Journal of Engineering Science and Technology
Special Issue on INCITEST2025 Vol. 21, No. 2 (2026) 113 - 120

K-NEAREST NEIGHBOR ALGORITHM FOR IDENTIFICATION OF HUMAN ACTIVITY BY UTILIZING A SMARTPHONE

WARTIKA WARTIKA[1,*], AGUS NURSIKUWAGUS[1], TONO HARTONO[1], WAHYUNI WAHYUNI[1], ZULFIKAR ZULFIKAR[1], NARGIZA TOXIROVNA AVULOVA[2]

[1]Universitas Komputer Indonesia, Bandung, Indonesia
[2]Samarkand State University, Uzbekistan
*Corresponding Author: wartika@email.unikom.ac.id

Abstract

This study is geared toward creating a human activity recognition (HAR) system with the use of internal sensors in smartphones and the K-Nearest Neighbor (K-NN) algorithm. Smartphones as data source have great potential in identifying activities without the need for other additional devices, thus being very suitable for health applications, daily activity monitoring, and context-based intelligent systems. The research methodology consists of the signal data collection from the accelerometer and gyroscope sensors while performing a number of the most common activities i.e. walking, running, sitting, standing, and climbing and descending stairs. Next, the obtained data undergoes pre-processing, then statistical feature extraction is performed in both time and frequency domains and finally, normalization is done to bring the range of values to the same scale. The K-NN algorithm was evaluated with different k values to determine the optimal classification setup. The experimental results demonstrated that this method was quite successful in identifying human activities, achieving an accuracy of 79.56%, thus K-NN can be considered a very effective algorithm for smartphone-based HAR systems. Although its performance depends on the selection of k parameters and the features' quality, K-NN is still capable of providing consistent results and has the advantage of high computational efficiency. This work demonstrates a straightforward yet fairly accurate method and it may become a viable tool for various mobile applications operating physical activity monitoring in real-time, particularly in digital health, fitness, and user-context-based smart systems.

Keywords: Human activity recognition, K-nearest neighbor (K-NN), Machine learning, Pattern recognition, Smartphone sensor.

1. Introduction

Technological progress in the field of mobile devices, especially smartphones, has significantly sped up the development of human activity recognition (HAR) systems using sensors. Nowadays, smartphones with built-in accelerometers, gyroscopes, and magnetometers are capable of continuously tracking human movement patterns in various situations, thus allowing the detection of activities such as walking, running, sitting, standing, and even climbing stairs at a very high level of accuracy. This paper titled "The K-Nearest Neighbor Algorithm for Human Activity Identification Using Smartphones" is centred on the use of the k-nearest neighbor (K-NN) algorithm as the main classification method for the identification of several human activities from the smartphone sensor data. The K-NN algorithm was selected as the method of choice due to its straightforwardness, the fact that it doesn't require specific parameter assumptions, and its proficiency in clustering data based on closeness [1].

Numerous pieces of research demonstrated that K-NN algorithm could provide performance at a competitive level with other models in the area of HAR. A study indicates that K-NN is able to get high accuracy in simple activity classification by employing statistical features from smartphone sensors [2]. Another research work also confirmed that K-NN performs well on three-axis acceleration data [3]. It has been revealed that even though K-NN is a simple algorithm, it may still outperform or at least match in performance with more complicated machine learning methods [4]. One research has demonstrated that the combined use of accelerometer and gyroscope sensors can enhance the K-NN classification accuracy of dynamic activities [5]. On the other hand, other research shows that K-NN is one of the algorithms that remains relatively stable when applied to multi-sensor-based smartphone data [6]. Moreover, feature management and pre-processing procedures play a significant role in enhancing the performance of HAR [7]. Smartphones can determine complicated activities through statistical feature extraction at an advanced level [8].

The significance of K-NN hyperparameter setting on the resultant model performance has been discussed [9]. In the meantime, smartphone-based HAR has tremendous potential in the healthcare and behavioural monitoring sectors [10]. Recent research offers an adaptive approach that increases model robustness to changing movement patterns [11]. Several other studies also confirm the effectiveness of using smartphone sensors in HAR systems. Another study compared a number of algorithms and concluded that K-NN offers an optimal balance between accuracy and computational load [12]. Signal stages that are able to improve the performance of K-NN-based models [13]. Combining features from the time and frequency domains can significantly improve the effectiveness of HAR systems [14]. In addition, the normalization and dimension reduction processes are important factors in improving the accuracy of K-NN on sensor data [15]. Based on this study, this study sets three main objectives: (1) building a human activity recognition model based on smartphone sensor data using the K-NN algorithm; (2) testing the model performance with various k values and feature types; and (3) comparing the findings of this study with previous studies to identify contributions to improving accuracy [16].

This study has utilized a human activity data gathering method that incorporates the use of accelerometer and gyroscope sensors on a smartphone to collect the data. Further on, statistical and frequency-based feature extraction, data normalization, and K-NN algorithm classification are done. Subsequently, the model performance evaluation is done using several performance metrics, such as accuracy, precision, recall, and F1 score. The main point of difference of this research is the simultaneous use of k-value adjustment strategies, selection of the most relevant features, and balancing the dataset obtained from real-world conditions, instead of laboratory data. It is anticipated that this method will yield a HAR model that is less sensitive to the variations in user behaviour and environmental changes, thereby being suitable for mobile applications based on activity monitoring.

2. Literature Review

A literature review is a part of a scientific paper where the author summarizes, analyses, and synthesizes previous relevant research. It mainly intends to show a thorough understanding of the research area and point out the gaps in research which the proposed work will fill.

2.1. Smartphone based human activity recognition (HAR)

Human activity recognition (HAR) is the process of identifying human physical activity based on data generated by sensors. These activities can include walking, running, sitting, standing, climbing stairs, sleeping, and even driving. HAR is important due to its diverse applications in everyday life, ranging from fitness tracking and patient monitoring in the medical field, security systems, to the development of smart homes and wearable devices [17, 18]. Figure 1 shows that modern smartphones are equipped with various built-in sensors.

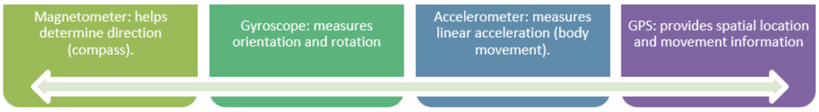

Fig. 1. Modern smartphones.

These sensors make smartphones ideal devices for HAR because they are practical, inexpensive, and always carried by users [15]. The general processes in a smartphone-based HAR System are data acquisition (data collection is performed in real-time from smartphone sensors), data preprocessing (includes time segmentation [windowing], noise filtering, and signal normalization), feature extraction (sensor data is transformed into statistical features, such as: mean, median, maximum, minimum, standard deviation, energy, entropy, and FFT), classification (using machine learning algorithms such as: Classical: K-nearest neighbor (K-NN), decision tree, SVM, random forest, deep learning: CNN, LSTM, GRU), evaluation (system performance is measured using metrics such as accuracy, precision, recall, and F1-score) [13, 14]. Benefits and applications of HAR include health (monitoring elderly patients, fall detection, activity reminders), sports and fitness (activity tracker, calorie counter), smart environment (smart home, smart office based on user context), transportation (detecting transportation modes (walking, cycling, car)).

Challenges in Smartphone-Based HAR are variation in smartphone position (in pocket, hand, bag), variation in movement between individuals, high battery consumption when sensors are continuously active, the need for precise segmentation and features, user data privacy and security.

2.2. K-nearest neighbour (K-NN) algorithm

K-Nearest Neighbor (K-NN) is one of the simplest classification algorithms in supervised machine learning. K-NN works by classifying test data based on the closest distance to training data with known labels. This algorithm does not require an explicit training phase, making it a lazy learner [3, 5]. How K-NN Works are to predict the class of a test dataset: calculate the distance between the test dataset and the entire training dataset, select the k closest training datasets (nearest neighbors), determine the majority class from these neighbors, and this majority class is the predicted result [15]. The performance of K-NN depends heavily on how the "distance" between the data is calculated. Some common methods as shown in Eq. (i-iii):

Euclidean Distance: $$d(p,q)=\sqrt{\sum_{i=1}^{n}(Pi-qi)^2} \tag{1}$$

Manhattan Distance: $$d(p,q)=\sum_{i=1}^{n}|Pi-qi| \tag{2}$$

Minkowski Distance: $$d(p,q)=\left(\sum_{i=1}^{n}|Pi-qi|^r\right)^{1/r} \tag{3}$$

Important Parameter K Value: the K value indicates the number of neighbors taken into account. If K is too small (e.g., K = 1), the model can be too sensitive to noise. If K is too large, classification decisions can be biased toward the majority class, therefore the ideal K value is usually found through cross-validation.

3. Research Method

This research follows an experimental approach to develop a Human Activity Recognition (HAR) system using smartphone sensor data. Figure 2 shows that the methodology consists of several structured stages [5].

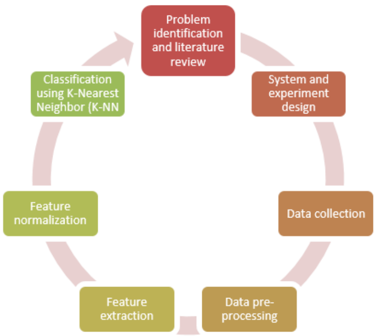

Fig. 2. Methodology.

4. Results and Discussion

4.1. Data retrieval mechanism

Primary data collection was carried out by 2 subjects using a Samsung S8 smartphone that had the aHar application installed. This application will record values from the accelerometer and gyroscope sensors when the user is active. Data collection mechanism using a smartphone is shown in Fig. 3.

Stand (1) → Rest →Sit(1) → Rest →Stand(2)→Rest →Lay Down(2) →Rest →Sit(2) → Rest →Laydown(2) →Rest →Walk(1) →Rest→Walk(2) →Rest→Walk downstairs(1)→Rest→Walks upstair(1)→Rest → Walk downstairs(2)→ Rest →Walks upstair(2)→Rest →Run(1)→Rest →Jump(1)→ Run(2)→Rest →Jump(2) :
Time (Seconds) 5,10,15

Fig. 3. Start standing position.

For each activity, the subject completed several steps. The first step was to get into position. Next, the subject pressed the start button (there was a 5-second delay before data collection began), before performing the experimental activity. After completing the data collection activity, the subject pressed the sync button to synchronize the data with the database server. This study used 2 types of smartphone positions when the subject conducted the experiment, namely in the right trouser pocket with placement using a portrait position and the smartphone held freely by the subject

4.2. Data processing mechanism for K-NN in human activity recognition

The data processing mechanism for K-NN in human activity recognition can be seen in Fig. 4. (1) Data Acquisition. Smartphones continuously capture sensor data such as accelerometers and gyroscopes. Data is recorded as a time series of values along the X, Y, and Z axes. User activities (e.g., walking, sitting, climbing stairs) are labelled according to the time of recording; (2) Pre-processing (Data Preparation). The purpose of pre-processing is to clean and prepare data to make it easier to process; (3) Feature Extraction. From each window, statistical features are calculated that represent activity patterns; (4) Feature and Label Dataset Formation. The feature extraction results for each window are converted into a single row of data. Each row is paired with a corresponding activity label. This dataset is used for: Training data, Testing data; (5) Classification with K-NN. During testing, new data (with extracted features) is compared with the training data.

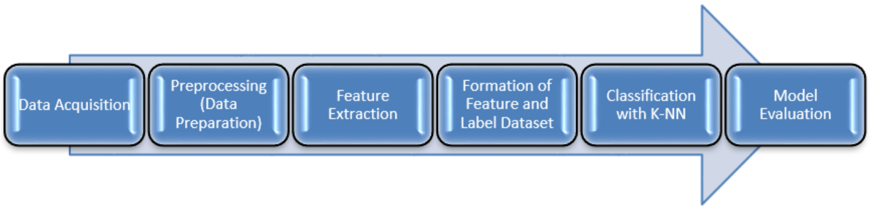

Fig. 4. The data processing mechanism for K-NN.

The distance between the test data and each training data set is calculated using a distance metric (usually Euclidean). Take the K nearest neighbours of the training data set. Predict the class (activity) based on the majority class of those K neighbours; (6) Model Evaluation. Calculates accuracy and other metrics by comparing predicted results with actual labels. Evaluation is performed to improve the model, for example by tuning the K value or selecting better features [5, 15].

Tables 1 and 2 show the accuracy results for human activity recognition using K-NN with experimental durations of 5 and 10 respectively. The values in bold represent the highest accuracy values at a given k value for each experimental duration, using both the Manhattan and Euclidean equations. It can be observed in the three tables that the position of the smartphone when held freely always has a higher accuracy than when the smartphone is placed in a trouser pocket, both using Manhattan and Euclidean calculations.

This may occur because when the smartphone is held freely, the values obtained by the accelerometer and gyroscope sensors have a more "definitive" characteristic than when the smartphone is placed in the right trouser pocket which may be “vaguer”. This shows that the longer the experiment time, the more data will be produced to be processed by the K-NN method so that it is able to obtain a better accuracy value.

Table 1. Activity accuracy results with 5 seconds time (in %).

	Manhattan		Euclidean	
K	Independent	Pocket	Independent	Pocket
1	76.1176	69.8413	75.2941	68.9342
2	70.7059	68.254	69.5294	67.2336
3	73.7647	68.254	71.8824	67.0068
4	72.8235	69.476	69.8824	67.8005
5	73.6471	67.6871	71.5294	66.6667
6	71.6471	67.1202	70.0	66.6667
7	71.5294	66.78	70.3529	64.966
8	70.4706	66.2132	69.0588	65.3061
9	71.7647	66.4399	70.0	64.7392
10	70.7059	65.5329	68.9412	64.6259

It can be observed that the longer the data collection or experiment, the longer the computation time. This applies to both types of distance calculations, Manhattan and Euclidean. The longer computation time occurs because calculations using the KNN algorithm require calculations between one testing object and all training data objects in the dataset.

Table 2. Activity accuracy results with 10 seconds time (in %).

	Manhattan		Euclidean	
K	Independent	Pocket	Independent	Pocket
1	77.352	64.1595	76.5756	63.0199
2	72.56	62.3292	72.56	61.1396
3	73.3965	62.5071	73.285	61.7664
4	72.783	62.849	73.3408	62.5071
5	73.7869	62.1652	73.0619	62.3932
6	72.5042	61.7094	71.7791	62.3362
7	72.5042	61.7094	71.8349	61.8803
8	72.1138	61.4245	71.2214	61.7664
9	71.5561	62.3932	71.2214	62.2222
10	71.5561	61.9373	71.1099	61.8803

5. Conclusion

The k-nearest neighbour (K-NN) technique has been recognized as one of the straightforward but efficient ways to find out what human beings are doing through data from phones equipped with sensors. Essentially, K-NN with the help of accelerometer and gyroscope data can differentiate a human's activity like whether he is walking, running, sitting, or standing by referring to the closest match of the data samples. The outcomes demonstrate that K-NN yields high precision in activity recognition if the parameter K-value is chosen properly. Moreover, K-NN is a lightweight algorithm that could be easily implemented in real-time on devices with limited computing power such as smartphones.

It is very practical and cheap to use a smartphone as an activity sensor because the phone already has sensors that can be used for this purpose, so there is no need to buy extra devices. An overall accuracy of 79.56% was achieved, with variations depending on the experiment duration, which also affects computation time. This study confirms that K-NN has a great potential to be the efficient activity classification method using a smartphone.

However, K-NN performance can be affected by dataset size and number of features. Sensor data quality (noise, sampling rate, etc.), and Inappropriate K value selection. All in all, K-NN is a viable option for human activity recognition (HAR) on smartphones not only with respect to the effectiveness but also the efficiency standpoint, and hence the applications range of this technology could be extended in the domains of health, fitness, security, and human-computer interaction.

References

1. Nurhayati, S.; Wicaksono, M.; and Syahrul, D.E. (2024). Prediction system for vitamin a requirement in West Java Indonesia using linear regression. *Journal of Engineering Science and Technology*, 19(1), 55-64.
2. Arifien, Z.; Bachtiar, F.A.; and Yudistira, N. (2022). Pengenalan aktivitas manusia menggunakan sensor akselerometer dan giroskop pada smartphone dengan metode k-nearest neighbor. *Jurnal Teknologi Informasi dan Ilmu Komputer,* 9(1), 33 - 40.
3. Kaghyan, S.; and Sarukhanyan, H. (2012). Activity recognition using k-nearest neighbor algorithm on smartphone with tri-axial accelerometer. *International Journal of Informatics Models and Analysis,* 1(2), 146 - 156.
4. Nursikuwagus, A.; Purwanto, H.; and Dewi, D.A. (2023). Feature classification based on heterogenous data using hybrid machine learning: A review. *Journal of Theoretical and Applied Information Technology*, 101(17), 6991-7010.
5. Shdefat, A.Y.; Mostafa, N.; Al-Arnaout, Z.; Kotb, Y.; and Alabed, S. (2024). Optimizing HAR systems: Comparative analysis of enhanced SVM and k-NN classifiers. *International Journal of Computational Intelligence Systems*, 17(1), 150-181.
6. Sassi-Hidri, M.; Hidri, A.; Alsaif, S.A.; Alahmari, M.; and AlShehri, E. (2025). Enhancing sensor-based human physical activity recognition using deep neural networks. *Journal of Sensor and Actuator Networks*, 14(2), 42-63.
7. Rezaee, K. (2024). An advanced deep learning structure for accurate student activity recognition and health monitoring using smartphone accelerometer data. *Health Management and Information Science*, 11(2), 85-97.

8. Attoumi, H.; Achroufene, A.; Saifi, R.; Souici, L.; and Boukredera, D. (2025). Smartphone sensor fusion for joint recognition of human activity and behavioral context. *Computing*, 107(10), 1-20.
9. San-Segundo, R.; Montero, J.M.; Moreno-Pimentel, J.; and Pardo, J.M. (2016). HMM adaptation for improving a human activity recognition system. *Algorithms*, 9(3), 60-77.
10. Straczkiewicz, M.; James, P.; and Onnela, J.P. (2021). A systematic review of smartphone-based human activity recognition methods for health research. *NPJ Digital Medicine*, 4(1), 148-163.
11. Sakalauskas, L.; and Vaiciulyte, I. (2025). Adaptive learning approach for human activity recognition using data from smartphone sensors. *Applied Sciences,* 15(14), 7731-7742.
12. Challa, S.; Krishnan, S.; and Patel, R. (2022). Comparative evaluation of machine learning algorithms for smartphone-based activity recognition. *IEEE Access,* 10, 92214 - 92226.
13. Bhattacharya, S.; Ghosh, D.; and Roy, A. (2023). Feature engineering approaches for smartphone-based human activity recognition. *Sensors,* 23(4), 2114-2126.
14. Kim, J.; Lee, D.; and Park, J. (2021). Enhanced human activity recognition using smartphone sensors with hybrid time - frequency features. *Applied Sciences,* 11(9), 4051-4059.
15. Zainuddin, M.; and Rahman, N. (2020). Improving k-nn performance for smartphone-based activity recognition using feature normalization. *Journal of Information Systems,* 16(2), 115 - 125.
16. Rafdhi, A.A.; Soegoto, E.S.; Luckyardi, S.; and Albar, C.N. (2023). Tree algorithm model on size classification data mining. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(4), 895-903.
17. Garcia-Gonzalez, D.; Rivero, D.; Fernandez-Blanco, E.; and Luaces, M.R. (2020). A public domain dataset for real-life human activity recognition using smartphone sensors. *Sensors*, 20(8), 2200-2214.
18. Abdullah, M.H.; and Ahmed, M.A. (2024). Human activity recognition using accelerometer and gyroscope smartphone sensor by extract statistical features. *Journal of Robotics and Control,* 5(2), 100-108.