UltaNet: An Antithesis Neural Network for Recognizing Human Activity Using Inertial Sensors Signals

Hamza Ali Imran

Department of Computing School of Electrical Engineering & Computer Science, National University of Sciences and Technology (NUST), Islamabad 46300, Pakistan

Manuscript received November 23, 2021; accepted December 30, 2021. Date of publication January 4, 2022; date of current version January 18, 2022.

Abstract—Human activity recognition (HAR) is an essential component of ambient assistive living. HAR has traditionally relied on computer vision techniques. However, it has several drawbacks, including lack of privacy, higher operational costs, and being constrained by the number of spaces available for cameras, so it cannot be used for applications that require long-term monitoring of people. The use of initial sensors is proving to be a vital solution for HAR. Smartphones and smartwatches have embedded accelerometers and gyroscope sensors that help deep neural networks for reliable activity recognition and classification problems. In this letter, we have presented a novel deep neural network model, which is opposite to the traditional models. It has a gated recurrent unit layer followed by different kernel-sized convolutional layers. Our model has been evaluated on an openly available dataset by the wireless sensor data mining lab, and comparison with previously existing models proves that it outperforms them.

Index Terms—Sensor signal processing, ambient assistive living, antithesis model, human activity recognition (HAR), inertial signal processing.

I. INTRODUCTION

Monitoring human activity is a critical component of ambient assisted living. This is important for elderly people who need to be monitored on a regular basis. Health care, sports, personal fitness, social networking, ambient supported living, human–computer interface, aged care, rehabilitation, entertainment, surveillance, and so on [1], [2] are among other applications. One of the goals of the body area network, which is an extension of wireless sensor network [3], is to recognize human activity. Traditionally, computer vision techniques have been utilized for human activity recognition (HAR) [4], [5], which has numerous issues, such as privacy, effect of ambient factors, limited portability, increased operational costs, obstruction, and so on. A new trend in the use of sensors, particularly inertial sensors, has lately arisen [6], [7]. Fig. 1 shows the categorization of technologies utilized for HAR. In the case of HAR based on sensors, there are two choices. Regular monitoring applications are hampered by the use of environment-based sensors.

There are various advantages of using sensor data instead of traditional computer vision algorithms. Almost all of the constraints of visualization techniques are solved by the utilization of sensor data. HAR employs a variety of sensors, including accelerometers, gyroscopes, and heart rate monitors. The use of DNN and ML approaches for categorization utilizing sensor data is well documented in the literature.

There are various examples of DNN that use a CNN–RNN model to handle inertial sensor data for HAR. However, we investigated into it, and offered a model that is the polar opposite. As a consequence, we decided to call it "Ulta," an Urdu term. "Reversed, contrary, or twisted" is what it means. The presented model is evaluated on WISDM 2011 dataset [7]. Various research have utilized this dataset to evaluate HAR models. A comparison with previously presented studies reveals that the current model outperforms the state of the art. Summarizing the key contributions of this letter as follows:

Corresponding author: H. A. Imran (e-mail: himran.mscs18seecs@seecs.edu.pk).

1) We have presented an antithesis nature model, which works on raw inertial sensors data for HAR.
2) The presented model has been evaluated on WISDM 2011 dataset [7], and comparison has been done with previous studies.

The rest of this letter is organized as follows. Section II presented a brief synopsis of previous studies. Section III shares details on architecture of presented approach. Section III-C presents details of the used dataset. Section IV discusses the results from experiments. Finally, Section V concludes this letter.

II. RELATED WORK

Imran and Latif [8] presented a novel model with inception-like modules, having three different kernel sizes. The feature maps generated by different layers are then concatenated. They call these modules InceptionDense. The presented model is evaluated on two datasets. Comparison with the study is presented in the comparison section. Peppas *et al.* [9] calculated a couple of statistical features and applied CNN type model on it. The WISDM dataset achieved 94.18% accuracy. Xia *et al.* [10] used a similar nature antithesis type of model, used LSTM units before CNN and have achieved 95.85% accuracy. Mehmood *et al.* [11] presented a densenet-inspired CNN, which concatenated the feature maps of all the CNN layers presented in the model. The model is evaluated on three different datasets, one of which is the WISDM dataset. They have achieved 94.65% accuracy. Ignatov [12] also presented a CNN type of model and evaluated it on the WISDM dataset and have achieved 93.32% accuracy. Zhang *et al.* [13] demonstrated the use of an attention mechanism for HAR using inertial sensors data. They have achieved 96.4% accuracy on it.

III. PROPOSED APPROACH

There are numerous examples of DNN that use a CNN–RNN model to handle inertial sensor data for HAR. However, we explored and offered a model that is the inverse of it. As a result, we named it "Ulta," which is an Urdu term. That means "reversed, opposite, or twisted." The presented model possesses the qualities of being the polar opposite of the traditional model.

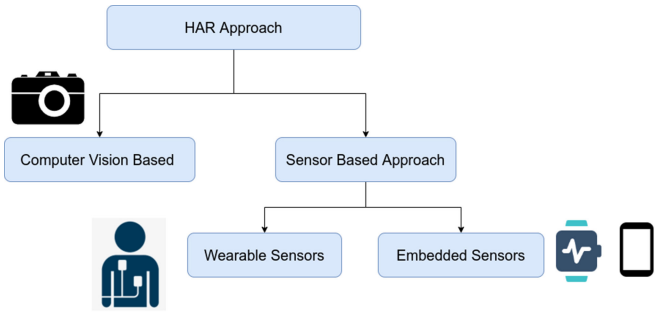

Fig. 1. HAR classification technologies.

A bidirectional gated recurrent unit (BiGRU) with 64 units is used in the model. The BiGRU's feature map is exhibited on an inception-like module. This uses (1×1, 1×3, 1×5, and 1×7) kernels with max-pooling. Before sending the input to (1×3, 1×5, and 1×7) sized kernels, 1×1 kernels were applied. After max-pooling, 1×1 kernels are also utilized to maintain the spatial dimensions of the feature maps. Imran and Latif [8] served as the inspiration for the inception-like model. A comparison with past studies is presented, with the results demonstrating that the provided model outperformed earlier experiments.

A. Model's Motivation

The motivation was that the GRU component of the model will extract features that affect consecutive segments, and then CNN will extract feature maps based on these snapshots. As a result, classification will be improved. One of the primary goals for building the model was to keep the number of parameters as low as possible in order to create a less complex, edge-suitable model. The provided model comprises a total of "35 492" parameters, of which "35 236" are trainable and "256" are nontrainable. It categorizes six distinct activities based on raw accelerometer measurements from the smartphone's integrated sensor. Fig. 2 shows the model diagram of the presented model. We used raw sensor data to keep the solution completely deep learning. The raw sensor data was segmented before being fed into the model.

B. Segmentation Size and Step Size

As a DNN has a fixed input size, therefore, segmentation of the input signal is necessary. The size of the segmentation window is also an important factor; we have evaluated its importance by empirical analysis and found a size of 128 with step size 32 to be the best performing one. The selection of these parameters was done once the model was finalized. For finalizing the model, we first fixed the segmentation window size to 128 and stepping size of 64. The overall pipeline of the project is shown in Fig. 3.

C. Dataset Used

The dataset used for the evaluation of the model is from the WISDM lab [7]. It has raw sensor values of the accelerometer of the smartphone. Fig. 4 shows the class distribution of the dataset. The dataset is unbalanced, with the "Walking" class dominating with 424 397 items, accounting for 38.6% of the total dataset. The class "Jogging" has 342 176 entries, accounting for 31.15% of the dataset, class "Upstairs" has 122 869 entries, accounting for 11.18%, class "Downstairs" has 100 427 entries, accounting for 9.14%, class "Sitting" has 59 939

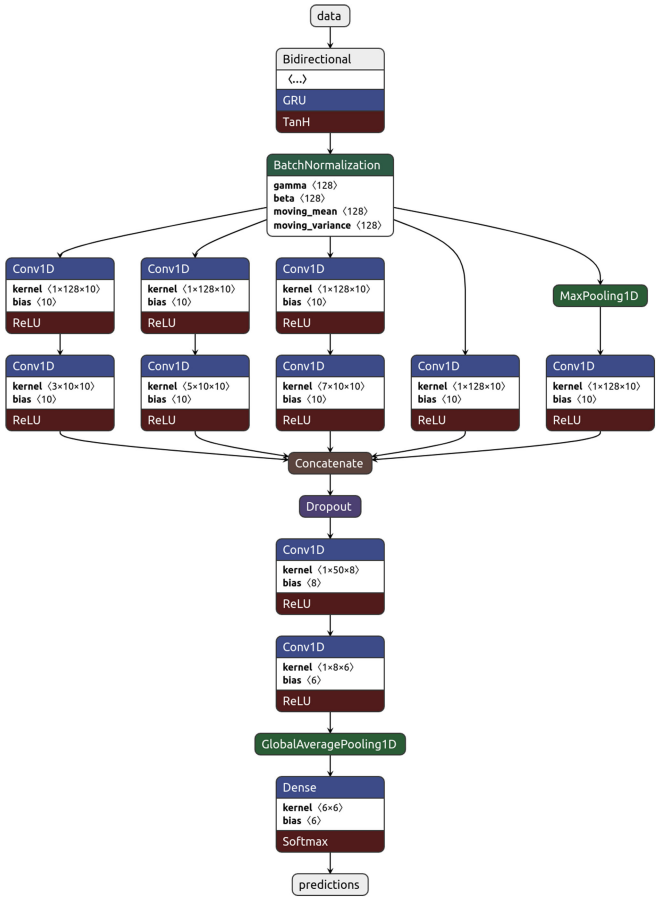

Fig. 2. UltaNet: The proposed antithesis model.

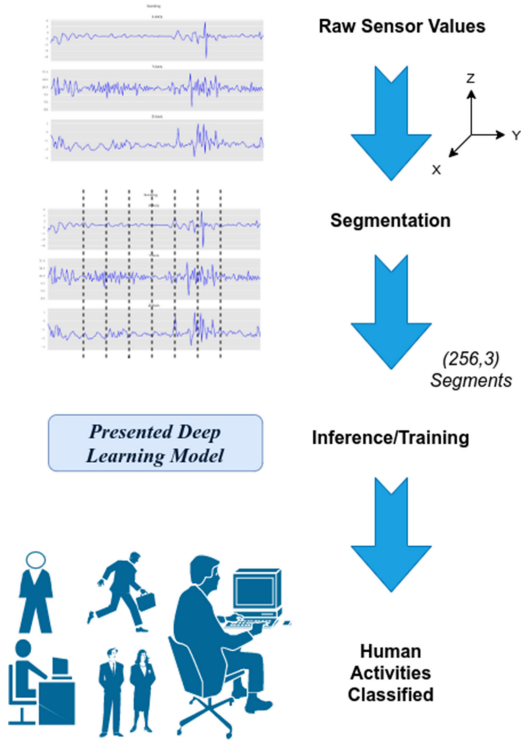

Fig. 3. Overall pipeline of the project.

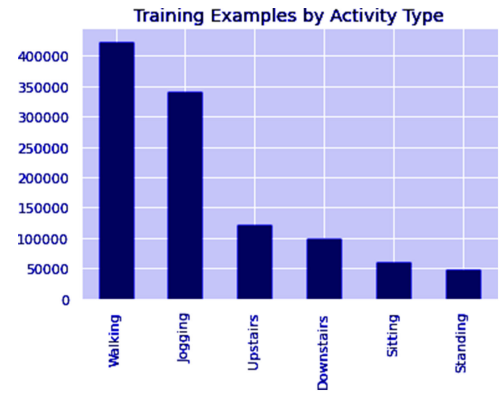

Fig. 4. Class distribution of the used dataset.

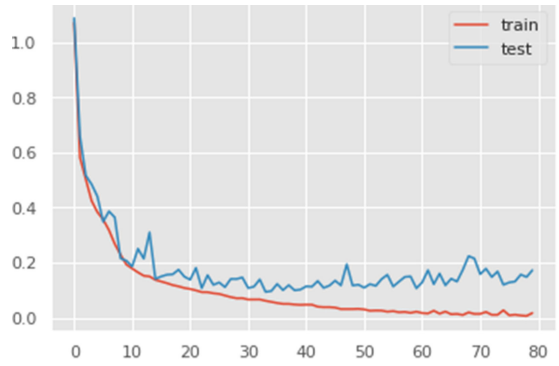

Fig. 5. Accuracies plot for test and training data.

entries, accounting for 5.45%, and class "Standing" has 48 395 entries, accounting for 4.40% of the dataset.

The dataset was divided into 70% set 15% testing and 15% cross-validation sets.

IV. RESULTS

The epoch size set was 80, and the batch size set was 120. These values were selected empirically by monitoring the loss and accuracy plots. Fig. 5 shows the accuracy plots. The optimizer used was "adam."

We have achieved a cross-validation accuracy of 97.7% and test accuracy of 97.2%. The confusion matrix for the dataset and trained model is shown in Fig. 6. The most confused classes are *Downstairs* and *Upstairs*, which have the confusion of just 2.3%. The reason of these two cases getting confused is that the movement of the body in performing them is very similar. The best performing case is of *Standing*, which has 100% classification accuracy. The reason is that this activity is different from all other activities, reason is that it does not require any movement, and hence is easiest to be classified among classes involving movements.

Because the dataset is unbalanced, we have reported F1, recall, and precision. Fig. 7 shows the detailed report. The average values for F1, recall, and precision all are 97%.

A. Comparison

The presented model is compared with other studies that used the same dataset for evaluating their solutions. Table 1 gives the comparison. The comparison showed that our model has outperformed the existing works.

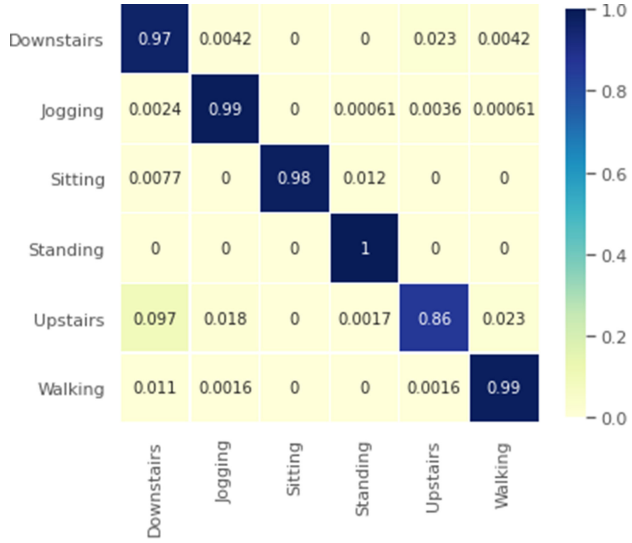

Fig. 6. Accuracies plot for test and training data.

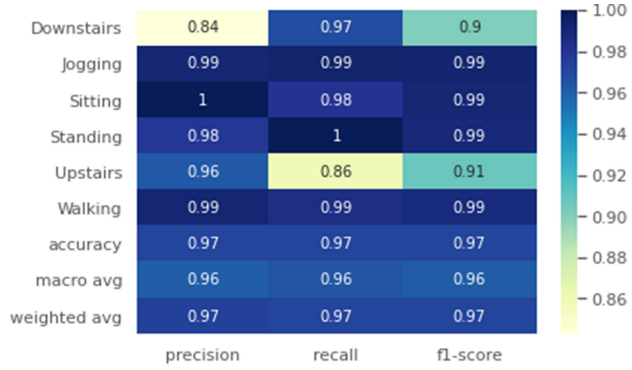

Fig. 7. Performance report for model classification.

TABLE 1. Performance Comparison.

Type & Reference	Accuracy
8 CNN	95.26%
10 LSTM-CNN	95.85%
11 CNN	94.65 %
12 CNN	93.32%
13 CNN with an attention mechanism	96.4%
Presented Model	97.2 %

V. CONCLUSION

For HAR, we offered an opposite-nature model. With a 97% F1 score on the WISDM dataset, it performed admirably. When compared to past research, it outperformed them. The major takeaway from this letter is that using RNN modules before applying CNNs can also yield reasonable results, despite the fact that this method is not often utilized or researched.

REFERENCES

[1] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," *Expert Syst. Appl.*, vol. 105, pp. 233–261, 2018.

[2] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino, "Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges," *Inf. Fusion*, vol. 35, pp. 68–80, 2017.

[3] H. A. Qazi, U. Jahangir, B. M. Yousuf, and A. Noor, "Human action recognition using SIFT and HOG method," in *Proc. IEEE Int. Conf. Inf. Commun. Technol.*, 2017, pp. 6–10.

[4] S.-R. Ke, H. L. U. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," *Computers*, vol. 2, no. 2, pp. 88–131, 2013.

[5] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using convolutional neural network," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, 2017, pp. 131–134.

[6] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Gener. Comput. Syst.*, vol. 81, pp. 307–313, 2018.

[7] J. W. Lockhart, G. M. Weiss, J. C. Xue, S. T. Gallagher, A. B. Grosner, and T. T. Pulickal, "Design considerations for the WISDM smart phone-based sensor mining architecture," in *Proc. 5th Int. Workshop Knowl. Discov. Sensor Data*, 2011, pp. 25–33.

[8] H. A. Imran and U. Latif, "HHARNet: Taking inspiration from inception and dense networks for human activity recognition using inertial sensors," in *Proc. IEEE 17th Int. Conf. Smart Communities: Improving Qual. Life Using ICT, IoT AI (HONET)*, 2020, pp. 24–27.

[9] K. Peppas, A. C. Tsolakis, S. Krinidis, and D. Tzovaras, "Real-time physical activity recognition on smart mobile devices using convolutional neural networks," *Appl. Sci.*, vol. 10, no. 23, 2020, Art. no. 8482.

[10] K. Xia, J. Huang, and H. Wang, "LSTM-CNN architecture for human activity recognition," *IEEE Access*, vol. 8, pp. 56855–56866, 2020.

[11] K. Mehmood, H. A. Imran, and U. Latif, "HARDenseNet: A 1D DenseNet inspired convolutional neural network for human activity recognition with inertial sensors," in *Proc. IEEE 23rd Int. Multitopic Conf.*, 2020, pp. 1–6.

[12] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Appl. Soft Comput.*, vol. 62, pp. 915–922, 2018.

[13] H. Zhang, Z. Xiao, J. Wang, F. Li, and E. Szczerbicki, "A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1072–1080, Feb. 2020.