



Ensemble residual network-based gender and activity recognition method with signals

Turker Tuncer¹ · Fatih Ertam¹ · Sengul Dogan¹ · Emrah Aydemir² · Paweł Pławiak^{3,4}

Published online: 22 February 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Nowadays, deep learning is one of the popular research areas of the computer sciences, and many deep networks have been proposed to solve artificial intelligence and machine learning problems. Residual networks (ResNet) for instance ResNet18, ResNet50 and ResNet101 are widely used deep network in the literature. In this paper, a novel ResNet-based signal recognition method is presented. In this study, ResNet18, ResNet50 and ResNet101 are utilized as feature extractor and each network extracts 1000 features. The extracted features are concatenated, and 3000 features are obtained. In the feature selection phase, 1000 most discriminative features are selected using ReliefF, and these selected features are used as input for the third-degree polynomial (cubic) activation-based support vector machine. The proposed method achieved 99.96% and 99.61% classification accuracy rates for gender and activity recognitions, respectively. These results clearly demonstrate that the proposed pre-trained ensemble ResNet-based method achieved high success rate for sensors signals.

Keywords Ensemble residual network · Gender identification · Daily sport activity recognition · Sensor signals · Machine learning

✉ Turker Tuncer
turkertuncer@firat.edu.tr

¹ Department of Digital Forensics Engineering, Technology Faculty, Firat University, Elazig, Turkey

² Department of Computer Engineering, Faculty of Engineering and Architecture, Kirsehir Ahi Evran University, Kirsehir, Turkey

³ Department of Information and Communications Technology, Faculty of Computer Science and Telecommunications, Cracow University of Technology, Warszawska 24 st., F-3, 31-155 Krakow, Poland

⁴ Institute of Theoretical and Applied Informatics, Polish Academy of Sciences, Bałtycka 5, 44-100 Gliwice, Poland

1 Introduction

Human activity recognition (HAR) or action recognition is a challenging but popular field of research in signal and image processing. HAR basically includes automatic detection, recognition and analysis of human actions from data from different sensor types such as range sensor, RGB camera, depth sensor or inertia sensor [1]. In recent years, the widespread use of mobile devices [2] has made HAR a new research area in artificial intelligence and pattern recognition based on wearable sensors [3]. The purpose of action recognition or analysis is to determine which action appears in the data. Sensors such as accelerometers, gyroscopes and magnetometers [4] built into mobile devices can generate time series data for HAR. Over the last few years, research on HAR has gained considerable popularity and is becoming increasingly vital in various disciplines. Information extraction using artificial intelligence applications is very important for HAR applications [1]. HAR has been successfully applied in many areas such as sports training, remote health monitoring, health self-management, military practice, play, home behavior analysis, gait analysis and gesture recognition [5, 6]. Various sensor methods have been presented to monitor people and their activities. HAR approaches can generally be divided into two main categories, depending on the type of sensors used. These categories are: vision-based HAR and inertial sensor-based HAR. Vision-based HAR has developed rapidly in recent years. Thanks to the use of cameras, activity classification was made by monitoring and recognition [7]. There are vision-based studies using Kinect camera to analyze indoor activities [8]. With the development of deep learning techniques, convolutional neural networks have been used for vision-based HAR studies and successful results have been obtained [9]. The major disadvantage of vision-based approaches is that poor results cannot be achieved in dark environments and are not suitable for installation in areas where personal privacy is to be protected. In addition, the fact that it is installed in a fixed location has the disadvantage that it allows only the activities to be determined in the area where it is established [10]. With the development of electronic systems, small size and light inertia detection devices, which have lower power consumption, are widely used in various digital products such as mobile phones and computers. Therefore, sensor-based HAR systems are used in most of today's studies [6, 11–13]. Sensor-based systems can be divided into single sensor-based systems and body network-based systems. Body sensor network-based systems have been used for crowd analysis, gait analysis or training of sports branches [14, 15]. Even if the body sensor network-based activity recognition system provides an increase in generalization performance, it is impossible to use it for a long time in real life. It is very difficult for the body to remain unchanged for a long time and to remain unchanged against external influences. In addition, the high cost is another disadvantage. Researches are widely developed on single sensor-based systems. In order to make the drop detection, the researchers used a single accelerometer and searched four positions in the human body to place this accelerometer [16]. In another study, a single accelerometer placed on the wrist and the performance of using

template matching method in activity recognition were analyzed [17]. In another study, the researchers compared the waist-mounted smartphone and the power of accelerometer and gyroscope used to recognize human physical activity [18].

Deep learning methods are phenomena of machine learning. For big datasets, deep learning methods have achieved high success rates. Therefore, they have been used in many areas and HAR is one of them. However, million parameters should be set in deep learning methods, and they need high-cost hardware, for instance graphical processing unit, tensor processing unit. One of the most important problems of the deep networks is weight assignment. Deep learning methods need big dataset to assign weight correctly, and this process has long execution time. In order to overcome this problem, pre-trained networks have been used, and these networks have generally been trained in ImageNet dataset. In the pre-trained networks, the calculated final weights are used. Therefore, high classification rates can be achieved in a short execution time by using pre-trained network. The main motivation of this study is to propose an ensemble feature extractor method by using pre-trained deep networks. HAR is a signal processing-based research area, but we converted all signals to image and we used pre-trained CNN. As we know from the literature, CNNs are very effective methods for computer vision. Our aims are to use this effectiveness of the CNNs on the HAR and propose a novel hybrid feature extractor by using pre-trained deep networks. A brief explanation of this method is given as below.

In this method, a novel ensemble ResNet is proposed. This method uses ResNet18, ResNet50 and ResNet101 together. Fully connected 1000 (FC1000) layers features are extracted from these networks, and these features are concatenated. Then, 1000 most distinctive features are selected. The selected feature set is forwarded to cubic SVM classifier. The main motivation of the proposed study is to obtain high classification accuracy from sensors signals by using ResNets. Therefore, we defined two cases, and the main aim of these cases is to classify daily sport activities and genders. By using these cases, highly classification capability of the proposed ensemble ResNet-based sensor signal classification method is shown. Our main motivation is to use ResNets as feature extractor and evaluate performance of them. ResNets have high classification capabilities. The used ResNets (ResNet18, ResNet50 and ResNet101) are lightweight networks (they have 18, 50 and 101 layers, respectively); hence, we selected these networks. They have both small number of layers and high classification ability. Therefore, an ensemble feature extractor is presented by using these three ResNets. The proposed ensemble feature extractor was tested on signal datasets. The key contributions of this research work are as follows:

- ResNets [19] have been used for computer vision and image classification. In the signal classification methods, recurrent neural networks have been widely used. The main problem of the recurrent neural networks is to set parameters. CNNs are very effective methods for computer vision, and there is no need parameters selection in the CNNs. Therefore, we used pre-trained CNNs (ResNet18, ResNet50 and ResNet101) as feature extractor, and effectiveness of this feature extractor is shown by using two cases.

- In the CNN-based signal classification methods, spectrograms of the signals are utilized as input of the CNNs. In this study, vector-to-matrix transformation is used and signals are converted to images by using this transformation.
- To demonstrate success of the proposed ensemble deep feature extractor, two cases were defined and these are gender and daily sport activities recognitions. These are two signal classification problems. The proposed ensemble ResNet-based feature extractor achieved 99.96% recognition rate for gender recognition. It also achieved 99.61% recognition rate for 19 daily sport activities recognition. These results prove success of the proposed method.
- Effectiveness of the proposed ensemble deep feature extractor is showed by using conventional classifiers.

2 Residual network

Nowadays, information extraction from big data has been very important research area. Therefore, many artificial intelligence and machine learning methods have been proposed in the literature. Deep learning is one of the machine learning methods developed with these parameters [20, 21]. Deep learning has become most popular branch of artificial intelligence and machine learning. Deep learning allows us to train a system to predict meaningful outputs from a large data set. Deep learning has been widely used in many research areas such as defense and security, medical researches and industrial systems [20, 22–24].

In deep learning, different architectures have been proposed such as convolution neural network (CNN), recurrent network (RN) and restricted Boltzmann machines (RBM) [25–29]. The basic logic of these architectures comes from the idea of creating structures that mimic human beings. The most widely used of these architectures is CNN, a mixture of biology and computer science. Like other architectures, CNN is based on neural networks. When defining an object, CNN tries to obtain properties that make it unique. While the curves and edges are first detected in a CNN object, the abstract concepts are created. CNN features convolutional, nonlinearity, pooling, flattening and fully connected layers to achieve features in an image. Commonly used networks in CNN architecture are LeNet [30], ResNet [19], GoogleNet [31], AlexNet [32], ImageNet, Visual Geometry Group Network (VGGNet) [33] and DenseNet. ResNet [19] has been widely used in the image and signal processing. The main aim of the ResNet is to solve vanishing gradient problem. Therefore, ResNets have many layers and use $F(x) + x$ equation. The block diagram of ResNets is shown in Fig. 1.

The widely used ResNets are ResNet18, ResNet50, ResNet101 and ResNet152, and ResNet-based deep learning methods have still been presented. For instance, ResNet18 has 18 layers; hence, it is called as ResNet18. By using ResNet-based deep learning methods, high success rates have been achieved. ResNet18, ResNet50, ResNet101 and ResNet152 have 11 M, 25.6 M, 44.5 M and 60.2 M parameters, respectively.

In this study, we used ResNet18, ResNet50 and ResNet101 together to propose ensemble ResNet.

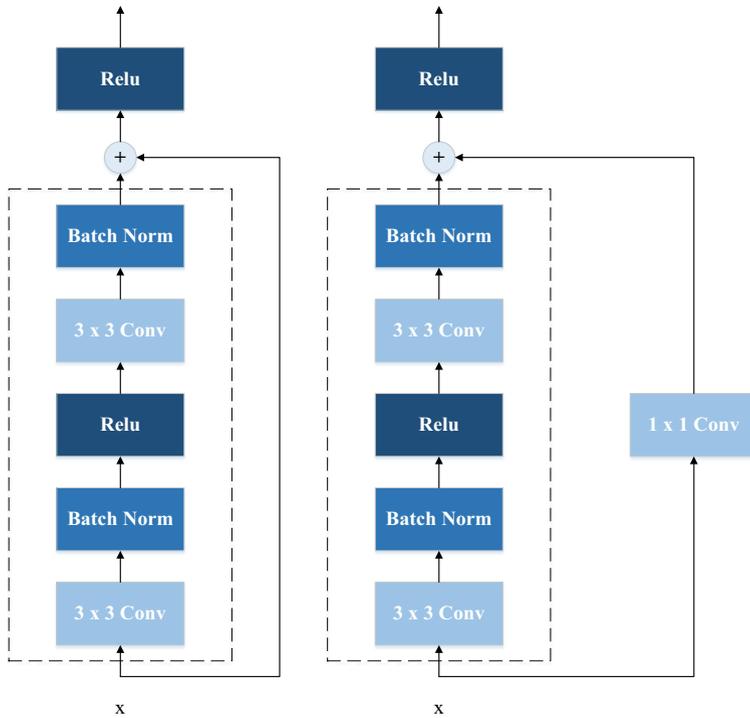


Fig. 1 Residual networks schematic explanation

3 Material

In this study, daily and sports activities data set were used. Data were used by four women and four men between the ages of 20 and 30, for a total of 19 different activities over five minutes [6]. The data obtained from torso, right arm, left arm, right leg and left leg units with the help of nine sensors includes 60 segments and eight subjects. In the data, the speed and width values of some activities differ due to the fact that subjects perform the activities in their own way. At the 25 Hz sampling frequency, data were obtained from the sensors and the five-minute signals were fragmented into five-second segments. The 19 activity types in the data are as given in Fig. 2.

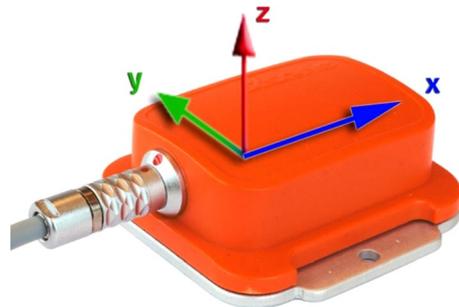
All the above data were obtained by the Xsens MTx sensor. This sensor has been developed for the orientation measurement of parts of the human body [34]. In Fig. 3, the used sensor is shown. The data obtained by the help of triaxial accelerometer, gyroscope and magnetometer within the sensors are programmed with MT manager interface.

The sensors are placed in the human body at five different locations, as shown in Fig. 4. Sensors placed around the knees, chest, wrists with duct tape are connected to a device called Xbus Master in the belt by means of cables. By using

Number	Activity	Number	Activity
A1	Sitting	A11	Walking on a treadmill with a speed of 4 km/h in 15 degree inclined positions
A2	Standing	A12	Running on a treadmill with speed of 8 km/h
A3	Lying on back side	A13	Exercising on stepper
A4	Lying on right side	A14	Exercising on cross trainer
A5	Stairs down	A15	Cycling on an exercise bike in horizontal position
A6	Stairs up	A16	Cycling on an exercise bike in vertical position
A7	Standing in elevator	A17	Rowing
A8	Moving around in elevator	A18	Jumping
A9	Walking parking lot	A19	Playing basketball
A10	Walking on a treadmill with a speed of 4 km/h in flat		

Fig. 2 The defined daily activities in the UCI HAR dataset

Fig. 3 Xsens MTx sensor and coordinates



a Bluetooth™ connection, data from the device to the receiver are obtained by connecting the receiver to the computer via USB. The data collected in accordance with the ethics committee were presented in the UCI machine learning repository [35]. This dataset consists of 9120 observations.

4 Ensemble ResNet-based signal recognition method

In this study, a novel ensemble signal recognition method is presented. We proposed a simple and effective learning method for signal classification. The main aim of this method is to achieve high classification accuracy both big and small datasets. Also, we used pre-trained networks. Any training phase was not used. We used optimal weights of the ResNets, and these weights were obtained by training ImageNet. The main aim of the proposed ensemble ResNet is to propose signal-to-matrix conversation (preprocessing), feature extraction using ensemble ResNet, feature selection by ReliefF [36] and classification phases. The graphical outline of the proposed is shown in Fig. 5.

Brief explanation of the proposed ensemble ResNet is given in Algorithm 1.

Algorithm 1. The proposed ensemble ResNet based signal processing method.

<p>Input: The raw 1D sensor signal (S) with size of 5625</p> <p>Output: Predicted results (PR).</p>
<p>0: Load signals</p> <p>1: Use vector to matrix transformation and obtain gray level image (Im).</p> <p>2: $feature_1 = ResNet18(Im)$; // $feature_1$ is FC1000 layer of the ResNet18.</p> <p>3: $feature_2 = ResNet50(Im)$; // $feature_2$ is FC1000 layer of the ResNet50.</p> <p>4: $feature_3 = ResNet101(Im)$; // $feature_3$ is FC1000 layer of the ResNet101.</p> <p>5: Concatenate $feature_1$, $feature_2$ and $feature_3$ to obtain final feature.</p> $feature^{final} = feature_1 feature_2 feature_3$ <p>6: Select the most 1000 discriminative features.</p> <p>7: Use cubic SVM with 10-fold cross validation and obtain PR.</p> <p>8: Calculate results using PR.</p> <p>10: Interpret these results.</p>

4.1 Preprocessing

In this study, sensors signals are utilized as input. However, we used pre-trained deep convolutional networks as feature extractor. Therefore, the used 1D sensor signals should be converted a 2D matrix or spectrograms of them should be extracted. Vector-to-matrix conversation is chosen as preprocessing method in this study. Pseudo-code of the used vector-to-matrix transformation is shown as Algorithm 2.

Algorithm 2. Pseudo code of the used preprocessing method.

Input: The raw 1D sensor signal (S) with size of 5625

Output: Graylevel image (Im) with size of 125×45

```

1:  $count = 1$ ;
2: for  $i=1$  to 125 do
3:   for  $j=1$  to 45 do
4:      $Im(i,j) = S(count)$ ;
5:      $count = count + 1$ ;
6:   end for  $j$ 
7: end for  $i$ 
8: Normalize  $Im$  by using min-max normalization.

```

As seen from Algorithm 2, the 1D raw signal is transformed to 125×45 sized image. The used dataset consists of text files and each text file has 125 rows and 45 columns. Therefore, we selected 125×45 as size of matrix. Mathematical explanation of this phase is also shown in below.

$$Im = \text{vec2mat}(S, [125 \times 45]) \quad (1)$$

Fig. 4 Connection type of the sensors on the body



$$I_m = \text{round} \left(\frac{I_m - I_{m_{\min}}}{I_{m_{\max}} - I_{m_{\min}}} \times 255 \right) \tag{2}$$

where $vec2mat(., .)$ is vector-to-matrix transformation, and Eq. 2 defines min–max normalization. We used min–max normalization to code calculate 8-bit gray-scale image from sensor signals.

4.2 Feature extraction by using the proposed ensemble ResNet

Deep learning methods have been widely used method in the literature. Especially, convolutional neural networks (CNN) are very hot topic for artificial intelligence. The CNNs are utilized as both learning method and feature extractor. One of the mostly used CNN is ResNet. ResNet has many variations, and widely used ResNets are ResNets18, ResNet50 and ResNet101.

In this study, pre-trained ResNets (ResNet18, ResNet50, ResNet101) are used. These networks pre-trained on ImageNet. These networks have 18 (72 sublayers), 50 (177

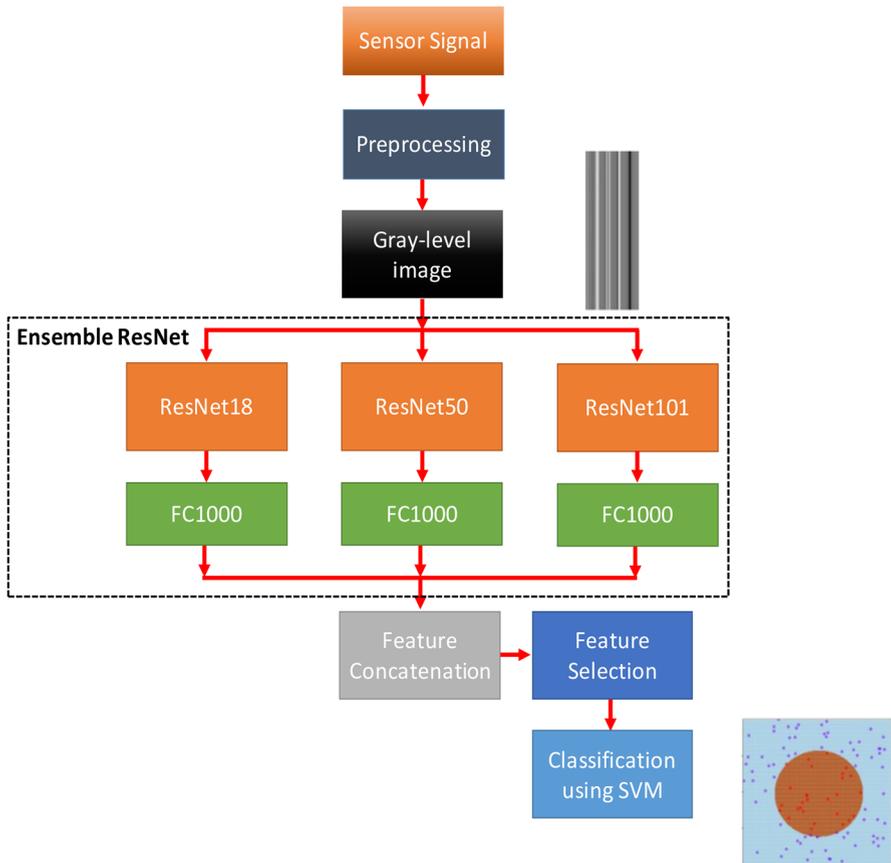


Fig. 5 Block diagram of the proposed ensemble ResNet-based signal recognition method

sublayers) and 101 (347 sublayers) layers, respectively. These networks are utilized as feature extractors in this study. Therefore, the softmax and classification layers of these networks are not used. All of these networks have FC1000 (1000 fully connected layer). The graphical explanation of the proposed ensemble ResNet is shown in Fig. 6.

As seen from Fig. 6, 1000 features are extracted. Then, these features are concatenated and 3000 final features are obtained. The mathematical explanation of this section is:

$$F_1 = \text{ResNet18}(\text{Im}) \tag{3}$$

$$F_2 = \text{ResNet50}(\text{Im}) \tag{4}$$

$$F_3 = \text{ResNet101}(\text{Im}) \tag{5}$$

where ResNet18, ResNet50 and ResNet101 feature extraction function of the feature extraction method. We used FC1000 layer of them, and each feature extraction function generates 1000 features. F_1 , F_2 and F_3 are feature vectors of the deep ResNet18, ResNet50 and ResNet101 feature generation methods, respectively. These features are concatenated using Eq. 6.

$$\text{feature} = F_1 | F_2 | F_3 \tag{6}$$

where feature is final feature with size of 3000 and | is concatenation operator.

4.3 Feature selection

In this section, the obtained 3000 features are used as input, and redundant 2000 features are eliminated by using ReliefF. ReliefF is one of the mostly used feature selector in the literature. It uses distance-based feature weighting and generates weights for all features. ReliefF generates weights by using distance metrics. In the ReliefF method, Euclidean distance is used but Manhattan distance is used in the ReliefF method. ReliefF method generates both negative and positive weights. The big weights imply distinctive features, and small weights describe redundant features [36, 37]. By using the generated weights, the feature selection is processed. Equations of weights are generation process of ReliefF.

$$W(ft_i) = W(ft_i) - \frac{\sum_{j=1}^k \text{dist}(A, R, H)}{n * k} + \frac{\sum_{C \neq \text{class}(R)} \left[\frac{P(C)}{1-P(R)} * \sum_{l=1}^k \text{dist}(A, R, M) \right]}{n * k} \tag{7}$$

$$\text{dist}(A, L_1, L_2) = \begin{cases} 0, & L_1 = L_2 \\ 1, & L_1 \neq L_2 \end{cases} \tag{8}$$

$$\text{dist}(A, L_1, L_2) = \frac{|L_1 - L_2|}{A_{\max} - A_{\min}} \tag{9}$$

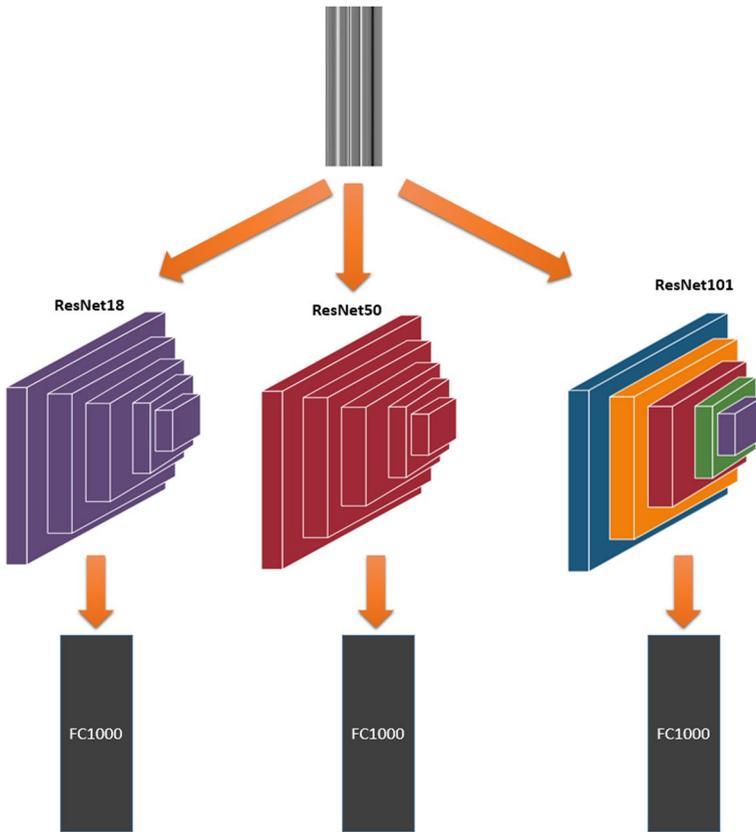


Fig. 6 Graphical explanation of the proposed ensemble ResNet

Equations 7–9 mathematically define weights generation process of ReliefF, where $W(f_i)$ is weights of i th feature, k is missing number of classes, R is selected data in cycle, $dist$ defines distance, H represents nearest class, n is number of cycles and P is probability.

After weights generation, the generated weights are sorted by descending and 1000 most weighted features are selected.

The steps of the used ReliefF-based feature selection are shown as below.

Step 1: Calculate weights of the concatenated features by using ReliefF and target values.

$$\text{weight} = \text{ReliefF}(\text{feature}, \text{target}) \tag{10}$$

Step 2: Sort the generated weights descending.

$$[\text{weight}_{\text{sorted}}, \text{indices}] = \text{sort}(\text{weight}) \tag{11}$$

Step 3: Select the 1000 most discriminative features by using Algorithm 3.

Algorithm 3. The used ReliefF based feature selection.

Input: Indices and feature with size of 3000.
Output: Selected features (<i>feat</i>) with size of 1000.
1: for $i=1$ to 1000 do
2: $feat(i) = feature(indices(i));$
3: end for i

4.4 Classification

Classification is the final phase of the proposed ensemble ResNet-based signal recognition method. In order to show strength of the proposed ensemble ResNet-based feature extraction method, a conventional classifier is used and this classifier is cubic SVM [38, 39]. SVM is one of the conventional classifiers and it is an optimization-based classification method. It uses various kernels for instance linear, quadratic, cubic and Gaussian. SVM can also be used in nonlinear classification tasks by implementing this kernel functions. Thus, each n —dimensional input vector $x_i = 1, 2, 3, \dots, M$, in which M represents the sample number, is mapped to a L dimensional property field $\Phi x = [\phi_1^x, \dots, \phi_L^x]$, where $K(x_i, x_j)$ is a kernel function. Cubic SVM uses the third-degree polynomial activation function. To obtain test results by using cubic SVM, stratified tenfold cross-validation is used. The attributes of the cubic SVM are given as follows. Box constraint level is 1, multiclass is selected as one-vs-all. The used cubic SVM calculates high-dimensional relationships, and equation of the used kernel is shown in Eq. 12.

$$(a \times b + r)^3 \quad (12)$$

where a and b refer to the two observations we want to calculate and r determines cross-validation.

5 Experimental results

In order to test the proposed method, we used a dataset which is explained in Sect. 2. By using this dataset, two cases are defined and these are explained in below.

Case 1 Daily sport activity recognition. In this case, 19 classes are defined.

Case 2 Gender recognition. In this case, two classes are defined.

To obtain numerical results from these cases, accuracy, *F1*-score and geometrical mean are used. The mathematical notation of these performance metrics is given as below.

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (13)$$

$$F1\text{-score} = \frac{2tp}{2tp + fp + fn} \quad (14)$$

$$\text{Geometric mean} = \sqrt{\frac{tp \cdot tn}{(tp + fn) \cdot (tn + fp)}} \quad (15)$$

where *tp*, *tn*, *fp* and *fn* are true positive, true negative, false positive and false negatives.

The proposed ensemble ResNet was repeated 1000 times to obtain comprehensive results. The obtained results for each case are given as below.

As shown from Table 1, the obtained maximum accuracy rates are 99.61% and 99.96% for Case 1 and Case 2, respectively. Confusion matrixes of the best result of the Case 1 and Case 2 are shown in Figs. 7 and 8.

As seen from Fig. 7, 100.0% accuracy rate was achieved for 10 (2nd, 3rd, 4th, 7th, 8th, 9th, 10th, 14th, 15th and 19th classes) of the 19 classes. The worst accuracy rate was 96.25% calculated for 18th class (Jumping).

In the Case 2, gender classification was performed. According to Fig. 8, male recognition was achieved 100.0% accuracy rate by using the proposed ensemble ResNet and 99.91% success rate was calculated for female recognition.

6 Discussion

In this study, sensors signals are classified using an ensemble CNN method (ResNet). As we know from the literature, ResNet is generally used for image recognition and classification. By using a basic vector-to-matrix transformation, the proposed ensemble ResNet is applied to sensors signals by using two cases. These cases are defined to recognize daily sport activities and genders. By using these cases, success of the proposed ensemble ResNet is proved. According to results, the proposed method achieved high results for all cases. As seen from Fig. 7, the proposed method achieved 100.0% accuracy rate in 10 classes (2nd, 3rd, 4th, 7th, 8th, 9th, 10th, 14th, 15th, 19th). The worst accuracy rate was calculated as 96.25% for 18th class. In the gender recognition, 100.0% accuracy rate is achieved for male recognition and it achieved approximately 99.96% accuracy. These results clearly demonstrate success of the proposed ensemble ResNet. To show effectiveness of the ensemble ResNet, the other deep ResNets (ResNet18, ResNet50, ResNet101) are used for comparisons and the comparatively results are shown in Table 2.

Output Class	1	4560 50.0 %	0 0.0 %
	2	4 0.0 %	4556 50.0 %
		1	2
		Target Class	

Fig. 8 Confusion matrix of the best result of the Case 2

proposed ensemble ternary ResNet achieved the best classification accuracy among them. The best of the couple ensemble ResNet is ResNet50-101 because it achieved 99.25% and 99.82% success rate for Case 1 and Case 2, respectively. Success rates of the proposed ensemble ResNet are 0.36% and 0.14% higher than ResNet50-101 for Case 1 and Case 2, respectively. Also, GoogLeNet, VGGNet16 and VGGNet19 were chosen as feature extractor. The proposed ensemble ResNet-based HAR method was also resulted higher than these networks.

The proposed ensemble ResNet uses ResNet18, ResNet50 and ResNet101 features together. In the case 1, the proposed ensemble ResNet uses 96, 496 and 408 features from ResNet18, ResNet50 and ResNet101, respectively. In the case 2, 177, 406 and 417 features are used from ResNet18, ResNet50 and ResNet101.

To clearly understand success of the proposed ensemble ResNet, the proposed method is compared to the previously presented state-of-art methods.

Kuncan et al. [43] proposed local binary pattern-based gender recognition methods using the sensors signals. The proposed Case 2 is compared to Kuncan et al.'s methods, and comparatively results are given in Table 4.

Table 4 clearly demonstrates that the proposed ensemble ResNet achieved 3.92%, 3.24% and 2.68% higher classification rate than 1D-LBP, 1D-RLBP and weighted 1D-LBP, respectively. In the case 2, only 4 of the 9120 observations are false predicted (see Fig. 4). In Tables 2, 3 and 4, the best results were shown by using bold font type.

Moreover, features of the proposed ensemble ResNet-based feature extractor are classified with deep neural network (DNN). DNN achieved 97.94% and 99.20% classification accuracies for Case 1 and Case 2, respectively.

We also applied the proposed ensemble ResNet method to MobiAct [44] dataset to generalize success of this deep ensemble feature extractor. The obtained results from MobiAct [44] dataset are listed in Table 5.

Table 2 Comparatively results of the ResNets, VGGNets, GoogLeNet and the proposed ensemble ResNet

Method	Case 1 (%)	Case 2 (%)
ResNet18 [19]	97.66	99.10
ResNet50 [19]	98.67	99.61
ResNet101 [19]	98.50	99.37
ResNet18-50	99.21	99.73
ResNet18-101	99.24	99.81
ResNet50-101	99.25	99.82
GoogLeNet [31]	96.94	98.76
VGGNet16 [33]	97.73	99.09
VGGNet19 [33]	97.01	98.76
Proposed ensemble ResNet	99.61	99.96

In this table, the calculated best results of these networks are presented by using Cubic SVM. These networks are used as feature extractor and cubic SVM is utilized as classifier

Table 3 Comparison results of the Case 1

Method	Accuracy (%)
Hu et al. [40]	98.51
Hammad and El-Sankaray [41]	99.26
Deeper LSTM [42]	93.27
Altun et al. [6]	99.16
Proposed ensemble ResNet	99.61

Table 4 Comparatively results of the Case 2

1D-LBP [43]	1D-RLBP [43]	Weighted 1D-LBP [43]	The proposed ensemble ResNet
96.04%	96.72%	97.28%	99.96%

Table 5 clearly shows that the proposed method is successful for HAR. We achieved 1.20% higher success rates than Ferrari et al.'s method [48] (the best of the others). Table 5 shows that ResNet is good solution for HAR because ResNet-based method achieved higher 90% classification accuracy. This test was implemented for activity recognition.

According to these results, the proposed ensemble ResNet-based method achieved best results.

The advantages of the proposed ensemble ResNet are given as below.

- A simple preprocessing algorithm is used in this work (see Algorithm 2) instead of spectrogram extraction. We used matrix-to-vector transformation as preprocessing. Time complexity of this method was calculated as $O(n)$ (n is length of the signal). Therefore, this method was selected for preprocessing.

- The proposed ensemble ResNet has high success rates (See Tables 1–5).
- The proposed ensemble ResNet uses ResNet18, ResNet50 and ResNet101 together, and it improves success rates all of them (See Table 2).
- In this work, a simple preprocessing method (vector-to-matrix transformation), pre-trained three networks, ReliefF feature selector and conventional classifier (SVM) are used together. The used methods are well known and basic methods. By using these, an effective learning method is proposed for signal classification (See Table 1).
- Two cases were defined in this paper. The proposed method achieved high success rates for these cases. This situation clearly demonstrates that the proposed method is a general signal recognition method (see Table 2).
- The presented ensemble ResNet-based HAR method achieved higher performance than other HAR methods (see Table 3).
- The extracted features are more suitable conventional classifier (SVM) than deep classifier (DNN). DNN achieved 97.94% and 99.20% classification accuracies for Case 1 and Case 2, respectively, while SVM achieved 99.61% and 99.96% success rates for Case 1 and Case 2.
- The proposed deep feature extractor was also applied to MobiAct [44] dataset and effectiveness of it shows (see Table 5). Table 5 clearly shows the general success of the proposed ensemble ResNet method for HAR.

The disadvantage of this method is deep networks and is not lightweight methods because millions parameters should be optimized. Therefore, computational complexity of these methods is high. To overcome this disadvantage, we used pre-trained networks and low-layered three effective ResNets.

7 Conclusion

In this article, a novel ensemble network is proposed by using pre-trained three networks. The used deep networks are ResNet18, ResNet50 and ResNet101. Therefore, the proposed method is called as ensemble ResNet. As we know from the literature and applications, ResNet-based networks are generally used for images. In this paper, a novel sensor signal recognition method is presented by using the proposed

Table 5 Mean accuracy rates of the proposed ensemble ResNet and other methods

Method	Accuracy (%)
Ferrari et al.'s method [45]	69.51
Chen et al.'s method (LSTM) [46]	92.00
Ajerla et al.'s method (LSTM) [47]	85.20
Ferrari et al.'s method (kNN) [48]	87.69
Ferrari et al.'s method (ResNet) [48]	92.98
Welhenge and Taparugssanagorn et al.'s method [49]	90.00
Proposed ensemble ResNet	94.18

ensemble ResNet. The proposed signal recognition method consists of preprocessing, feature extraction, feature selection and classification. In the preprocessing, a basic vector-to-matrix transformation is used. Then, pre-trained ResNet18, ResNet50 and ResNet101 are used for feature extraction. FC1000 layers of these networks are selected as output and 1000 features are extracted from each network. These features are concatenated and 3000 features are obtained. In the feature selection phase, ReliefF is used and 1000 most discriminative features are selected. The selected features are forwarded to cubic SVM, and results are obtained by using tenfold cross-validation. To test performance of this method, daily sport activities and gender recognition cases are defined. The proposed method achieved 99.61% and 99.96% accuracy rates for daily sport activity recognition and gender recognition, respectively. The proposed ensemble ResNet was also compared to ResNet-based networks and the state-of-art methods. Results clearly demonstrated that the proposed ensemble ResNet increased success rates of the used ResNets, and it achieved the best results among the selected state-of-art methods. The proposed ensemble ResNets based was also tested on MobiAct dataset, and it was shown that this method increased success rate of the ResNet for HAR.

In the future work, the proposed ensemble ResNet can be used for images. Novel mobile health monitoring applications can be proposed using the proposed method. In the literature, many deep learning methods have been proposed. By using these methods, novel ensemble networks can be proposed to solve real-world recognition problems for instance image classification, gender identification, audio classification and facial expressions recognition.

References

1. Munoz-Organero M (2019) Outlier detection in wearable sensor data for human activity recognition (HAR) based on DRNNs. *IEEE Access* 7:74422–74436
2. Chen Y, Shen C (2017) Performance analysis of smartphone-sensor behavior for human activity recognition. *IEEE Access* 5:3095–3110
3. Wang K, He J, Zhang L (2019) Attention-based convolutional neural network for weakly labeled human activities recognition with wearable sensors. *IEEE Sens J* 19:7598–7604
4. Cornacchia M, Ozcan K, Zheng Y, Velipasalar S (2016) A survey on activity detection and classification using wearable sensors. *IEEE Sens J* 17(2):386–403
5. Wang J, Chen Y, Hao S, Peng X, Hu L (2019) Deep learning for sensor-based activity recognition: a survey. *Pattern Recognit Lett* 119:3–11
6. Altun K, Barshan B, Tunçel O (2010) Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit* 43(10):3605–3620
7. Song B, Kamal AT, Soto C, Ding C, Farrell JA, Roy-Chowdhury AK (2010) Tracking and activity recognition through consensus in distributed camera networks. *IEEE Trans Image Process* 19(10):2564–2579
8. Huynh-The T et al (2018) Hierarchical topic modeling with pose-transition feature for action recognition using 3D skeleton data. *Inf Sci* 444:20–35
9. Mario M-O (2018) Human activity recognition based on single sensor square HV acceleration images and convolutional neural networks. *IEEE Sens J* 19(4):1487–1498
10. Tian Y, Wang X, Chen L, Liu Z (2019) Wearable sensor-based human activity recognition via two-layer diversity-enhanced multiclassifier recognition method. *Sensors* 19(9):2039

11. Jiang W, Yin Z (2013) Human activity recognition using wearable sensors by deep convolutional neural networks. In: Proceedings of the 23rd ACM International Conference on Multimedia, ACM, pp 1307–1310
12. Chernbumroong S, Cang S, Atkins A, Yu H (2013) Elderly activities recognition and classification for applications in assisted living. *Expert Syst Appl* 40(5):1662–1674
13. Nweke HF, Teh YW, Al-Garadi MA, Alo UR (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. *Expert Syst Appl* 105:233–261
14. Wang Z, Guo M, Zhao C (2016) Badminton stroke recognition based on body sensor networks. *IEEE Transa Hum Mach Syst* 46(5):769–775
15. Gravina R, Alinia P, Ghasemzadeh H, Fortino G (2017) Multi-sensor fusion in body sensor networks: state-of-the-art and research challenges. *Inf Fusion* 35:68–80
16. Cheng W-C, Jhan D-M (2012) Triaxial accelerometer-based fall detection method using a self-constructing cascade-AdaBoost-SVM classifier. *IEEE J Biomed Health Inf* 17(2):411–419
17. Margarito J, Helaoui R, Bianchi AM, Sartor F, Bonomi AG (2015) User-independent recognition of sports activities from a single wrist-worn accelerometer: a template-matching-based approach. *IEEE Trans Biomed Eng* 63(4):788–796
18. Wang A, Chen G, Yang J, Zhao S, Chang C-Y (2016) A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sens J* 16(11):4566–4578
19. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 770–778
20. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436
21. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge
22. Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning. In: International Conference on Machine Learning, pp 1139–1147
23. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18(7):1527–1554
24. Bengio Y, Goodfellow I, Courville A (2017) Deep learning. Citeseer, Princeton
25. Oh J, Guo X, Lee H, Lewis RL, Singh S (2015) Action-conditional video prediction using deep networks in atari games. In: Advances in Neural Information Processing Systems, pp 2863–2871
26. Abdel-Hamid O, Mohamed A-R, Jiang H, Deng L, Penn G, Yu D (2014) Convolutional neural networks for speech recognition. *IEEE/ACM Trans Audio Speech Lang Process* 22(10):1533–1545
27. Cao W, Wang X, Ming Z, Gao J (2018) A review on neural networks with random weights. *Neurocomputing* 275:278–287
28. Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: Proceedings of the 24th International Conference on Machine Learning, ACM, pp 791–798
29. Le Roux N, Bengio Y (2008) Representational power of restricted Boltzmann machines and deep belief networks. *Neural Comput* 20(6):1631–1649
30. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
31. Szegedy C et al (2015) Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 1–9
32. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2016) SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. arXiv preprint [arXiv:1602.07360](https://arxiv.org/abs/1602.07360)
33. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
34. Xsens N (2010) MTi and MTx user manual and technical documentation
35. Barshan B, Yükses MC (2014) Recognizing daily and sports activities in two open source machine learning environments using body-worn sensor units. *Comput J* 57(11):1649–1667
36. Robnik-Šikonja M, Kononenko I (2003) Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn* 53(1–2):23–69
37. Kononenko I, Šimec E, Robnik-Šikonja M (1997) Overcoming the myopia of inductive learning algorithms with RELIEFF. *Appl Intell* 7(1):39–55
38. Rowland R et al (2019) A simple burn wound severity assessment classifier based on spatial frequency domain imaging (SFDI) and machine learning. In: Photonics in Dermatology and Plastic Surgery 2019, 2019, vol. 10851: International Society for Optics and Photonics, p 1085109

39. Liapis A, Katsanos C, Sotiropoulos D, Xenos M, Karousos N (2015) Recognizing emotions in human computer interaction: studying stress using skin conductance. In: IFIP Conference on Human–Computer Interaction. Springer, pp 255–262
40. Hu C, Chen Y, Hu L, Peng X (2018) A novel random forests based class incremental learning method for activity recognition. *Pattern Recognit* 78:277–290
41. Hammad I, El-Sankary K (2019) Practical considerations for accuracy evaluation in sensor-based machine learning and deep learning. *Sensors* 19(16):3491
42. Wang L, Liu R (2019) Human activity recognition based on wearable sensor using hierarchical deep LSTM networks. *Circuits Syst Signal Process* 39:837–856
43. Kuncan F, Kaya Y, Kuncan M (2019) New approaches based on local binary patterns for gender identification from sensor signals. *J Faculty Eng Architect Gazi Univ* 34(4):2173–2185
44. Vavoulas G, Chatzaki C, Malliotakis T, Pediaditis M, Tsiknakis M (2016) The MobiAct dataset: recognition of activities of daily living using smartphones. In: *ICT4AgeingWell*, pp 143–151
45. Ferrari A, Mobilio M, Micucci D, Napoletano P (2019) On the homogenization of heterogeneous inertial-based databases for human activity recognition. In: *2019 IEEE World Congress on Services (SERVICES)*, vol. 2642. IEEE, pp 295–300
46. Chen Y, Zhong K, Zhang J, Sun Q, Zhao X (2016) LSTM networks for mobile human activity recognition. In: *2016 International Conference on Artificial Intelligence: Technologies and Applications*. Atlantis Press
47. Ajerla D, Mahfuz S, Zulkernine F (2019) A real-time patient monitoring framework for fall detection. *Wirel Commun Mob Comput*. <https://doi.org/10.1155/2019/9507938>
48. Ferrari A, Micucci D, Mobilio M, Napoletano P (2019) Hand-crafted features vs residual networks for human activities recognition using accelerometer. In: *2019 IEEE 23rd International Symposium on Consumer Technologies (ISCT)*. IEEE, pp 153–156
49. Welhenge AM, Taparugssanagorn A (2019) Human activity classification using long short-term memory network. *SIViP* 13(4):651–656

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.