Deep Transfer Learning Based Human Activity Recognition By Transforming IMU Data To Image Domain Using Novel Activity Image Creation Method

Mohammed hashim B.A (✉ mohammedhashim.cahcet@gmail.com)

  C.Abdul Hakeem College of Engineering and Technology   https://orcid.org/0000-0002-6548-6334

Amutha R

  Sri Sri Sivasubramaniya Nadar College of Engineering

Research Article

Deep Transfer Learning based Human Activity Recognition by Transforming IMU Data to Image Domain using Novel Activity Image Creation Method

Mohammed Hashim B A[1], Amutha R[2]

[1]Department of Electronics and Communication Engineering, C.Abdul Hakeem College of Engineering and Technology, Hakeen Nagar, Melvisharam, Ranipet, TamilNadu, India.
[2]Department of Electronics and Communication Engineering, Sri Sri Sivasubramaniya Nadar College of Engineering, Kalavakkam, Chennai, TamilNadu, India.

ABSTRACT

Human Activity Recognition is the most popular research area in the pervasive computing field in recent years. Sensor data plays a vital role in identifying several human actions. Convolutional Neural Networks (CNNs) have now become the most recent technique in the computer vision phenomenon, but still it is premature to use CNN for sensor data, particularly in ubiquitous and wearable computing. In this paper, we have proposed the idea of transforming the raw accelerometer and gyroscope sensor data to the visual domain by using our novel activity image creation method (NAICM). Pre-trained CNN (AlexNet) has been used on the converted image domain information. The proposed method is evaluated on several online available human activity recognition dataset. The results show that the proposed novel activity image creation method (NAICM) has successfully created the activity images with a classification accuracy of 98.36% using pre trained CNN.

Key words: Human Activity Recognition; CNN; Pervasive computing; NAICM; Transfer Learning.

1. INTRODUCTION

Human activity recognition is a popular research area in ubiquitous computing, human-computer interaction and human behavior analysis. Activity recognition is very helpful in areas like Ambient Assisted Living, Elderly care etc (Mohammed & Sangavi, 2019). Recent research on wearable and ubiquitous sensors shows that these sensors can be used to sense data for human actions such as construction workers activity recognition (Akhavian &

Behzadan, 2016), fall detection (Alveraz de la Concepción et al., 2017; Jansi & Amutha, 2020) etc. Enormous amount of data have been obtained by the presence of sensors in pervasive computing environments. Sensors like accelerometer and gyroscope play a major role for human action recognition (Mohammed Hashim & Amutha, 2020). Machine learning and data mining methods have been used to extract vital information from raw sensor data (Chetty et al., 2015) and this vital information is nothing but features and it is usually numerical statistical data. These features are always reliable in classification, segmentation and recognition.

Deep learning is extremely successful in many domains. Convolutional neural networks (CNNs) have now been used for practical tasks and it has gained popularity after achieving exceptional accuracy on tasks related to image classification (He et al., 2016; Szegedy et al., 2017; Zhang et al., 2017). Deep CNN have been used to recognize human activities also (Long et al., 2015). One of the limitations of these techniques is the requirement of big labeled data to carry out the training of the deep neural network. The computer vision community have created large labeled datasets like ImageNet (Russakovsky et al., 2015) for object recognition and classification. Only few labeled datasets are available since the scope is limited compared to general images. The limitation mentioned above can be overcome by using transfer learning. Transfer learning uses a pre-trained CNN, the activations of the last hidden layer are taken as input datasets feature descriptors by eliminating the last fully connected layer. The emerged feature descriptors are taken for training the classification model. The Pre-trained convolutional neural networks like AlexNet (Krizhevsky et al., 2012) and VGGNet (Simonyan & Zisserman, 2014) are transferred by fine-tuning the semantic segmentation task. Transfer learning removes the urge to create a big dataset required for training the CNNs. Additional costs like time and computational resources can be saved by using transfer learning.

Conventionally, the transfer learning is used on visually interpretable domain. Images such as flowers, vegetables, animals, biological medical images like X-ray, MRI etc., will come under the category of images which are easily visually interpretable. The application of deep transfer learning on these images is mainly possible because of two reasons. First, it is feasible because of being visually interpretable and secondly, there is a very huge number of datasets available to carry out the tasks. However, the sensor data is not visually interpretable and it is unclear whether it can be visually interpretable. These data are called not visually interpretable data (Singh et al., 2017). Only a few have worked on converting the IMU sensor data to images for activity recognition and the scope is wide open in this area.

In this paper, we introduce the concept of using transfer learning on a not visually interpretable domain.

We have proposed a conversion technique called novel activity image creation method (NAICM), which involves the conversion from a non suitable domain for transfer learning to a suitable domain on which the CNNs can been trained. The raw sensor data is converted into a group of visual images using novel activity image creation method (NAICM) and then it is used as an input dataset for a pre-trained convolutional network model (AlexNet).

2. ACTIVITY DATASETS

We have used 3 activity recognition datasets namely SKODA, UCI, and USC-HAD to identify the adaptability of our proposed algorithm. For comparison with other works, we have chosen the USC-HAD dataset since most of the works have used USC-HAD. We have discussed the USC-HAD dataset elaborately compared to the other two datasets because of the aforementioned reason.

SKODA dataset

The Skoda Mini Checkpoint dataset contains the activity of assembly line laborers in a vehicle maintenance environment. The dataset is generated by making a laborer wear 20 accelerometers on both arms. The dataset contains 10 different activities and some of the activities are "write on notepad", "close hood", "check trunk gaps", etc. We limit our experiments to only 10 sensors placed on the right arm.

UCI HAR dataset

The UCI HAR dataset (Mohammed Hashim & Amutha, 2020) consists of labeled data from thirty human subjects. Totally 6 different activities are performed. Activities like "Standing", "Lying down", "Walking upstairs", etc are a few of the total 6 activities. The accelerometer and gyroscope sensor data is collected using the smartphone. The dataset consists of extracted features from the sensor data.

USC HAD dataset

The information about human activity dataset USC-HAD (Zhang & Sawchuk, 2012) have been discussed here. In order to construct an efficient recognition system, it has to be trained on a big diverse set of individuals. In point of human activities, the diversity of the enrolled subjects in the dataset is based on the following 4 factors: (1) Gender (2) Age (3) Weight and (4) Height. Based on these guidelines, 14 human subjects (7 female, 7 male) have been chosen to participate in the process of data collection. In Table 1, the statistics of height, age and weight are listed. A wider range of population has been covered by the diversity in each of these 4 factors.

Table 1 USC HAD dataset - Statistics of human subjects participated.

Parameters	Age	Height in (Centimeter)	Weight in (Kilogram)
Range (Subjects)	21 to 49	160 to 185	43 to 80
Mean	30.1	170	64.6
Standard Deviation	7.2	6.8	12.1

There are totally 12 activities performed which are the most common human activities in daily lives of people. All the activities and its description are shown in Table 2.

Table 2 USC HAD dataset - Activities and its descriptions.

S.No	Activity	Description
11	Walk Forward (WF)	The subject walks in forward direction in a straight line
2	Walk Left (WL)	In a full circle, the subject walks in counter-clockwise direction
3	Walk Right (WR)	In a full circle, the subject walks in clockwise direction
4	Walk Upstairs (WU)	The subject walks up multiple flights in stairs
5	Walk Downstairs (WD)	The subject walks down multiple flights in stairs
6	Run Forward (RF)	In a straight line, the subject runs forward
7	Jump (J)	The subject continuously jumps up and down staying at the same position
8	Sit (St)	The subject sits on chair either resting or working. Fidgeting is also considered to belong to this class
9	Stand (Sd)	The subject stands still and talks to someone
10	Sleep (Sl)	The subject lies or sleeps down on a bed
11	In Elevator (IE)	The subject moves in an elevator

3. PRE-TRAINED CNN

AlexNet shown in Figure1 is much larger than the previous CNNs utilized for computer vision. It consists of 6 crore parameters and 6.5 lakh neurons and it took more than 5 days for training on 2 GTX 580 3GB GPUs. AlexNet is consists of cascaded stages. They are convolution, pooling, ReLU and fully connected layers. In particular, AlexNet consists of five convolutional

layers, the first to fourth layer and it is followed by the pooling layer and 3 fully connected layers after the fifth layer. For the AlexNet architecture, the convolutional kernels are extracted while the back propagation optimization process by optimizing the whole cost function with the stochastic gradient descent algorithm. Usually, the convolutional layer along with sliding convolutional kernels acts on input feature maps to produce convolved feature maps. The pooling layer groups the information within the neighborhood window with a max or average pooling layer by acting on the convolved feature maps.
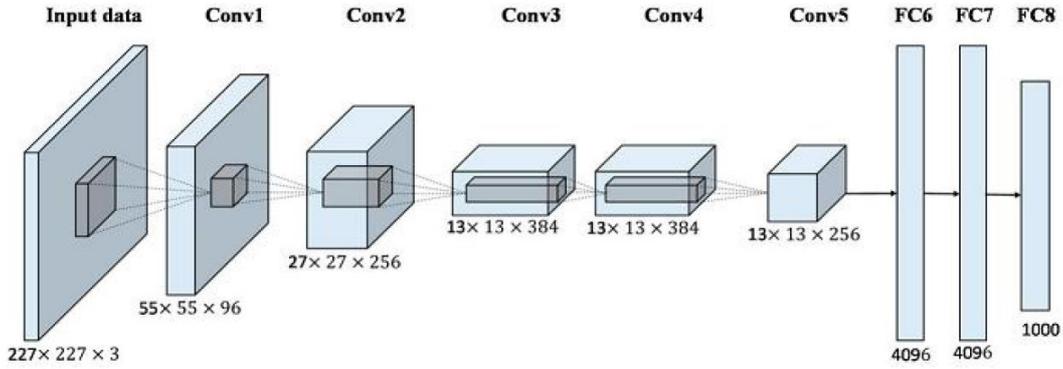
Figure 1 AlexNet (Russakovsky et al., 2015)

4. PROPOSED METHOD

An accelerometer and a gyroscope are used in the wearable device as inertial measurement unit which measures the strength of angular velocity and tri-axis acceleration, respectively. In our work, we have used both accelerometer and gyroscope sensor data from the datasets. In previous works, features which are used for activity recognition have been mostly independently extracted from several time series sensor signals (Anguita et al., 2013). The correlations among different signals are mostly overlooked. We are proposing to transfer all the time series signals from gyroscope and accelerometer into a new activity image using our proposed novel activity image creation method (NAICM) (Table 3).

We take the inputs, Empty Image $E(i,j)$, Classes (C_n), Assumed Value (A_n), Accelerometer & Gyroscope Data $(D(k,m))$ and Random Value (R), where $i = 1:100$, $j = 1:100$, n= Number of Classes, $A = \{20, \ldots, A_{n-1} + 20\}$, $k = 1:100$, $m = 1:6$. First of all, we find out the minimum value from (L) from given $D(k,m)$ using eqn (1)

$$L = min(D(k,m)) \qquad (1)$$

The next step is to calculate the intensity values for Empty Image $E(i,j)$ using eqn (2) & (3).

$$E(k,m) = D(k,m) + L + A_n \qquad (2)$$

$$E(i, j = m + 1, .., 100) = R \qquad (3)$$

where R varies from A_n to $A_n + 5$

This step converts the negative values to positive values of sensor data. Further allocating some random values in the allotted range for each class. Discrete Fourier transform has been used to

calculate the frequency components of E using equation (4)

$$F = \frac{1}{\sqrt{(PQ)}} \sum_{p=1}^{100} \sum_{q=1}^{100} E(p,q)e^{-2j\pi(tp/P+lq/Q)} \qquad (4)$$

$$t = 1,\dots,100 \quad l = 1,\dots,100$$

The image F with frequency components has been resized to (227×227×3) to make it suitable for AlexNet.

Table 3 Proposed Novel Activity Image Creation Method.

Novel Activity Image Creation Method (NAICM)
Input: Empty Image $E(i,j)$, Classes (C_n), Assumed Value (A_n), Accelerometer & Gyroscope Data $(D(k,m))$ Random Value (R). where $i = 1:100$, $j = 1:100$, n= Number of Classes, $A = \{20,\dots,A_{n-1}+20\}$, $k = 1:100$, $m = 1:6$
Output: Output Image $O(227 \times 227 \times 3)$
Step1: Find the minimum value (L) from given $D(k,m)$ $L = min(D(k,m))$
Step2: Calculate intensity values for E $E(k,m) = D(k,m) + L + A_n$ $E(i,j = m+1,..,100)= R,$ where R varies from A_n to A_n+5;
Step3: Calculate Discrete Fourier transform for E $$F = \frac{1}{\sqrt{(PQ)}} \sum_{p=1}^{100} \sum_{q=1}^{100} E(p,q)e^{-2j\pi(tp/P+lq/Q)}$$ $t = 1,\dots,100; \quad l = 1,\dots,100;$
Step4: Resize F to $(227 \times 227 \times 3)$ $O = Resize(F(i,j));$
Step5: Repeat step 1 to step 4 for all the classes.

Few activity images have been shown in Figure 2 corresponding to different activities. Deep CNN can extract enough features to differentiate different activities from these images.
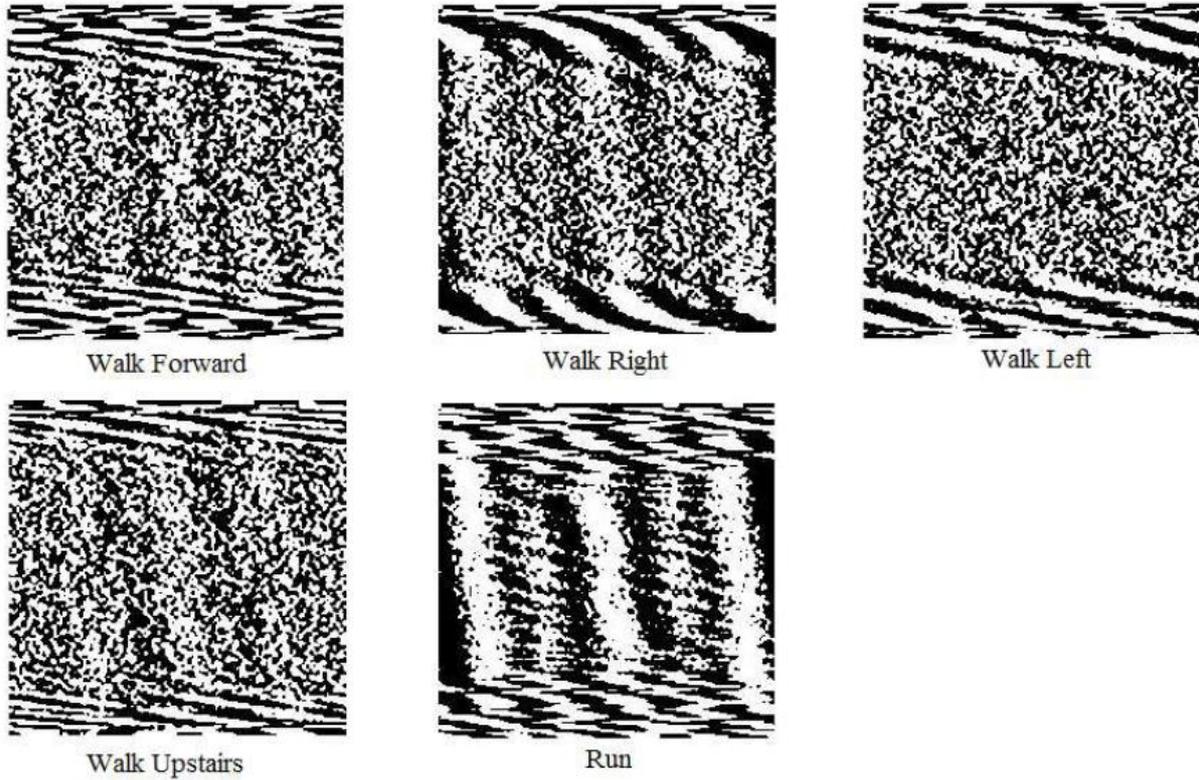
Figure 2 Sample Activity Images

5. RESULTS

We have used a 64-bit operating system, 32GB RAM, Intel Core i5 processor, GPU, MATLAB 2019b for our simulation work. The accuracy obtained for all the datasets using our proposed method have been mentioned in Table 4. The results show that the position of the sensor plays a vital role in getting better accuracy. From the results, it can be observed that the sensor placed in hands gives better accuracy compared to sensors placed in other positions like hip and waist. We have used three different train and test ratios like 7:3, 8:2, and 9:1.

Table 4 Accuracy obtained from all three datasets using NAICM.

Datasets	Position of sensor	Accuracy (%) - Train:Test		
		7:3	8:2	9:1
SKODA	Hands	97.67	98.02	98.57
UCI	Hip	95.75	96.05	96.20
USC- HAD	Waist	97.23	97.89	98.36

Comparison of our proposed NAICM with a similar image conversion algorithm

Table 5 shows the accuracy comparison between our proposed NAICM and a similar image conversion algorithm (Jiang, W., & Yin, Z. (2015)). The results show that our proposed method has outperformed the existing method by nearly 0.57% with UCI dataset and 0.22% with USC HAD dataset.

Table 5 Comparison of accuracy between NAICM with a similar image conversion algorithm (Jiang, W., & Yin, Z. (2015)).

Datasets	Image conversion Algorithm + DCNN (Jiang, W., & Yin, Z. (2015))	Proposed NAICM + Pretrained CNN model
UCI	95.18%	95.75%
USC-HAD	97.01%	97.23%

Comparison of results using USC HAD dataset with other works

In this section, we have discussed the results obtained for our proposed method using USC HAD dataset and a comparison with other works using the same USC HAD dataset. Several quantitative metrics have been used for evaluation. They are 1. Sensitivity 2. Precision 3. F-Score 4. Accuracy 5. Specificity

Table 6 USC HAD dataset - Confusion Matrix for classification of different activities.

Activities	WF	WL	WR	WU	WD	J	St	Sd	Sl	IE
WF	903	0	0	0	7	9	4	5	0	0
WL	4	634	3	1	0	1	0	0	0	0
WR	3	1	677	4	0	0	0	0	0	0
WU	1	3	4	568	5	2	0	4	0	0
WD	0	9	1	2	524	1	0	0	0	0
RF	0	0	0	1	0	395	0	0	0	0
J	0	0	0	0	0	0	0	0	0	0
St	0	0	0	0	0	0	620	10	0	0
Sd	1	4	0	1	3	2	8	580	0	2
Sl	0	0	0	0	0	0	1	0	900	0
IE	0	0	0	0	0	0	0	0	0	411

The following parameters can be calculated from the confusion matrix. True Positives (TP): The total number of positive instances that were correctly predicted as positive. True Negatives (TN): The total number of negative instances that were correctly predicted as negative. False Positives (FP): The total number of negative instances that were correctly predicted as positive. False Negatives (FN): The total number of positive instances that were correctly predicted as negative. The most common standard metrics used to analyze classification performances are sensitivity, precision, F-score, accuracy and specificity (Jansi, R., & Amutha, R. (2018)).

These are obtained using the following equations.

Sensitivity S measures the proportion of correctly classified positive instances.

$$S = \frac{TP}{TP+FN} \qquad (5)$$

Specificity SP measures the proportion of correctly classified negative instances.

$$SP = \frac{TN}{TN+FP} \qquad (6)$$

Precision P measures the proportion of classified positive instances to the total number of positive instances.

$$P = \frac{TP}{TP+FP} \qquad (7)$$

Table 7 USC HAD dataset - Comparison of specificity, precision and sensitivity for proposed method and DRNN (Simonyan & Zisserman, 2014).

Activities	Specificity (%)		Precision (%)		Sensitivity (%)	
	Proposed	**DRNN (Murad & Pyun, 2017)**	**Proposed**	**DRNN (Murad & & Pyun, 2017)**	**Proposed**	**DRNN (Murad & & Pyun, 2017)**
WF	99.84	99.72	99.01	98.38	97.31	98.07
WL	99.71	99.8	97.39	98.24	98.6	97.7
WR	99.86	99.75	98.83	97.86	98.83	98.29
WU	99.85	99.89	98.44	98.78	96.6	98.02
WD	99.75	99.77	97.22	97.33	97.58	97.45
RF	99.75	99.99	96.34	99.86	98.26	98.21
J	99.82	99.84	95.85	96.04	100	96.26
St	99.76	99.85	97.79	98.53	98.41	98.44
Sd	99.76	99.24	97.64	92.61	96.51	95.05
Sl	100	100	100	100	99.89	99.94
IE	99.97	99.64	99.52	94.47	100	94.33

The F-score F is determined by finding the harmonic mean of recall and positive predictive value (precision). F-score is computed using the following equation.
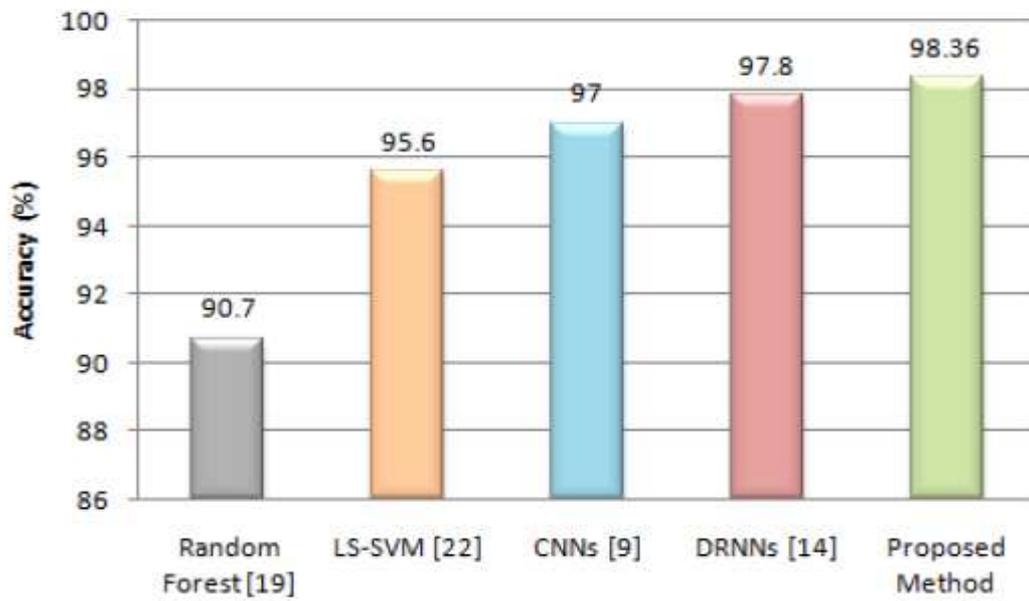
$$F = \frac{2 \times TP}{2 \times TP+FP+FN} \qquad (8)$$
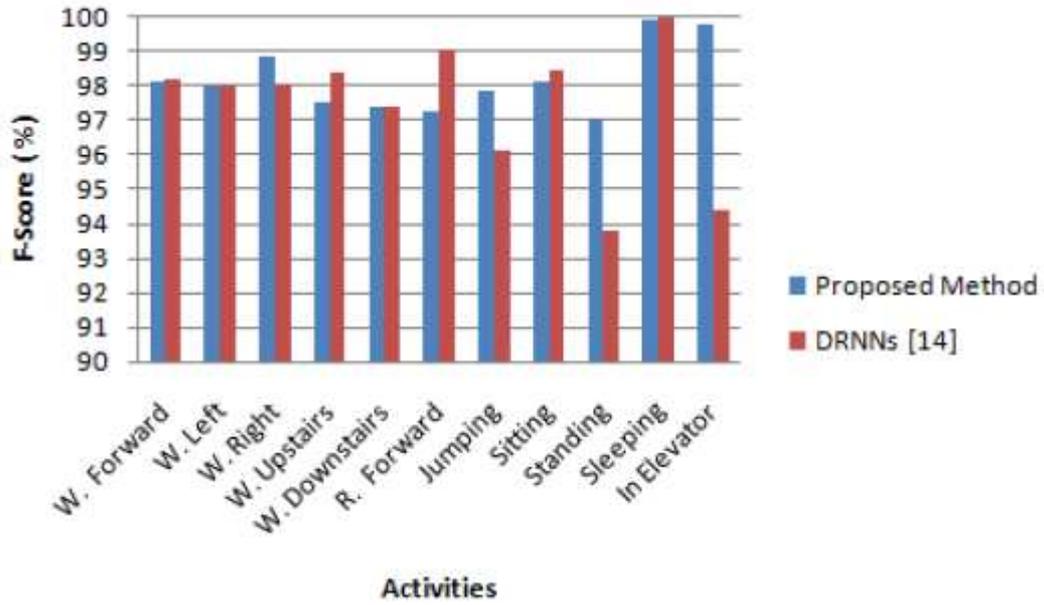
Figure 3 USC HAD dataset - Relative F-scores using DRNNs (Murad & Pyun, 2017) and Proposed method

The accuracy A is the mostly used performance evaluation measure to find out the overall classification performance of all classes.

$$A = \frac{TP+TN}{TN+FP+FN+TN} \qquad (9)$$

Figure 4 USC HAD dataset - Comparative accuracy of proposed model against other methods

Table 6 shows the obtained confusion matrix for classification. From the confusion matrix, we can calculate all the performance metrics. Table 7 compares the values of sensitivity, specificity, precision of proposed method with DRNNs ((Murad & Pyun, 2017)). The comparison shows the performance metrics of proposed method is better than the DRNNs (Murad & Pyun, 2017). The average specificity, precision and sensitivity of the proposed

method is 99.82%, 98% and 98.36% which is 0.05%, 0.54% and 0.93% greater than the existing method ((Murad & Pyun, 2017)). The F-score values (%) calculated for each activity is shown in Figure 3. The bar graph shows that the proposed method outperforms the existing DRNNs ((Murad & Pyun, 2017)) in most of the activities.

The average F-score of the proposed method is 98.17% which is 0.73% greater than the existing DRNNs ((Murad & Pyun, 2017)). Figure 4 depicts the comparison of accuracy of different techniques with proposed method. The accuracy of proposed method is 98.36% and the accuracy of Random forest [19], LS-SVM (Zheng, Y. (2015)), CNNs (Jiang, W., & Yin, Z. (2015)) and DRNNs (Murad & Pyun, 2017) is 90.7%, 95.6%, 97% and 97.8%. It clearly shows the proposed method outperformed all other techniques with considerable margin in terms of accuracy.

6. CONCLUSION

In this work, we have proposed a novel activity image creation method to convert the sensor data into an image. We have used three activity recognition datasets available online. Raw sensor signals from accelerometer and gyroscope have been taken and the novel activity image creation method is used to get a newly-defined activity image suitable for pre-trained CNN. We compared our proposed method with the existing works. The proposed method achieves excellent performance in terms of recognition precision, specificity, sensitivity, and F-score compared to the existing method ((Murad & Pyun, 2017)). The proposed method shows a good performance in terms of accuracy compared to existing methods by a margin of 7.66%*(Vaka, P., Shen, F., Chandrashekar, M., & Lee, Y. (2015)), 2.76%(Zheng, Y. (2015)), 1.36%(Jiang, W., & Yin, Z. (2015)), and 0.56%((Murad & Pyun, 2017)) respectively. Furthermore, the proposed method performed better than a similar image conversion algorithm (Jiang, W., & Yin, Z. (2015)) by a narrow margin. In the future, we can use our method to other IMU sensor data and study the performance.

Conflict of Interest

Authors declare no conflict of interests

REFERENCES

[1] **Akhavian, R., & Behzadan, AH. (2016).** Smartphone-based construction workers activity recognition and classification. Autom Constr 71:198–209

[2] **Alveraz de la Concepción, M. Á., Morillo, L. M. S., García, J. A. Á., & González-Abril, L. (2017).** Mobile activity recognition and fall detection system for elderly people using Ameva algorithm. Pervasive and Mobile Computing, 34, 3-13.

[3] **Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J. L. (2013).** A public domain dataset for human activity recognition using smartphones. In Esann (Vol. 3, p. 3).

[4] **Chetty, G., White, M., & Akther, F. (2015).** Smart phone based data mining for human activity recognition. Procedia Computer Science, 46, 1181-1187.

[5] **Han, X., Zhong, Y., Cao, L., & Zhang, L. (2017).** Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. Remote Sensing, 9(8), 848.

[6] **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).

[7] **Jansi, R., & Amutha, R. (2018).** A novel chaotic map based compressive classification scheme for human activity recognition using a tri-axial accelerometer. Multimedia Tools and Applications, 77(23), 31261-31280.

[8] **Jansi, R., & Amutha, R. (2020).** Detection of fall for the elderly in an indoor environment using a tri-axial accelerometer and Kinect depth data. Multidimensional Systems and Signal Processing, 31(4), 1207-1225.

[9] **Jiang, W., & Yin, Z. (2015).** Human activity recognition using wearable sensors by deep convolutional neural networks. In Proceedings of the 23rd ACM international conference on Multimedia (pp. 1307-1310).

[10] **Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).** Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25, 1097-1105.

[11] **Long, J., Shelhamer, E., & Darrell, T. (2015).** Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[12] **Mohammed, B. A., & Sangavi, S. (2019).** Human Activity Recognition for Ambient Assisted Living. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-4). IEEE.

[13] **Mohammed Hashim, B.A., & Amutha, R. (2020).** Human activity recognition based on smartphone using fast feature dimensionality reduction technique. Journal of Ambient Intelligence and Humanized Computing, 1-10.

[14] **Murad, A., & Pyun, J. Y. (2017).** Deep recurrent neural networks for human activity recognition. Sensors, 17(11), 2556.

[15] **Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015).** Imagenet large scale visual recognition challenge. International journal of computer vision, 115(3), 211-252.

[16] **Simonyan, K., & Zisserman, A. (2014).** Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[17]   **Singh, M. S., Pondenkandath, V., Zhou, B., Lukowicz, P., & Liwickit, M. (2017).** Transforming sensor data to the image domain for deep learning—An application to footstep detection. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 2665-2672). IEEE.

[18]   **Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017).** Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

[19]   **Vaka, P., Shen, F., Chandrashekar, M., & Lee, Y. (2015).** PEMAR: A pervasive middleware for activity recognition with smart phones. In 2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops) (pp. 409-414). IEEE.

[20]   **Zhang, M., & Sawchuk, A. A. (2012).** USC-HAD: a daily activity dataset for ubiquitous activity recognition using wearable sensors. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (pp. 1036-1043).

[21]   **Zhang, X., Li, Z., Change Loy, C., & Lin, D. (2017).** Polynet: A pursuit of structural diversity in very deep networks. In Proceedings of the IEEE Conference on Computer   Vision and Pattern Recognition (pp. 718-726).

[22]   **Zheng, Y. (2015).** Human activity recognition based on the hierarchical feature selection and classification framework. Journal of Electrical and Computer Engineering, 2015.