Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

WSAMLP: Water Strider Algorithm and Artificial Neural Network-based Activity Detection Method in Smart Homes

Jamile Barazandeh and Nazbanoo Farzaneh*

Computer Engineering Department, Imam Reza International University, Mashhad, Iran.

Article Info

Article History:
Received 29 April 2021
Revised 29 June 2021
Accepted 20 September 2021

DOI:10.22044/JADM.2021.10781.2215

Keywords:
IoT, Smart Home, Users' Activities, Data Mining, Water Strider Algorithm.

**Corresponding author:*
nazbanou.farzaneh@imamreza.ac.ir
(N. Farzaneh).

Abstract

One of the crucial applications of IoT is to develop smart cities via this technology. Smart cities are made up of smart components such as smart homes. In such homes, a variety of sensors are used in order to make the environment smart, and the smart things in such homes can be used to detect the activities of the people inside them. Detecting the activities of the smart homes' users may include the detection of the activities such as making food or watching TV. Detecting the activities of the residents of smart homes can tremendously help the elderly or take care of the kids or even promote the security issues. The information collected by the sensors could be used to detect the kind of activities; however, the main challenge is the poor precision of most of the activity detection methods. In the proposed method, for reducing the clustering error of the data mining techniques, a hybrid learning approach is presented using the water strider algorithm. In the proposed method, this algorithm can be used in the feature extraction phase, and exclusively extract the main features for machine learning. The analysis of the proposed method shows that it has a precision of 97.63%, an accuracy of 97.12%, and an F1 index of 97.45%. It, in comparison with similar algorithms (such as butterfly optimization algorithm, Harris hawks optimization algorithm, and black widow optimization algorithm) has a higher precision while detecting the users' activities.

1. Introduction

IoT is a vast network with numerous sensor nodes and smart things. IoT is, in fact, the interaction among the sensors and the everyday-used devices for connecting the physical things to virtual things through smart networks [1]. A smart home is a complicated research field in smart automation systems whose total aim is to increase the users' welfare and guarantee their security at a minimum cost [2][3]. A smart home provides its residents with a smart and automatic environment; therefore, it is able to track, detect, and record the daily activities using various kinds of sensors and communicational technologies. The users' daily activities create patterns that play an important role in the smart home [4][5]. These patterns are used to support the detection of the user's activity; it is useful for improving the smart home programs such as energy efficiency and management, hygiene and health care, and security [6]. Detection of the users' activities in smart homes can be performed with the help of the information collected by the sensors in the smart home. Through detecting the users' activities in the smart home, it would be possible to record and control them through a remote control process. The remote control of the residents' activities is required in most of the elderly houses or any place in which people live alone [7][8]. For instance, at the time of pandemics like Covid-19, it is necessary to control the people, and specially, the sick people's activities. In such cases, it is possible to monitor the old people or the sick people infected with Covid-19 through the remote control process and hospitalize them if they require more health care in the hospital [9]. Activity detection can also be useful for the smart home security; if a stranger enters the house, the system detects him and the alarm will go-off [10].

Using body sensors attached to the people's bodies and using other sensors, the information can be collected and given to the machine learning algorithm as the input to detect the kind of activity. Moreover, the information is processed by the system in order to detect the user's behavior disorders. Therefore, if any unwanted behavior is detected, it is possible to help the user through the remote control process [11]. One of the key points of such monitoring systems is the ability to respond via recognizing the normal behavior of the user. For instance, if the system detects that the user has fallen down, an emergency call is made or if the system detects a stranger in the house, an automatic police call will be made. Up to the present time, several methods have been proposed for detection of activities in smart homes such as deep learning [12] , artificial neural network [13], support vector machine [14], artificial intelligence [15], evidence theory based method [16] and meta-heuristic methods [17]. Unfortunately, in most of the studies carried out, no control has been done at the learning machine input and learning has been done on all features; however, in the present work, learning is exclusively done on the extracted features. In the proposed method, for detecting the users' activities in smart homes, the water strider algorithm, introduced in 2020, is used [18]. The main goal of the present work is to decrease the neural network input via smart extraction of features as well as performing the learning process on the important features in order to decrease the clustering error as well as the user's activity detection error.

The structure of the present paper is as what follows. In Section 2, the water strider algorithm is introduced. In Section 3, a review of the related works done in this field is performed. In Section 4, the proposed method for detecting the users' activities by the water strider algorithm and Artificial Neural Network (ANN) is explained. In Section 5, the performance and analysis of the proposed method are presented. Finally, in Section 6, the conclusion is stated.

2. Water Strider Algorithm

Water striders are a group of insects that live on the water of rivers and ponds. These insects have an amazing ability to live on water and maintain their weight on water with the surface tensile force of water surface molecules. The studies show that the males take the lead in mating and are a major factor in reproduction, while monitoring food resources and food-related activities is more closely related to the female species. In each territory, usually an adult male and several species of females live. The studies show that if an annoying insect and another male enter their territory, then a fight takes place between the male nest and the stranger male. The studies show that these insects can use water waves in order to communicate with each other as a result of hitting the water surface. The Water Strider Algorithm (WSA) [18] is a novel meta-heuristic method, and has phases of birth, territory establishment, mating, feeding, and death. The studies show that WSA has a simple structure but at the same time, due to the intelligent search of the problem space, it can calculate the optimal answers with a high accuracy. WSA is at least more accurate than the metaheuristic methods such as GA and PSO that can find the optimal answers with a high accuracy. The details of the operation of WSA and the corresponding mathematical equations are available in [18].

3. Related Works

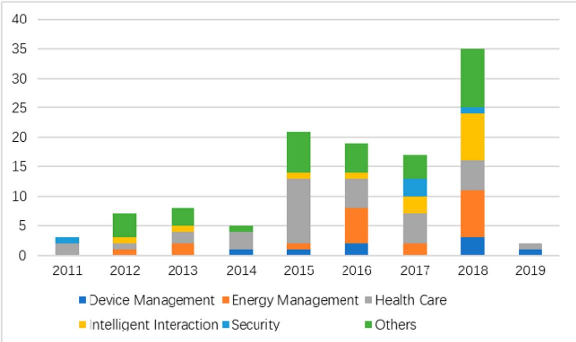
There are several techniques for detecting the activities of people in the smart homes. A review of the previous studies in the field of human activity detection in smart homes shows that most methods use the machine learning and deep learning techniques. One of the important and practical methods used in detecting the user activity is the use of knowledge discovery and machine learning methods, and with the help of these methods, the users' information and activities can be analyzed, and the behavioral pattern of the smart home users can be identified, and the necessary control signals are provided, if necessary.

In the recent years, many studies have been conducted in various fields in the smart homes, for which solutions based on artificial intelligence have been proposed.

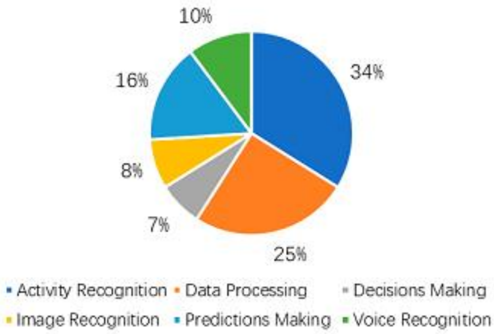
3.1 An Overview of Previously Presented Methods

[15] divides the smart home applications into 5 functions and 6 clusters (each function uses a set of listed clusters). The functions include device management, energy management, security, healthcare and intelligent interaction, and clusters including activity recognition, data processing,

decision making, image recognition, prediction making, and voice recognition. Figure 1 shows the extent to which each one of the above fields is discussed in the articles and research works [15].
According to Figure 1, which was presented in the early 2019s, it is clear that smart homes have been used in various fields such as security, energy, object management, and health, and from 2015 onwards, a number of studies in this field, showing that smart homes and the related technology have been considered, and the development of methods to apply this technology such as the proposed method is of great importance.



(a)



(b)

Figure 1. Smart home applications for which AL methods are provided (a) Functions (b) Clusters [15].

One of the applications and necessities of smart home development is the use of AL methods in order to detect the activities of people inside the home, and with the help of this technology, the issues such as health or safety of the home users can be increased by predicting and detecting the activities. Since the proposed method is in the same category, in the following, some methods of detecting the user activity using knowledge discovery and machine learning techniques will be mentioned. In [19], the probabilistic neural network has been used to detect the activities in the smart homes. If the number of identified activities is more or less than normal, the condition is considered abnormal. [20] has introduced a knowledge-based approach to multi-resource collaboration in the smart homes from activity detection and provides suitable services, in particular, a smart home open-layer architecture, which combines ontology and open-layer technologies, designed to automatically acquire a semantic knowledge and support the heterogeneity and interoperability services. In such an architecture, a general inference algorithm based on unordered actions and activity timing is proposed to infer both the continuous compound activity and real-time personalized services. In [21], delay in decision-making has been used in order to prevent wrong decisions that are made due to quick conclusions. A multiple incremental fuzzy temporal windows is used to extract the features of sensor activity information and preceding. A knowledge extraction method has been used in [22]. In the first stage, using an unsupervised learning method, the information hidden is extracted, and in the second stage, the information of the first stage and the basic information of the activities are given to a classifier to make the final decision. In [23], in order to achieve a better result, four classifiers named Bernoulli naïve Bayes, decision tree, logistic regression, and K-nearest neighbor have been used for a multi-label classification. In order to get the final result, the majority voting method has been used to combine the results of the classifier. A graph-based semi-supervised learning method for labeling the activities in the health smart home has been presented in [24]. Building the activity recognition models usually requires large amounts of labeled data, which imposes a heavy burden on the manual labeling. In this paper, the authors propose an activity labeling method based on the semi-supervised learning algorithm. This method can split the raw sensor event sequence into appropriate segments without any labeled information. In [25], a clustering method has been used to identify two-resident activities. In the first step, a de-noising method is used to extract the dataset features. Then in the second step, clustering is done based on the start and end time, and finally, the similarity matching method is used to complete the activity detection process. [6] has used a fuzzy method, and ontology to diagnose activity in the home medical system. The data in this system is unlabeled. Using the fuzzy rules used in the ontology-based activity detection system, an attempt has been made to present a clear representation of the data. In [26], a convolutional neural network (CNN) method with bidirectional long short-term memory (Bi-LSTM) units has been used to forecast the activity. In [27], a method with three stages has been used to detect the user activities. These stages include home activity definition (uses the RFID tag to identify the user's connection to the device), activity recognition based on weighted usage data (to determine the accuracy of the sensor in detecting activity), and activity prediction with long short-term memory (to specify user activity).

In [28], a method for detecting the human personal activities based on an integrated wearable sensor and transfer learning has been proposed. In order to solve the personal identification of the user activities, they proposed an algorithm for learning, which is an adaptive method of common probability domain with quasi-improved tags (IPL-JPDA).

In [29], they introduced a system of personalization of human personal activity based on intelligent sensors. The proposed framework uses the Deep Recurrent Neural Network (DRNN) for an extensive training dataset. In [30], they identified the activities of the individuals in the smart homes by weighing the local characteristics and types of classifiers of the nearest neighbors. A Local Feature Weighting (LFW) method that assigns weight based on the importance between a class and within a feature in an activity is presented. For classification, they used two types of K-nearest methods including KNN (ETKNN) and fuzzy KNN (FKNN). In [31], they presented activity detection for the smart home systems with four computational methods. They introduced four computational methods for generating natural language descriptions of smart home data. In [12], the sensor data classification presented several imbalance modes for detecting the human activities in a smart home using deep learning.

3.2 Existing Challenges and Innovations of Proposed Method

The major challenge of data mining methods is that they learn more than data and information from the smart homes without any special preprocessing, while for a better learning, it is necessary that feature selection is considered as an important data mining process in order to reduce the error of recognizing people's activities in the smart homes.

Some methods use deep learning techniques in order to detect the activities of the smart home users but these methods require a lot of training time. Some methods use machine learning techniques in order to detect the human activities in the smart homes but these methods require an intelligent feature selection. Machine learning methods alone have little ability to detect the people's activities so it is necessary to use the feature selection methods intelligently. The innovation of our research work is that the WSA algorithm, introduced in 2020 and has a high accuracy, has been used to select the feature before learning artificial neural networks. Feature selection has two important roles in detecting the user activities. First, it increases the accuracy of classifying the type of user activity, and then increases its speed by reducing the input dimensions of the neural network. Our contribution to this research work is that a binary version of the WSA algorithm is presented, and this binary version is used to select a feature to detect the activities of the smart home users. The advantage of our proposed method is the combination of swarm intelligence and machine learning methods to detect the types of activities of the smart home users.

4. Proposed Method

In the proposed method, called WSAMLP, for detecting the people’s activities in smart homes, the methods based on machine learning and data mining (ANN) are used. WSAMLP has two layers. In this method, the smart home data is received, and then the most important features related to the smart home are extracted. Next, ANN uses the meta-heuristic algorithm in order to reduce the estimation error. In the proposed method, for the feature extraction phase and reduction in the ANN error, WSA 16 is used, which has been recently introduced and proved to work better than the famous algorithms such as GA and PSO, and to decrease the detection and estimation error of ANN to a suitable extent; therefore, in the proposed method, WSA is used to decrease the ANN error in detecting the people’s activity states in smart homes. Figure 2 shows an overview of the proposed method.

The proposed method (Figure 2) briefly includes the following steps:

- A random feature vector set is created as the initial population for WSA, and it is used to train ANN to select the most appropriate feature vector.
- Using an ANN whose input is optimized by WSA, it is possible to predict the activities of smart homes and detect them with a small error.
- In each iteration, using WSA, the feature vectors are updated and used to train the multi-layer ANN.

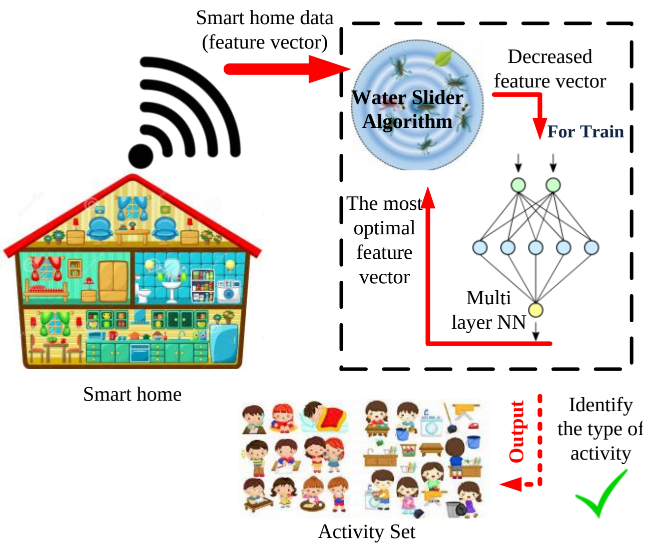


Figure 2. An overview of the proposed method.

The proposed method will be explained in detail below.

4.1. Problem Formulation of WSAMLP

The proposed method steps for detecting the kinds of the people's activities using the feature extraction by WSA and multi-layer neural network, as an activity clustering method in smart homes, are shown in Figure 3. In the proposed method, for the feature extraction phase and reduction of ANN error, WSA is used; then ANN is used for activity clustering and detection. Based on the problem formulation of the proposed method, it can be said that, firstly, a set of data related to the smart home is considered as the input. Any of these samples show one state of the smart home owning a number of features. The mentioned dataset can be pre-processed and normalized, and then divided into two groups: train group and test group. Finally, this data can be used for training ANN.

In the proposed method, the following steps could be considered for detecting the users' activities in smart homes:

- The data related to smart home and the kinds of the user's activities are considered as the input;
- The data is pre-processed and normalized;
- The data is divided into the train and test groups;
- A feature vector is considered as the member of WSA, and a number of such feature vectors are randomly created with the 0 and 1 values;
- Each feature vector owning 0 component shows that this feature is not important for learning in neural network; instead, component 1 is considered important, and is used for learning in the multi-layer ANN;
- Each feature vector is a member of WSA, and can be evaluated by goal function. Goal function uses two components: the mean error of activity detection and the number of extracted features. Each feature vector able to simultaneously reduce these two components is considered the optimal feature vector;
- Feature vectors can, after evaluation by goal function, be updated in WSA via the equations available;
- Each feature vector is required to be revised after being updated and is needed to be transformed to a binary space with the help of a conversion function.

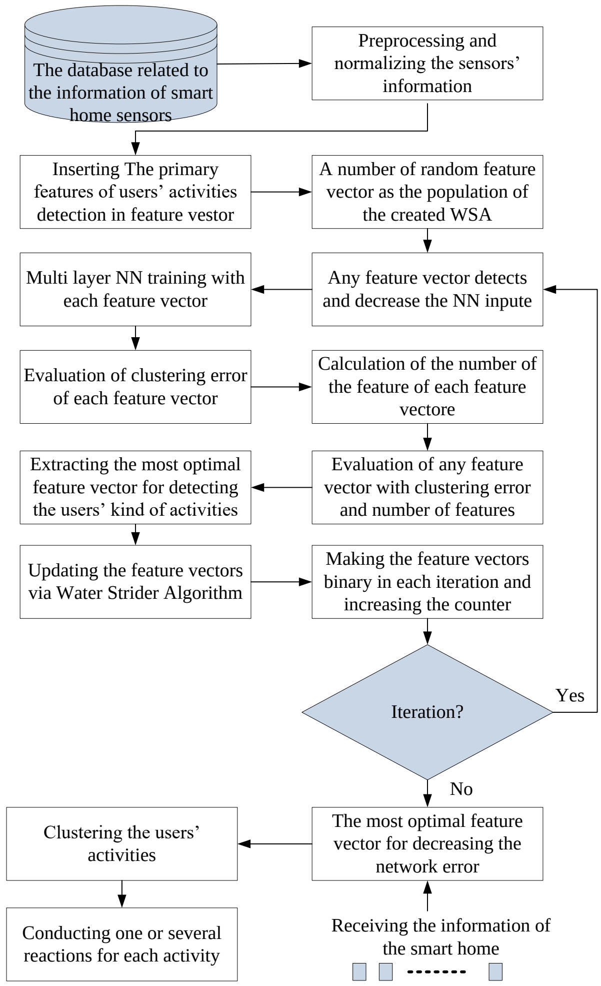


Figure 3. Problem Formulation of Proposed Method for Detecting User's Activities in Smart Homes.

4.2. Feature Vector Codding

In the proposed method, every feature vector can be considered similar to (1), which is, in fact, a member of Water Strider Optimization Algorithm:

$$F_i = (F_i^1, F_i^2, \dots, F_i^D) \qquad (1)$$

In Eq. 1, Fi is a feature vector or a Water Strider, which has the dimension (D). On the other hand, F_{ji} is and indicator of the 'j'th feature in the 'i'th vector. A random population of Water Striders or feature vectors can be randomly created by (2):

$$F=(F_1,F_2,...,F_n) \tag{2}$$

where n is the total number of feature vectors with the members of WSA. Here, the aim is to find the optimal feature vector or $F_i=(F_i^1,F_i^2,...,F_i^D)$ which is, in fact, a vector calculated by goal function; in other words, finding any feature vector with a value less than the goal function. It is possible to define the primary population of the feature vectors in a matrix like pop as (3), in which every row is a feature vector, and its columns are the features used in a feature vector:

$$Pop=\begin{bmatrix} F_1=(F_{11},F_{12},...,F_{1d}) \\ F_2=(F_{21},F_{22},...,F_{2d}) \\ F_3=(F_{31},F_{32},...,F_{3d}) \\ ... \\ F_n=(F_{n1},F_{n2},...,F_{nd}) \end{bmatrix} \tag{3}$$

4.3. Goal Function

In order to evaluate any feature vector, the two important following factors could be taken into account, and one can re-write a feature extraction goal function based on these two factors:

1. The error mean of classification and detection of the kind of activity performed by the smart home user via a feature vector;
2. The number of extracted features for detecting the kind of activity performed by the smart home user.

If one feature vector, in detecting the kind of the activity performed by the smart home user, can minimize this index, one can say that an optimal feature vector is used for training the multi-layer ANN in detecting the kind of activity performed by the smart home user. Each feature vector can be used for training ANN, and can be evaluated based on evaluation or goal function of (4):

$$f=\alpha\times mse+(1-\alpha)\times\frac{\|S\|}{\|A\|} \tag{4}$$

In (4), $\|A\|$ is the total number of features that can be considered for detecting the smart home activities, $\|S\|$ is the number of features extracted in a feature vector, which can be considered equal to the number of 1s used in a feature vector, MSE is the error mean of estimation of activity clustering in a smart home, and α is a random number in the 0-1 interval.

4.4. Optimization of Neural Network Input

In the proposed method, every solution is a feature vector. First, based on (5), a number of them are created randomly as the population of WSA. These feature vectors are randomly created in the question space between a top and bottom threshold:

$$WS_i^0=Lb+(Ub-Lb)\times rand(0,1) \tag{5}$$

In (5), *i* is the digit of a solution or a Water Strider in the '0'th iteration or the start of the algorithm. *Lb* is the bottom threshold of any solution, *Ub* is the top threshold of any solution, and rand (0,1) is the random number between 0 and 1. In WSA, there exist some areas and territories; if the total population of the feature vector is *nws* and *nt* territories exist, the *ntl/nws* territories exist in every area. In every territory, the best strider is a female, and the other members are male members; in other words, several feature vectors exist in one area, and the best vector is a female, and other vectors with less competence are considered the male vectors. Feature vectors can be updated based on various phases of WSA. Any feature vector can participate in mating phase and update itself accordingly. For the feature vector mating, there is a p probability; the probability of lack of mating phase for feature vector is considered 1-p. Modeling the success or failure behavior in mating of feature vectors can be updated based on (6) for detecting the kind of activity of the smart home user:

$$WS_i^{t+1}=\begin{cases} WS_i^t+R\times rand & Mating \\ WS_i^t+R\times(1+rand) & No\ mating \end{cases} \tag{6}$$

In (6), WS_i^t is the situation of a feature vector or a Water Strider in iteration t, and WS_i^{t+1} is the situation of Water Strider or its corresponding feature vector in the new iteration or t+1.

$$R=\left\|WS_F^t-WS_i^t\right\| \tag{7}$$

In (7), *R* is the distance between a male Water Strider like WS_i^t and a female Water Strider like WS_F^t. The nutrition equation of Water Striders, for updating the feature vectors, can be calculated based on (8):

$$WS_i^{t+1}=WS_i^t+2rand\times(WS_{BL}^t-WS_i^t) \tag{8}$$

In (8), a Water Strider or corresponding feature vector investigates the space between itself (WS_i^t) and the best solution (WS_{BL}^t).

4.5. Making Feature Vector Binary

In the proposed method, in every iteration of WSA, the feature vectors can be updated but there exists the probability that the problem space converts from binary into continuous.

$$T(F_i) = \frac{1}{1+\exp(-a, F_i)} \tag{9}$$

$$T(F_i) = \frac{|F_i|}{\sqrt{a+F_i^2}} \tag{10}$$

In order to convert a continuous space to a feature (binary) space, the conversion functions can be used. In other words, by updating the feature vectors, we can convert them into the 0-1 form, and to do this, we can use the conversion functions such as S and V, the rules of which are formulated below in (9) and (10), respectively, and their representations are illustrated in Figures 4 and 5, respectively.

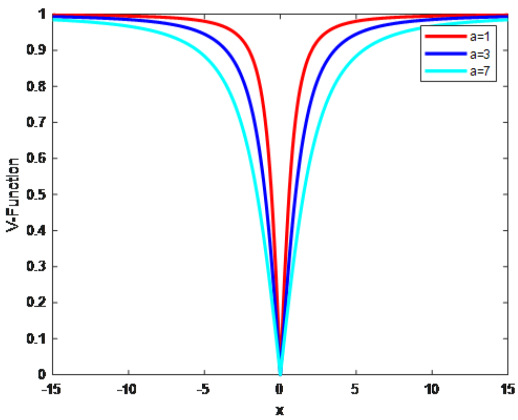


Figure 4. Conversion functions V.

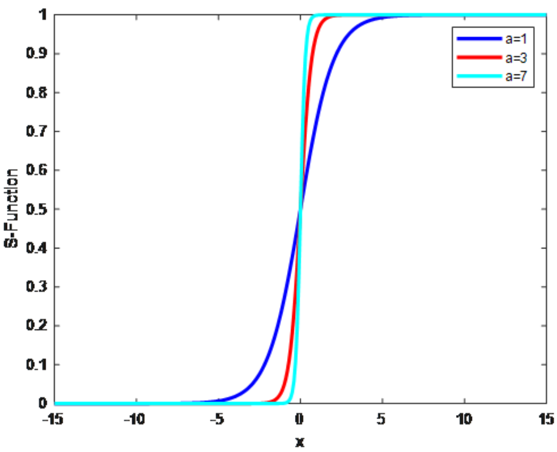


Figure 5. Conversion functions S.

Using the conversion function V, the feature vector can be taken out of the probable non-0-1 state, and it can be converted to the 0-1 vectors. In this situation, (9) can be used to have the feature space in a binary form. Using the conversion function V, the feature vector can be taken out of the non-0-1 state, and it can be converted to the 0-1 vector. In this situation, (11) can be used in order to convert the feature space to a binary form:

$$F_i^j = \begin{cases} 0 & rand < \dfrac{1}{1+e^{-F_i^j}} \\ 1 & rand \geq \dfrac{1}{1+e^{-F_i^j}} \end{cases} \tag{11}$$

Using this instruction, we can map any feature vector component that is gone out of the binary form so as to get it back to binary again, and the feature vector will maintain its binary state.

The pseudo-code of the proposed method is as follows:

Proposed method pseudo-code

- *Setting basic parameters such as number of neural network layers and number of hidden neurons.*
- *Input data of smart home and pre-processing and normalization of data*
- *Divide the data into two categories: training and testing*
- *Adjustment of multi-layer ANN and its parameters*
- *A number of feature vectors are set as members of WSA*

WSAMLP Algorithm (smart home data)

```
begin
    Random creation of feature vectors
    for i=1 to n do
    for i=1 to D do
    WSA_{i,j} = L + (U - L).rand(0,1)
    end for
    end for
    Neural network training with feature vector
  evaluation of the initial population by CostFunction()
    for i=1 to n do
    Cost(WSA_i) = f = a.mse + (1 - a).||S|| / ||A||
    end for
    it=1;
    While(it<=MaxIt) do
      WS_BL^t = min(Cost(pop))
    Mating behavior;
    for i=1 to n do
    if rand<=p
      WS_i^{t+1} = WS_i^t + R.rand
    Else
      WS_i^{t+1} = WS_i^t + R.(1+rand)
      endif
    end for
    Feeding behavior;
    for i=1 to n do
```

$$WS_i^{t+1} = WS_i^t + 2rand(WS_{BL}^t - WS_i^t)$$

end for
Create new solutions
Eliminate inappropriate solutions
Binary with conversion functions S, V
it=it+1
end while
return the best solution and neural network training in detecting the activities of smart home users
Neural network evaluation to detect the activity of smart home users
end

5. Simulation of Results

In order to implement the proposed method that combines a multi-layer ANN and Water Strider Optimization Algorithm (WSAMLP) for activity detection in the smart homes, the Matlab software was employed. The data was collected from the smart homes, and the experiments were analyzed and evaluated by this software. Later in this section, the issues related to datasets, normalization, and evaluation parameters are discussed, and then analysis and evaluations are performed, and comparisons are made with other similar methods.

5.1. Dataset

The data is collected from the USA's University of Washington database, and then subjected to the necessary pre-processing.
The Tulum 2010 dataset [25] was collected at the Washington State University, belonging to smart homes from 2009 to 2010, and people have lived in these homes for some time and their activities have been studied. A sample map of the smart home, provided by the University of Washington, can be seen as shown in Figure 6 [25]. This figure represents a map of two homes, each of which has some sensors collecting the environmental information including instance, motion sensors information represented by the letter M or thermal sensor information shown by the letter T.

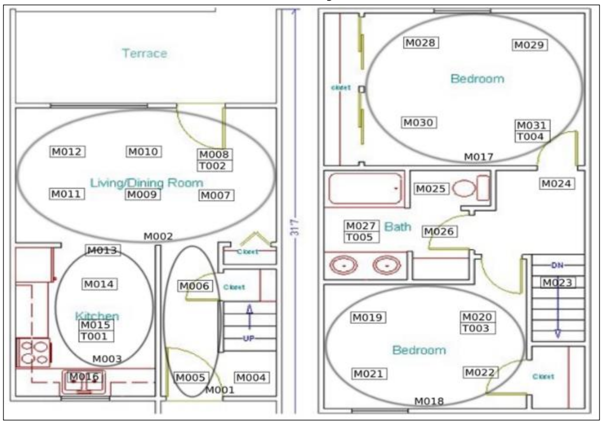


Figure 6. A smart home layout in Tulum 2010 dataset [25]**.**

Several scenarios can be considered and used in order to evaluate the data mining algorithms. Here, we can use the data and datasets, the output of which has multi-classes, with each class presenting the information of one single activity in the house. In fact, we can determine the type of activity class by predicting this output. This online database for smart homes has several laboratories that provide online data to the researchers, and the data may be used to evaluate the data mining algorithms. In order to mention some sensors related to the smart homes, we can name the motion sensors, kitchen-related sensors, humidity meter, etc. Using the information provided by the sensors and their values, it is possible to evaluate the smart home condition to some extent, and identify the kind of activity being done. In this dataset, 5 different activities are presented including calling, washing hands, cooking, eating, cleaning, and so on.

5.2. Pre-processing

The range of attributes and features that can be used to analyse the activities of the smart homes has many variations, and it is necessary that the range of their changes is considered as the same. For this purpose, normalizing is being done. If the value of a non-normalized feature in the dataset is $f(n)$ for attribute n, and the normalized value of that attribute is assumed to be $f_n(n)$, the normalizing process can be done according to (12):

$$f_n(n) = \frac{f(n) - \min}{\max - \min}(b-a) + a \tag{12}$$

In (12), [a,b] is considered the normalization range in the dataset, and Max and Min are, respectively, considered to be the maximum and minimum values of a feature in the dataset. By normalization, the range of changes in all attributes becomes the same, and the learning becomes more precise.

5.3. Evaluation Criteria

In order to evaluate the proposed method (WSAMLP) for detecting the activities in smart homes, the indicators such as accuracy, precision, and F1 (F-score with parameter 1) can be used, which are calculated as indicated in (13) to (16):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$recall = \frac{TP}{TP + FN} \tag{15}$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{16}$$

In these equations, TP, TN, FP, and FN are the number of True Positive, True Negative, False Positive, and False Negative samples, respectively.

5.4. Analysis of Goal Function with Variable Population

The reduction of the feature extraction goal function in the proposed method is due to the following two reasons:

- Reduction in the error mean of activity kind detection at smart home;
- Reduction in the number of data and features.

In order to evaluate the quality of the proposed method, a multi-layer ANN with two layers is taken into consideration, where each layer has 20 artificial neurons, and, on the other hand, the number of primary features of the dataset is equal to the total number of features of the dataset and the output also shows 5 different classes. According to the evaluations performed in this section, the population size is considered to be 10, and the number of iterations is assumed to be 50. In the experiments, the training data constitute 70%, while the test data constitute 30% of the total data.

WSA or feature vector population changes from 10 to 20 with feature vectors to clearly demonstrate the effect of this parameter, as shown in Figure 7.

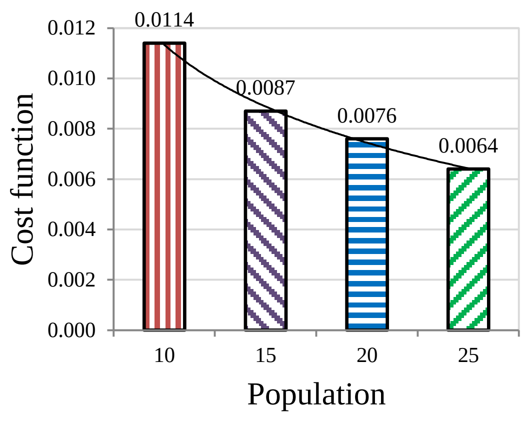


Figure 7. Reduction of feature extraction function based on feature vector population.

According to Figure 7, by increasing the population size from 10 to 25, the value of the feature extraction goal function decreases from 0.0114 to 0.0064, and this decrease is about 43.85%. In fact, the role of population increase of feature vectors in WSA is to minimize the value of the feature extraction goal function, and it has two indicators, namely the error mean and the number of features. The decrease in the value of the goal function is due to a better feature extraction while increasing the population of feature vectors. In other words, as the number of feature vectors increases, the chances of finding a better feature vector increase.

5.5. Clustering Error Analysis

The activity detection problem is a clustering problem, and therefore, for analysing the kind of activity, the indicators such as the error mean of the activity kind detection and the error square of activity kind detection at the smart home can be used, as in (17) and (18), respectively:

$$mse = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y_i})^2 \tag{17}$$

$$rmse = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y_i})^2} \tag{18}$$

In (17) and (18), y_i is the actual class of the ‘i’th activity, $\overline{y_i}$ is the estimated value of that activity class, and N is the number of evaluation samples. The experiments suggest that the error trend in terms of iteration of WSA due to feature extraction is a descending trend. In other words, in iteration number one of water strider optimization algorithm, only the neural network is learning, and its error in this state is like when water strider optimization algorithm has not been run. However, in the last iteration, the error reduction is due to the existence of WSA. Therefore, it can be concluded that the error in the first iteration is only related to neural network without WSA, and that the error in the last iteration is concerned with ANN with WSA. Table 1 represents a comparison between the states with and without WSA.

Table 1. Comparison of activity detection error with and without butterfly algorithm.

Method	RMSE	MSE
Neural network without feature extraction	0.085	0.0074
Neural network with feature selection of water strider algorithm (WSAMLP)	0.054	0.0036

It can be found out that the WSAMLP algorithm has less error in detecting the kind of activity of the users in a smart home than ANN without WSA.

5.6. Clustering Indicators Analysis

In order to evaluate the proposed method, its value of precision, accuracy, and F1 can be compared with the other data mining methods and with the results of the study [25] in 2020.

In Table 2, the mean values of precision, accuracy, and F1 indicators of the proposed method have been evaluated and compared with the KNN, NB, Ripper, C 4.5, RF, BOA, HHO, and BWO methods.

Table 2. Comparison of proposed method in terms of precision, accuracy, sensitivity, and F1 in detecting the kind of activities of users of smart homes.

Method	Accuracy	Precision	F1
KNN	77.10	77.20	77.10
RIPPER	80.30	80.50	80.30
C4.5	80.90	81.00	80.90
RF	83.60	88.00	87.90
CASAS	88.10	88.00	87.90
BOA	96.82	96.49	96.18
HHO	96.76	96.44	97.22
BWO	97.28	96.85	97.14
WSAMLP	97.63	97.12	97.54

The analysis and comparison of the proposed method with respect to precision, accuracy, and F1 with the methods containing and excluding feature extraction like Butterfly, Hawk, and Black Widow optimization algorithms are shown in Figures 8, 9, and 10.

The experiments show that the activity detection precision at a smart home using KNN, Ripper, C 4.5, RF, CASAS, BOA, HHO, BWO and the WSAMLP are 77.10%, 80.30%, 80.90%, 83.60%, 88.10%, 96.82%, 96.76%, 97.28%, and 97.63%, respectively. The comparisons indicate that the proposed method has a more precision compared to the other methods; the worst performance is produced by the KNN method, and among the meta-heuristic methods, HHO, proposed in 2019, has the least precision.

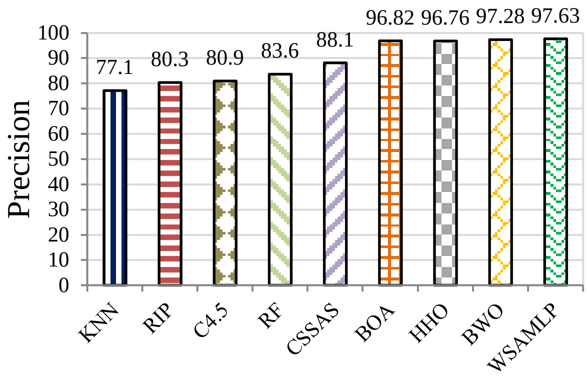


Figure 8. Comparing proposed method with other methods with respect to precision.

In Figure 9 the activity detection accuracy indicators at a smart home using KNN, NB, Ripper, C 4.5, RF, BOA, HHO, BWO, and the proposed method are 77.20%, 80.50%, 81.00%, 88.00%, 96.49%, 96.85%, and 97.12%, respectively, and the proposed method provides the highest accuracy.

In Figure 11, the sensitivity index in KNN, Ripper, C 4.5, RF, CASAS, BOA, HHO, BWO, and WSAMLP are compared.

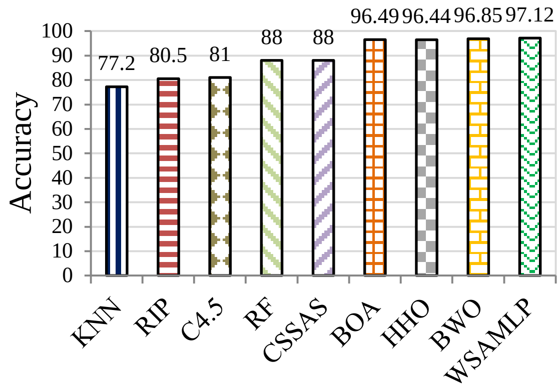


Figure 9. Comparing proposed method with other methods with respect to accuracy.

Figure 10 shows the mean values of F1 for the users' activity detection in the mentioned methods are 77.10%, 80.30%, 80.90%, 87.90%, 96.18%, 97.22%, 97.14%, and 97.54%, respectively. The analysis indicates that the meta-heuristic methods in combination with the data mining methods, compared to other methods with no feature extraction such as KNN, Support Vector Machine, Social Network Optimization, NB, C 4.5, and RF can detect the type of users' activities at smart homes more precisely because they provide important features to learn in clustering tools.

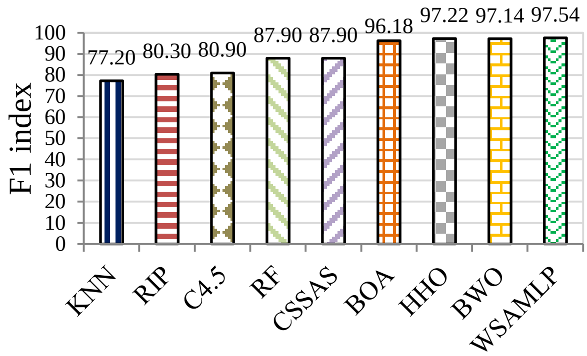


Figure 10. Comparing proposed method with other methods with respect to F1.

Figure 11 indicates that the sensitivity index values in KNN, Ripper, C 4.5, RF, CASAS, BOA, HHO, BWO, and WSAMLP for detecting the smart home user activities are 76.44%, 79.68%, 8036%, 82.56%, 87.38%, 97.02%, 96.14%, 97.12%, and 97.54%, respectively.

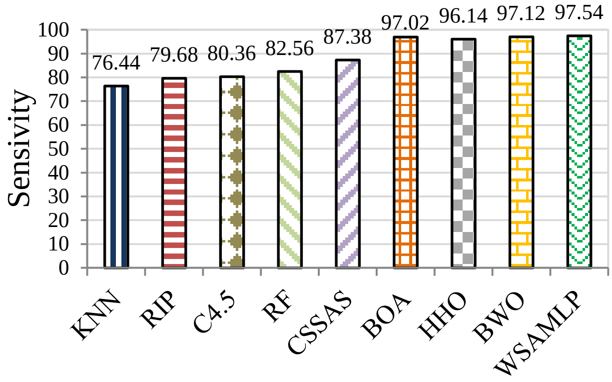


Figure 11. Comparing proposed method with other methods with respect to sensitivity.

Among the methods compared, the proposed method ranks first in the four indicators of accuracy, sensitivity, precision, and F1. In the accuracy, sensitivity, and precision index, the BWO algorithm is in the second place, and in the F1 index, the HHO algorithm is in the second place. The reasons why the proposed method has a better performance in detecting the activities of the smart home users than similar metaheuristic algorithms such as BOA, HHO, and BWO are as follows:

1- The ability to generate solutions in terms of competence exists only in this algorithm but not in the algorithms such as HHO and BOA. The BWO algorithm has this mechanism but it is similar to the genetic algorithm, and the two solutions ultimately create only two solutions.

2- In the WSA algorithm, there is a search for the optimal solution but not in BWO. In the HHO and BOA algorithms, only one optimal solution is searched per iteration. In the WSA algorithm, a search is performed around several optimal solutions, and therefore, more problem space is searched.

3- There is no mortality mechanism in the HHO and BOA algorithms but there is in WSA and BWO. In WSA, unsuitable solutions are removed more intelligently than BWO.

5.7. Execution Time Analysis

In order to better evaluate the proposed method, its execution time is compared with the BOA, BWO, and HHO algorithms in detecting the user activities in Figure 12.

The execution time in detecting the type of activity in the BOA, HHO, BWO algorithms, and the proposed method and terms of seconds is equal to 1.36, 2.89, 2.16, and 1.28, respectively. Among the methods compared, the proposed method has the shortest time to detect the type of activity, and the worst performance is related to the HHO algorithm.

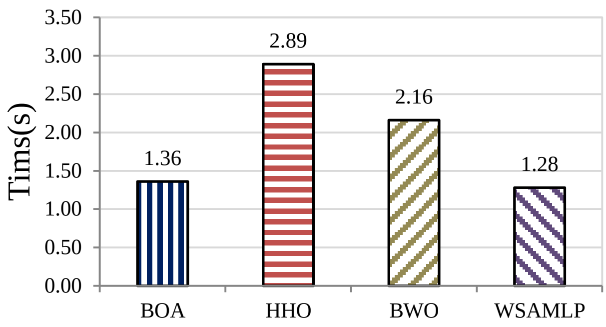


Figure 12. Comparison of time to detect type of activities in s.

5. Conclusion

There are sets of smart things in the smart homes that can collect environmental information at a house and send it to the users and owners of the house. Smart objects at home can process and analyse the environmental information and report the activities done at home to the users. Sometimes, in some cases, the smart home activities are required to be taken care of more precisely, especially if the users intend to monitor children or the elderly inside home, and increase their security. One of the major challenges at the smart homes is how to use the information provided by the sensors in order to detect the type of activity and make decisions accordingly. In this paper, a hybrid method based on group intelligence of water striders and clustering by ANN was presented in order to detect the kind of activities of the users at a smart home. In the proposed method, the role of WSA is to extract the optimized feature vector for learning a multi-layer ANN. The experiments indicate that the proposed method provides a mean precision indicator of 97.63%, an accuracy indicator of 97.12%, and an F1 indicator of 97.54%. The proposed method has a better performance in the users' activity detection with respect to the mentioned indicators, compared to the KNN, Ripper, C 4.5, RF, CASAS, BOA, HHO, and BWO methods. In a future research work, deep learning will be employed as a method to cluster and extract the features in order to detect the users' kind of activities.

References

[1] M. Lupión, J. L. Redondo, J. F. Sanjuan, and P. M. Ortigosa, "Deployment of an IoT Platform for Activity Recognition at the UAL's Smart Home," in *International Symposium on Ambient Intelligence*,

2020, pp. 82–92.

[2] M.-D. González-Zamar, E. Abad-Segura, E. Vázquez-Cano, and E. López-Meneses, "IoT technology applications-based smart cities: Research analysis," *Electronics*, vol. 9, no. 8, p. 1246, 2020.

[3] P. Agarwal and M. Alam, "Investigating IoT middleware platforms for smart application development," *Smart Cities—Opportunities and Challenges; Springer: Singapore*, pp. 231–244, 2020.

[4] N. Hossein Motlagh, M. Mohammadrezaei, J. Hunt, and B. Zakeri, "Internet of Things (IoT) and the energy sector," *Energies*, vol. 13, no. 2, p. 494, 2020.

[5] A. Mousavi, A. Sheikh Mohammad Zadeh, M. Akbari, and A. Hunter, "A New Ontology-Based Approach for Human Activity Recognition from GPS Data," *J. AI Data Min.*, vol. 5, no. 2, pp. 197–210, 2017.

[6] K. S. Gayathri, K. S. Easwarakumar, and S. Elias, "Fuzzy ontology based activity recognition for assistive health care using smart home," *Int. J. Intell. Inf. Technol.*, vol. 16, no. 1, pp. 17–31, 2020.

[7] B. Reeder, J. Chung, K. Lyden, J. Winters, and C. M. Jankowski, "Older women's perceptions of wearable and smart home activity sensors," *Informatics Heal. Soc. Care*, vol. 45, no. 1, pp. 96–109, 2020.

[8] H. M. Do, M. Pham, W. Sheng, D. Yang, and M. Liu, "RiSH: A robot-integrated smart home for elderly care," *Rob. Auton. Syst.*, vol. 101, pp. 74–92, 2018.

[9] N. E. Kogan *et al.*, "An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time," *Sci. Adv.*, vol. 7, no. 10, p. eabd6989, 2021.

[10] D. Agrawal *et al.*, "Enhancing smart home security using co-monitoring of iot devices," in *Companion of the 2020 ACM International Conference on Supporting Group Work*, 2020, pp. 99–102.

[11] J. Kang, M. Kim, and J. H. Park, "A reliable TTP-based infrastructure with low sensor resource consumption for the smart home multi-platform," *Sensors*, vol. 16, no. 7, p. 1036, 2016.

[12] A. A. Alani, G. Cosma, and A. Taherkhani, "Classifying imbalanced multi-modal sensor data for human activity recognition in a smart home using deep learning," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8.

[13] P. N. Huu, Q. T. Minh, and others, "An ANN-based gesture recognition algorithm for smart-home applications," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 5, pp. 1967–1983, 2020.

[14] S. B. ud din Tahir, A. Jalal, and M. Batool, "Wearable sensors for activity analysis using SMO-based random forest over smart home and sports datasets," in *2020 3rd International Conference on Advancements in Computational Sciences (ICACS)*, 2020, pp. 1–6.

[15] X. Guo, Z. Shen, Y. Zhang, and T. Wu, "Review on the application of artificial intelligence in smart homes," *Smart Cities*, vol. 2, no. 3, pp. 402–420, 2019.

[16] V. Ghasemi, A. Pouyan, and M. Sharifi, "A sensor-based scheme for activity recognition in smart homes using dempster-shafer theory of evidence," *J. AI Data Min.*, vol. 5, no. 2, pp. 245–258, 2017.

[17] M. Kaur, G. Kaur, P. K. Sharma, A. Jolfaei, and D. Singh, "Binary cuckoo search metaheuristic-based supercomputing framework for human behavior analysis in smart home," *J. Supercomput.*, vol. 76, no. 4, pp. 2479–2502, 2020.

[18] A. Kaveh and A. D. Eslamlou, "Water strider algorithm: A new metaheuristic and applications," in *Structures*, 2020, vol. 25, pp. 520–541.

[19] L. G. Fahad and S. F. Tahir, "Activity recognition and anomaly detection in smart homes," *Neurocomputing*, vol. 423, pp. 362–372, 2021.

[20] Y. Zhang, G. Tian, S. Zhang, and C. Li, "A knowledge-based approach for multiagent collaboration in smart home: From activity recognition to guidance service," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 2, pp. 317–329, 2019.

[21] R. A. Hamad, A. S. Hidalgo, M.-R. Bouguelia, M. E. Estevez, and J. M. Quero, "Efficient activity recognition in smart homes using delayed fuzzy temporal windows on binary sensors," *IEEE J. Biomed. Heal. informatics*, vol. 24, no. 2, pp. 387–395, 2019.

[22] S. Yan, K.-J. Lin, X. Zheng, and W. Zhang, "Using latent knowledge to improve real-time activity recognition for smart IoT," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 574–587, 2019.

[23] M. Jethanandani, A. Sharma, T. Perumal, and J.-R. Chang, "Multi-label classification based ensemble learning for human activity recognition in smart home," *Internet of Things*, vol. 12, p. 100324, 2020.

[24] Y. Hu, B. Wang, Y. Sun, J. An, and Z. Wang, "Graph-Based Semi-Supervised Learning for Activity Labeling in Health Smart Home," *IEEE Access*, vol. 8, pp. 193655–193664, 2020.

[25] J. Guo, Y. Li, M. Hou, S. Han, and J. Ren, "Recognition of daily activities of two residents in a smart home based on time clustering," *Sensors*, vol. 20, no. 5, p. 1457, 2020.

[26] H. Yang, S. Gong, Y. Liu, Z. Lin, and Y. Qu, "A multi-task learning model for daily activity forecast in smart home," *Sensors*, vol. 20, no. 7, p. 1933, 2020.

[27] Y. Du, Y. Lim, and Y. Tan, "A Novel Human Activity Recognition and Prediction in Smart Home Based on Interaction," *Sensors*, vol. 19, no. 20, 2019.

[28] Z. Fu, X. He, E. Wang, J. Huo, J. Huang, and D. Wu, "Personalized Human Activity Recognition Based on Integrated Wearable Sensor and Transfer Learning," *Sensors*, vol. 21, no. 3, p. 885, 2021.

[29] A. R. Javed, R. Faheem, M. Asim, T. Baker, and

M. O. Beg, “A smartphone sensors-based personalized human activity recognition system for sustainable smart cities,” *Sustain. Cities Soc.*, vol. 71, p. 102970, 2021.

[30] L. G. Fahad and S. F. Tahir, “Activity recognition in a smart home using local feature weighting and variants of nearest-neighbors classifiers,” *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 2, pp. 2355–2364, 2021.

[31] D. Das *et al.*, “Explainable Activity Recognition for Smart Home Systems,” *arXiv Prepr. arXiv2105.09787*, 2021.

WSAMLP: تشخیص فعالیت در خانه های هوشمند مبتنی بر الگوریتم حشره آبسوار و شبکه عصبی مصنوعی در شهر هوشمند

جمیله برازنده و نازبانو فرزانه*

گروه کامپیوتر، دانشکده مهندسی، دانشگاه بین المللی امام رضا(ع)، مشهد، ایران.

ارسال ۲۰۲۱/۰۴/۲۹؛ بازنگری ۲۰۲۱/۰۶/۲۹؛ پذیرش ۲۰۲۱/۰۸/۲۰

چکیده:

یکی از کاربردهای مهم اینترنت اشیاء، توسعه شهرهای هوشمن توسط این فناوری است. شهرهای هوشمند متشکل از اجزای هوشمند نظیر خانه‌های هوشمند است. در خانه‌های هوشمند انواع حسگرها برای هوشمندسازی محیط وجود دارد و می‌توان از این اشیاء هوشمند برای شناسایی فعالیتهای افراد درون خانه‌ها استفاده نمود. شناسایی فعالیتهای کاربران خانه هوشمند، می‌توان شامل مواردی مانند شناخت فعالیت‌هایی مانند تهیه غذا یا تماشای تلویزیون که توسط یک ساکن خانه هوشمند در نظر گرفته شود. شناسایی فعالیتها افراد در خانه‌های هوشمند می‌تواند به زندگی افراد سالمند یا مراقب از بچه‌ها یا حتی در موارد امنیتی کمک زیادی نماید. اطلاعات گردآوری شده توسط توسط حسگرها می‌تواند برای تشخیص نوع فعالیت کاربران مورد استفاده قرار گرفته شود اما چالش اصلی دقت اندک تشخیص بیشتر روش‌های تشخیص فعالیت کاربران است. در روش پیشنهادی برای کاهش دادن خطای طبقه‌بندی تکنیک‌های داده‌کاوی یک رویکرد یادگیری ترکیبی با استفاده از الگوریتم حشره آبسوار ارایه شده است. در روش پیشنهادی الگوریتم حشره آبسوار می‌تواند درفاز انتخاب ویژگی استفاده شود و فقط ویژگی‌های مهم را برای یادگیری ماشین انتخاب نماید. تجزیه و تحلیل روش پیشنهادی نشان می‌دهد روش پیشنهادی دارای دقت برابر۹۷,۶۳٪، صحت برابر ۹۷,۱۲٪ و شاخص F1 (شاخص امتیاز) برابر۹۷,۵۴٪است. نسبت به الگوریتم‌های مشابه مانند الگوریتم بهینه‌سازی پروانه، الگوریتم بهینه‌سازی شاهین، و الگوریتم بهینه‌سازی بیوه سیاه دارای دقت بیشتری در تشخیص فعالیت کاربران است.

کلمات کلیدی: اینترنت اشیاء، خانه هوشمند، فعالیت کاربران، داده‌کاوی، الگوریتم حشره آبسوار.