Journal of Advanced Zoology

ISSN: 0253-7214
Volume 44 Issue S-6 Year 2023 Page 1024:1030

Human Activity Recognition Using CNN and Lstm Deep Learning Algorithms

Vinaya R M[1], Geeta C Mara[2]

[1,2]School of Computing & IT, REVA University
Email: [1]vinnumegs@gmail.com, [2]geetac.mara@reva.edu.in

**Corresponding author's E-mail: vinnumegs@gmail.com*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 30 Nov 2023

Abstract

Human Activity Recognition recognizes and classifies the activities performed by the users or people based on the data collected from the sensors of special devices such as smart-watches, smartphones etc. It has become easy to collect a huge amount of data from inertial sensors that are embedded in wearable devices. An accelerometer and gyroscope sensors are most commonly used inertial sensors. There are various already available datasets, in our paper, we are using the Wireless Sensor Data Mining dataset which contains 1,098,207 data of 6 physical activities performed. In this paper, the activities we aim to classify are walking, jogging, going up and downstairs, standing, and sitting. There are various algorithms applied on the various datasets. In our paper, we use Convolutional Neural Network and Long Short-Term Memory deep learning algorithm on the data set, we split the data into training data [80%] and testing data [20%]. By using a confusion matrix, we recognize and classify the activities performed using maximum accuracy.

Keywords: *Human Activity Recognition, Sensors, Convolution Neural Network, Long Short-Term Memory, Confusion Matrix.*

1. Introduction

Human Activity Recognition [HAR] identifies and classifies the activities done by humans or users. It helps people to understand the type of action or motion performed by the users and also extract the details that a user's body needs to convey by the information extracted from physical activities. HAR is wide research area and are used in different fields such as analysing behaviour, monitoring patient's health, video recognition, recognition of gestures, image recognition etc.

HAR can be categorized based on where the sensors are placed in wearable devices. The data is in the form of signals are acquired from the various sensors that are mounted in the devices. Accelerometer and Gyroscope sensors are most frequently used among all other sensors that are embedded to collect the signals for recognizing physical activities performed.

Sensors are affordable, consumes less power, and are less dependent on surroundings. Therefore, activity recognition using devices such as mobile sensors and smart watches has become more popular due to its portability nature. Various researches has in this particular field were been explored as they are portable and easily available.

Deep learning [DL] is a part of Machine Learning [ML] inspired by the working and functioning of human brain. Our human brain is very complex and it contains basic and functional cells called neurons that form a complex network (a neural network), that helps us in coordination with our surroundings. These models are called as ANN tries to represent the working of brain to solve complicated problems. Several neural networks have been implemented for several different problems.

Deep learning models produces better performance compared to ML models in performing difficult problems such as recognizing speech, images, detecting objects, face recognition and many more. These problems may look very simple to us (humans), but for a system, they are extremely complex.

Motivation

The cloud computing and sensor technologies for the sensor-based HAR is growing very fast, has become more popular and is widely used in many applications. Researchers have researched and explored on various different types of sensing technologies in the recognition of the activities performed by the humans to improve the classification and prediction using accuracy rates.

Contributions

In our paper, we use WISDM dataset to recognize human activity. There are various ML and DL algorithms used for recognizing activity and classifying them. Here, we use LSTM and CNN neural network models to perform the recognition.

Organization

The paper is sorted as follows, Section II Related work which contains the previous works done by the authors in HAR field. Section III. Is the proposed methodology which consists of the algorithm used and the dataset description, Section IV is about our results obtained through our proposed model followed by conclusion and references done in our paper.

RELATED WORK

In paper [1], the authors proposed deep multi-channel learning architecture which is a combination of CNN and LSTM on two datasets i.e., WISDM and MHEALTH, the features extracted from the multi-modal sensing devices to recognize the activities accurately. The convolution layers maps the features directly and the LSTM is used on the features to enhance the identification of the activity. Their proposed model gave an accuracy of 97%.

The authors proposed a system capable of recognizing regular activities of patients undergoing dialysis to monitor their exercise [2]. The wearable device is embedded with inertial sensors. They normalized the motion signals, and used feature learning method on the data. They have used 1-D CNN to obtain and classify from raw data. The accuracy obtained from the training and testing sample to recognise activities were 95.99% and 93.77%, respectively.

A wireless multi-sensor activity monitoring system [3] was used on dataset by applying three neural network algorithms that is CNN, ConvLSTM, and LSTM. The activities are categorized as basic, complex. From their proposed model, Convolution LSTM outperforms the other two networks that is CNN and LSTM for each activity category.

A combination of convolutional with LSTM was proposed in the paper [4]. The raw data was collected from the mobile sensors and was fed to the LSTM followed by the convolutional layers. They evaluated on three available datasets and obtained 95.78%, 95.85% and 92.63% accuracy. LSTM-CNN model showed better performance and generalization rate was also better compared to other models.

A deep learning hybrid network was proposed for HAR [5]. They have used Convolution LSTM and the features were aggregated which were obtained from convolutional layer with the help of the Multi-layer Feature Aggregation Model [MFAM] on 2 datasets- OPPORTUNITY and UniMiB-SHAR. 95% accuracy was obtained.

CNN-LSTM classifier was trained on the PAMAP2 and HARD datasets [6]. They proposed an approach to improve cross-subject performance. It was noticed that the performance gains reduced the use of large dataset and annotated procedures compared to other existing models.

A combination of 3-Dimensional CNN and LSTM was fed by optical flow and by auxiliary information over video frames for HAR [7]. The proposed model exhibited good prediction success rate compared to other DL models to recognize human activity.

Feed forward network of a single layer, with extracted features, is used with LSTM [8]. LSTM is used to learn the features of raw data from the sensors, and feed forward network to learn how it generalizes the data. The combination of deep LSTM and feed forward network was able to outperform other models. The accuracy for the recognition of human activity is 97.7% with a smartphone sensor.

A system to recognize action using accelerometer and gyroscope was proposed [9] on the ETEXWELD dataset and UCI-HAR dataset. Deep Learning algorithms like CNN and LSTM with machine learning was used on the data. 97.4% accuracy was obtained by using the model on recognizing activity using sensor.

The data extracted with the help of Short-Time Fourier Transformation is generally converted into a 2D spectrogram [10]. The radar spectrograms was used as a time sequencing channel and 1-D CNN and LSTM algorithms are trained on the dataset. The proposed model could extract features of the data from radar and achieved good recognition accuracy rate and complexity is less compared to the existing methods.

The main objective of this paper is about classifying and recognizing the human activity. There are various researchers who approached the classification of human activities using various deep learning algorithms on various datasets. In this paper, we approach this problem by using the CNN and the LSTM deep learning algorithm.

2. Materials And Methods

To perform our proposed model [Fig. 1] we follow the steps:

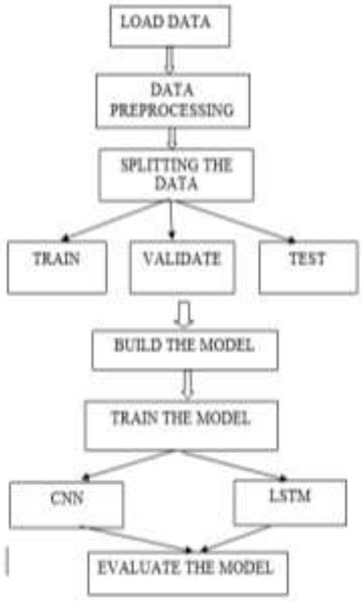

Fig. 1 Proposed model

A. *Steps involved*

Load data and Pre-processing the data: The data is loaded and the raw text data is converted into a comma separated file. The data is then pre-processed by normalizing, transforming and one-hot encoding. We normalize the data to the values zeroes and ones such that the system can understand. We then transform the data and the data is chunked before feeding it to the model. Then onehot encoding method is applied on the dataset. One-hot encoder labels the samples or the data as the neural network model understands 0 and 1.

Splitting the data: The dataset is split into train dataset, validation dataset and test dataset. The 80% data is taken as test dataset, validation dataset is taken to check how accurate the model can classify the unknown data and 20% of the data is used as test dataset.

Build and train the model: We are building and training the model using 2 deep learning algorithms that is the CNN and LSTM. We are using ADAM Optimizer to optimize our model.

Evaluating the model: The performance of the model is observed by plotting the graphs and acquiring the loss as well as the accuracy rates of the model we built. Accuracy shows how well our model performs when we train and test model and how accurate our model is which can be used to classify and recognize the activities performed. Loss represents the error rate, lower the loss rate better is the model, if the loss is high then, we need to back-propagate and adjust the weights corresponding to the inputs to get the desired output by minimizing the loss.

B. *Algorithms*

1) Convolutional Neural Network [CNN]: CNN is a neural network, most commonly used for analysing images. Generally, the image is divided in regions [Fig. 2], and each region is assigned to different nodes of the neural network. The 1st hidden layer performs convolution it consists of feature maps. The 2nd hidden layer performs local averaging and subsamples the data or image loaded to the model. In the hidden layer all the neurons are connected as the result of the previous neuron and is taken as the input to the next neuron and continues in this manner. The final hidden layer performs pooling and final convolution is performed in the output layer.

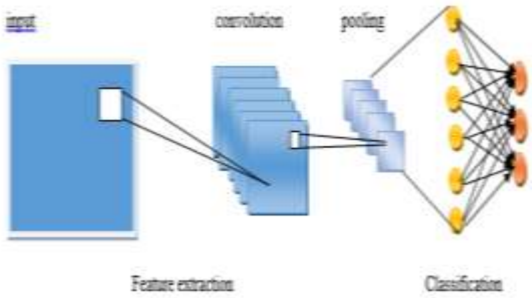

Fig. 1 Convolution Neural Network

2) Long Short Term Memory [LSTM]: LSTM is a recursive network model [Fig. 3], where the units are trained in a supervised manner using an optimization algorithm which is used to reduce the error and maximize the accuracy, in our paper we make use of ADAM optimizer with back-propagation to alter each weight of the input in the LSTM network to get less error in the last layer to corresponding weights. The difference between the expected output and the target output should be less. The error values are back-propagated from the output layer, the errors remain in the LSTM unit cells. We keep back-propagating until we get less error and the desired output by adjusting the corresponding weights of the respective inputs.

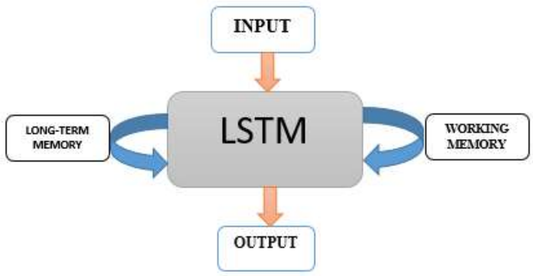

Fig. 2 LSTM

C. Dataset Description

The WISDM dataset consists of 1,098,207 [Fig. 4] data of 6 physical activities such as walking, jogging, going up and downstairs, sitting and standing. It has no missing value. This dataset is widely used for classification purposes. The signals are collected from various sensor based devices such as smart-phones and smart watches etc.

The dataset allows us to classify and predict or identify the physical activities performed by users or humans as they are measured from the accelerometer which is embedded in the smartphones. These signals are collected and stored in a raw form in a text file which is later converted to a csv file. These signals can be plotted against x, y, and z axis.

	user	activity	time	x	y	z
0	33	Jogging	49105962326000	-0.6946377	12.680544	0.50395286
1	33	Jogging	49106062271000	5.012288	11.264028	0.95342433
2	33	Jogging	49106112167000	4.903325	10.882658	-0.08172209
3	33	Jogging	49106222305000	-0.61291564	18.496431	3.0237172
4	33	Jogging	49106332290000	-1.1849703	12.108489	7.205164
5	33	Jogging	49106442306000	1.3756552	-2.4925237	-6.510526
6	33	Jogging	49106542312000	-0.61291564	10.56939	5.706926
7	33	Jogging	49106652389000	-0.50395286	13.947236	7.0553403
8	33	Jogging	49106762313000	-8.430995	11.413852	5.134871

Fig. 4 WISDM dataset

3. Results and Discussion

In our model, we have implemented two deep neural networks on the WISDM dataset. One is the convolutional neural network and the other is the Long Short-Term Memory. We have obtained an accuracy of 98.39% [Fig. 5] and loss of 0.2324% [Fig. 6] for the CNN model and an accuracy of 94.17% [Fig. 8] and a loss of 0.462% for the LSTM model.

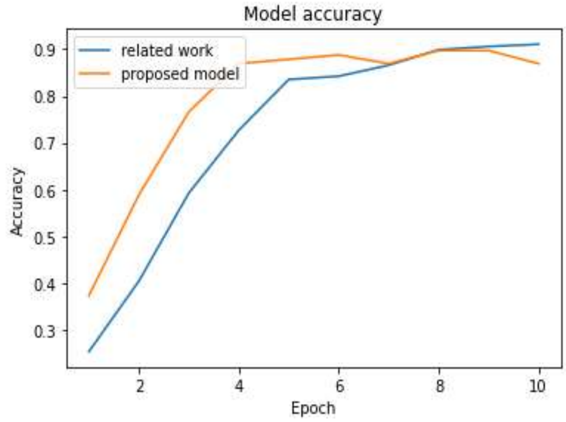

Fig. 5 Accuracy model on CNN

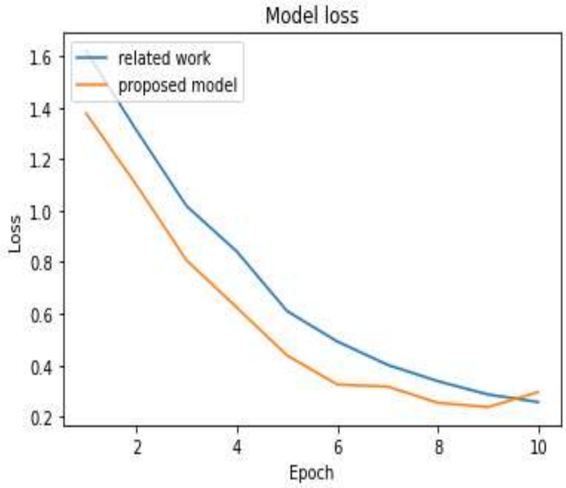

Fig. 6 Loss model on CNN

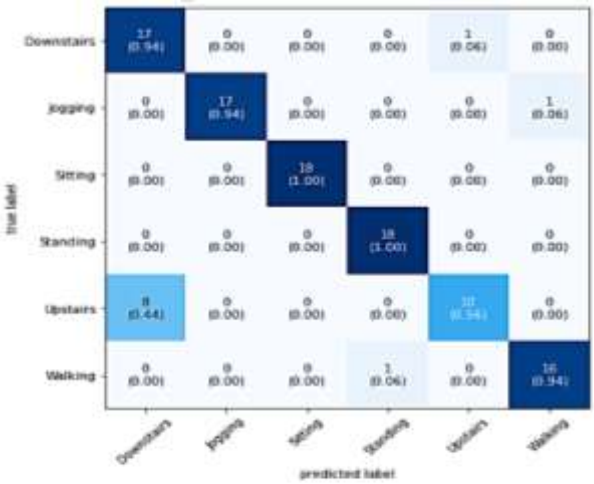

Fig. 7 Confusion Matrix

Here, we have plotted a confusion matrix [Fig. 7] true labels vs predicted labels for the human activities performed. Other than the activities downstairs and upstairs having similar signals, the rest of the activities can be identified with 100% accuracy. A confusion matrix is used to check the model if it is confused among two classes that is to see if the model is mislabelled one as another.

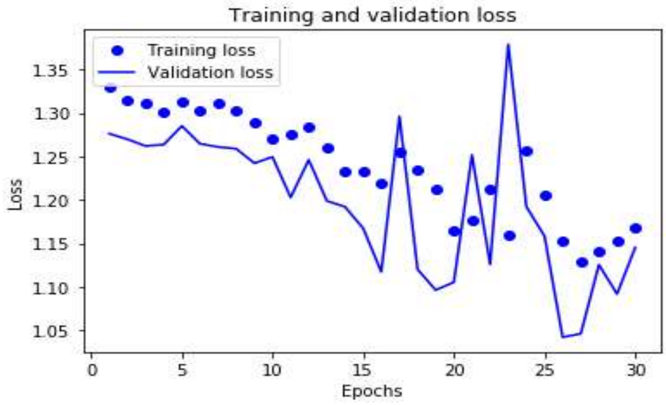

Fig. 8 Training and validation loss of LSTM

The Table I gives the accuracy obtained in our paper using the CNN and LSTM model.

TABLE I Accuracy of cnn and lstm

	Algorithm	**Classification Accuracy (%)**
1	Convolutional Neural Network	98.39
2.	Long Short Term Memory	94.17

The TABLE II gives the comparison with previous work. We observe that our proposed model has more accuracy than the related [2] work. In the related work they have implemented CNN deep learning algorithm on UCI dataset and obtained 95.99% accuracy but in our proposed model we have obtained 98.39% for CNN model.

TABLE III comparison table

Reference Paper	**Algorithm**	**Accuracy**
Human Daily Activity Recognition Performed Using Wearable Inertial Sensors Combined With Deep Learning Algorithms [2]	CNN	95.99%
Proposed Model	CNN LSTM	98.39% 94.17%

4. Conclusion

HAR is a very challenging problem.Various ML and DL algorithms have been implemented on various already available datasets to recognize and predict the activities performed by humans. Here, in our work we have taken WISDM dataset which consists of 1,098,207 data of 6 activities performed by the subjects and implemented CNN and LSTM and obtained good results. When CNN is compared with LSTM we find that the CNN produces good accuracy of 98.39% whereas, LSTM produces an accuracy of 94.17%. By looking at the confusion matrix [Fig. 7] we can deduce that the activities can be recognized with maximum accuracy.

We can also implement various other deep learning algorithms on the other readily available human activity datasets and compare the accuracy between them.

References:

1. Ihianle, A. Nwajana, S. Ebenuwa, R. Otuka, K. Owa and M. Orisatoki, "A Deep Learning Approach for Human Activities Recognition From Multimodal Sensing Devices", *IEEE Access*, vol. 8, pp. 179028-179038, 2020. Available: 10.1109/access.2020.3027979.
2. Yen, J. Liao and Y. Huang, "Human Daily Activity Recognition Performed Using Wearable Inertial Sensors Combined With Deep Learning Algorithms", *IEEE Access*, vol. 8, pp. 174105-174114, 2020. Available: 10.1109/access.2020.3025938.

3. Ascioglu and Y. Senol, "Design of a Wearable Wireless Multi-Sensor Monitoring System and Application for Activity Recognition Using Deep Learning", *IEEE Access*, vol. 8, pp. 169183-169195, 2020. Available: 10.1109/access.2020.3024003.
4. Xia, J. Huang and H. Wang, "LSTM-CNN Architecture for Human Activity Recognition", *IEEE Access*, vol. 8, pp. 56855-56866, 2020. Available: 10.1109/access.2020.2982225.
5. Lv, X. Wang, L. Jin, Y. Xiao and M. Song, "A Hybrid Network Based on Dense Connection and Weighted Feature Aggregation for Human Activity Recognition", *IEEE Access*, vol. 8, pp. 68320-68332, 2020. Available: 10.1109/access.2020.2986246.
6. Leite and Y. Xiao, "Improving Cross-Subject Activity Recognition via Adversarial Learning", *IEEE Access*, vol. 8, pp. 90542-90554, 2020. Available: 10.1109/access.2020.2993818.
7. Tanberk, Z. Kilimci, D. Tukel, M. Uysal and S. Akyokus, "A Hybrid Deep Model Using Deep Learning and Dense Optical Flow Approaches for Human Activity Recognition", *IEEE Access*, vol. 8, pp. 19799-19809, 2020. Available: 10.1109/access.2020.2968529.
8. Chen, Le Zhang, Z. Cao and J. Guo, "Distilling the Knowledge From Handcrafted Features for Human Activity Recognition", *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4334-4342, 2018. Available: 10.1109/tii.2018.2789925.
9. Tufek, M. Yalcin, M. Altintas, F. Kalaoglu, Y. Li and S. Bahadir, "Human Action Recognition Using Deep Learning Methods on Limited Sensory Data", *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3101-3112, 2020. Available: 10.1109/jsen.2019.2956901.
10. 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. Available: 10.1109/access.2020.2971064.
11. Lu, Y. Wu, L. Feng and J. Song, "Deep Learning for Fall Detection: Three-Dimensional CNN Combined With LSTM on Video Kinematic Data", *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 314-323, 2019. Available: 10.1109/jbhi.2018.2808281.
12. Rustam et al., "Sensor-Based Human Activity Recognition Using Deep Stacked Multilayered Perceptron Model", *IEEE Access*, vol. 8, pp. 218898-218910, 2020. Available: 10.1109/access.2020.3041822.
13. Osayamwen and J. Tapamo, "Deep Learning Class Discrimination Based on Prior Probability for Human Activity Recognition", *IEEE Access*, vol. 7, pp. 14747-14756, 2019. Available: 10.1109/access.2019.2892118.