ISSN: 2710-2165

Classification of Human Activity Recognition Using Machine Learning on the WISDM Dataset

Sarah W. Abdulmajeed*
*Computer Engineering Department, College of Engineering, Al-Iraqia University, Iraq
Email: sara.waleed@aliraqia.edu.iq
https://orcid.org/0000-0001-6222-4905

Abstract

The significance of human activity recognition (HAR) is rising as it seeks to improve everyday life and healthcare through better technology access and efficiency. Its objective is to transform industries by enabling smart homes, improving robots, bolstering security, and improving human-computer interactions. HAR works to improve well-being, which is essential to health, wellness, and sports.
While the complexity of human behavior poses challenges, machine learning advancements offer hope for solutions. Continuous research in accurately detecting a wide array of human activities underscores the significant impact of HAR on technological development and its broad applications.
In this work, a convolution neural network CNN algorithm and random forest RF algorithms were produced for human recognition activity classification using WISDM-51 dataset that contains 18 human activities. The CNN achieved an accuracy of 89.36%, whereas the RF algorithm reached a slightly higher accuracy of 93.46%. The results suggest that the proposed algorithms offer promising potential.

Keywords- HAR, CNN, RF, machine learning.

I. INTRODUCTION

Researching the use of body-worn sensors to identify people's activities is a topic that has gotten much attention. Recording and knowing someone's everyday actions helps us understand their normal activity levels and how well they can function[1]. Machine learning plays a crucial role in tracking human activities because it helps keep an eye on someone's health and daily living habits, and it is also helpful in medical settings for tracking diseases and other health signs. Recently, there has been notable value in efficiently allocating resources for medical care. Using machine learning concepts for human activity recognition is a far more economical option than investing in other initiatives. Human activity recognition capabilities could surpass the knowledge of individuals in the medical and health sectors if combined with the notion that it can be based explicitly on data gathered from simply accessible devices like smartphones and smartwatches [2].

There are two main types of HAR systems: sensor-based and video systems. The sensor-based is used more commonly. Sensor-based HAR tracks and detects the surrounding environment, bodily motions, and other important information using a range of sensors. Sensor-based systems may, however, have certain drawbacks. For instance, they might not correctly depict some tasks requiring precise or complicated motions, like writing or playing an instrument. Furthermore, sensor-based systems could be more vulnerable to mistakes or inefficiencies as a result of signal noise, calibrating, or other technical concerns. Despite these drawbacks, Sensor-based HAR systems are widely utilized in many applications [3].

Our objective is to classify the human activity using the WISDM dataset with 18 activities (Walking, Running, Using Stairs, Sitting, Standing, Typing, Brushing Teeth, Eating Soup, Eating chips, Having Pasta, Swallowing From a Cup, Taking a Sandwich, Soccer Ball

Kicking, Catching a Tennis Ball, Basketball Dribbling, Writing, Clapping, Folding Laundry) using the two classification algorithms random forest and CNN which gives 93% and 89% of accuracy respectively.

II. LITERATURE REVIEW

[1] Explored a novel approach for selecting features for recognizing human behaviors. Numerous complex classification techniques were evaluated, including SVM, Bayes classifier, MLP, KNN, MLM, and MLM-NN.

[2] reveals the outcomes of various models that were specifically trained with designated sensors. It details the development of these models by applying algorithms such as Random Forest, k-Nearest Neighbor, or Support Vector Machine. The effectiveness of each model is evaluated and compared by investigating the impact of various configurations of mobile sensors on their precision in recognizing patterns or activities.

In [3], four types of machine learning models: Decision Tree Classifiers (DTC), Artificial Neural Networks, K-Nearest Neighbors, and Random Forest Classifiers were deployed to precisely classify various human activities, like ascending stairs, descending stairs, regular walking, walking on toes, walking on heels, and performing sit-ups.

[4] Valued using data from the CogAge dataset, which is available to the public and includes basic and complex human actions. This data was collected using three types of movable technology: smartphones, smartwatches, and smart glasses. Additionally, the effectiveness of different classification algorithms in recognizing these complex activities was evaluated.

[5] Present a detailed analysis and a precise definition of the system's lifecycle for developing an efficient Human Activity Recognition (HAR) system in smart homes, requiring little manual intervention. This lifecycle outlines the different phases in developing the HAR system, guided by scenarios typically found in home settings.

[6] Suggests a tailored and intuitive structure for identifying human activities, leveraging an automatic approach to machine learning via Neural Architecture Search. It adeptly analyzes 3D video data, encompassing RGB, depth, skeleton, and environmental object information, by utilizing separate 2D convolutional neural networks for each type of data stream.

[7] Thoroughly organizes and evaluates prior research utilizing deep learning approaches for HAR through wearable devices, providing an extensive analysis of the latest developments, future directions, and major hurdles.

[8] Developed a framework that utilizes Long Short-Term Memory (LSTM) networks, designed to maintain the training of deep learning models on human activity information gathered from controlled and real-world environments.

[9] proposes a deep neural network architecture incorporating residual bidirectional Long Short-Term Memory (LSTM) units.

[10] presents a deep learning tool to identify exercises for seniors, overcoming the data shortage by creating the Routine Exercise Dataset (RED) with 19 exercises for the elderly and 14,440 samples from 19 participants. Using new feature extraction methods and a complex LSTM network, the approach was evaluated on 16 different datasets, from popular public ones to their RED dataset, including versions with added noise.

[11] leveraged the complementary strengths of Autoencoders (AEs), Convolutional Neural Networks (CNNs), and Long Short-Term Memory networks (LSTMs) by integrating them into a single architecture called "ConvAE-LSTM." on four regular public datasets: WISDM, UCI, PAMAP2, and OPPORTUNITY, This approach was tested.

[12] Simplifies recognizing human activities through a web API designed for wearables with limited processing abilities. It processes raw data directly, avoiding complex pre-processing by employing advanced image recognition on plot images created from the raw sensor data.

[13] Outlines a method for detecting human activities using a Convolutional Neural Network (CNN) model, which leverages smartphone data. This model recognizes everyday actions like running, sitting, strolling, standing up, and moving up or down stairs.

[14] It focuses on creating a cost-efficient and fast human activity recognition system that can process video and image data to determine the activities depicted. It is designed to benefit end-users across various applications, such as surveillance and aid-related tasks.

In [15] the well-regarded WISDM dataset for activity detection was employed. By conducting a multivariate analysis of covariance (MANCOVA), identified a statistically significant difference ($p < 0.05$) in the sensor data obtained from smartphones and smartwatches.

[16] focused on training machine learning models with a wide range of human activities. Techniques such as K-Nearest Neighbor (KNN), Naive Bayes (NB), and Support Vector Machine (SVM) were applied to the analysis, leading to the creation of a new model termed the feature-based fused SVM-KNN approach.

[17] looks at different ways to identify human activities, focusing on how smartphones and smartwatches use their sensors. It also explores how Machine Learning and newer deep learning techniques are applied. The review is organized around four main topics: the types of sensors used, their application, the Machine Learning (ML) and Deep Learning (DL) techniques involved, and what successes and problems have been encountered.

[18] reviews emotion recognition via EEG signals in BCIs, highlighting key developments, data sources, and techniques for eliciting emotions. It explores EEG feature handling, machine and deep learning methods (like SVM, ANNs, CNNs, and RNNs with LSTM), and how these relate to EEG patterns and emotions. The study also evaluates and compares emotion recognition efforts and algorithms from 2015 to 2021.

[19] focuses on Human Activity Recognition (HAR) through wearable sensor data for health monitoring. It highlights the limitations of current Artificial Neural Networks (ANNs) due to high computational demands and poor temporal feature handling. Proposes Spiking Neural Networks (SNNs) as a more efficient alternative inspired by the functioning of biological neurons, addressing the drawbacks of ANNs in HAR.

III. METHODOLOGY

The proposed method used in this paper is illustrated in Figure 1 below, where the data was collected and preprocessed and then classified by the two algorithms (CNN and RF) to find the results and effectiveness of each algorithm.

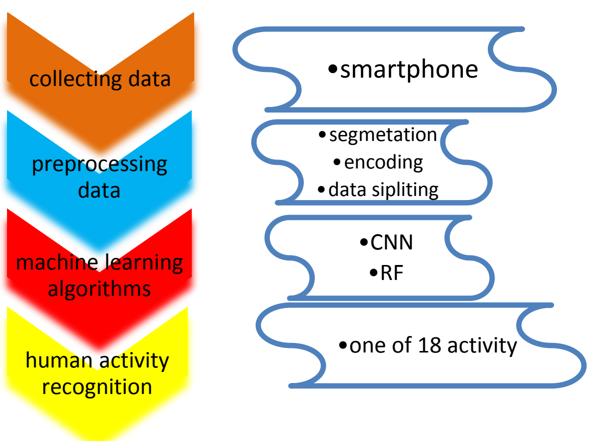


Figure 1: the proposed method

IV. RESULTS AND DISCUSSION

A. Data Preprocessing

First, the WISDM-51 data set[20] were downloaded from the Kaggle website. The data set contains 18 human activities, as explained i n Table 1 below. The dataset needs some preprocessing to be used for CNN and RF algorithms. The first preprocessing step is the seg mentation, in which the continuous stream of accelerometer data is segmented into smaller, fixed-size windows. Label encoding and o ne hot encoding were then done to make the data suitable for the machine learning algorithms. Data splitting is important to evaluate t he model's performance on unseen data. Then, reshaping is done for the input to CNN, which is flattened to the RF algorithm.

TABLE1: HUMAN ACTIVITY FOR THE WISDM-51 DATA SET

Activity	Code
Walking	A
Running	B
Stairs	C
Sitting	D
Stand-up	E
Typing	F
Brushing Teeth	G
Eating Soup	H
Eating Chips	I
Eating Pasta	J
Drinking from Cup	K
Eating Sandwich	L
Kicking (Soccer Ball)	M
Playing Catch w/Tennis Ball	O
Dribbling (Basketball)	P
handwriting	Q
Clapping	R
Folding Clothes	S

B. Hyperparameter tuning

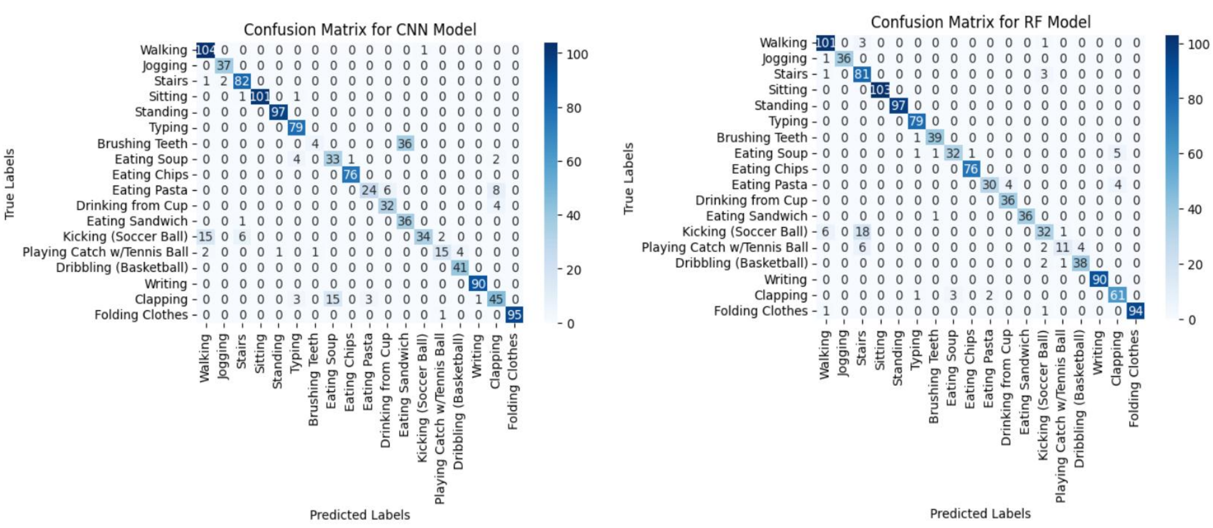
The hyperparameter search for CNN is completed using a random search, shown in Table 2.

TABLE 2: THE HYPERPARAMETER TUNING

Hyperparameter	number
The optimal filters number in the first Conv1D layer	128
The optimal size of the kernal for the Conv1D layer	5
The optimal units number in the Dense layer	150
The optimal optimizer's learning rate	0.0003244843025329892

C. Performance metrics

- *Confusion matrix*



(a) (b)

Figure 2: confusion matrix for (a) convolution neural network algorithm (b) random forest algorithm

Figure 2 shows the confusion matrix for both CNN and RF algorithms; the diagonal elements were compared since they represent the true positives for each activity category. The higher value indicates that the model identifies the activity more correctly.

The two algorithms seem to give good behavior. In contrast, CNN is better at detecting walking, running, staring, eating soup, kicking, Playing Catch w/Tennis Ball, Dribbling (Basketball), and Folding Clothes activities.

RF classifies a bit better in sitting, brushing teeth, eating pasta, drinking from a cup, and clapping activities. They behave the same way for the rest of the activities (standing, typing, eating chips, eating sandwiches, and writing).

- *Classification report*

Precision measures the frequency at which the model's predictions for an activity are accurate, based on all instances it has made such predictions.
Recall: This assesses the model's efficiency in accurately identifying occurrences of each activity.
F1-score merges precision and recall into a single measure, balancing them. A higher F1-score indicates effective model performance in both precision and recall.
Table 3 clarifies the performance for each algorithm using precision, recall, F1-score, and support score for each activity:
RF is better at classifying Walking, typing, eating soup, eating pasta, drinking from a cup, playing catch w, and clapping without confusion with other activities. The RF is superior in identifying brushing teeth activity compared to CNN. Although the RF model has perfect precision in eating sandwich activity and the CNN is better at recall, the RF still has a higher F1-score
CNN is better at recognizing stairs without confusion with other activities and has much better precision in kicking (soccer ball). However, it gives a little recall, like RF in this activity.
The two models seem to be the same in sitting, standing, and eating chips, but RF has perfect precision in running and perfect F1-score in writing activity.
For dribbling (basketball) and folding clothes, the two models perform perfectly with CNN's slight edge
Generally, the Random Forest (RF) model shows superior precision in identifying a wide array of activities, underscoring its dependability in recognizing specific behaviors. However, the Convolutional Neural Network (CNN) model is not far behind and even outperforms the RF model in some cases, notably in Basketball Dribbling, where it achieves a more even mix of precision and recall. It is noteworthy that both models encounter more difficulty with activities that entail intricate movements, like Soccer Ball Kicking or Playing Catch, compared to more stationary activities, such as Sitting or Standing.

TABLE 3: THE PERFORMANCE METRICS FOR BOTH CONVOLUTION NEURAL NETWORK AND RANDOM FOREST ALGORITHMS

Human activity		**precision**		**recall**		**f1-score**		**support**	
name	***label***	***CNN***	***RF***	***CNN***	***RF***	***CNN***	***RF***	***CNN***	***RF***
Walking	A	0.85	0.92	0.99	0.96	0.92	0.94	105	105
running	B	0.95	1.00	1.00	0.97	0.97	0.99	37	37
Stairs	C	0.91	0.75	0.96	0.95	0.94	0.84	85	85
Sitting	D	1.00	1.00	0.98	1.00	0.99	1.00	103	103
Standing	E	0.99	1.00	1.00	1.00	0.99	1.00	97	97
Typing	F	0.91	0.96	1.00	1.00	0.95	0.98	79	79
Brushing Teeth	G	0.80	0.95	0.10	0.97	0.18	0.96	40	40
Eating Soup	H	0.69	0.91	0.82	0.80	0.75	0.85	40	40
Eating Chips	I	0.99	0.99	1.00	1.00	0.99	0.99	76	76
Eating Pasta	J	0.89	0.94	0.63	0.79	0.74	0.86	38	38
Drinking from Cup	K	0.84	0.90	0.89	1.00	0.86	0.95	36	36
Eating Sandwich	L	0.50	1.00	0.97	0.97	0.66	0.99	37	37
Kicking (Soccer Ball)	M	0.97	0.78	0.60	0.56	0.74	0.65	57	57
Playing Catch w/Tennis Ball	O	0.83	0.85	0.65	0.48	0.73	0.61	23	23
Dribbling (Basketball)	P	0.91	0.90	1.00	0.93	0.95	0.92	41	41
Writing	Q	0.99	1.00	1.00	1.00	0.99	1.00	90	90
Clapping	R	0.76	0.87	0.67	0.91	0.71	0.89	67	67
Folding Clothes	S	1.00	1.00	0.99	0.98	0.99	0.99	96	96

From Table 4, it can be observed that RF outperforms CNN in accuracy. RF model also shows superior precision, recall, and F1-score performance in both macro and weighted averages.

TABLE 4: ACCURACY OF CONVOLUTION NEURAL NETWORK AND RANDOM FOREST ALGORITHMS

convolution neural network	**random forest**
val_accuracy: 0.8936 accuracy 0.89 1147 macro avg 0.88 0.85 0.84 1147 weighted avg 0.90 0.89 0.88 1147	Test accuracy: 0.934612031386225 accuracy 0.93 1147 macro avg 0.93 0.90 0.91 1147 weighted avg 0.94 0.93 0.93 1147

V. Conclusion

In conclusion, both models exhibit strengths in different areas when recognizing human activities. The Random Forest model presents a more consistent performance across various activities, highlighted by its superior precision and recall metrics. Conversely, the Convolutional Neural Network model has its advantages; it tends to face more challenges in accurately distinguishing between certain activities.

VI. References

[1] L. B. Marinho, A. H. de Souza Junior, and P. P. R. Filho, "A new approach to human activity recognition using machine learning techniques," Adv. Intell. Syst. Comput., vol. 557, no. January, pp. 529–538, 2017, doi: 10.1007/978-3-319-53480-0_52.

[2] W. Z. Tee, R. Dave, N. Seliya, and M. Vanamala, "Human Activity Recognition models using Limited Consumer Device Sensors and Machine Learning," Proc. - 2022 Asia Conf. Algorithms, Comput. Mach. Learn. CACML 2022, pp. 456–461, 2022, doi: 10.1109/CACML55074.2022.00083.

[3] N. G. Nia, E. Kaplanoglu, A. Nasab, and H. Qin, "Human Activity Recognition Using Machine Learning Algorithms Based on IMU Data," BioSMART 2023 - Proc. 5th Int. Conf. Bio-Engineering Smart Technol., no. June, pp. 1–8, 2023, doi: 10.1109/BioSMART58455.2023.10162095.

[4] F. Amjad, M. H. Khan, M. A. Nisar, M. S. Farid, and M. Grzegorzek, "A comparative study of feature selection approaches for human activity recognition using multimodal sensory data," Sensors, vol. 21, no. 7, pp. 1–21, 2021, doi: 10.3390/s21072368.

[5] S. K. Hiremath and T. Plötz, "The Lifespan of Human Activity Recognition Systems for Smart Homes," Sensors, vol. 23, no. 18, 2023, doi: 10.3390/s23187729.

[6] A. -C. Popescu, I. Mocanu and B. Cramariuc, "Fusion Mechanisms for Human Activity Recognition Using Automated Machine Learning," in *IEEE Access*, vol. 8, pp. 143996-144014, 2020, doi: 10.1109/ACCESS.2020.3013406.
https://dblp.org/rec/journals/access/PopescuMC20
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9153764&isnumber=8948470

[7] S. Zhang et al., "Deep Learning in Human Activity Recognition withWearable Sensors: A Review on Advances," Sensors, vol. 22, no. 4, 2022, doi: 10.3390/s22041476.

[8] S. Chung, J. Lim, K. J. Noh, G. Kim, and H. Jeong, "Sensor data acquisition and multimodal sensor fusion for human activity recognition using deep learning," Sensors (Switzerland), vol. 19, no. 7, 2019, doi: 10.3390/s19071716.

[9] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors," Math. Probl. Eng., vol. 2018, 2018, doi: 10.1155/2018/7316954.

[10] S. H. H. Shah, A. S. T. Karlsen, M. Solberg, and I. A. Hameed, "An efficient and lightweight multiperson activity recognition framework for robot-assisted healthcare applications," Expert Syst. Appl., vol. 241, no. June 2023, p. 122482, 2024, doi: 10.1016/j.eswa.2023.122482.

[11] D. Thakur, S. Biswas, E. S. L. Ho and S. Chattopadhyay, "ConvAE-LSTM: Convolutional Autoencoder Long Short-Term Memory Network for Smartphone-Based Human Activity Recognition," in IEEE Access, vol. 10, pp. 4137-4156, 2022, doi: 10.1109/ACCESS.2022.3140373.
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9668957&isnumber=9668973

[12] A. I. Alexan, A. R. Alexan, and S. Oniga, "Real-Time Machine Learning for Human Activities Recognition Based on Wrist-Worn Wearable Devices," Appl. Sci., vol. 14, no. 1, p. 329, 2023, doi: 10.3390/app14010329.

[13] P. Seelwal, C. Srinivas, and ..., "Human Activity Recognition using WISDM Datasets," J. …, no. December, 2023, doi: 10.52783/joee.v14i1s.92.

[14] D. Parmar, M. Bhardwaj, A. Garg, A. Kapoor, and A. Mishra, "Human activity recognition system," 2023 Int. Conf. Comput. Intell. Commun. Technol. Networking, CICTN 2023, vol. 13, pp. 533–535, 2023, doi: 10.1109/CICTN57981.2023.10141250.

[15] B. Oluwalade, S. Neela, J. Wawira, T. Adejumo, and S. Purkayastha, "Human activity recognition using deep learning models on smartphones and smartwatches sensor data," Heal. 2021 - 14th Int. Conf. Heal. Informatics; Part 14th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2021, pp. 645–650, 2021, doi: 10.5220/0010325906450650.

[16] V. S. Devi, K. Sumathi, M. Mahalakshmi, A. J. Anand, A. Titus, and N. N. Saranya, "Machine Learning Based Efficient Human Activity Recognition System," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 5s, pp. 338–346, 2024.

[17] R. A. Lateef and A. R. Abbas, "Human Activity Recognition using Smartwatch and Smartphone: A Review on Methods, Applications, and Challenges," Iraqi J. Sci., vol. 63, no. 1, pp. 363–379, 2022, doi: 10.24996/ijs.2022.63.1.34.

[18] E. H. Houssein, A. Hammad, and A. A. Ali, Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review, vol. 34, no. 15. Springer London, 2022.

[19] Y. Li, R. Yin, Y. Kim, and P. Panda, "Efficient human activity recognition with spatio-temporal spiking neural networks," Front. Neurosci., vol. 17, 2023, doi: 10.3389/fnins.2023.1233037.

[20] https://www.kaggle.com/datasets/ruhallahahmadian/WISDM-51.