Deep Analysis for Smartphone-based Human Activity Recognition

Abstract— **Wearable-based approach and vision-based approach are two of the most common approaches in human activity recognition. However, the concern of privacy issues may limit the application of the vision-based approach. Besides, some individuals are reluctant to wear sensor devices. Hence, smartphone-based human physical activity recognition is a popular alternative. In this paper, we propose a deep analysis to interpret and predict accelerometer data captured using a smartphone for activity recognition. The proposed deep model is able to extract deep features from both spatial and temporal domains of the inertial data. The recognition accuracy of the proposed model is assessed using UCI and WISDM accelerometer data. Empirical results exhibit a promising performance.**

Keywords—deep learning, smartphone, human activity recognition, inertial data, accelerometer

I. INTRODUCTION (*HEADING 1*)

World Health Organization's (WHO) statistics reveal that a quarter of people is not sufficiently active [1]. Frank et al. highlighted in their work [2] that physical inactivity could lead to a rise in chronic diseases. Chronic diseases include but not limited to heart diseases, high blood pressure, stroke, and cancers. The mass increase of chronic diseases in the community will instigate dramatic challenges to the healthcare system.

With the splendid development of the Internet of Things (IoT), innovative ambient assisted living (AAL) facilities are invented for smart environments, sport training, healthcare applications etc. [3][4]. Regular monitoring and recognition of physical activity could support healthcare management and encourage health-enhancing physical activity. At the fundamental level, an intelligent human activity recognition, or coined as HAR, system is fostered and introduced in AAL solutions to recognize physical inactivity of the users and recommend them appropriate exercise activities based on their age, gender, BMI level, and health level, etc.

Wearable-based approach and vision-based approach are two of the most common approaches in HAR. Both approaches have produced promising results, reaching above 80% accuracy. However, there are privacy issues revolving around the vision-based approach. Placing a surveillance camera in public places may violate the law and required extensive justification to obtain permits. On the other hand, some individuals are reluctant to wear sensor devices.

Henceforth, human physical activity recognition using a smartphone is a popular alternative [5][6][7]. Smartphones are packed with high-end hardware and features. Numbers of sensors are embedded on smartphones such as accelerometer and gyroscope meter. These sensors are adequate to measure the triaxial acceleration as well as orientation and angular velocity of our body movement, detecting human activities.

Kwapisz et al. popularizes the adoption of smartphone-based accelerometers HAR [8]. In their work, time series motion data is aggregated into examples that encapsulate user activity. A predictive model is trained for activity recognition. In recent years, deep learning has been a hot topic in HAR [9][10][11]. Deep learning methods remove the dependency of the time-consuming hand-crafted feature that many researchers face [12]. Furthermore, the key component of deep learning models is the layered architecture that enables deeper extraction of features, allowing informative feature representation and accurate classification.

Zeng et al. employs Convolutional Neural Network (CNN) to capture local dependencies and spatial domain of activity signals [13]. They utilize multichannel time series data to recognize the user's activity and hand gestures. Further, Lee et al. implements a one-dimensional CNN approach on the triaxial acceleration data to recognition activity [14]. The potential of CNN could be limited to spatial domain features; while motion data signal is a time series data containing temporal features.

In viewing of that, Chen et. al. utilizes a temporal deep learner, that is Long-Short-Term-Memory (LSTM), on triaxial accelerometers data for human activity recognition [15]. LSTM uses past information to predict the outcome of a HAR model. However, Yu and Qin claims that some information may not be captured since human motion is continuous [16]. Yu and Qin then adopt Bidirectional-LSTM (Bidir-LSTM) in HAR. Different from LSTM that only considers the past information, Bidir-LSTM employs both past information and future information, as shown in Fig 1. Bidir-LSTM is stacked into layers horizontally and vertically. A single LSTM node takes in information from the horizontal layer for both past and future information, as well as from the vertical layer (i.e. the lower hidden layer). Promising result is obtained.

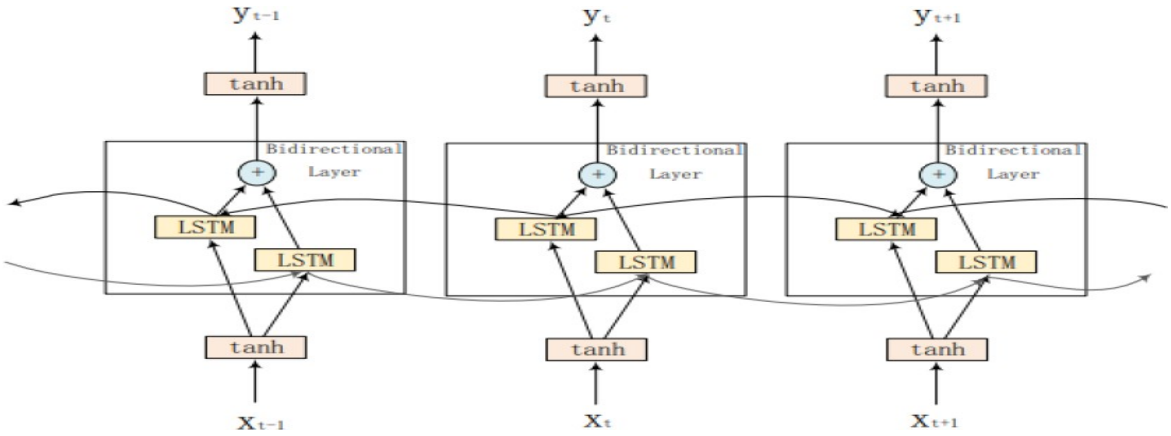


Fig. 1 Bi-directional LSTM architecture (illustration is from [16])

Inspiring from the works of [13] and [16], we propose a temporal deep learner by amalgamating the concepts of CNN and Bidir-LSTM. The proposed approach is evaluated using two publicly available datasets: WISDM and UCI databases. The empirical results exhibit encouraging human activity recognition performance with accuracy above 90% in WISDM database and 87% in UCI database.

II. The Proposed Approach

In our proposed deep learner architecture, there are four convolutional-max-pooling layers and one layer of Bidir-LSTM layer, following with one dropout layer and two dense layers (one fully-connected layer and softmax). We term the learner as Conv4-Max4-BidirLSTM, coined as C4M4BL. Fig. 2 illustrates the architecture of C4M4BL.

Convolution operation in the convolutional layers can apprehend the neighborhood dependency of data points. This feature enables those spatial patterns such as shapes of motion data to be identified. The inclusion of Bidir-LSTM is to overcome the limitation of CNN layers for not being able to extract the temporal information in time-series data.

To avoid overfitting, a dropout layer is included after the Bidir-LSTM layers to turn off on a selective region of nodes in a neural network. It makes training neurons not depend on any specific neurons, allowing learning to be more reliable.

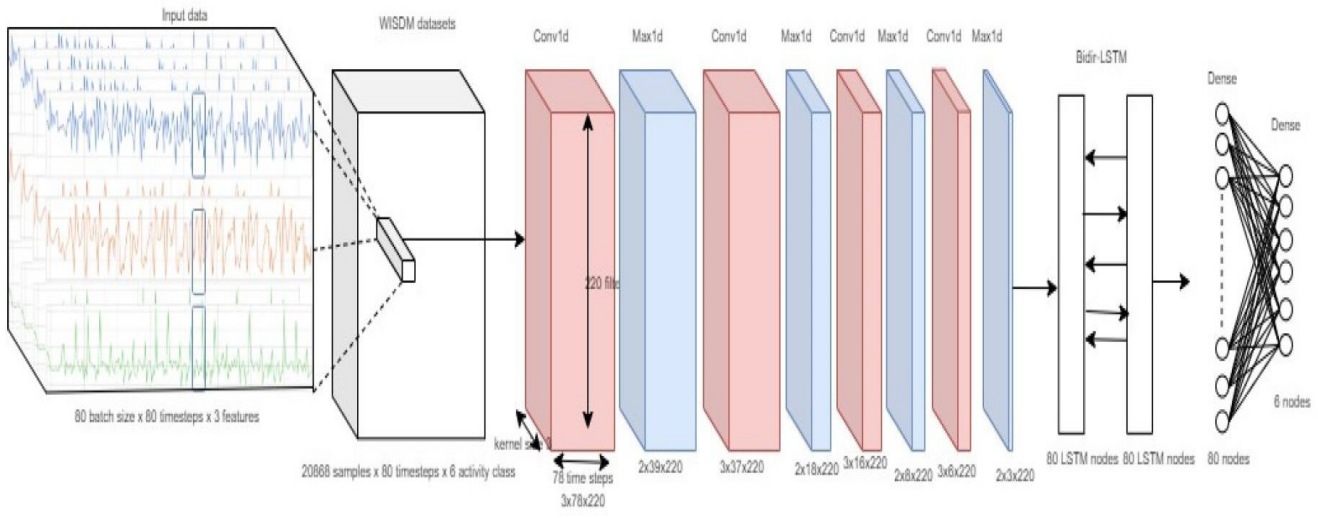


Fig. 2 The architecture of C4M4BL

III. Hardware and Software Specifications

The proposed C4M4BL model is developed and run under the computing environment as recorded in Table I.

TABLE I. Hardware and Software Information

Software	Hardware
Jupyter notebook	Lenovo Yoga 530
Anaconda Python	Intel Core i7-8550u processor, 1.85GHz
Keras, Numpy, Scipy, Seaborn and Pandas library	16GB RAM
	64-bit Microsoft Operating System, x86-based processor

IV. Databases and Performance Measures

UCI and WISDM datasets are used to assess the effectiveness of the proposed C4M4BL model. In UCI dataset, six activity classes, i.e. walking, walking upstairs, walking downstairs, sitting, standing and laying down, are performed by 30 subjects, while carrying a waist-mounted smartphone [5]. Both triaxial acceleration and triaxial angular velocity data are collected. But, in this paper, only acceleration data is considered. The acceleration data is transformed to triaxial gravity acceleration and triaxial body acceleration data for activity recognition. The data is segmented using 50% overlapping sliding windowing with 2.56 seconds time step, consisting of 128 samples in each time step.

Similar to UCI dataset, WISDM dataset contains 6 activity classes, but the activities are walking, jogging, stair-up, stair-down, sitting and standing [8]. 29 subjects perform the activities with smartphone placing in the front leg pocket. In this dataset, triaxial gravity acceleration data is collected from the accelerometer. The data signal is segmented using a 50% sliding windowing method with 10 seconds. 200 samples are in each time step.

In our experiments, training-testing split experimental protocol is implemented, where 70% of each dataset is used for training and the remaining 30% is used for testing. Among the training data, 20% is selected for system validation and the remaining is for model training.

A confusion matrix is used to analyze the recognition performance of the proposed model. Besides, the measures of accuracy, recall, and prevision are employed as well:

$$accuracy = \frac{True\ class}{Total\ number\ of\ samples} \quad (1)$$

$$recall = \frac{True\ class}{Actual\ class} \quad (2)$$

$$precision = \frac{True\ class}{predicted\ class} \quad (3)$$

True class means the model predicts a class correctly. From the above equation, we can understand that accuracy shows how often the model's predictions are true; recall is associated with model sensitivity, i.e. if the actual label is correct, how often the model's predictions are true; precision is about if the predicted label is correct, how often the model's predictions are true.

V. Experiments

We tested the proposed C4M4BL model with various hyperparameter settings, as illustrated in Fig. 3. Due to the limitation of our computing machine, only a maximum of 50 epochs is set. Fig. 4 depicts the validation performance of different hyperparameter configurations. From the empirical results, we can observe that c0 obtains the best results in UCI dataset and c9 shows superiority in WISDM dataset. Hence, these configurations will be used in our subsequent experiments. Table II and III record the confusion matrices of UCI and WISDM datasets.

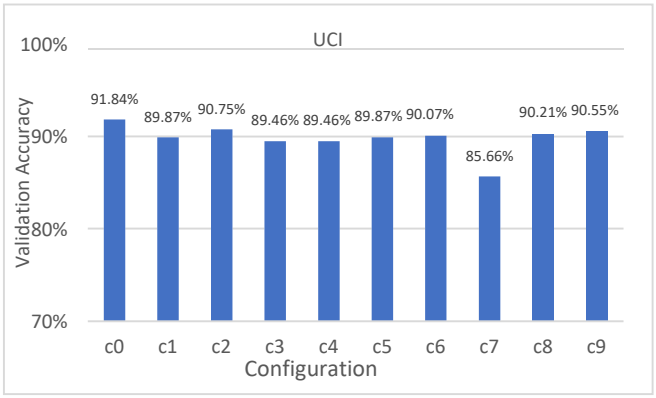
	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
Feature map	32	64	64	88	96	128	128	164	196	230
Kernel size	3	2	3	3	2	5	3	3	3	2
Bidirlstm	128	128	64	128	186	64	128	78	64	88
Dropout	0.1	0.4	0.5	0.4	0.5	0.3	0.5	0.7	0.2	0.5
1st Dense	128	128	128	128	128	128	128	128	128	128
2nd Dense	6	6	6	6	6	6	6	6	6	6
Epochs	30	25	50	50	25	40	40	30	30	30
Learning rate	0.001	0.005	0.005	0.003	0.001	0.001	0.001	0.002	0.004	0.001
Decay	0.0001	0.001	0.0001	0	0	0	0	0.00005	0.001	0.001
Batch size	128	64	128	64	32	32	64	64	64	32

(a)

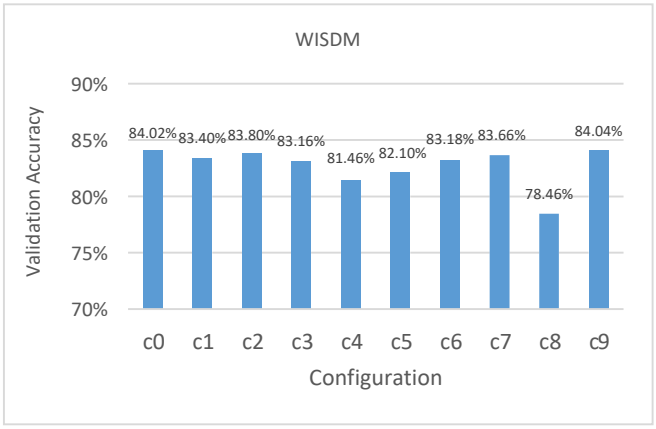
	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9
Feature map	20	40	50	60	80	80	100	140	200	220
Kernel Size	3	2	5	2	2	2	2	3	3	3
Bidirlstm	50	100	100	100	80	160	100	200	100	80
Dropout	0.2	0.5	0.3	0.4	0.5	0.5	0.3	0.5	0.7	0.1
1st Dense	80	80	80	80	80	80	80	80	80	80
2nd Dense	6	6	6	6	6	6	6	6	6	6
Epochs	30	25	40	25	30	30	50	30	30	30
Learning rate	0.004	0.001	0.001	0.005	0.004	0.001	0.003	0.001	0.001	0.001
Decay	0.001	0	0	0.001	0.0001	0.001	0.00005	0	0	0.0001
Batch size	40	200	50	50	80	100	200	80	50	100

(b)

Fig. 3 hyperparameter settings of (a) UCI dataset and (b) WISDM dataset



(a)



(b)

Fig. 4 Validation performance of (a) UCI and (b) WISDM databases with different hyperparameter settings

TABLE II. CONFUSION MATRIX OF UCI DATASET

		Predicted Class						*Recall (%)*
		W	***WU***	***WD***	***Si***	***St***	***L***	
Actual Class	***W***	492	2	2	0	0	0	99.19
	WU	37	416	18	0	0	0	88.32
	WD	1	3	416	0	0	0	99.05
	Si	0	25	0	413	53	0	84.11
	St	0	1	0	109	422	0	79.32
	L	0	27	0	0	0	510	94.97
Precision (%)		92.83	87.76	95.41	79.12	88.84	100	**90.57**

Note: W=walking, WU=walking-upstair, WD=walking-downstair, Si=sitting, St=standing, L=laying

TABLE III. CONFUSION MATRIX OF WISDM DATASET

		Predicted Class						*Recall (%)*
		WD	***J***	***Si***	***St***	***WU***	***W***	
Actual Class	***WD***	456	9	0	0	136	49	70.15
	J	18	1902	0	0	22	48	95.58
	Si	0	0	412	0	1	39	91.15
	St	0	0	41	328	1	0	88.89
	WU	73	20	2	0	592	38	81.66
	W	38	231	1	0	48	2079	86.73
Precision (%)		77.95	87.97	90.35	100	74.00	92.28	**87.62**

Note: WD=walking-downstair, J=jogging, Si=sitting, St=standing, WU=walking-upstair, W=walking

From the above tables, we can notice that static postures such as standing and sitting are the two most easily confused activities in UCI dataset. However, in WISDM dataset, the proposed model has a hard time differentiating walking downstairs and upstairs activities, albeit it can recognize static postures accurately. Empirical results show that the proposed C4M4BL model is able to attain 90% performance accuracy in UCI dataset and 87% accuracy in WISDM dataset. Next, additional experiments are conducted to address performance comparison with other approaches as well as to investigate the effect of hyperparameters to the performance of the proposed model. These experiments are carried out by using WISDM dataset

A. Additional Experiment: Performance Comparison with the existing approaches

In this experiment, several existing models are run for performance comparison. However, due to the limitation of the current computing specifications, our computer is not able to train models up to hundreds or thousands of epochs as what the authors did using their advanced computer with high-end GPU(s). Hence, all the models are trained with 30 epochs based on their architecture in this work. From the empirical results, we can observe that our proposed model is slightly superior to other models.

TABLE IV. RESULT COMPARISON OF UCI DATASET

Model	**Architecture**	**Accuracy (%)**
CNN [ronao 2016]	C(96)-MAX-C(96)-MAX-C(96)-MAX-DL(1000)-DL(6)	89.41%
CNN [Lee SM 2017]	C(128)-MAX-C(128)-MAX-DL(384)-DL(6)	89.24%
LSTM [Chen, Chong 2016]	L(100)-L(100)-DL(128)-DL(6)	89.79%
Bidir-LSTM [Yu 2018]	B(L(28))-B(L(28)-B(L(28))-DL(128)-DL(6)	89.07%
Bidir-LSTM [Hernandez2019]	B(L(175))-B(L(175)-B(L(175))-DL(128)	87.41%
C4M4BL	C(32)-MAX-C(32)-MAX-C(32)-MAX-C(32)-MAX-B(L(128))-DL(128)-DL(6)	**90.57%**

TABLE V. RESULT COMPARISON OF WISDM DATASET

Model	**Architecture**	**Accuracy (%)**
CNN [9]	C(96)-MAX-C(96)-MAX-C(96)-MAX-DL(1000)-DL(6)	74.79%
CNN [14]	C(128)-MAX-C(128)-MAX-DL(384)-DL(6)	85.98%
LSTM [15]	L(100)-L(100)-DL(80)-DL(6)	79.62%
Bidir-LSTM [16]	B(L(28))-B(L(28)-B(L(28))-DL(80)-DL(6)	72.98%
Bidir-LSTM [17]	B(L(175))-B(L(175)-B(L(175))-DL(80)-DL(6)	82.09%
C4M4BL	C(220)-MAX-C(220)-MAX-C(220)-MAX-C(220)-MAX-B(L(80))-DL(80)-DL(6)	**87.62%**

B. *Additional Experiment: Effect of the Number of Kernels*

This experiment is set up using a range of 5 to 200 numbers of kernels/ filters. Fig. 5 illustrate the performance of the proposed model. The obtained result demonstrates that the number of feature maps is directly proportional to the performance accuracy of the model. This is justifiable that the more the feature maps are, the more abstract representations can be retrieved. Table IV and V reveal the performance of the models.

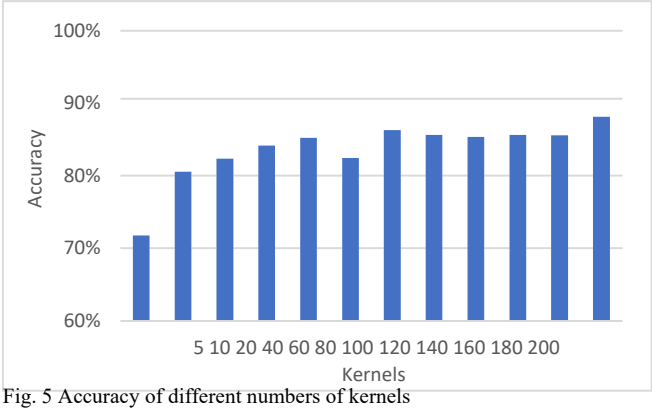


Fig. 5 Accuracy of different numbers of kernels

C. *Additional Experiment: Effect of the Kernel Size*

This experiment is set up using various kernel sizes with the range from 1 to 5. Kernel size or sometimes calls "filter size" is the dimension of the filters used in the convolutional layer. Fig. 6 illustrates the performance of the proposed model with different kernel sizes. The finding implies that the employment of a higher range of temporal local dependency, derived from the increasing kernel size, provides more meaningful features. However, the excessive kernel size might overlook some essential details in the data.

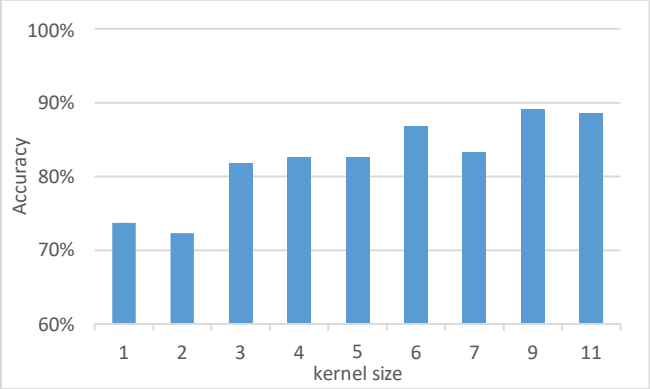


Fig. 6 Accuracy of different kernel sizes

CONCLUSION

In this paper, a deep learning architecture that adopts CNN with Bi-directional LSTM is proposed for smartphone-based human physical activity recognition. CNN layers are able to extract deep features of the data neighborhood dependency, revealing spatial patterns such as shapes of motion data. On the other hand, Bi-directional LSTM extracts the temporal information in the inertial data. The capability of the proposed model in recognizing human activity is validated using publicly available databases: UCI and WISDM datasets. Empirical results show a promising performance.

ACKNOWLEDGMENT

Authors would like to express their appreciation to CUCC research centre for the hardware support to this final year project.

REFERENCES

[1] "WHO | Prevalence of insufficient physical activity," WHO, 2018.

[2] F. W. Booth, C. K. Roberts, and M. J. Laye, "Lack of exercise is a major cause of chronic diseases," Compr. Physiol., vol. 2, no. 2, pp. 1143–1211, Apr. 2012.

[3] M. Memon, S. R. Wagner, C. F. Pedersen, F. H. Aysha Beevi, and F. O. Hansen, "Ambient Assisted Living healthcare frameworks, platforms, standards, and quality attributes," Sensors (Switzerland). 2014.

[4] E. Borelli et al., "HABITAT: An IoT solution for independent elderly," Sensors (Switzerland), 2019.

[5] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in ESANN 2013 proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2013.

[6] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," Appl. Soft Comput. J., vol. 62, pp. 915–922, 2018.

[7] R. A. Voicu, C. Dobre, L. Bajenaru, and R. I. Ciobanu, "Human physical activity recognition using smartphone sensors," Sensors (Switzerland), vol. 19, no. 3, Feb. 2019.

[8] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," ACM SIGKDD Explor. Newsl., 2011.

[9] C. A. Ronao and S. B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," Expert Syst. Appl., vol. 59, pp. 235–244, 2016.

[10] X. Shi, Y. Li, F. Zhou, and L. Liu, "Human Activity Recognition Based on Deep Learning Method," 2018 Int. Conf. Radar, RADAR 2018, 2018.

[11] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," Pattern Recognit. Lett., 2019.

[12] H. Friday Nweke, T. Ying Wah, and U. Alo, "Deep Learning Algorithms for Human Activity Recognition using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges Mobile Cloud Computing View project Novel Deep Learning Architecture for Physical Activities assessment, mental Res," vol. 105, pp. 233–261, 2018.

[13] M. Zeng et al., "Convolutional Neural Networks for human activity recognition using mobile sensors Article," pp. 381–388, 2014.

[14] S. M. Lee, S. M., Cho, H., & Yoon, "Human Activity Recognition From Accelerometer Data Using Convolutional Neural Network," IEEE Int. Conf. Big Data Smart Comput. (BigComp)., vol. 62, pp. 131–134, 2017.

[15] Y. Chen, K. Zhong, J. Zhang, Q. Sun, and X. Zhao, "LSTM Networks for Mobile Human Activity Recognition," no. Icaita, pp. 50–53, 2016.

[16] S. Yu and L. Qin, "Human activity recognition with smartphone inertial sensors using bidir-LSTM networks," Proc. - 2018 3rd Int. Conf. Mech. Control Comput. Eng. ICMCCE 2018, pp. 219–224, 2018.

[17] F. Hernández, F. Luis Suárez, Javier Villamizar, and Miguel Altuve. "Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network." In 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), pp. 1-5. IEEE, 2019.