Enhancing Healthcare Monitoring: A Deep Learning Approach to Human Activity Recognition using Wearable Sensors

Sami Aziz Alshammari
Information Technology Department, Faculty of Computing and Information Technology, Northern Border University Rafha, Saudi Arabia
Sami.Alshammari@nbu.edu.sa (corresponding author)

Nasser S. Albalawi
Department of Computer Sciences, Faculty of Computing and Information Technology, Northern Border University, Rafha, Saudi Arabia
nasser.albalawi@nbu.edu.sa

Received: 13 October 2024 | Revised: 29 October 2024 and 2 November 202X | Accepted: 3 November 202

 DOI: https://doi.org/10.48084/etasr.9255

ABSTRACT

Wearable devices and deep learning methods for Human Activity Recognition (HAR) have attracted a lot of interest because they could change healthcare monitoring. This study presents a CNN-LSTM model to accurately and reliably detect human movements from smartphone sensor data. The proposed model takes advantage of both the strengths of Long Short-Term Memory (LSTM) networks for modeling time and Convolutional Neural Networks (CNNs) for extracting features from space. This enables determining how the input data change over time and space. This study examines whether this method can work and is practical in real-life healthcare settings, focused on uses such as watching patients from distance, caring for the elderly, and therapy. The proposed model was evaluated on publicly accessible standard datasets. Various architectural configurations and hyperparameters were examined to determine their performance. The proposed CNN-LSTM model performed well and has great potential for practical use in activity tracking and environment understanding systems.

Keywords-deep learning; healthcare monitoring; smartphone sensor data; convolutional neural network (CNN); human activity recognition (HAR); long short term memory (LSTM)

I. INTRODUCTION

Recently developed Human Activity Recognition (HAR) technology has emerged as a game changer in healthcare by automatically and accurately monitoring patients' physical activities [1, 2]. Wearable sensors, advanced machine learning algorithms, and more people with chronic diseases and older populations make HAR a more important part of modern healthcare services [3, 4]. The goal of HAR systems is to collect, process, and make sense of data from different sensors, such as accelerometers and gyroscopes built into smartphones or wearable technology, to identify and group different human activities. Walking, running, sitting, standing, and sleeping are just a few of the actions and habits included in this list. The fact that HAR systems can correctly find and group these tasks into the right categories shows great potential to improve healthcare in many areas. What makes HAR important in healthcare tracking is that it could help solve some of the biggest problems that both healthcare workers and patients face. HAR technology has changed many things, such as remote patient monitoring. It allows doctors and nurses to see what their patients are doing and check their vital signs in real time, so they can help them sooner and prevent them from having to go to the hospital as often. HAR also has great potential in eldercare, where it can be very helpful in keeping an eye on seniors' Activities of Daily Living (ADL), providing important information about their health with more freedom and the ability to move. As shown in [3], noticing changes in normal behavior patterns early can help identify possible health problems and make it easier to take precautions to reduce risks. HAR is also used in rehabilitation settings, where it helps therapists keep track of their patient's progress during physical therapy sessions, ensure that approaches are tailored to each patient's needs, and get the best results from treatment.

Wearable sensor technologies and HAR algorithms have come together to provide new solutions. These solutions give people more control over their health and make their daily lives less stressful. Despite the significant advances in HAR technology, there are still several issues that need resolution. Concerns about data security, algorithm dependability, and user approval are among them. HAR can be useful in healthcare tracking. Patients, doctors, and hospitals can reap its benefits, reducing healthcare spending and improving clinical results. Feature construction from raw sensor data and conventional machine-learning algorithms was the backbone of traditional HAR systems [5]. These methods may work in some situations, but they are not always capable of capturing the complex spatial and temporal dynamics of human behavior. These systems aren't practical since they cannot be changed or expanded because they depend on fixed qualities. In recent years, the HAR sector has been backed up by deep learning approaches, especially CNNs, RNNs, and LSTM networks. CNNs are great at extracting spatial features from sensor data because they can conduct convolutions across several input signal dimensions [6]. However, when it comes to sequential data, LSTM networks excel at finding long-range correlations and temporal connections [7]. An efficient HAR technique is to combine CNN with LSTM networks combining the best features of both designs. These hybrid models can provide a hitherto unseen degree of granularity in their depictions of human actions by combining spatial and temporal data. As a consequence, the classification results can be more reliable and precise.

This study uses mobile phone sensor data and presents a novel CNN-LSTM method for human activity detection. The goal is to provide reliable activity classification by integrating the capabilities of CNNs for feature extraction and LSTMs for connection modeling in space and time. The proposed model was compared to current methods to evaluate its accuracy, efficiency, and generalizability using the UCI HAR dataset [8, 9]. Furthermore, extensive experiments were carried out to determine the effects of different architectural configurations and hyperparameters on the model's performance, helping to illuminate the optimization and design of HAR systems [10-12]. A brief overview of this study's main points includes:

- Introducing a novel CNN-LSTM model designed for HAR using smartphone sensor data [13, 14].
- Leveraging CNNs to extract spatial features and employing LSTMs for temporal modeling to achieve precise activity recognition [15, 16].
- Analyzing different architectural configurations and hyperparameters to optimize the performance of the CNN-LSTM model [17, 18].
- Highlighting potential applications of the CNN-LSTM model in the smart healthcare system, including remote patient monitoring, eldercare, rehabilitation, and activity monitoring [19].
- Carrying out an extensive evaluation of the proposed CNN-LSTM model on the UCI HAR dataset, with performance comparison against existing approaches.

II. HUMAN ACTIVITY RECOGNITION BASED ON A DEEP LEARNING APPROACH

The proposed deep learning-based HAR system uses wearable sensors. Due to their robust nature and ability to extract pertinent features from sequential and spatial input, CNN and LSTM work very well on HAR issues [20]. The proposed method combines CNN and LSTM to provide accurate and robust HAR using data from smartphone sensors [6, 21]. CNNs process raw data collected from sensors in smartphones and wearables. Using convolutional filters on the input data, a CNN can independently construct hierarchical feature representations that capture patterns in the sensor readings for both space and time. Contrary to popular belief, CNNs do not need human feature engineering to accurately identify complicated behaviors in unprocessed sensor data.

As an ideal fit for collecting temporal changes in time-series data, LSTMs differ from typical RNNs in that they include gated mechanisms that enable them to selectively retain or discard information over time. Processing successive sensor input, such as data received over time from a gyroscope or accelerometer, is a common use of LSTM networks in HAR. To correctly recognize and categorize intricate patterns of behavior, LSTM can imitate the sequential structure of human actions by storing temporal correlations. Classifying ADLs, identifying falls, and recognizing gestures are all possible with the use of LSTM networks applied to time-series sensor data.

A. Overall Framework for HAR Task

Figure 1 shows the proposed HAR task structure using CNN-LSTM models. Accelerometers and gyroscopes, among other wearable sensors, can be used to track individuals while performing a range of physical activities. To ensure consistency and eliminate noise and outliers, preprocessing is performed on raw sensor data before training the model, including data cleaning, normalization, and feature extraction. To enable model training on labeled activity sequences, data segmentation methods were used to separate continuous sensor streams into discrete activity segments. The combination of CNNs with LSTM networks was important in the development of the HAR deep learning model. CNN layers improved the model's robustness and accuracy in activity detection by extracting spatial information from sensor data, while LSTM layers recorded temporal correlations in successive activity sequences.

The model was trained using stochastic gradient descent with specific values of learning rate and batch size. To avoid overfitting, dropout regularization was also used. Early stopping criteria were used to end the training when the model reached a plateau in its validation set performance. To make the model more applicable to different scenarios, data augmentation methods were used, including adding random noise and modifying signals in the training data. Confusion matrices and classification metrics, including recall, accuracy, precision, and F1-score, were used to evaluate the HAR model's ability to discriminate between activity classes on an independent test set. When evaluating the model's readiness for deployment, qualitative criteria such as inference time and energy efficiency were also considered.

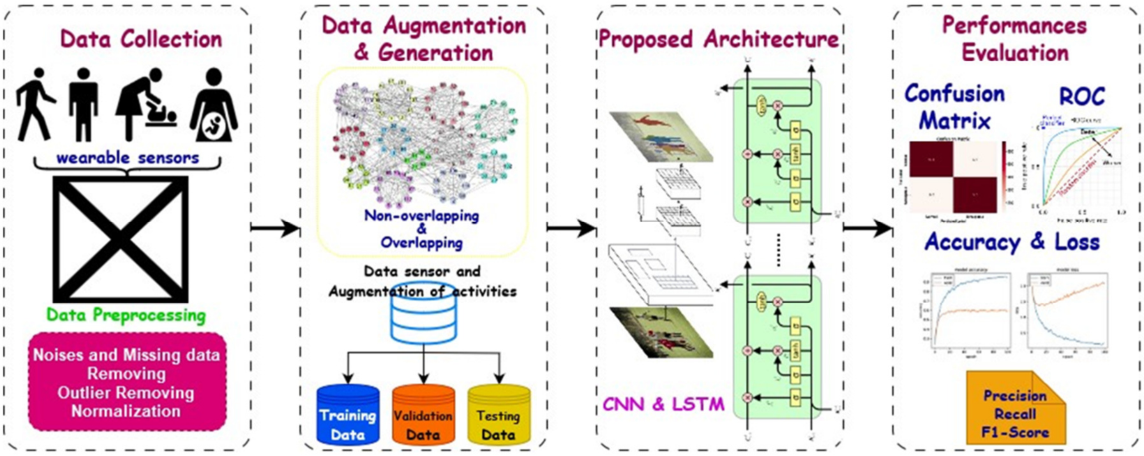

Fig. 1. Proposed framework for HAR.

B. Combined Networks for HAR: CNN-LSTM

Figure 2 shows a CNN-LSTM architecture that includes an input layer, two dense layers, two LSTM layers, two pooling layers, a flattened layer, two dropout layers, two 1-D convolutional layers layered on top of each other, and finally an output layer. Before using the ReLU activation function, a 1D convolutional operation with 128 filters and a kernel size of 4 is performed. The temporal dimension is preserved while analyzing incoming sequences of sensor data. The feature sequences are then compressed by max pooling along the time dimension with a pool size of 2. To avoid overfitting, a dropout layer is used, and the convolutional and pooling layers are repeated to collect deeper spatial characteristics. Next, the output from the convolutional layers is flattened, preparing it for input into the LSTM layers. The LSTM layers, each consisting of 100 units, capture temporal dependencies in the data sequence, with dropout regularization introduced to prevent overfitting. Following the LSTM layers, a dense layer with 100 units and a ReLu activation function processes the extracted features. Finally, the output layer consists of six neurons, each dedicated to classifying one of the six human activities using a softmax activation function. The Adam optimization algorithm was employed to optimize the network, iterating over 100 epochs with a batch size of 32. Unlike traditional optimization techniques that use a single learning rate for all weights, Adam dynamically adjusts the learning rate for each weight parameter. These rates are adapted based on both the mean and uncentered variances, with moving averages for gradients and squared gradients calculated by the Adam optimizer. In this case, the parameters that control how fast these moving averages lose value are changed. To get even better results, a loss function was used to direct the model's parameters for learning toward minimizing loss. Given the availability of many output labels, categorical cross-entropy was selected for the loss function. If the output labels are given in integer format, they are transformed to categorical encoding, and then one-hot encoding is applied to them.

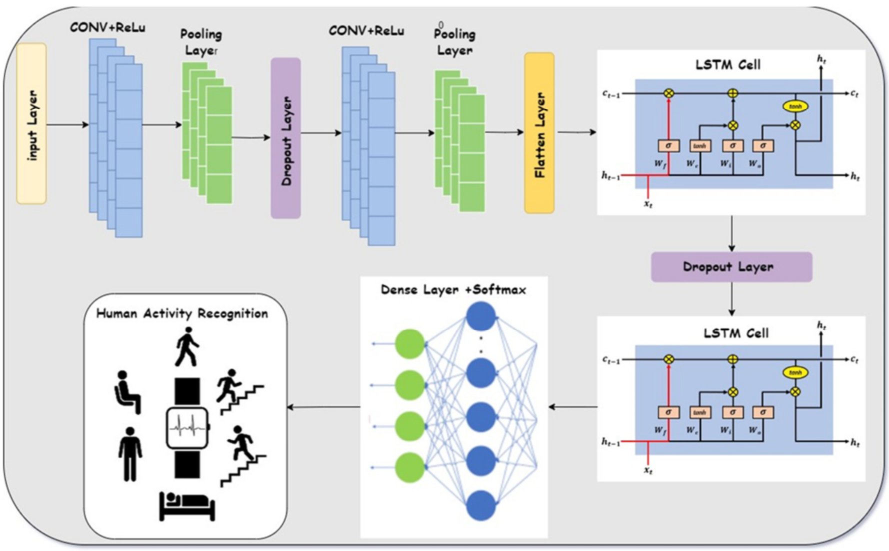

Fig. 2. Proposed combined networks for HAR.

III. EXPERIMENTAL RESULTS AND DISCUSSION

The UCI Human Activity Recognition (HAR) dataset consists of smartphone sensor data collected from a group of 30 volunteers performing six different activities while carrying a waist-mounted smartphone with embedded sensors [18, 22]. Activities include walking, walking upstairs, walking downstairs, sitting, standing, and laying. Sensor data was recorded using smartphone accelerometers and gyroscopes at a constant rate of 50 Hz. Each sample in the dataset comprises 561 features extracted from raw sensor data, which were preprocessed and normalized to have zero mean and unit standard deviation. In addition, each sample is labeled with the corresponding activity performed by the volunteer. The dataset was divided into two subsets: a training set (70%) and a test set (30%). It contains a total of 10,299 instances, with 7,352 instances used in the training set and 2,947 instances in the test set. Figure 3 illustrates various human activities along with their respective counts in both the training and testing datasets. The TSNE clustering visualization demonstrates distinct clusters for all activities, except for Standing and Sitting, which exhibit some overlap, as shown in Figure 4. This overlap suggests that these two activities may share similar features in the dataset, making their differentiation difficult. To improve the model's ability to differentiate between these activities, further research is needed to identify the precise characteristics that cause them to overlap.

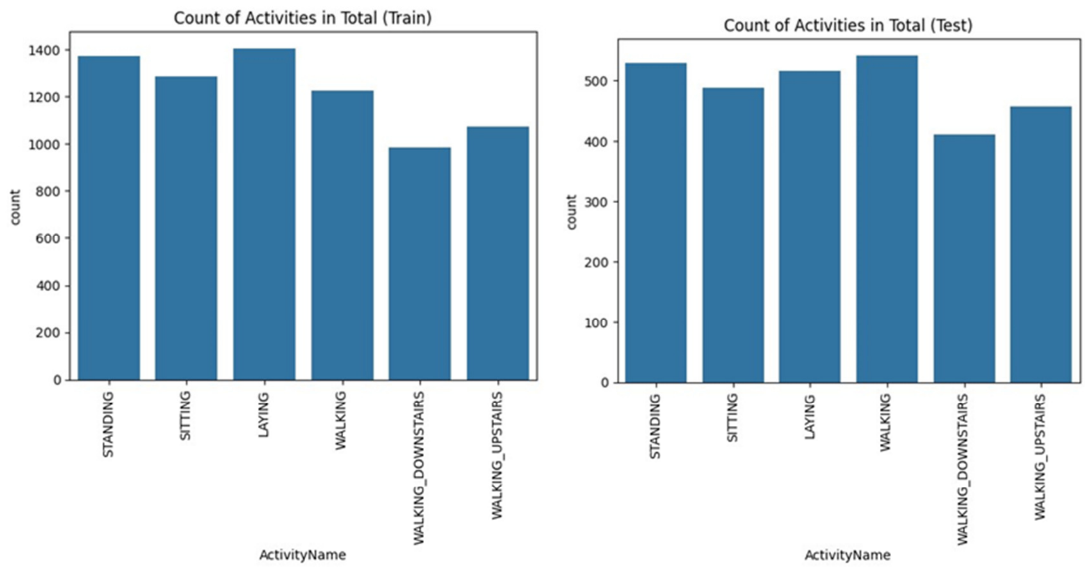

Fig. 3. Distribution of activities in both training and testing datasets.

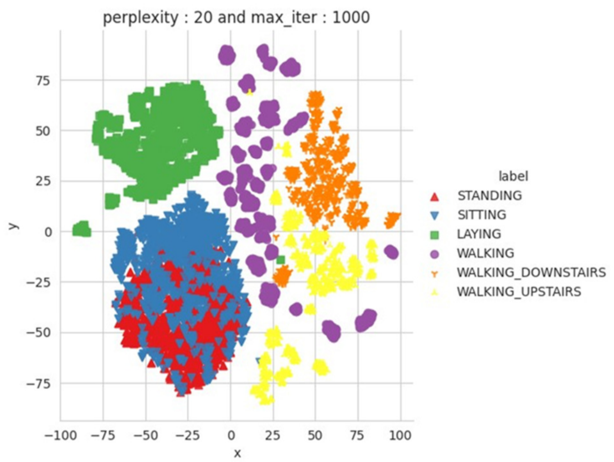

Fig. 4. TSNE visualization.

A. Evaluation Metrics

The accuracy, precision, recall, and F1-score metrics were used to evaluate the model. Accuracy shows the proportion of accurate predictions to the total predictions on the test dataset (1). Precision is defined as the ratio of the number of true predictions for a certain class activity to the total number of cases for that class in the testing dataset. Recall, also known as sensitivity, is the proportion of correctly classified positives for a given class relative to the total number of those class activities in the test dataset. The F1-score determines the harmonic mean of the recall and accuracy metrics.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1-score = 2 \times \frac{(Precision \times Recall)}{Precision+Recall} \quad (4)$$

B. Results and Discussion

Figure 5 illustrates the accuracy and loss curves on the training and validation sets, demonstrating the model's performance during the training process. These curves provide insights into how well the model is learning from the training data and generalizing to unseen validation data. The training and validation accuracies were 98.8% and 95.6%, respectively. The loss rate steadily decreased as the number of epochs increased. For the training set, the loss rate decreased to 0.12 after 100 epochs, while for the validation set, it started at 0.9 and decreased to 0.36. Table I presents the evaluation metrics for precision, recall, and F1 score of the proposed approach, assessing its performance in recognizing activities from the UCI HAR dataset. The classified activities demonstrate an average precision of 97%, an average recall of 96%, and an average F1 score of 96%.

Fig. 5. Accuracy and loss curves on training and test sets.

TABLE I. PERFORMANCE EVALUATION OF THE PROPOSED APPROACH

Activity	Precision	Recall	F1-score
WALKING	1.00	0.98	0.99
WALKING_UPSTAIRS	0.96	0.94	0.93
WALKING_DOWNSTAIRS	0.95	1.00	0.95
SITTING	0.94	0.93	0.91
STANDING	0.99	0.94	0.97
LAYING	1.00	0.95	1.00

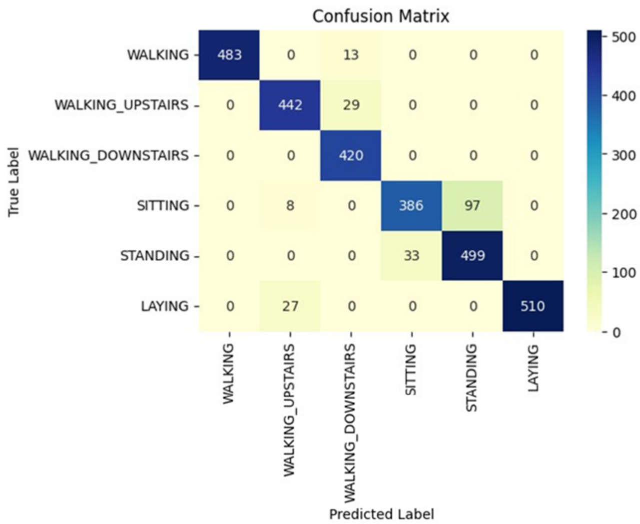

Fig. 6. Confusion matrix of CNN-LSTM model.

Figure 6 illustrates the confusion matrix generated by the CNN-LSTM network, which compares the predicted to the true labels in the test dataset. The diagonal values represent the accuracy of the classification, while values above and below the diagonal indicate classification errors. Specifically, the confusion matrix identifies 483, 442, 420, 386, 499, and 510 instances as true positives for the activities walking, walking upstairs, walking downstairs, sitting, standing, and laying, respectively.

The proposed CNN-LSTM model was compared with recent state-of-the-art methods in the field, using the same dataset. The approaches in [23-25] use a variety of architectures and methods to classify activities, as shown in Table II. The accuracy achieved in [23] using a CNN architecture on the UCI HAR dataset was 91.98%. This approach used CNN's spatial feature extraction capabilities to classify actions from sensor data. Using the same UCI HAR dataset, a better accuracy of 95.09% was achieved in [24] using an iSPLInception network, which is a variation of the Inception design to identify regional and worldwide trends in the collected sensor data using multiscale feature extraction. The AM-DLFC model in [25] achieved 94.25% accuracy, on the UCI HAR dataset.

TABLE II. COMPARISON WITH STATE-OF-THE-ART METHODS IN HAR

Approach	Network	Dataset	Accuracy (%)
[23]	CNN	UCI HAR	91.98
[24]	iSPLInception	UCI HAR	95.09
[25]	AM-DLFC	UCI HAR	94.25
Proposed	CNN-LSTM	UCI HAR	98.80

IV. CONCLUSION

This study investigated HAR technology, which is an essential element of contemporary healthcare as it monitors the physical activities of patients using machine learning methods and wearable sensors. A new CNN-LSTM model was developed for HAR with data collected from smartphone sensors. This model uses LSTM networks for temporal modeling and CNNs for spatial feature extraction, achieving an impressive 98.8% accuracy on the UCI HAR dataset. The results show that the proposed CNN-LSTM model outperformed the state-of-the-art approaches, highlighting its potential to completely restructure healthcare administration and monitoring. This method can improve healthcare delivery and patient outcomes by precisely identifying human activities, allowing proactive treatment and enhancing eldercare.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through project no. NBU-FFR-2024-2119-02.

REFERENCES

[1] L. Bibbò and M. M. B. R. Vellasco, "Human Activity Recognition (HAR) in Healthcare," *Applied Sciences*, vol. 13, no. 24, Jan. 2023, Art. no. 13009, https://doi.org/10.3390/app132413009.

[2] K. M. Gritsyuk and V. I. Gritsyuk, "Convolutional and Long Short-Term Memory Neural Networks Based Models for Remaining Useful Life Prediction," *International Journal on Information Technologies & Security*, vol. 14, no. 1, 2022.

[3] A. Nouriani, R. A. McGovern, and R. Rajamani, "Deep-Learning-Based Human Activity Recognition Using Wearable Sensors," *IFAC-PapersOnLine*, vol. 55, no. 37, pp. 1–6, Jan. 2022, https://doi.org/10.1016/j.ifacol.2022.11.152.

[4] M. Alanazi, R. S. Aldahr, and M. Ilyas, "Human Activity Recognition through Smartphone Inertial Sensors with ML Approach," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12780–12787, Feb. 2024, https://doi.org/10.48084/etasr.6586.

[5] L. Bao and S. S. Intille, "Activity Recognition from User-Annotated Acceleration Data," in *Pervasive Computing*, Vienna, Austria, 2004, pp. 1–17, https://doi.org/10.1007/978-3-540-24646-6_1.

[6] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Systems with Applications*, vol. 59, pp. 235–244, Oct. 2016, https://doi.org/10.1016/j.eswa.2016.04.032.

[7] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, May 2017, pp. 1578–1585, https://doi.org/10.1109/IJCNN.2017.7966039.

[8] K. Gusain, A. Gupta, and B. Popli, "Transition-Aware Human Activity Recognition Using eXtreme Gradient Boosted Decision Trees," in *Advanced Computing and Communication Technologies*, 2018, pp. 41–49, https://doi.org/10.1007/978-981-10-4603-2_5.

[9] I. N. Yulita and S. Saori, "Human Activities and Postural Transitions Classification using Support Vector Machine and K-Nearest Neighbor Methods," *IOP Conference Series: Earth and Environmental Science*, vol. 248, no. 1, Nov. 2019, Art. no. 012025, https://doi.org/10.1088/1755-1315/248/1/012025.

[10] C. L. Liu, W. H. Hsaio, and Y. C. Tu, "Time Series Classification With Multivariate Convolutional Neural Network," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 6, pp. 4788–4797, Jun. 2019, https://doi.org/10.1109/TIE.2018.2864702.

[11] J. Zhu, H. Chen, and W. Ye, "A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," *IEEE Access*, vol. 8, pp. 24713–24720, 2020, https://doi.org/10.1109/ACCESS.2020.2971064.

[12] J. Huang, S. Lin, N. Wang, G. Dai, Y. Xie, and J. Zhou, "TSE-CNN: A Two-Stage End-to-End CNN for Human Activity Recognition," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 292–299, Jan. 2020, https://doi.org/10.1109/JBHI.2019.2909688.

[13] K. Xia, J. Huang, and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020, https://doi.org/10.1109/ACCESS.2020.2982225.

[14] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, Mar. 2011, https://doi.org/10.1145/1964897.1964918.

[15] F. J. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors*, vol. 16, no. 1, Jan. 2016, Art. no. 115, https://doi.org/10.3390/s16010115.

[16] H. S. Saad, J. F. W. Zaki, and M. M. Abdelsalam, "Employing of machine learning and wearable devices in healthcare system: tasks and challenges," *Neural Computing and Applications*, vol. 36, no. 29, pp. 17829–17849, Oct. 2024, https://doi.org/10.1007/s00521-024-10197-z.

[17] S. S. Pushpalatha and S. Math, "Human activity recognition using hybrid model," *International Journal on Information Technologies & Security*, vol. 14, no. 4, pp. 55–66, 2022.

[18] S. W. Pienaar and R. Malekian, "Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture," in *2019 IEEE 2nd Wireless Africa Conference (WAC)*, Pretoria, South Africa, Aug. 2019, pp. 1–5, https://doi.org/10.1109/AFRICA.2019.8843403.

[19] N. Y. Hammerla, S. Halloran, and T. Ploetz, "Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables." arXiv, Apr. 29, 2016, https://doi.org/10.48550/arXiv.1604.08880.

[20] R. Chavarriaga *et al.*, "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition," *Pattern Recognition Letters*, vol. 34, no. 15, pp. 2033–2042, Nov. 2013, https://doi.org/10.1016/j.patrec.2012.12.014.

[21] S. Mohsen, A. Elkaseer, and S. G. Scholz, "Industry 4.0-Oriented Deep Learning Models for Human Activity Recognition," *IEEE Access*, vol. 9, pp. 150508–150521, 2021, https://doi.org/10.1109/ACCESS.2021.3125733.

[22] D. Anguita, A. Ghio, L. Oneto, X. Parra Perez, and J. L. Reyes Ortiz, "A public domain dataset for human activity recognition using smartphones," presented at the Proceedings of the 21th International European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2013, pp. 437–442.

[23] F. Cruciani *et al.*, "Feature learning for Human Activity Recognition using Convolutional Neural Networks," *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 1, pp. 18–32, Mar. 2020, https://doi.org/10.1007/s42486-020-00026-2.

[24] M. Ronald, A. Poulose, and D. S. Han, "iSPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021, https://doi.org/10.1109/ACCESS.2021.3078184.

[25] M. Akter, S. Ansary, M. A.-M. Khan, and D. Kim, "Human Activity Recognition Using Attention-Mechanism-Based Deep Learning Feature Combination," *Sensors*, vol. 23, no. 12, Jan. 2023, Art. no. 5715, https://doi.org/10.3390/s23125715.