A Novel Deep Learning Model for Smartphone-Based Human Activity Recognition

Nadia Agti[1,2(✉)], Lyazid Sabri[1,3], Okba Kazar[4], and Abdelghani Chibani[3]

[1] Maths and Informatics Faculty, Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria
nadia.agti@univ-bba.dz
[2] LMSE, The Laboratory of Materials and Electronic Systems, Mohamed El Bachir El Ibrahimi University, Bordj Bou Arreridj, Algeria
[3] LISSI-The Laboratory of Images, Signals and Intelligent Systems, University Paris-Est Vitry-sur-Seine, Vitry-sur-Seine, Île-de-France, France
[4] University of Kalba, Sharjah, United Arab Emirates

Abstract. Traditional pattern recognition methods have shown improvements in recent years. However, these techniques relied on human intervention to identify crucial information within the data. Deep learning has revolutionized this scenario by enabling computers to learn from data autonomously. This capability is precious for comprehending how individuals interact with mobile and wearable technology. The growing popularity of this method stems from its capacity to function effectively with minimal or no human involvement. This study introduces a novel hybrid deep learning network that merges Convolutional Neural Network (CNN) and Multi-Layer Perceptron (MLP) layers. While maintaining the ability to perform accurate activity identification, this CNN-MLP approach captures the temporal features from sensors, facilitating, therefore, multi-class classification. We also explore various machine learning models, such as Random Forests (RF), K-Nearest Neighbors (KNN), Logistic Regression (LR), Long Short-Term Memory (LSTM), and CNN-LSTM, on two well-established datasets: the WISDM and UCI HAR datasets. Through extensive testing on these datasets, we showcase the superior accuracy of our proposed CNN-MLP model compared to other competing machine learning and deep learning models. Our research opens up new possibilities for precise and effective human activity recognition on smartphones.

Keywords: Human activity recognition · convolution neural network · multiLayer perceptron · smartphone · spatial-temporal knowledge

1 Introduction and Related Work

Human activity recognition (HAR) is a critical task in various domains, such as health [16], behavior analysis [20], transportation [6], and security [14].

A. Zaslavsky et al. (Eds.): MobiQuitous 2023, LNICST 594, pp. 231–243, 2024.
https://doi.org/10.1007/978-3-031-63992-0_15

Improving human-machine interactions and raising quality of life are significantly impacted by the capacity to identify and comprehend human behaviors automatically. Since they can develop hierarchical representations from unprocessed sensor readings, deep learning models have demonstrated impressive performance in tackling HAR issues. One of the most significant drawbacks of conventional methods employed in human activity recognition is their dependence on manual feature extraction. This approach demands substantial human labor and introduces subjectivity, potentially leading to the omission of crucial details. Moreover, these methods often need help to generalize effectively across diverse contexts, limiting their practical applicability.

Smartphone sensor-based human activity recognition harnesses the capabilities of diverse built-in sensors within cell phones to capture various facets of human activity, such as movements and gestures. Many researchers use accelerometers [5,8] to record acceleration across multiple directions. Gyroscopes [3] assess angular velocity, aiding in the detection of rotation and orientation changes. Magnetometers [7] gauge magnetic field intensity, facilitating orientation determination. Barometers [4] measure atmospheric pressure and enable the identification of elevation changes, such as ascending or descending stairs or navigating uphill. The Global Positioning System (GPS) [17] tracks outdoor activities and locations.

Numerous research endeavors have been conducted to explore diverse deep-learning approaches for HAR. Wan et al. [21] focused on comparing five algorithms for human behavior recognition: CNN, LSTM, BLSTM, MLP, and SVM. They underscored CNN's sustained significance as a premier classification and recognition system by presenting promising outcomes on two datasets. In a separate study, Moradi et al. [11] undertook a comprehensive evaluation of multiple algorithms for HAR, encompassing CNN, LSTM, Bidirectional LSTM, CNN-LSTM, as well as Logistic Regression and SVM with RBF kernel. Among these, CNN-LSTM and Bidirectional LSTM emerged as the most accurate techniques, while SVM with RBF kernel exhibited superior training efficiency. A balanced compromise between accuracy and training duration was identified in LSTM. Oluwalade et al. [15] used artificial neural networks (LSTM, Bi-LSTM, CNN, and Convolutional LSTM) to classify accelerometer and gyroscope data from smartphones and smartwatches, achieving over 91% average accuracy across 15 diverse activities, signifying substantial generalization improvements. Addressing real-time applications, Mutegeki et al. [12] introduced a CNN-LSTM approach for real-time Human Activity Recognition (HAR). Their integration of CNN and LSTM models resulted in superior performance compared to alternatives, with a notable accuracy increase of over 1% and a nearly 2% reduction in Softmax loss, demonstrating significant computational efficiency improvements.

To cater to the demand for wearable sensor-based HAR, Xu et al. [23] introduced the InnoHAR model, featuring max-pooling layers and concatenated convolution kernels. Their research consistently outperformed competitors across three datasets, highlighting its potential for real-time applications. In their paper, Xia et al. [22] proposed a unique LSTM-CNN model, combining convolu-

tional layers with LSTM. It achieved robust F1 scores on UC-HAR, WISDM, and OPPORTUNITY datasets, demonstrating superior performance and generalization. Mekruksavanich et al. [10] introduced a 4-layer CNN-LSTM network for HAR, surpassing basic LSTM networks with remarkable accuracy. Their study highlights the potential of CNN-LSTM models in human behavior identification tasks.

Our study represents a substantial contribution to the field of deep learning-based human activity identification. We harness the potential of deep learning to autonomously decipher how individuals interact with mobile and wearable devices, eliminating the need for human intervention in data pattern analysis. Our hybrid model, skillfully amalgamating Convolutional Neural Network (CNN) and Multi-Layer Perceptron (MLP) architectures, not only elevates activity identification accuracy but also establishes a new benchmark for hybrid models in this domain. By outperforming well-established machine learning and deep learning models on widely recognized datasets like WISDM and UCI HAR, our work underscores the ongoing progress in deep learning-based activity identification models.

The rest of this paper is organized as follows: Sect. 2 introduces our methodology, including dataset details, data preprocessing, and our proposed model. Section 3 presents experimental results, and Sect. 4 concludes the paper and outlines future work.

2 Methodology

In this section, we introduce the methodological framework utilized in this study, covering data collection and preprocessing, as well as our proposed architecture.

2.1 Data Collection and Preprocessing

We utilize in our study, two widely recognized and openly accessible datasets for sensor-based human activity recognition: the UCI HAR dataset [1] and WISDM dataset [2] for human activity recognition using smartphones.

Table 1 provides a comparison of the two datasets in terms of instances, attributes, subjects, activities, characteristics, tasks, sampling rate, sensors used, challenges, and file type.

We applied the data preparation principle to the WISDM dataset, such as label encoding, linear interpolation, normalization, segmentation, and one-hot encoding data preparation techniques. We incorporate encoded labels from the "activity" column into an "activity label" column. The labels are converted into numerical equivalents as detailed in Table 2.

We employ linear interpolation with missing data points (NaN). This technique fills in gaps by estimating values between existing data points, even though our dataset has few NaN values. Then, we normalize the characteristics of the training data to a range of 0 to 1 using the normalization equation:

$$Y^{(i)}_{\text{normalized}} = \frac{Y^{(i)} - Y^{(i)}_{\text{min}}}{Y^{(i)}_{\text{max}} - Y^{(i)}_{\text{min}}} \quad \text{for} \quad i = 1, 2, \ldots, n \tag{1}$$

Table 1. Comparison between UCI HAR Dataset and WISDM Dataset [13]

Factors	Datasets	
	UCI HAR Dataset	WISDM Dataset
# examples	10299	1098209
# attributes	561	6
# subjects	30	36
# activity types	6	6
Characteristics	Multivariate/Time series	Multivariate/Time series
Tasks	Classification/Clustering	Classification
Sampling Rate	50 Hz	20 Hz
Sensors types	Accelerometer, Gyroscope	Accelerometer
Challenges	Multimodal	Class Imbalance
File Type	csv	txt

Table 2. Activity labels for WISDM datset

Activity	Label
Downstairs	0
Jogging	1
Sitting	2
Standing	3
Upstairs	4
Walking	5

where n represents the number of channels, and $Y_{\max}$ and $Y_{\min}$ correspond to the maximum and minimum values of the i-th channel, respectively.

We also reshape the data frame; each record's length is equal to 80 steps. We define a function called *segments* that takes the data frame, label names, and three input parameters. The *x_train* and *y_train* functions then segregate features and labels into separate sets. We employ the *reshape* function to convert the two-dimensional data into a list, preparing it for model input.

Additionally, all elements are converted to the *float32* format to ensure compatibility with the model's requirements.

One-hot encoding enhances machine learning algorithms by transforming categorical variables into a suitable format. Each category is represented by a binary vector, where only one element is set to 1 (hot), and the rest are 0 (cold). The length of the binary vector matches the number of distinct categories. The labels undergo one-hot encoding and are stored as *y_train_hot*, concluding the data preparation process.

The UCI HAR dataset is also pre-processed. We start by looking for duplicates and empty fields. There are no duplicates or null values in the data collection. After receiving the train data from X_train.txt and subject_train.txt, and the test data from X_test.txt and subject_test.txt, respectively, we save the train and test data in two CSV files. Feature names are then stripped of any extra chanting.

The two datsets are split, with 80% for training, and 20% for testing the model.

2.2 Proposed Model Architecture

The architecture of our CNN-MLP model is designed to effectively capture temporal attributes from the input data, facilitating multi-class classification. The model consists of distinct layers, each serving a specific feature extraction and classification role. This structure is well-suited for tasks requiring concurrent sequential data processing, such as time-series sensor data. The complete structure of our approach is illustrated in Fig. 1, which visually portrays its crucial components.

Our CNN-MLP model, designed for smartphone-based human activity recognition, utilizes a series of three 1D Convolutional layers to efficiently extract temporal features. In the realm of time series analysis, 1D convolutions are well-known for identifying intricate temporal patterns. Within our model, these 1D convolutions serve a central role, progressively unveiling patterns within data segments and constructing a hierarchical representation that mirrors task complexity. The identified temporal features are then seamlessly fed into the fully connected Multi-Layer Perceptron (MLP) layers for classification.

Our MLP layers, consisting of two dense tiers enriched with Rectified Linear Unit (ReLU) activation functions, meticulously analyze dynamic temporal patterns within sensor data. This process enables the recognition of intricate interdependencies across successive time steps. We've incorporated Global Average Pooling (GAP) and Batch Normalization. GAP consolidates temporal dynamics, enhancing generalization and diminishing sensitivity to minor sensor fluctuations. Meanwhile, Batch Normalization ensures consistent activations within mini-batches, addressing gradient challenges during training.

The output layer, activated with softmax, predicts human behavior based on the extracted temporal features. Each neuron within this densely connected output layer represents a distinct activity class, generating a probability distribution signifying the likelihood of each activity.

To facilitate training convergence and optimization, our model implements learning rate scheduling techniques, adjusting the learning rate during training epochs. This dynamic adjustment enhances the model's efficiency and promotes effective convergence. The choice of the Adam optimizer and categorical cross-entropy loss function is aimed at multi-class classification, ensuring the model minimizes the loss while assessing performance through accuracy.

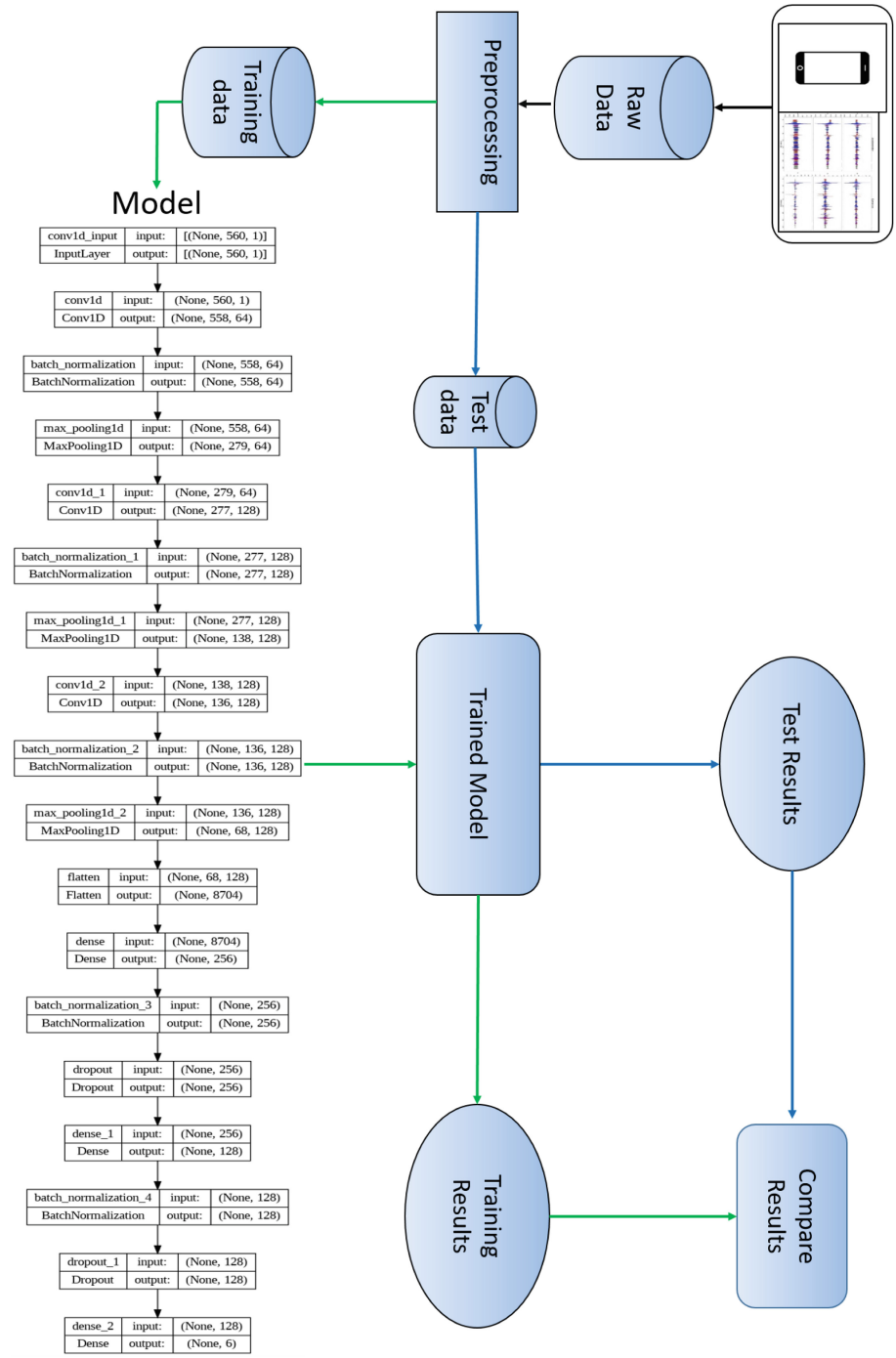

Fig. 1. Architecture of the proposed approach

3 Experimental Results and Discussion

This section showcases test outcomes and engages in insightful discussions. We commence by elucidating the hardware and software configurations employed to construct our model. Subsequently, we introduce the measurement metrics harnessed for evaluating our model's performance. Proceeding from there, we provide comprehensive explanations and discussions of the outcomes stemming from our two conducted experiments, and finally, we present a comparison of our model with state of the art machine learning models. The Code for our experiment is available online[1].

3.1 Model Implementation

We leveraged Python's Keras, a versatile neural network API, for architectural development. Keras seamlessly integrates Tensorflow or Theano as backends, and we deliberately chose Tensorflow for its capacity to accelerate computations and utilize GPU resources, greatly enhancing computational efficiency.

Our hardware setup was a robust server hosted on Google Colab, featuring a single-core 2.20 GHz Intel(R) Xeon(R) CPU with 16 GB of RAM. This powerful environment was instrumental in our experiments and analysis.

3.2 Evaluation Metrics

The confusion matrix and classification reports serve as the assessment tools within this study. The confusion matrix comprehensively evaluates the model's performance for each class by comparing predicted and actual values. Its components include True Positive, True Negative, False Positive, and False Negative. Table 3 depicts a generic representation of a confusion matrix's architecture.

Metrics such as precision (Eq. 2), recall (Eq. 3), F1 score (Eq. 4), and accuracy (Eq. 5) are featured in the classification report.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

Table 3. Simplified binary classification confusion matrix

	Positive Class	Negative Class
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

[1] https://github.com/NadiaAGTI/New-CNN-MLP-model-for-Human-Activity-Recognition.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

3.3 Results and Discussion

We tested our model using four epoch values: 10, 50, 100, and 200 for both datasets. To ensure fairness and consistency in the subsequent comparison, we validated the outcomes from previous steps using the Accuracy metric.
The primary reason for the efficacy of our model lies in its thoughtful architecture. To achieve these remarkable results, the combination of a Convolutional Neural Network (CNN) and a Multi-Layer Perceptron (MLP) has proven to be essential. The CNN's ability to accurately identify human actions critically depends on its capacity to detect temporal patterns in sequential data, particularly time series data. This architecture excels in identifying intricate patterns, which is a fundamental requirement for robust activity recognition.

Furthermore, the excellent results are, in large part, due to careful data preparation. Methods like batch normalization and data normalization have been applied with care. These processes enhance the quality of the input data, reduce noise, and ensure consistency, all of which are essential for the model to successfully identify human activity.

The robust performance of our model is also a result of careful hyperparameter optimization. Parameters such as the number of filters, kernel sizes, activation functions, and dropout rates, among other factors, have all been finely tuned. Additionally, the inclusion of L2 regularization has played a crucial role in preventing overfitting and ensuring the model's effective generalization to new data.

Another essential element of the design involves the incorporation of batch normalization layers. They expedite both the model's convergence and training stability. Consequently, the training process becomes more resilient and efficient, and it is less dependent on parameter initialization.

Pooling layers, especially the max-pooling within the CNN component, effectively reduce the spatial dimensions of the data while retaining crucial features. This is highly beneficial when dealing with time series data, as it allows the model to focus on the most significant information. Finally, overfitting has been effectively mitigated through the cautious implementation of dropout layers within the MLP component. This technique randomly deactivates neurons during training, encouraging the model to capture more robust and transferable features, ultimately enhancing its overall performance.

Collectively, these design principles and components have yielded a model that consistently delivers outstanding performance on the datasets WISDM and UCI HAR. It achieves high recall, accuracy, F1-scores, precision, and overall

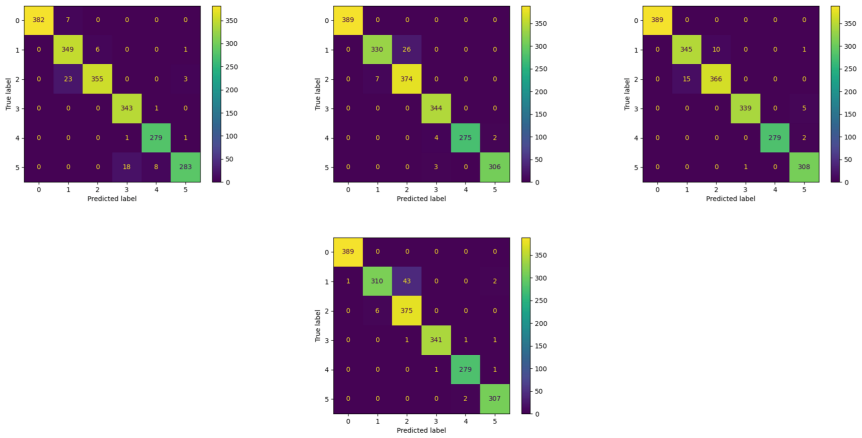

Fig. 2. Confusion matrix for the proposed CNN-MLP model with different epoch values with UCI dataset

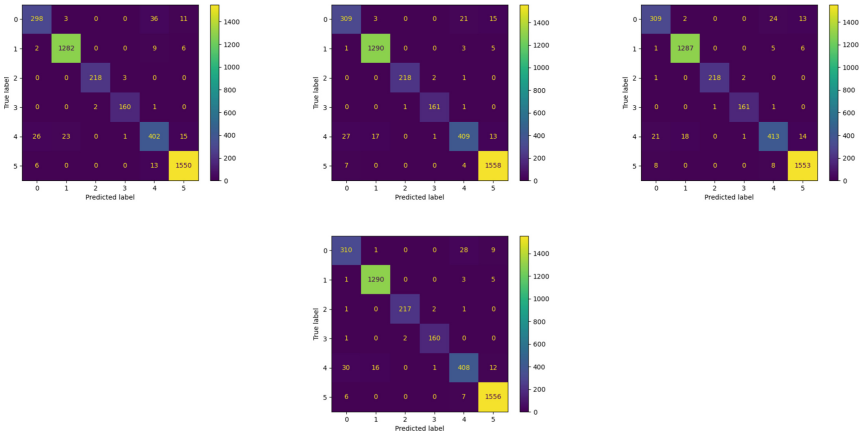

Fig. 3. Confusion matrix for the proposed CNN-MLP model with different epoch values with WISDM dataset

accuracy across various activity classes. Consequently, our model offers a versatile and adaptable solution for human activity recognition, with potential applications in diverse industries, including wearable technology and healthcare.

Figures 2 and 3 illustrates the confusion matrix of the model for four distinct epoch values using the UCI HAR dataset and WISDM dataset respectively.

The outcomes of the model with both datasets are presented in Table 4.

Table 4. The proposed CNN-MLP model results using different epochs values

Dataset	Metrics(%)				
	Epochs	Precision	Recall	F-Score	Accuracy
WISDM Dataset	10	96.09	96.14	96.11	96.14
	50	96.96	97.00	96.97	97.00
	100	96.86	96.90	96.88	96.90
	200	96.87	96.90	96.88	96.90
UCI HAR Dataset	10	89.81	85.92	85.37	85.92
	50	95.99	95.63	95.61	97.96
	100	97.24	97.18	97.18	97.18
	200	97.29	97.14	97.12	97.14

3.4 Comparison with Baselines and State-of-the-art

As illustrated in Table 5, the performance comparison results of the study provide compelling insights into the effectiveness of our proposed CNN-MLP model for human activity recognition. The comparison encompasses a range of established baseline models, including Random Forests, Logistic Regression, K-Nearest Neighbors, Multi-Layer Perceptron, Long Short-Term Memory, and CNN-LSTM, conducted on both the WISDM and UCI HAR datasets.
Our proposed CNN-MLP model demonstrated superior performance to all other baseline models, achieving an impressive accuracy of 96.9% on the WISDM dataset. This achievement stands out notably in contrast to RF, which attained an accuracy of 38.52%, highlighting the substantial distinction between conventional techniques and modern deep learning methodologies. Furthermore, the suggested model significantly outperformed both LR (90%) and KNN (75.04%), underscoring its ability to discern intricate temporal patterns within the sensor data.

Table 5. Comparison of our model with baseline models on both WISDM and UCI datasets(Accuracy in %)

Methods	WISDM dataset	UCI HAR dataset
RF [19]	38.52	86
LR [11,24]	90	96.1
KNN [18,24]	75.04	89.1
MLP [9,21]	91.7	86.83
LSTM [12,19]	87.7	91.28
CNN-LSTM [12,19]	86.3	92.13
Proposed model	**96.9**	**97.14**

Turning our attention to the UCI HAR dataset, a similar trend emerges. Our CNN-MLP model continues to outshine all other methods, boasting an exceptional accuracy of 97.14%. KNN achieved an accuracy of 89.1%, and LR reached 96.1%. Impressively, our model surpassed LSTM (91.28%), and CNN-LSTM (92.13%), which inherently consider sequential patterns. This notable advantage underscores our proposed architecture's effectiveness in effectively capturing local and global information.

4 Conclusion

The distinctive CNN-MLP model proposed in this research is tailored for smartphone sensor-based human activity recognition. This innovative approach integrates convolutional neural network (CNN) layers and multi-layer perceptron (MLP) layers, skillfully leveraging different convolutional kernel sizes and incorporating well-considered max-pooling layers. Our suggested approach consistently exhibits impressive performance advantages in comprehensive comparisons against baseline models and state-of-the-art techniques. Moreover, its robust generalization capabilities are demonstrated across two widely recognized public datasets, further affirming its potential for real-world applications.

In our forthcoming research, we aim to enhance the compatibility of our system with mobile contexts. This endeavor will involve exploring hardware-efficient implementations and robust cloud access mechanisms to ensure seamless usability on mobile devices.

References

1. Human activity recognition using smartphones dataset. https://archive.ics.uci.edu/dataset/240/human+activity+recognition+using+smartphones. Accessed 05 July 2023
2. Wisdm smartphone and smartwatch activity and biometrics dataset. https://www.cis.fordham.edu/wisdm/dataset.php. Accessed 05 July 2023
3. Biswas, D., et al.: Recognizing upper limb movements with wrist worn inertial sensors using k-means clustering classification. Hum. Mov. Sci. **40**, 59–76 (2015)
4. Chernbumroong, S., Cang, S., Atkins, A., Yu, H.: Elderly activities recognition and classification for applications in assisted living. Expert Syst. Appl. **40**(5), 1662–1674 (2013)
5. Chernbumroong, S., Cang, S., Yu, H.: A practical multi-sensor activity recognition system for home-based care. decision support systems **66**, 61–70 (2014)
6. Choujaa, D., Dulay, N.: Activity recognition from mobile phone data: state of the art, prospects and open problems. Imp. Coll. Lond. **5**, 32 (2009)

7. Gjoreski, H., Gams, M.: Activity/posture recognition using wearable sensors placed on different body locations. Proc. (738) Signal Image Process. Appl. Crete Greece **2224**, 716724 (2011)
8. Guo, J., Zhou, X., Sun, Y., Ping, G., Zhao, G., Li, Z.: Smartphone-based patients' activity recognition by using a self-learning scheme for medical monitoring. J. Med. Syst. **40**, 1–14 (2016)
9. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. ACM SIGKDD Explor. Newsl. **12**(2), 74–82 (2011)
10. Mekruksavanich, S., Jitpattanakul, A.: LSTM networks using smartphone data for sensor-based human activity recognition in smart homes. Sensors **21**(5), 1636 (2021)
11. Moradi, B., Aghapour, M., Shirbandi, A.: Compare of machine learning and deep learning approaches for human activity recognition. In: 2022 30th International Conference on Electrical Engineering (ICEE), pp. 592–596. IEEE (2022)
12. Mutegeki, R., Han, D.S.: A CNN-LSTM approach to human activity recognition. In: 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 362–366. IEEE (2020)
13. Nayak, S., Panigrahi, C.R., Pati, B., Nanda, S., Hsieh, M.Y.: Comparative analysis of HAR datasets using classification algorithms. Comput. Sci. Inf. Syst. **19**(1), 47–63 (2022)
14. Niu, W., Long, J., Han, D., Wang, Y.F.: Human activity detection and recognition for video surveillance. In: 2004 IEEE International Conference on Multimedia and Expo (ICME) (IEEE Cat. No. 04TH8763), vol. 1, pp. 719–722. IEEE (2004)
15. Oluwalade, B., Neela, S., Wawira, J., Adejumo, T., Purkayastha, S.: Human activity recognition using deep learning models on smartphones and smartwatches sensor data. arXiv preprint arXiv:2103.03836 (2021)
16. Preuveneers, D., Berbers, Y.: Mobile phones assisting with health self-care: a diabetes case study. In: Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 177–186 (2008)
17. Reddy, S., Mun, M., Burke, J., Estrin, D., Hansen, M., Srivastava, M.: Using mobile phones to determine transportation modes. ACM Trans. Sens. Netw. (TOSN) **6**(2), 1–27 (2010)
18. Semwal, V.B., Lalwani, P., Mishra, M.K., Bijalwan, V., Chadha, J.S.: An optimized feature selection using bio-geography optimization technique for human walking activities recognition. Computing **103**(12), 2893–2914 (2021)
19. Trabelsi, I., Françoise, J., Bellik, Y.: Sensor-based activity recognition using deep learning: a comparative study. In: Proceedings of the 8th International Conference on Movement and Computing, pp. 1–8 (2022)
20. Vepakomma, P., De, D., Das, S.K., Bhansali, S.: A-Wristocracy: deep learning on wrist-worn sensing for recognition of user complex activities. In: 2015 IEEE 12th International conference on wearable and implantable body sensor networks (BSN), pp. 1–6. IEEE (2015)
21. Wan, S., Qi, L., Xu, X., Tong, C., Gu, Z.: Deep learning models for real-time human activity recognition with smartphones. Mob. Netw. Appl. **25**, 743–755 (2020)

22. Xia, K., Huang, J., Wang, H.: LSTM-CNN architecture for human activity recognition. IEEE Access **8**, 56855–56866 (2020)
23. Xu, C., Chai, D., He, J., Zhang, X., Duan, S.: InnoHAR: a deep neural network for complex human activity recognition. IEEE Access **7**, 9893–9902 (2019)
24. Zaki, Z., Shah, M.A., Wakil, K., Sher, F.: Logistic regression based human activities recognition. J. Mech. Contin. Math. Sci. **15** (2020)