished by ELM classifier with 20 hidden neurons. The training has been performed for 20 epochs and with batch size 32. Categorical Cross entropy is employed for loss calculation and Adam optimizer with a learning rate of 0.004 is used for optimization.

The confusion matrix has been calculated for both classification static and dynamic activities and the same is showed Tables 2 and 3.



Fig. 7 Training and validation loss

| Layer (type) | Output Shape |
|-------------------|-----------------|
| conv1d_1 (Conv1D) | (None, 122, 64) |
| conv1d_2 (Conv1D) | (None, 120, 32) |
| dropout (Dropout) | (None, 120, 32) |
| max_pooling1d | (None, 40, 32) |
| flatten (Flatten) | (None, 1280) |
| ELM Classifier | (10, Tanh) |

Fig. 8 Architecture of CNN-ELM classifier

Table 2 Confusion matrix for static activities

| Actual class | Predicted class | | | | Recall (%) |
|---------------|-----------------|-------|-------|-------|------------|
| | | SIT | STAND | LYING | |
| SIT | 534 | 3 | 0 | 99.44 | |
| STAND | 0 | 450 | 41 | 91.64 | |
| LYING | 0 | 27 | 505 | 94.04 | |
| Precision (%) | 100 | 93.75 | 92.49 | 95.44 | |

Table 3 Confusion matrix for dynamic activities

| Actual class | Predicted class | | | Recall (%) | |
|--------------------|-----------------|-------|------------------|------------|--------------------|
| | | Walk | Walking_upstairs | | Walking_downstairs |
| Walk | | 486 | 0 | 10 | 97.98 |
| Walking_upstairs | | 3 | 415 | 3 | 98.57 |
| Walking_downstairs | | 3 | 2 | 465 | 98.93 |
| Precision (%) | | 98.78 | 99.52 | 97.28 | 98.48 |

5 Conclusion

The proposed work focuses on the hierarchical multilevel CNN-ELM classifiers for the recognition of static and dynamic activities. This work concludes that the differentiation of static activities, i.e., standing and lying was a difficult task as it is evinced by the comparatively large number of the misclassified instances (41 and 27 respectively). The precision and recall values are better for dynamic activity recognition compares to static activities, which concludes that the classification of dynamic activities is more accurate than the static activities. This classification achieves accuracy of 95.44% for static activities and 98.48% for dynamic activities. The proposed method accomplished the overall accuracy of 96.86% on the UCI-HAR dataset. This work can be extended by fine-tuning the parameters of CNN and by varying the ELM parameters like activation functions and the number of hidden neurons.

References

1. Bevilacqua A et al (2019) Human activity recognition with convolutional neural networks. In: Lecture notes in computer science, pp 541–552
2. Irvine N, Nugent C, Zhang S, Wang H, Ng WWY (2020) Neural network ensembles for sensor-based human activity recognition within smart environments. *Sensors* 20:216
3. Wang J et al (2019) Deep learning for sensor-based activity recognition: a survey. *Pattern Recogn Lett* 119:3–11
4. Bulling A, Blanke U, Schiele B (2013) A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys*
5. Nehal S, Abu-Elkheir M, Atwan A, Hassan S (2018) Current trends in complex human activity recognition. *J Theor Appl Inf Technol* 96(14):4564–4583
6. Lara OD, Labrador MA (2013) A Survey on human activity recognition using wearable sensors. *IEEE Commun Surv Tutor* 15(3):1192–1209
7. Yang J (2009) Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In: Proceedings of the 1st international workshop on interactive multimedia for consumer electronics, Beijing, China, 23 October 2009, pp 1–9
8. Peng L et al (2018) AROMA: a deep multi-task learning based simple and complex human activity recognition method using wearable sensors. *IMWUT* 2: 74:1–74:16
9. Chen K, Zhang D, Yao L, Guo B, Yu Z, Liu Y (2020) Deep learning for sensor-based human activity recognition: overview, challenges and opportunities. [arXiv:2001.07416v1](https://arxiv.org/abs/2001.07416v1) [cs.HC] 21 Jan 2020

10. Cruciani F, Vafeiadis A, Nugent C et al (2020) Feature learning for human activity recognition using convolutional neural networks. *CCF Trans Pervasive Comp Interact* 18–32
11. Clouthier A, Ross G, Graham R (2020) Sensor data required for automatic recognition of athletic tasks using deep neural networks. *Front Bioeng Biotechnol* 7:473. <https://doi.org/10.3389/fbioe.2019.00473>
12. Xiao F, Pei L, Chu L, Danping Z, Yu W, Zhu Y, Tao L (2020) A deep learning method for complex human activity recognition using virtual wearable sensors. [arXiv:2003.01874v2](https://arxiv.org/abs/2003.01874v2) [cs.CV] 6 Mar 2020
13. Hammerla NY, Halloran S, Plötz T (2016) Deep, convolutional, and recurrent models for human activity recognition using wearables. *IJCAI. AAAI Press*, pp 1533–1540
14. Anguita D et al (2013) A public domain dataset for human activity recognition using smartphones. *ESANN*
15. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
16. Jayaweera CD, Aziz N (2018) Development and comparison of extreme learning machine and multi-layer perceptron neural network models for predicting optimum coagulant dosage for water treatment. *IOP Conf Ser J Phys Conf Ser* 1123
17. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine theory and applications. *Neuro Comput* 70:489–501
18. Silva DNG, Pacifico LDS, Ludermitr TB (2011) An evolutionary extreme learning machine based on group search optimization. In: 2011 IEEE congress of evolutionary computation (CEC), New Orleans, LA, pp 574–580