Human activity recognition for static and dynamic activity using convolutional neural network

Agus Eko Minarno, Wahyu Andhyka Kusuma, Yoga Anggi Kurniawan
Universitas Muhammadiyah Malang, Malang, Indonesia

Article Info	ABSTRACT
	Evaluated activity as a detail of the human physical movement has become a leading subject for researchers. Activity recognition application is utilized in several areas, such as living, health, game, medical, rehabilitation, and other smart home system applications. An accelerometer was popular sensors to recognize the activity, as well as a gyroscope, which can be embedded in a smartphone. Signal was generated from the accelerometer as a time-series data is an actual approach like a human actifvity pattern. Motion data have acquired in 30 volunteers. Dynamic actives (walking, walking upstairs, walking downstairs) as DA and static actives (laying, standing, sitting) as SA were collected from volunteers. SA and DA it's a challenging problem with the different signal patterns, SA signals coincide between activities but with a clear threshold, otherwise the DA signal is clearly distributed but with an adjacent upper threshold. The proposed network structure achieves a significant performance with the best overall accuracy of 97%. The result indicated the ability of the model for human activity recognition purposes. *This is an open access article under the [CC BY-SA](#) license.*

Corresponding Author:

Wahyu Andhyka Kusuma
Department of Informatics
Universitas Muhammadiyah Malang
Raya Tlogomas St. No. 246, Third Campus Universitas Muhammadiyah Malang, East Java, Malang, 65144, Indonesia
Email: kusuma.wahyu.a@gmail.com

1. INTRODUCTION

Human activity recognition (HAR) is the use of knowledge and image models for modeling activity and sensor data [1]. Human activity recognition has the availability of complete sensors so that they can recognize human activities such as sitting, walking, sleeping, running, and standing. HAR can also be used as a tool to diagnose a disease [2], activity recognition [3], [4], and be used in the military field [5]. A pioneer in HAR research using an accelerometer was published in the 90s [6]. However, the most widely cited research was able to produce satisfying results with many sensors simultaneously and using various algorithms carried out by Bao and Intile [7]. Classification of the introduction of human activities using sensors that vary from the device is a classic problem. It is, therefore, important to find a method for the proper recognition of human activity from device sensors [8].

HAR using smartphone sensors is a classic multi-variate time series classification problem, which utilizes 1D sensor signals and extracts features to be able to recognize activities by utilizing classification. Very little research on HAR uses in-depth learning techniques and automatic feature extraction mechanisms. The latest breakthrough in image and sound recognition has resulted in a new field of research that attracts enthusiastic researchers called deep learnin [9]. The convolutional neural network (CNN) neural network, in

particular, is a suitable algorithm for image and sound recognition. But not only images and sounds can be processed by CNN but the HAR dataset is also a good implementation when processing it using time-series data types.

Previous studies used HAR from various types of sensors except for cameras and accelerometers, gyroscope sensors using electromyograph, infrared audio, and other sensors [10]. An accelerometer has several advantages, low grass, cheaper. With the small dimensions embedded in a smartphone, the accelerometer can easily measure human movements. It can be used in a variety of different positions such as arms, waist, head, shoulders, pockets [11].

Fuentes and colleagues in a study entitled "online motion recognition using an accelerometer in a mobile device" uses neural networks in recognition of human body motion [12], while khan uses the decision tree method in recognizing human body movements from Wii remote data [13]. Other studies using a combination of CNN and machine learning methods appear in Table 1. One of the researchers using the University of California Irvine (UCI) HAR dataset which has 128 features is Ronao in 2016 in Table 2 in his study entitled "human activity recognition with smartphone sensors using deep learning neural networks" produces an accuracy value of 93.75% using the CNN and multilayer perceptron (MPL) algorithm [14].

In this research work, we proposed convolution neural network approaches for human body motion recognition with static activity and dynamic datasets. The main contribution of this paper is a model of CNN achieved good accuracy compared to previous research. The model provided variable parameters that match and have a high accuracy value for the dataset dynamic actives (DA) and static actives (SA). Combination of DA and SA models to classify HAR with 6 classes. The remaining section of this paper is organized as follows: materials and methods, data acquisitions which the information about proposed methods are given in section 2. The obtained results and discuss how the proposed method solved the problem are given in section 3. The conclusion about problem results is given in section 4.

Table 1. Previous CNN research

No	Method	Accurate	Activity	Number of Features
1	Decision Tree (DT) [4]	93.44%	Laying, Standing, Sitting, Walking, Walking Upstairs, Walking Downstairs	561
	Random Forest (RF)	96.73%		
	Extra Tree (XT)	96.68%		
	K-Nearest Neighbor (KNN)	96.21%		
	Logistic Regresion (LR)	98.40%		
	Support Vector Classifier (SVC)	93.86%		
	Ensemble Vote Classifier	97.60%		
2	Support Vector Machine (SVM) [12]	93%	Stopping, Walking, Standing-up, Sitting-down	100
3	CNN+MPL [14]	93.75%	Walking, Upsatrair, Downstair, Sitting, Standing, Laying	128
	Human Crafted Features (HCF)+Artificial Neural Network (ANN) HCF+SVM	82.27% 77.66%		
4	Deep convolutional neural network (DCNN) [15]	94%	walking, jogging, jumping, and go upstairs and go downstairs, sitting, standing, laying to the left and right side, and laying supine and prone	248
	Fast and robust deep convolutional neural network structure (FRDCNN)	95%		
5	CNN+LSTM [16]	99%	Arm wave, Hammer, Forward punch, High trow, Handclap, Bend, Tennis serve, Pickup and Trow	200
6	CNN [17]	99%	Fitness, Walking	350
	Convolution Auto Encoder (CAE)	94%		
	Empirical Mode Decomposition (EMD)	100%		
7	CNN [18]	90.42%	Jogging, Walking, Upsatrair, Downstair, Sitting, Standing, Laying	200
8	CNN+KNN [19]	71%	15 minutes of walking	1.000
	CNN+SVM	94%		
9	CNN 2D [20]	79.73%	Standing still, sitting and relaxing, laying down, walking, climbing stairs, waist bends forward, the frontal elevation of arms, knees bending, cycling, jogging, running, jumping front and back	10.000
	CNN-pf (partial full)	66.65%		
	CNN-pff (partial full weight)	99.66%		
10	CNN [21]	88%	clockwise-draw circle, counterclockwise, upper-right-draw, upper-left-draw, straight-draw from left to right, straight-draw from right to left.	20,000

Table 2. Distribution of activity data

Category	Activity	Train	Test
Dynamic	Walking	1226	496
	Walking Upstair	1073	471
	Walking Downstair	987	420
Static	Standing	1423	556
	Laying	1413	545
	Sitting	1293	508

2. METHOD

2.1. Dataset

UCI dataset is a popular dataset used in machine learning and deep learning [22]. This dataset was obtained from 30 volunteers who carried out various activities by wearing their waistline while doing six activities (standing, sitting, laying, walking, walking downstair, walking upstairs) Figure 1. The gadgets used in data collection can record the activity using the help of the accelerometer and gyroscope sensors that have been installed in the gadget, on the side of the accelerometer and gyroscope data collection process recorded using the video to label it manually. By using the gyroscope and accelerometer sensors in the gadget, they obtain three-axis linear acceleration (XYZ) data from the accelerometer sensor, and the gyroscope sensor generates three-axis angular velocity (XYZ). The sensor signal is then processed using noise filters and then in the sample in fixed containers (sliding windows) at intervals of 2.56 seconds with an overlap of 50%. The processed dataset is divided into 70% as training data, and 30% of test data is shown in Table 2.

Signal data from dynamic and static activities has a very significant difference, as seen in Figure 2 with 6 static and dynamic activities. Figure 3 (a) shows that there is a problem that occurs in the HAR that is the similarity of static signal data with standing and sitting activities. The similarities of standing and sitting activity data result in deep learning errors in classifying activities, and this can lower the level of accuracy in the overall HAR processing. In this article, we use t-SNE, which can display a high-dimensional data spread by reducing its dimensionality to two dimensions. We use configuration 1000 iterations and perplexities 2, 5, 10, 20, and 50 Figure 3 (b). Figure 3 (b) result t-SNE with perplexity 2 and 5 using 1000 iteration shows all group activity with the same type of data, but there are standing and sitting activities that have the same type of data.
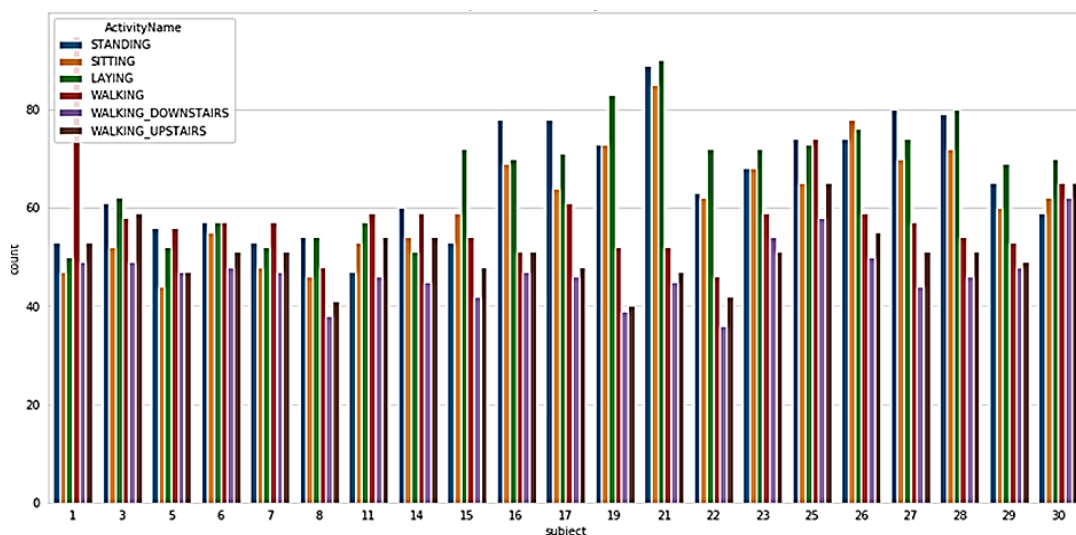

Figure 1. Volunteer activity data for standing, sitting, laying, walking, walking downstairs, walking

2.2. Hyperparameter

Hyperparameter is a method in the neural network that allows users to obtain a combination of parameters that have the best accuracy value from a number of previous neural network computing steps [23]. The combination of parameters obtained by using hyperparameter includes the number of layers used, the mapping feature, size convolution filter, size pooling dataset [24]–[30]. The parameters used on the proposed CNN model before tuning were seen in Table 3.
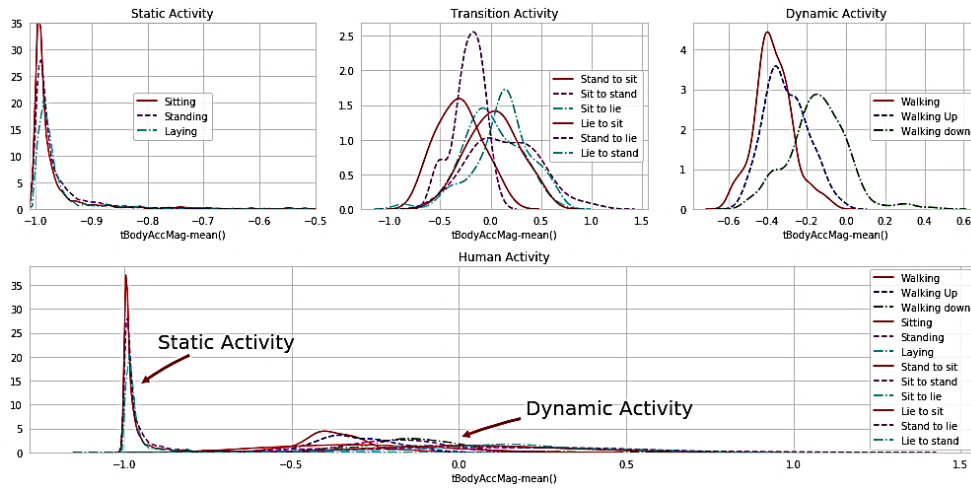
Figure 2. Static and dynamic activity: standing, sitting, laying, walking, walking downstairs, and walking upstairs
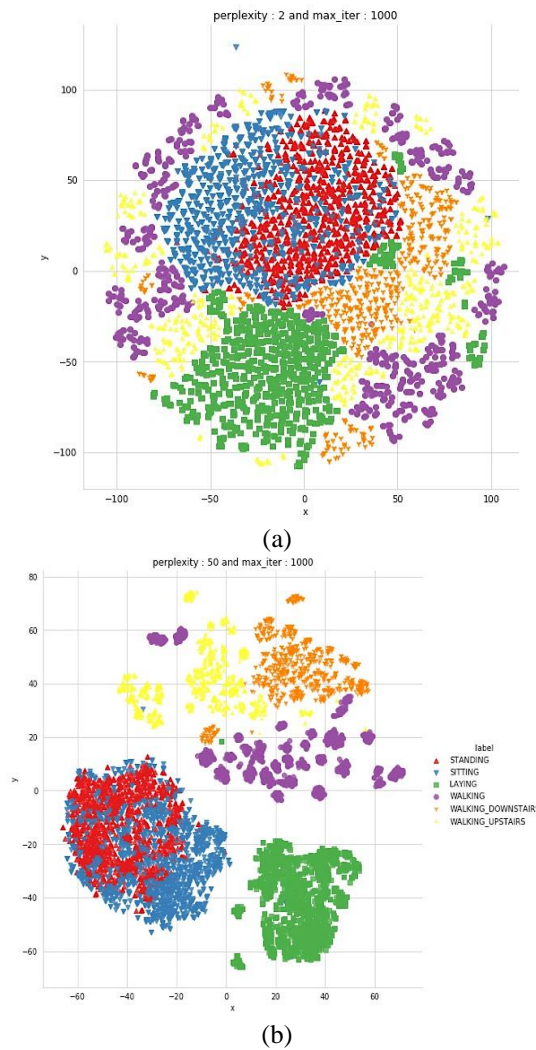

(a)

(b)

Figure 3. Spread data standing, sitting, laying, walking, walking downstairs, walking upstairs: (a) data similarity standing and sitting activities result in deep learning errors in classifying activity and (b) configuration 1000 iterations and perplexities 2, 5, 10, 20, and 50, standing and sitting activities that have the same type of data

Table 3. CNN used in 2 class

Layer	Parameter	Score
Layer 1	Filter	32
	Kernel size	3
	Activation	ReLu
	kernel_initializer	he_uniform
	input_shape	(128, 9)
Layer 2	Filter	32
	Kernel size	3
	Activation	ReLu
	kernel_initializer	he_uniform
Dropout	uniform	0.6
MaxPooling1D	pool_size	2
Flatten		Flatten
Dense	Act: ReLu	50
Dense	Act: softmax	2
Keras.optimizers	Adam	0.001
	nb_epoch	100

3. RESULTS AND DISCUSSION

Evaluation of predicted results from each model using a confusion matrix. The confusion matrix is a method used to perform accuracy calculations on a predictive system. Confusion matrix contains actual information and predictions on the classification system. To find accuracy, precision, and recall sequentially using (3)-(5).

$$Accuracy = \sum_{i=1}^{i} \frac{TP_i + Tn_i}{TP_i + TP_i + FN_i + TN_i} * 100\% \tag{3}$$

$$Precision = \frac{\sum_{i=1}^{i} TP_i}{\sum_{i=1}^{i} (TP_i + FP_i)} * 100\% \tag{4}$$

$$Recall = \frac{\sum_{i=1}^{i} TP_i}{\sum_{i=1}^{i} (TP_i + FN_i)} * 100\% \tag{5}$$

In this article, the author performs testing using CNN on the HAR dataset into 2 classes, namely the dynamic class and the static class with the parameters shown in Table 4. The use of hyperparameter tuning to get the best parameter combinations generates the highest accuracy on each static and dynamic dataset viewable in Table 5. Hyperparameter can provide the configuration of the parameters needed for CNN models of the selected dataset by randomly creating a combination of parameters. On the first layer, the hyperparameter will select the filter values between (28, 43, or 42), similarly the values of the kernel size, max pooling id, batch size, epoch, and dense parameters. While the configuration of the dropout parameter will be determined using a value between 0.45-0.7. The optimizer will be used between Adam and RMSprop with a value between 0.00065-0.004. Hyperparameter tuning is executed by the number of models to be generated as many as 100 models. The overall configuration can be seen in Table 4. 100 combinations are executed using Hyperparameter tuning, SA datasets get an accuracy of 97% in data train and 96% in the validation data shown in Figure 4, while the accuracy of the DA Dataset generates an accuracy value of 100% on the data train and 97.4% in the validation data shown in Figure 5.

Table 4. CNN tuning hyperparameter preparation

| Layer	Parameter	Score	Annotation	Layer	Parameter	Score	Annotation
Layer 1	Filter	28, 32, 42	Choice	Dropout	uniform	0.45 - 0.7	Range
	Kernel size	3 ,5, 7	Choice				
	Activation	Relu	Fixed				
	kernel_initializer	he_uniform	Fixed	MaxPooling1D	pool_size	2, 3 ,5	Cgoice
	kernel_regularizer l2	0,3	Fixed	Flatten		Flatten	Fixed
	input_shape	(128, 9)	Fided	Dense	Act: relu	16, 32, 64	Choice
Layer 2	Filter	16, 24, 42	Choice	Dense	Act: softmax	3	Fixed
	Kernel size	3, 5, 7	Choice				
	Activation	Relu	Fixed				
	kernel_initializer	he_uniform	Fixed	keras.optimizers	Adam, rmsprop	0.00065 - 0.004	Range
	kernel_regularizer l2	(0, 2)	Fixed	batch_size		16, 32, 64	Choice
				nb_epoch		35, 40	Choice
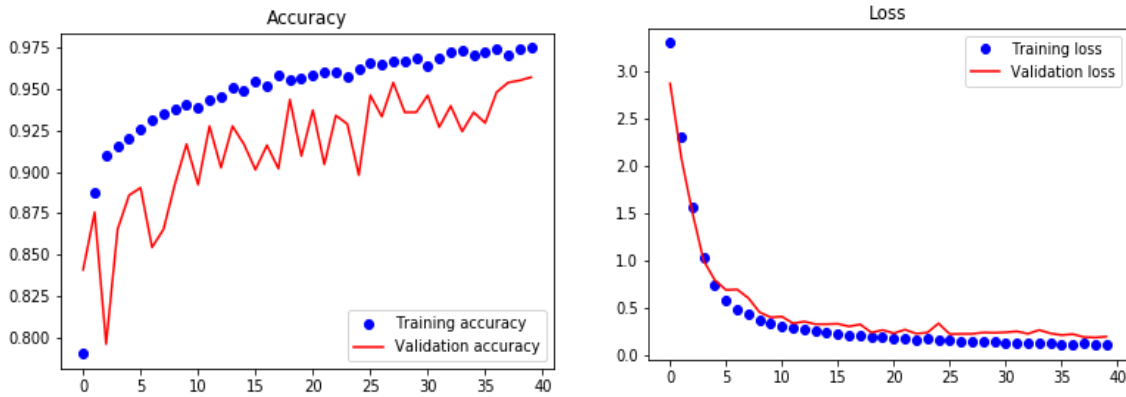
Figure 4. Accuracy and loss SA
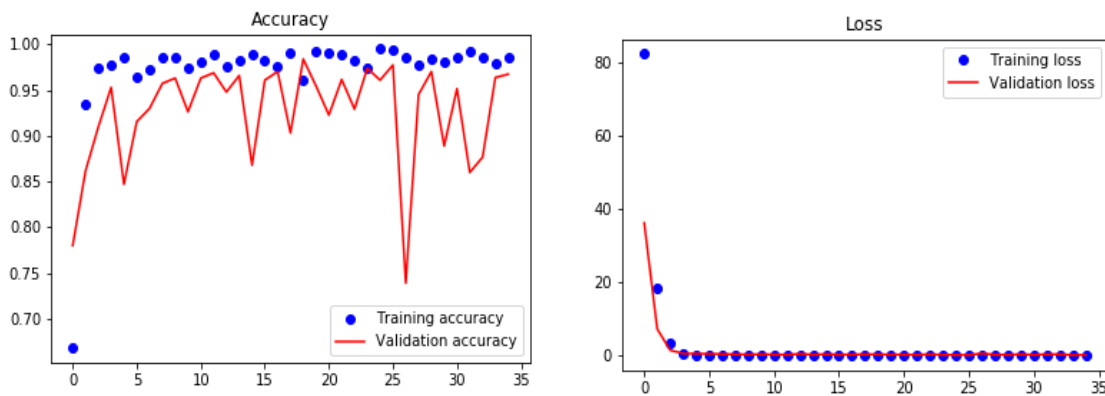
Figure 5. Accuracy and loss DA

Models obtained from three datasets are combined into the main model using the divide and conquer methods that can identify all human activities (walking, walking up, walking down, standing, laying, and sitting). This model generates an accuracy of 98.3% in data train and 97% in validation data. The model architecture IS seen in Figure 6. The comparison of this study with previous research titled "Human activity recognition with smartphone sensors using deep learning neural networks" [14] using the same dataset as well as tuning of the learning rate=0.006 resulted in an accuracy of 93.75%. In comparison, the proposed CNN model achieved a 97% accuracy can be seen in Table 6.

Table 5. Confusion matrix CNN research results in hyperparameter use divide and conquer

		Predicted class						Recall
		W	WU	WD	Si	St	L	
Actual Class	Walking	494	0	0	3	0	0	99.3%
	W. Up	0	447	0	2	0	0	99.5%
	W. Down	2	24	420	0	0	0	94.2%
	Sitting	0	0	0	460	29	0	94.1%
	Standing	0	0	0	22	501	0	95.8%
	Laying	0	0	0	4	2	537	98.9
	Prediction	99.5%	95%	100%	94%	94%	100%	97%

Table 6. Comparison of previous studies

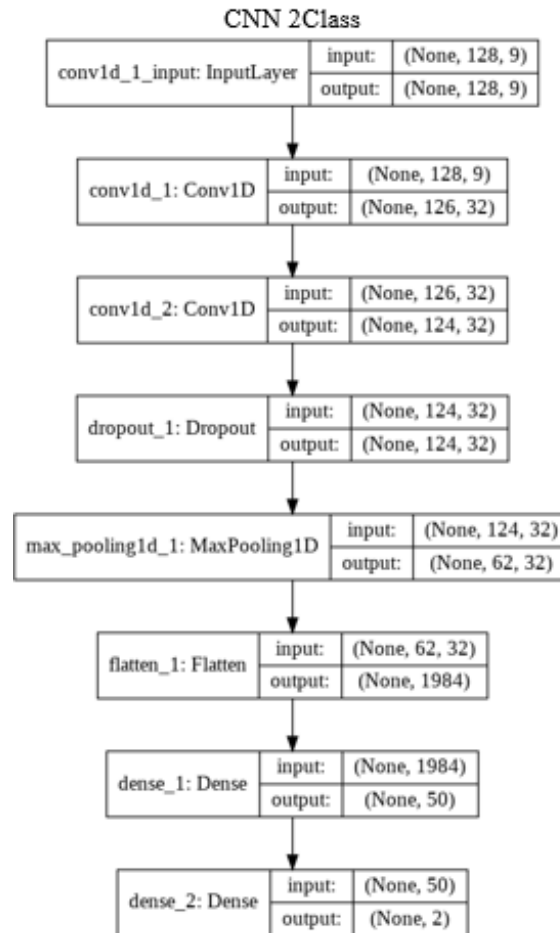| Method	Accuracy
HCF+NB [14]	79.43%
HCF+J48 [14]	82.62%
HCF+ANN [14]	82.27%
HCF+SVM [14]	77.66%
Convnet [14]	93.75%
CNN Proposed method	97%

CNN 2Class

Figure 6. SA and DA combined models

4. CONCLUSION

This paper proposed a human activity recognition using CNN with 6 classes (walking, walking upstairs, walking downstair, sitting, standing, and laying). Based on the results and discussion, the divide and conquer method combine with CNN tuning hyperparameter in each category of datasets achieved 97% accuracy. The proposed model also solved the similarity problems with confusing static (sitting and standing) datasets. The highest accuracy reached 100% in the study was shown on the activities of walking downstair and laying, while the lowest accuracy of 94% is owned by the activities of sitting and standing.

ACKNOWLEDGMENTS

REFERENCES

[1]   L. Chen and C. Nugent, "Human activity recognition and behaviour analysis," *Springer,* 2019.
[2]   Y. Xia, J. Zhang, Q. Ye, N. Cheng, and Y. Lu, "Evaluation of deep convolutional neural networks for detection of freezing of gait in Parkinson's disease patients," *Biomedical Signal Processing and Control*, vol. 46, pp. 221–230, 2018, doi: 10.1016/j.bspc.2018.07.015.
[3]   A. E. Minarno, W. Kusuma, and H. Wibowo, "Performance Comparisson Activity Recognition using Logistic Regression and Support Vector Machine," *2020 3rd International Conference on Intelligent Autonomous Systems (ICoIAS)*, 2020, pp. 19–24, doi: 10.1109/ICoIAS49312.2020.9081858.
[4]   A. Minarno, W. Kusuma, H. Wibowo, D. Akbi, and N. Jawas, "Single triaxial accelerometer-gyroscope classification for human activity recognition," *2020 8th International Conference on Information and Communication Technology (ICoICT)*, 2020, pp. 1–5, doi: 10.1109/ICoICT49345.2020.9166329.

[5]   C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698–703, 2019, doi: 10.1016/j.procs.2019.08.100.

[6]   F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in human behavior*, vol. 15, no. 5, pp. 571–583, 1999, doi: 10.1016/S0747-5632(99)00037-0.

[7]   L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *International conference on pervasive computing*, vol. 3001, 2004, pp. 1–17, doi: 10.1007/978-3-540-24646-6_1.

[8]   T. Plötz, N. Hammerla, and P. Olivier, "Feature learning for activity recognition in ubiquitous computing," *Twenty-second international joint conference on artificial intelligence*, 2011, doi: 10.5591/978-1-57735-516-8/IJCAI11-290.

[9]   T. Gonzalez, "Handbook of approximation algorithms and metaheuristics," *CRC Press*, 2007.

[10] O. C. Kurban and T. Yıldırım, "Daily motion recognition system by a triaxial accelerometer usable in different positions," *IEEE Sensors Journal*, vol. 19, no. 17, pp. 7543–7552, 2019, doi: 10.1109/JSEN.2019.2915524.

[11] M. Yang, H. Zheng, H. Wang, and S. McClean, "iGAIT: an interactive accelerometer based gait analysis system," *Computer methods and programs in biomedicine*, vol. 108, no. 2, pp. 715–723, 2012, doi: 10.1016/j.cmpb.2012.04.004.

[12] D. Fuentes, L. Gonzalez-Abril, Angulo. C, and J. Ortega, "Online motion recognition using an accelerometer in a mobile device," *Expert systems with applications*, vol. 39, no. 3, pp. 2461–2465, 2012, doi: 10.1016/j.eswa.2011.08.098.

[13] A. M. Khan, "Recognizing physical activities using Wii remote," *International Journal of Information and Education Technology*, vol. 3, no. 1, 2013, doi: 10.7763/IJIET.2013.V3.234.

[14] C. Ronao and S. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert systems with applications*, vol. 59, pp. 235–244, 2016, doi: 10.1016/j.eswa.2016.04.032.

[15] W. Qi, H. Su, C. Yang, G. Ferrigno, E. De Momi, and A. Aliverti, "A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone," *Sensors*, vol. 19, no. 17, 2019, doi: 10.3390/s19173731.

[16] J. Nunez, R. Cabido, J. Pantrigo, A. Montemayor, and J. Velez, "Convolutional neural networks and long short-term memory for skeleton-based human activity and hand gesture recognition," *Pattern Recognition*, vol. 76, pp. 80–94, 2018, doi: 10.1016/j.patcog.2017.10.033.

[17] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018, doi: 10.1016/j.asoc.2017.09.027.

[18] M. Gadaleta and M. Rossi, "IDNet: Smartphone-based gait recognition with convolutional neural networks," *Pattern Recognition*, vol. 74, pp. 25–37, 2018, doi: 10.1016/j.patcog.2017.09.005.

[19] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 381–388, doi: 10.1109/IJCNN.2016.7727224.

[20] J. Kim, G. Hong, B. Kim, and D. Dogra, "deepGesture: Deep learning-based gesture recognition scheme using motion sensors," *Displays*, vol. 55, pp. 38–45, 2018, doi: 10.1016/j.displa.2018.08.001.

[21] Y. Yoo, "Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches," *Knowledge-Based Systems*, vol. 178, pp. 74–83, 2019, doi: 10.1016/j.knosys.2019.04.019.

[22] M. Zhang, H. Li, J. Lyu, S. H. Ling, and S. Steven, "Multi-level CNN for lung nodule classification with Gaussian Process assisted hyperparameter optimization," 02-Jan-2019. [Online]. Available: http://arxiv.org/abs/1901.00276. [Accessed: 12-Oct-2021].

[23] S. H. Gupta, A. Sharma, M. Mohta, and A. Rajawat, "Hand movement classification from measured scattering parameters using deep convolutional neural network," *Measurement*, vol. 151, 2020, doi: 10.1016/j.measurement.2019.107258.

[24] W. Kusuma, Z. Sari, A. Minarno, H. Wibowo, and D. Akbi, "Performance Comparisson Human Activity Recognition using Simple Linear Method," *KINETIK*, vol. 5, no. 1, pp. 1–10, 2020, doi: 10.22219/kinetik.v5i1.1025.

[25] A. Galán-Mercant, A. Ortiz, E. Herrera-Viedma, M. Tomas, B. Fernandes, and J. Mora;-Munoz, "Assessing physical activity and functional fitness level using convolutional neural networks," *Knowledge-Based Systems*, vol. 185, 2019, doi: 10.1016/j.knosys.2019.104939.

[26] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys*, vol. 54, no. 4, Jul. 2021, doi: 10.1145/3447744.

[27] Z. Khan and J. Ahmad, "Attention induced multi-head convolutional neural network for human activity recognition," *Applied Soft Computing*, vol. 110, 2021, doi: 10.1016/j.asoc.2021.107671.

[28] S. K. Challa, A. Kumar, and V. B. Semwal, "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data," *The Visual Computer*, pp. 1–15, 2021, doi: 10.1007/s00371-021-02283-3.

[29] S. Mekruksavanich and A. Jitpattanakul, "Biometric user identification based on human activity recognition using wearable sensors: An experiment using deep learning models," *Electronics*, vol. 10, no. 3, 2021, doi: 10.3390/electronics10030308.

[30] 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, 2021, doi: 10.3390/s21030885.