International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

Complex Human Activity Recognition Using LSTM

Rituja Nerkar, Dr. P.J. Kulkarni

RAJARAMBAPU INSTITUTE OF TECHNOLOGY, RAJARAMNAGAR

Email: - ritujanerkar6@gmail.com

ABSTRACT:-

In the disciplines of ubiquitous computing and body area networks, Human Activity Recognition (HAR) based on sensor networks is a significant research direction. With the advancement of deep neural networks, manual feature extraction is no longer necessary. the use of HAR for security, monitoring, and surveillance applications in the sphere of healthcare. Deep neural networks can perform better in difficult issues involving the identification of human activities. An essential area of application for computer vision is human activity recognition. Its main objective is to properly characterize human behaviors and their relationships from a succession of never-before-seen sensor data. The development of several significant applications, including intelligent surveillance systems, human-computer interfaces, health care, security, and military applications, is made possible by the capacity to identify, comprehend, and anticipate complicated human activities. The computer vision community has focused particularly on deep learning in recent years. This study provides a summary of the state-of-the-art in action recognition using deep learning and video analysis. The most significant deep learning models for human action recognition are presented, and their strengths and weaknesses are highlighted in order to show the current state of deep learning algorithms used to address human action recognition issues in realistic videos. This study identifies state-of-the-art deep architectures in action recognition based on the quantitative analysis employing recognition accuracies published in the literature, and then presents current trends and open difficulties for further efforts in this field.

Keywords—: Human Activity Recognition, Deep Learning, Machine Learning

I. INTRODUCTION

A significant topic of study in body area networks and ubiquitous computing is Human Activity Recognition (HAR) based on sensor networks. Statistical machine learning techniques are extensively used in existing research to manually create and extract characteristics of different movements. Yet, due to the extraordinarily rapid growth of waveform data and the lack of any discernible rules, conventional feature engineering techniques are becoming less and less effective. With the development of Deep Learning technology, we can better perform in complicated human activity recognition challenges without manually extracting information. By transferring deep neural network experience in image recognition, we present a deep learning model (InnoHAR) based on the combination of Inception Neural Network and recurrent neural network. The model receives end-to-end waveform data from multi-channel data. Inception-like modules use a variety of kernel-based convolution layers to extract multidimensional

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

Characteristics. Combining modelling for time series features with GRU allows classification jobs to fully take use of data properties. For three of the most popular public HAR datasets, our suggested technique consistently beats state-of-the-arts in terms of performance and generalization.

Due to their simplicity of use, high accuracy, low power consumption, and other qualities, wearable sensors are also commonly utilized in applications for human activity detection and motion capture. For instance, biosensors are frequently used to track vital indicators including blood pressure, heart rate, temperature, electrocardiography (ECG), and electromyography (EMG). With physiological monitoring, it is possible to identify and treat asthma, dysthymias, hypertension, seizures, and dysmenorrhea. Inclineometers and goniometers are additional sensors for measuring the kinematics of the upper and lower limbs.

Technique for Human Activity Recognition

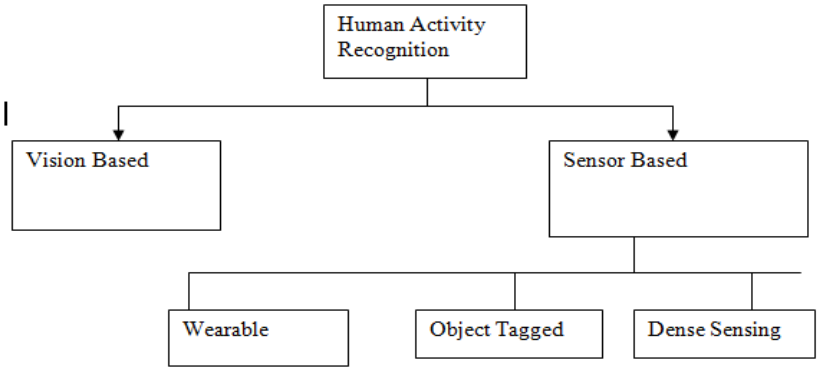


Figure1 Technique for human activity recognition

Vision Based

Vision-based techniques in the area of action recognition. Deep neural network-based solutions and representation-based solutions are the two categories of vision-based solutions. Subsets of representation-based solutions include holistic and local displays, as well as aggregation techniques. Multiple stream networks, temporal coherency networks, generative models, and spatiotemporal networks are some of the subcategories of deep neural network-based solutions.

Wearable

This approach uses sensor-based research to identify human activity. This survey categorises the ongoing studies into two main categories**:**

1. Research based on eyesight as opposed to research based on sensors.

2. Research that is informed by data as opposed to expertise.

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

Proposed System Objective

To create the most advanced algorithms for detecting complicated human behavior to do research on the created algorithm.

II. LITERATURE REVIEW

In [1], the author asserted that multi-sensor data fusion is a well-established research area and that there is a substantial body of work addressing sensor fusion at many levels and employing a range of approaches, even in the BSN industry. The author of this study focused on framework comparison and activity identification within the studied publications.

An organized survey of earlier research on healthcare recommender systems is provided by the author of. Unlike recent pertinent overview publications, this study provides insights into recommendation situations and approaches. Due to the wealth of medical information, medical professionals have had great difficulty making judgments that are patient-centered (such as information on pharmaceuticals, medical tests, and suggested treatments). [2]

The model proposed in this work [3] has a large difference between the peak and median performance and necessitates a significant expenditure in parameter research. For a practitioner, it is presumably the most useful option. Consequently, practitioners shouldn't give up on the model even if a first analysis reveals poor recognition performance.

While sensing human motion in [4], the author decides to take into consideration the geometric connection between joints and limbs, which help to some extent minimize sensor drift inaccuracy. The PCRLB of the human motion process is obtained from the proposed geometrical model. Simulation findings show better performance when the suggested model is applied and both distance and IMU data are taken into account. In field studies, the model is used to identify human lower limb motions. The testing outcomes show a substantially lower localization error and acceptable energy usage, which is significantly more efficient than the traditional technique ZUPT.

The author of [5] suggested a small, wireless wearable sensor node that combines an IMU sensor with an air pressure sensor. Both features with and without air pressure data are used to train the HAR model. The findings demonstrate that the HAR model with air pressure data training outperforms the model trained without air pressure data in terms of recognition performance. Also, we discovered that the performance of the transfer learning-based HAR model is more susceptible to the absence of air pressure data. The IPL-JPDA method suggested in this research has the best recognition performance in the comparison experiment of nine HAR models, and the average recognition accuracy of various subjects is 93.2%.

Deep learning models, according to the author of this paper [6], have radically changed how sensor data is analyzed and interpreted. Due to the accuracy benefits these approaches offer, they are intriguing for the next generation of mobile, wearable, and embedded sensing applications. Modern deep learning techniques, however, frequently need a lot of device and computing resources, even just for the inference stages, which are necessary to separate high-level classes from low-level data. Because of this, one of the main challenges to the broad adoption of these effective learning strategies is the limited memory, processing, and energy that are accessible on embedded and mobile systems. [6]

Table1. Comparative Analysis of Existing System

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

Approach	Technology	Advantage
Vision Based	Surveillance camera	High Accuracy
Depth Sensor	Kinect	High Accuracy
Wearable Sensor	Gloves, Bracelet, Smart Watch	Low Cost

III. PROPOSED SYSTEM

To determine the current approaches utilized for creating novel sophisticated human activity identification systems and to conduct a comparative study. Choose the present system's most successful approach, and then create a conversational system. In the domains of ubiquitous computing and body area networks, Human Activity Recognition (HAR) based on sensor networks is a significant study topic. Our goal in the proposed study is to identify human activity using live video surveillance rather than a network of body sensors. It will be able to tell whether there is human activity on a live broadcast or in a picture. Show the proportion of known activity.

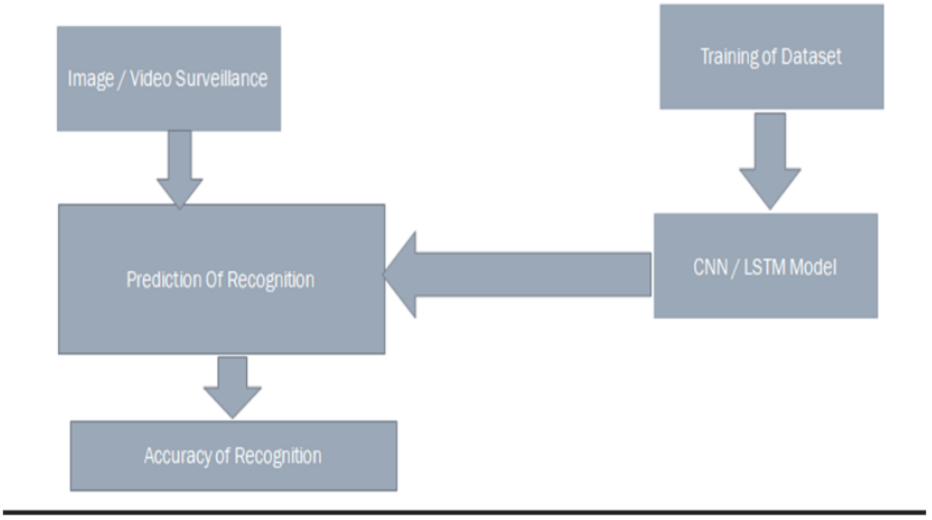


Figure1. System Architecture of Proposed work

The figure1 shows the proposed system architecture

Input Output Process of System

1. First start the video camera
2. Training of Algorithm with the dataset with different class labels
3. Pass the input image to system or model
4. Complex activity prediction is done with the testing of models. Complex activity like open door close door open drawer toggle switch.

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

5. Accuracy Prediction with multiple datasets and algorithms.

LSTM Algorithm

When someone shows someone an image and then asks them about it two minutes later, they will likely remember the image's contents. However, if they inquire about the same image a few days later, the information may have been lost or faded somewhat. For the first circumstance, recurrent neural networks (RNNs) are required. In the second scenario, LSTMs are required for large memory capacity. Long Term Short Memory is the term for such.

How LSTM Process Data

1. Import the data

The train and test data are both retrieved from the internet. We'll make predictions using the open price.

2. Feature Scaling

The train and test data are both retrieved from the internet. We'll make predictions using the open price.

3. Data Structure Creation

Creating Sliding Window of data

4. Data Reshaping

To make predictions, we use Open pricing. We only have one indicator or feature, specifically. However, using the same data processing techniques, we can add more indicators. To accomplish that, a new dimension for the number of indicators must be added.

5. Model Building

In essence, we are using LSTM to construct a NN regressor for continuous value prediction. Initialize the model first.

6. Model Compiling

7. Model Fitting

8. Data preprocessing for test dataset

9. Model Prediction

10. Result Visualization

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

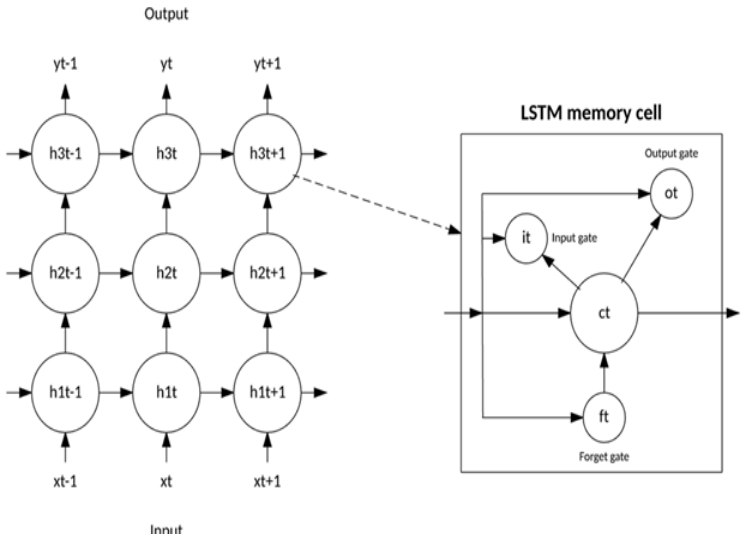


Figure2 LSTM Architecture

Confusion Matrix is the visual illustration of the particular VS foretold values. It measures the performance of our Machine Learning classification model and appears sort of a table-like structure.

This is. However, a Confusion Matrix of a binary classification downside sounds like

Precision: It may be outlined because of the range of correct outputs provided by the model or, out of all positive categories appropriately foretold by the model, what number of them was valid. It may be calculated as mistreatment by the below formula.

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

Recall: - It is outlined because the out of total positive categories, however our model foretold properly. The recall should be as high as doable

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

Recurrent neural networks (RNNs) of the LSTM (Long Short-Term Memory) variety have shown to be effective at recognizing complicated human behavior.

Handling Sequential Data: Human activities frequently include a series of events that happen one after another across time. LSTM is ideally suited for activity recognition tasks since it is specifically created to model sequential data and can efficiently capture temporal connections between various steps in the activity.

Memory Retention: LSTM cells can retain information for extended periods of time, which enables them to identify long-term dependencies in complex processes. When actions require several steps that may be spread out over time, memory

retention is essential.

LSTM automatically learns pertinent features from the input sequences, eliminating the requirement for manually engineered feature extraction. This is especially helpful for complicated activities where it could be difficult to pinpoint the right features.

Managing Sequences of Variable Length: Human activities can be of varied lengths and comprise a variety of phases or actions. By modifying its internal memory and altering the hidden state, LSTM can accommodate sequences of different lengths.

Robustness to Noise: LSTMs can handle uncertainties and noisy sensor readings that are prevalent in real-world activity recognition settings because they are resilient to noise and fluctuations in the input data.

Recognizing Hierarchical Activities: Hierarchical actions and sub-actions are frequently present in complex human activities. These hierarchical structures can be identified using LSTM, which will then be able to provide a more thorough picture of the total activity.

Transfer Learning: In particular activity recognition tasks, LSTM models that have been calibrated using a large dataset may also serve as the foundation for transfer learning. This enables handling complex actions even with a dearth of labeled data by utilizing pre-trained models.

Processing in real-time is possible with LSTMs thanks to their effective implementation and execution, which makes them appropriate for real-time activity identification applications including surveillance, health monitoring, and human-computer interaction systems.

The mathematical equations for complex human activity recognition using LSTM. We'll represent the input sequence as X = {X_1, X_2, ..., X_T}, where X_t represents the input vector at time step t. Similarly, the hidden state sequence will be H = {h_1, h_2, ..., h_T}, and the cell state sequence will be C = {C_1, C_2, ..., C_T}.

Input gate (i_t):

i_t = sigmoid(W_{xi} * X_t + W_{hi} * h_{t-1} + W_{ci} * C_{t-1} + b_i) ……..eq(1)

Forget gate (f_t):

f_t = sigmoid(W_{xf} * X_t + W_{hf} * h_{t-1} + W_{cf} * C_{t-1} + b_f)…….eq(2)

Output gate (o_t):

o_t = sigmoid(W_{xo} * X_t + W_{ho} * h_{t-1} + W_{co} * C_{t-1} + b_o)……..eq(3)

Candidate cell state (C~t):

C~t = tanh(W{xc} * X_t + W{hc} * h_{t-1} + b_c)….eq(4)

New cell state (C_t):

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

C_t = f_t * C_{t-1} + i_t * C~_t

New hidden state (h_t):

h_t = o_t * tanh(C_t)

Output layer (for classification):

y_t = softmax(W_{hy} * h_t + b_y)

Where:

X_t is the input vector at time step t.

h_t is the hidden state at time step t.

C_t is the cell state at time step t.

i_t, f_t, and o_t are the input, forget, and output gates, respectively, at time step t.

W_{xi}, W_{xf}, W_{xo}, W_{xc}, W_{hi}, W_{hf}, W_{ho}, W_{hc}, W_{ci}, W_{cf}, W_{co}, W_{hy} are the weight matrices for different connections in the LSTM.

b_i, b_f, b_o, b_c, b_y are the bias terms for the respective gates and output layer.

sigmoid() is the sigmoid activation function.

tanh() is the hyperbolic tangent activation function.

softmax() is the softmax activation function used in the output layer for classification.

During training, you would use labeled data to optimize the parameters (weights and biases) of the LSTM network using techniques like backpropagation through time (BPTT) and gradient descent. The final output y_t represents the predicted probabilities of different human activities at each time step t.

Dataset Used

OPPORTUNITY DATASET

The opportunity activity recognition dataset contains a significant amount of atomic activities (more than 27,000) that are complicated naturalistic behaviours that were gathered by a sensor. It includes recordings of 12 subjects made with 15 networked sensor systems, 72 sensors across 10 modalities, embedded in the surroundings, objects, and the subject's body. These qualities make it a good candidate to compare different activity identification methods.

PAMAP2 DATASET

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

It consists of recordings from 9 participants (8 males and 1 female) instructed to carry out 18 lifestyle activities, including household activities (lie, sit, stand, walk, run, cycle, Nordic walk, iron, vacuum clean, rope jump, ascend and descend stairs) and a variety of leisure activities (watch TV, computer work, drive car, fold laundry, clean house, play soccer).

Smart Phone Dataset

The recordings of 30 participants carrying a waist-mounted Smartphone with incorporated inertial sensors while engaging in activities of daily living (ADL) are used to create a Smartphone database.

IV. Experimental Result Analysis

1. Network Training Details

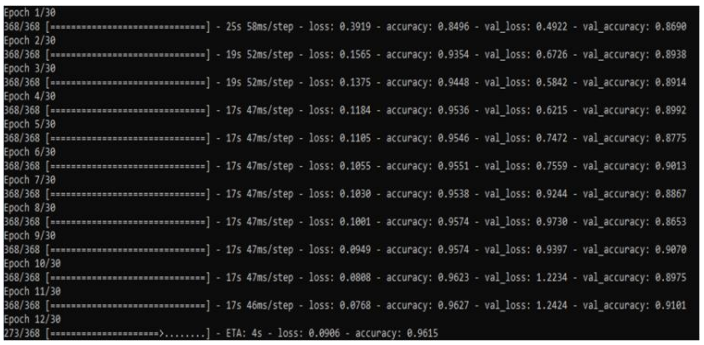
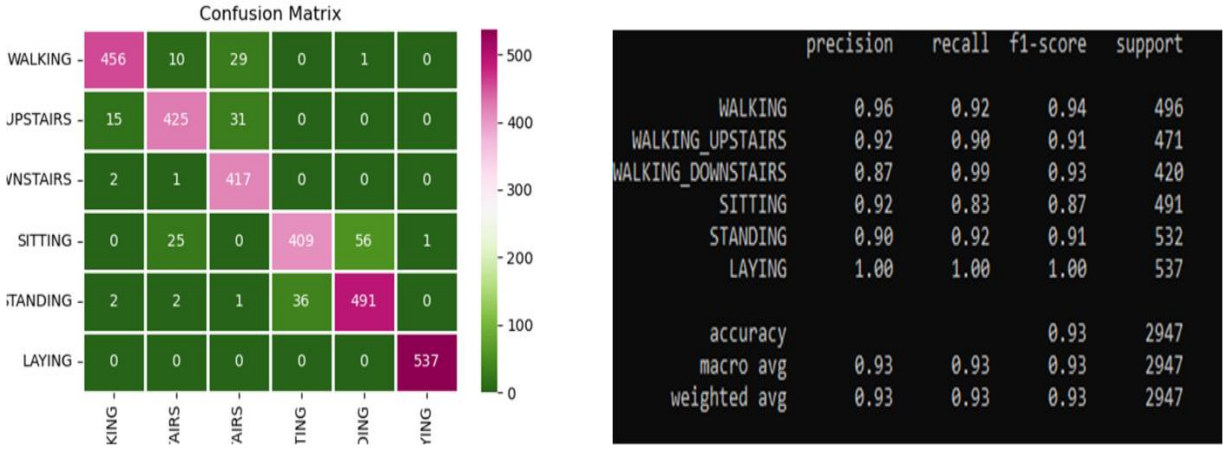


Figure5. Network Training details

Figure5 shows the network training details.

2. Confusion Matrix



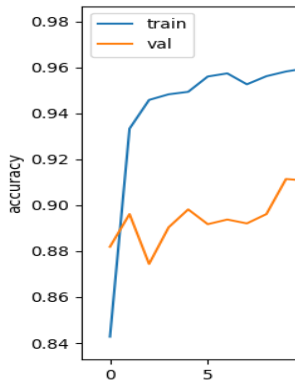
	precision	recall	f1-score	support
WALKING	0.96	0.92	0.94	496
WALKING_UPSTAIRS	0.92	0.90	0.91	471
WALKING_DOWNSTAIRS	0.87	0.99	0.93	420
SITTING	0.92	0.83	0.87	491
STANDING	0.90	0.92	0.91	532
LAYING	1.00	1.00	1.00	537
accuracy			0.93	2947
macro avg	0.93	0.93	0.93	2947
weighted avg	0.93	0.93	0.93	2947

Figure6 Confusion Matrix

The figure6 shows the confusion matrix.

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

3. Accuracy Graph



Epoch	Algorithm	Accuracy

Figure7 Accuracy Graph

The figure7 shows the training vs validation accuracy graph.

4. Training vs Validation Loss Graph

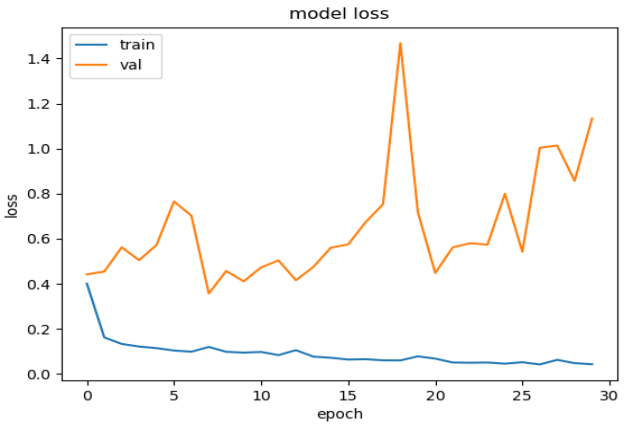


Figure8 Train & Validation Graph

The figure 8 shows the train & validation graph.

The Above graph shows that the generated model gives us accuracy above 98%. As you can see in the diagram, the accuracy increases rapidly in the first two epochs, indicating that the network is learning fast. Afterward, the curve flattens indicating that not too many generations are required to train the model further. Generally, if the training data accuracy keeps improving while the validation data accuracy gets worse, you are encountering over fitting. It indicates that the model is starting to memorize the data.

5. Result Analysis

International Journal of Electrical and Electronics Engineers
Volume 15, Issue No. 02, July-Dec 2023
ISSN (O) 2321-2055
ISSN (P) 2321 -2045

50	**CNN**	**90.91 %**
50	**LSTM**	**91.27 %**

CONCLUSION

In this project, we have implemented the LSTM Model for complex human activity recognition. In this we have created the LSTM model and also implemented metric for accuracy. This framework loosely helps for recognizing the activities from the live video camera surveillance.

REFERENCES

1. Gravina, R., Alinia, P., Ghasemzadeh, H., & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. Information Fusion, 35, 68-80.

2. Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, convolutional, and recurrent models for human activity recognition using wearables. arXiv preprint arXiv:1604.08880.

3. Fortino, G., Giannantonio, R., Gravina, R., Kuryloski, P., & Jafari, R. (2012). Enabling effective programming and flexible management of efficient body sensor network applications. IEEE Transactions on Human-Machine Systems, 43(1), 115-133.

4. Wang, W., Li, Y., Zou, T., Wang, X., You, J., & Luo, Y. (2020). A novel image classification approach via dense-MobileNet models. Mobile Information Systems, 2020.

5 Xu, C., He, J., Zhang, X., Yao, C., & Tseng, P. H. (2018). Geometrical kinematic modeling on human motion using method of multi-sensor fusion. Information Fusion, 41, 243-254.

6. Xu, C., He, J., Zhang, X., Yao, C., & Tseng, P. H. (2018). Geometrical kinematic modeling on human motion using method of multi-sensor fusion. Information Fusion, 41, 243-254.

7. Xu, C., He, J., Zhang, X., Yao, C., & Tseng, P. H. (2018). Geometrical kinematic modeling on human motion using method of multi-sensor fusion. Information Fusion, 41, 243-254.