Hypermeters Optimization in Recurrent Neural Networks-LSTM Approach for Human Activity Recognition

Ayşenur Topbaş
Department of Computer Engineering
Sabahattin Zaim University
Istanbul, Turkey
aysenur.gencdogmus@izu.edu.tr

Alaa Ali Hameed
Department of Computer Engineering
Sabahattin Zaim University
Istanbul, Turkey
alaa.hameed@izu.edu.tr

Akhtar Jamil
Department of Computer Engineering
Sabahattin Zaim University
Istanbul, Turkey
akhtar.jamil@izu.edu.tr

***Abstract*— Human activity recognition (HAR) has become more popular with the increase of applications involving human-computer interaction. The problem of recognizing and classifying people's daily life activities is a very important and challenging issue in the field of artificial intelligence. In recent years, deep learning based techniques have achieved high accuracy for HAR. However, these methods require obtaining a optimal number parameters. In this study, we investigated Long Short-Term Memory (LSTM) for HAR from videos. In addition, hyperparameters of the model were obtained thourh a thorough search to optimize ther performance of the model. Specifically, we obtained optimal values for number of layers, batch size, epochs to obtained best accuracy for the model. Experiments were performed to evaluate the performance of the proposed model on WISDM dataset. The proposed model produced an overall 97.18% overall accuracy indicates that LSTM is an effective technique for HAR.**

Keywords*— *Human Activity Recognition, RNN, LSTM Neural Networks, Deep Learning

I. INTRODUCTION

Human Activity Recognition problem is an active research area as it can be applied in assorted applications such as human machine interactions and robotics. In particular, the widespread use of human-computer interactive applications has made the HAR subject even more interesting. The daily movements of people can be tracked and recorded, and activities can be classified for this activity data. Especially in the public health field, HAR is very useful. Monitoring daily activities will help predict many diseases. One of the most common sensors used to monitor human activities are accelerometers in smart phones. It is possible to record activity through the sensors in smart phones, which have become an integral part of our day. Human activities can be captured in three axes (x,y,z axes) through these accelerometers. This tri-axial data is used to classify people's daily activities like sitting, jogging, walking, downstairs, upstairs, standing.

Nowadays, recognition of human activities is a very common research problem. A lot of work has been done on this subject in recent years. In particular, artificial neural networks and deep learning methods are highly preferred methods. Researchers have also proposed a neural network, a network that combines convolutional layers with long short-term memory (LSTM). [1] Through this model, human activities can be automatically extracted and classified through a few model parameters. The performance of the proposed model has been evaluated on 3 public datasets. (UCI, WISDM, and OPPORTUNITY datasets.) Similarly, we trained the model we proposed in our study using WISDM dataset. Others have proposed a model named EnsemConvNet for human activietis recognition [2]. This model used three classifiers: Encoded-Net, CNN, and CNN with LSTM. Three benchmark data sets are used to evaluate their proposed model: WISDM, MobiAct and UniMiB SHAR. Finally, they have also compared their method with some well know deep learning technique including hybrid of CNN, Multi Headed CNN, and LSTM models.

Although deep learning algorithms show great success in HAR, but these are expensive in terms of processing and it become difficulty to carry out on edge devices. [3] proposes a novel model making it suitable on edge devices by requiring less computational power for HAR. This models name is Lightweight Deep Learning Model. And they developed their models on the RNN-LSTM model similar to our study.

In this paper, we propose the RNN-LSTM model, which is one of the deep learning methods to recognize human activities. We trained our model on WISDM dataset, which is a public dataset. We tested it to find the optimum values of the hyperparameters to improve the performance of the model. We have shown that with optimum hyperparameters, our model achieves higher accuracy than previous studies.

The rest of this paper can be summarized as follows. Section 2 will examine previous studies using deep learning methods of human activity recognition. In section 3, we will give detailed information about the dataset used in our study. In the 4th section, the model and the hyperparameters of the proposed model will be examined. Section 5 will show the experiment results. And we will summarize this research with a brief summary in the final section.

II. RELATED WORK

HAR has been an important and challengin research problems for years. Many researchers are working to reach the best classification model on this subject. There has been a lot of work on HAR before. This section mainly examines the studies that used deep learning methods before.

The advent of deep learning era has lead to development of innovative methods to solve various problems in HAR. For instance, in [4], authors proposed deep neural network using bi-directional residual cells with LSTM. The technique has several advantages, such as it can be combined in the two-way connection with negative time direction (reverse state) and positive time direction (forward state). Considering that the Deep LSTM neural network can automatically learn the feature representations and model long-term transient dependencies, the experimental results showed that the method was effective for HAR [5]. In [6], authors proposed mobile and wearable sensor-based human activity as a compilation that provides an in-depth summary of deep learning methods.

Nowadays, recent advances in deep learning methods have made automatic high-end feature extraction possible and performed promising in many areas. This study, which summarizes the literature in terms of deep learning model, deep learning applications and sensor method, also examines the latest developments in sensor-based deep learning activity recognition. [7]. In another study [8] where deep convolutional neural network is proposed to realize effective and efficient human activity recognition application, an almost perfect classification was obtained, especially for activities very similar to those already considered to be very difficult to classify.

In another study [9] in which the deep learning model was applied independently of the user, Convolutional Neural Networks were proposed for feature extraction. Similarly, in another study, the Long Term Short Term Memory (LSTM) architecture is designed to apply human activity recognition. Over 94% accuracy and less than 30% loss were achieved in the first 500 training epochs [10]. Some researchers have proposed a hierarchical deep learning method for the LSTM neural network. They named this model H-LSTM. After the smoothing and denoising preprocesses, it will be featured with the time-frequency domain method [11].

A single layer stacked LSTM model was used in a study [12] where the LSTM model was proposed for the recognition of data including six different human movements obtained via smart phones. The proposed network have tried on network UCI data. The performance of the model is calculated in terms of precision-call and average accuracy.

Recognizing human activities has become a very important problem in terms of human health as it gives information about the movement intensity of people. LSTM model was used in this study [13], which deals with both the intensity of movement and the classification of movements problem. In another study, many methods used in human activity recognition were examined and these methods were tested on 10 publicy datasets [14]. In particular, CNN, gated recurrent unit networks (GRU), LSTM, bidirectional LSTM (biLSTM), and deep belief networks (DBN) models were evaluated on their dataset.

The data obtained from the sensors in smart phones are used especially by machine learning methods to recognize human activities. A bidirectional long short-term memory (LSTM) network has been proposed to classify this data collected by gyroscopes and accelerometers in smartphones [15].

In our study, we optimized the hyperparameters of the LSTM model to improve the accuracy performance obtained from the above methods. We evaluated it on the WISDM data set, one of the most widely used public data sets. We demonstrated the effect of parameters in the model on accuracy and determined the optimum parameters.

III. DATASET DESCRIPTION

The WISDM dataset is used in this study which consists of a total of 1,098,209 samples. The percentage of total samples present in the data set for each activity is summarized in Table 1. There are six different human activities which include walking (Walk), jogging (jog), upstairs (Up), downstairs (Down), sitting (Sit), standing (Std).

It is worth noting that data distribution is not balanced. For instance, walking and jogging accounts for more than 30% of all activities while rest of the classes account for less than 12% each. This can adversly effect the classification accuracy of the classifiers as model can be biased towards the class with donimant samples.

The data in the dataset were collected from 36 different subjects. Each subject carried out daily activities by putting a smartphone in their front leg pockets. This sensor used while creating the data set is an accelerometer. This accelerometer has a sampling frequency of 20 Hz and is also a motion sensor built-in smartphones.

TABLE I: SUMMARY OF THE WISDM DATASET

Activities	*No. of Samples*	*Percentage*
Walking	424400	38.6%
Jogging	342177	31.2%
Upstairs	122689	11.2%
Downstairs	100427	9.1%
Sitting	59939	5.5%
Standing	48397	4.4%

IV. DEEP LEARNING ARCHITECTURE

A. Deep Learning Methodology

The deep learning model consists of artificial neural networks. ANNs are composed of neuron cells just like the human brain structure. Neurons are all interconnected in the network, and the way they connect affects the output. There are basically 3 layers in a neuron. These layers; input layer, hidden layer (s) and output layer.

- **Input Layer:** It is the layer that contains the input data. It transmits input data from the input layer to the first hidden layer.
- **Hidden Layer(s):** It is the layer on which mathematical calculations are made on the input data. The expression "deep" in the concept of deep learning indicates that the neural network has multiple hidden layers. One of the biggest challenges in creating artificial neural networks is determining the number of hidden layers as well as determining the number of neurons for each layer.
- **Output Layer:** It is the layer from which the output data is obtained.

One of the most difficult tasks in deep learning methods is training the artificial neural network. The reason for this is that a lot of computational power and a very large data set are needed in order to apply the deep learning methods. In our study, we conducted LSTM (Long Short-Term Memory) neural network training for human activity data obtained from sensors.

B. Creating the LSTM Neural Network

We created the LSTM model, which is a specialized method of RNN, which is one of the deep learning methods commonly used in HAR. A neuron; LSTM consists of point processes and layers. These are several layers, data input and output layers that act as gates to forget to feed the cell state. What holds long-term memory and context for the inputs and the network is called the cell state.

An LSTM consists of several neurons arranged in layers. It consists of several gates which are terms as input, output, and forget. It also maintain a cell state. This cell state resembles the long-term memory which is maintanied across the network.

LSTM is a type of RNN. The main feature of such networks is that these architectures have feedback links in their architecture. This makes them more suitable for processing sequential data. In addition, these model have also been used to process image data.

The input layer of the network must have same number of neurons as the dimension of features in the data set, the size of each segment of the segmented data, and batch size properties. The weights and biases of for each layer in the network were initialized with normal distribution in the LSTM. In this neural network, a single output is reached from multiple inputs.

The tensors of the input and output in our model are defined separately. In the model, the L2 regulator will be used and this regulator we use should also specify the lost. While creating the model, Adam optimizer was used.

AdamOptimizer: This optimizer has become more popular and has been widely used for it high efficiency. The weight update is done iteratively in the network during the training dataset. It calculates learning rates from the first and second estimates of the gradients for the different parameters.

The activation function we use in our model is the softmax activation function.

Softmax Activation Function: it transforms the input vectors into probabilities. These probabilities are directly proportional to the scale of each element in the input vector [17]. The softmax activation function used in machine learning algorithms of applications is that it commonly acts as a neural network model activation. In cases where more than two classifications are required, this function has produced highly optimal results for neural networks. It ensures that the probability of the input belonging to a certain class is determined by producing values in the range of 0-1.

The parameters used in the proposed LSTM neural network and whose effects on accuracy are measured are as follows.

Epoch: To train the models, it is necessary to repeat the training not just once, but over and over again. Weights are updated for each training round. In order to find the most suitable weight value for the model, weight calculation is made for each new training data. The most suitable weight values for the model are thus calculated. Each of these steps is called an epoch.

Batch Size: It is the memory used by the model while it is being trained. It can be in the form of 2 and its multiples (4, 8, 16, 32, 64, ..., 512).

Number of Layers: Especially in complex problem examples, the most important feature that makes deep learning methods different from other artificial neural networks is the number of layers.

V. EXPERIMENTAL RESULTS

In this section, the results of the tests performed to examine the effect of the values of the hyperparameters (epoch, batch size, numbers of LSTM layers) on the accuracy of the proposed model are presented.

A. Effects of Number of LSTM Layers on Accuracy

Our aim in this section is to determine the effect of the number of LSTM layers in the model on accuracy. Layer number is very important especially in complex problems. In order to determine the appropriate number of layers for the dataset we use in our model, the number of epochs was kept constant as 200 and batch size as 32 in the tests in this section. Tests were carried out as LSTM layer numbers 2, 4 and 6. The accuracy and loss values of the tests are shown in Table 2.

TABLE 2: EFFECT OF LSTM LAYER SIZE ON ACCURACY AND LOSS (EPOCHS=200, BATCH SİZE=32)

LSTM Layers	*Accuracy*	*Loss*
2	95.99%	27.33%
4	**97.18%**	**27.33%**
6	95.66%	30.29%

The number of layers was initially set to 2 and the behavior of the model was observed by increasing it (4 and 6 layers). The dataset was divided into train (70%) and test (30%) data subset. As a result of our tests, when the effect of LSTM layer number on accuracy is examined, increasing the number of layers from 2 to 4 increased the accuracy and decreased the loss. However, when we built the model with 6 layers, the accuracy decreased and the loss increased. Fig. 1 shows the visualization of results for (a) train loss, (b) train accuracy, (c) test loss and (d) test accuracy.

B. Effects of Number of Epochs on Accuracy

Our aim in this section is to determine the effect of the number of epochs in the model on accuracy. In order to determine the appropriate number of epochs for the dataset we use in our model, the number of LSTM layers was kept constant as 2 and batch size as 32 in the tests in this section. Tests were carried out as number of epochs 300, 350 and 400. The accuracy and loss values of the tests are shown in Table 3.

TABLE 3: THE EFFECT OF EPOCHS ON ACCURACY AND LOSS

LSTM Layer=2, Batch size=32		
Epochs	*Accuracy*	*Loss*
300	96.30%	26.67%
350	96.54%	25.93%
400	**96.57%**	**25.53%**

The number of epochs was initially set to 300 and the behavior of the model was observed by increasing it (350 and 400 epochs). The train and test data in the LSTM neural network constituted 70% training data and 30% test data. As a result of our tests, when the effect of the number of epochs on accuracy is examined, with the increase of the number of epochs, the accuracy increased and the loss decreased. Fig. 2 shows the optimal results obtained for epochs (epochs number=400). It shows (a) train loss, (b) train accuracy, (c) test loss and (d) test accuracy.

C. *Effects of Batch Size on Accuracy*

Our aim in this section is to determine the effect of the batch size in the model on accuracy. In order to determine the appropriate batch size for the dataset we use in our model, the number of LSTM layers was kept constant as 2 and number of epochs as 200 in the tests in this section. Tests were carried out as batch size 32, 64 and 128. The accuracy and loss values of the tests are shown in Table 4.

TABLE 4: THE EFFECT OF NUMBER OF EPOCHS ON ACCURACY AND LOSS

LSTM Layer=2, Number of Epochs=200		
Epochs	***Accuracy***	***Loss***
32	96.30%	26.67%
64	96.54%	25.93%
128	**96.57%**	**25.53%**

The batch size was initially set to 32 and the behavior of the model was observed by increasing it (64 and 128). The train and test data in the LSTM neural network constituted 70% training data and 30% test data. As a result of our tests, when the effect of the batch size on accuracy is examined, with the increase of batch size, the accuracy increased and the loss decreased. Fig. 3 shows highest accuracy obtained with batch size (batch size = 128). It shows (a) train loss graph, (b) train accuracy graph, (c) test loss graph and (d) test accuracy graph.

D. *Optimized Hyperparameters Results*

We evaluated different values for the hyperparameters in the model and tried to find the optimum values. These the number of epochs, number of LSTM layers and the batch size. We empirically obtained the best performance of the model using 4 LSTM layers, batch size = 128 and number of epochs = 400.

With the optimum hyperparameters, we achieved the highest accuracy as 97.18% and the loss value as 27.33%. Fig. 4 shows the visual results obtained for these optimal paperaemters: (LSTM layer number=4, epochs number=400, batch size = 128). It shows (a) train loss, (b) train accuracy, (c) test loss and (d) test accuracy.

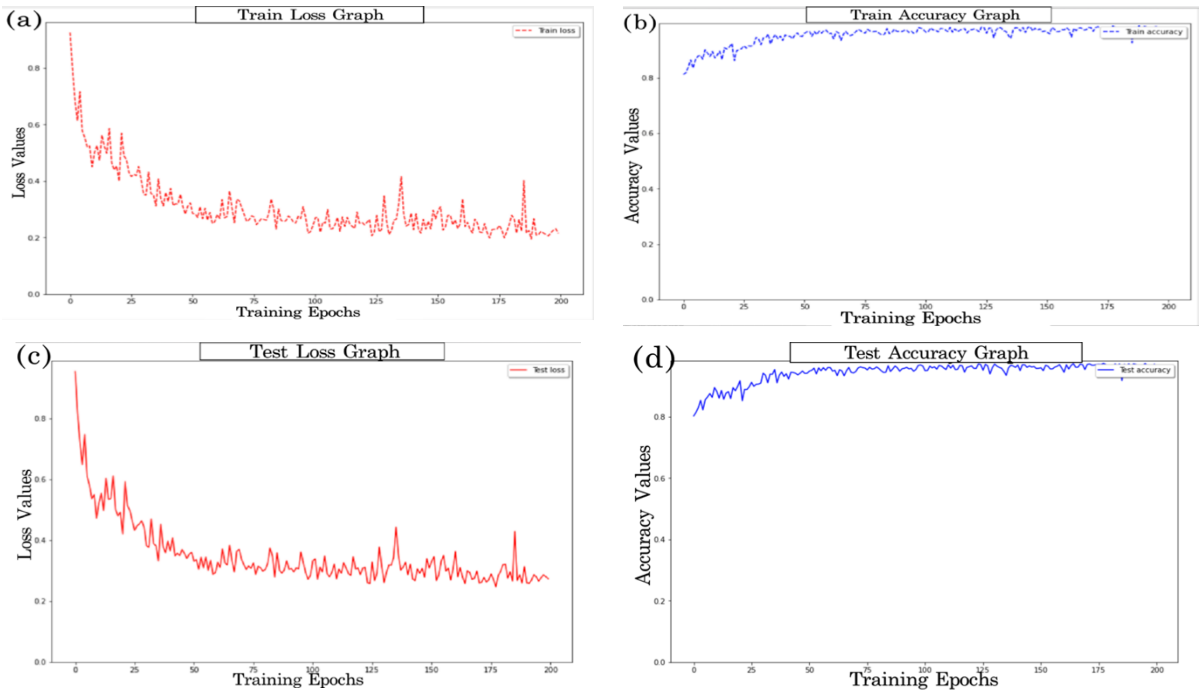

Fig 1. Train, test loss and train, test accuracy for LSTM with 4 layers.

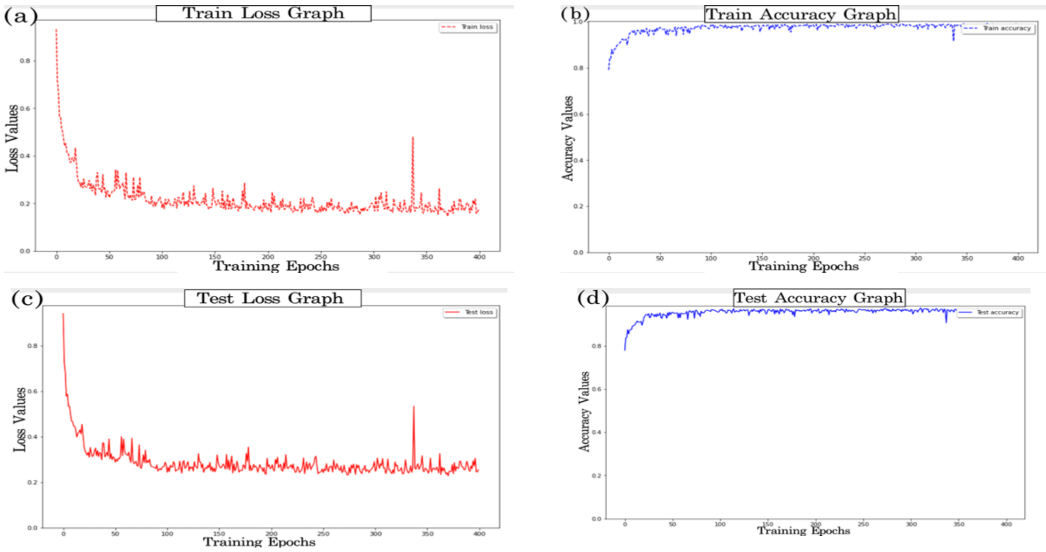

Fig 2. Train, test loss and train, test accuracy for LSTM with number of 400 epochs.

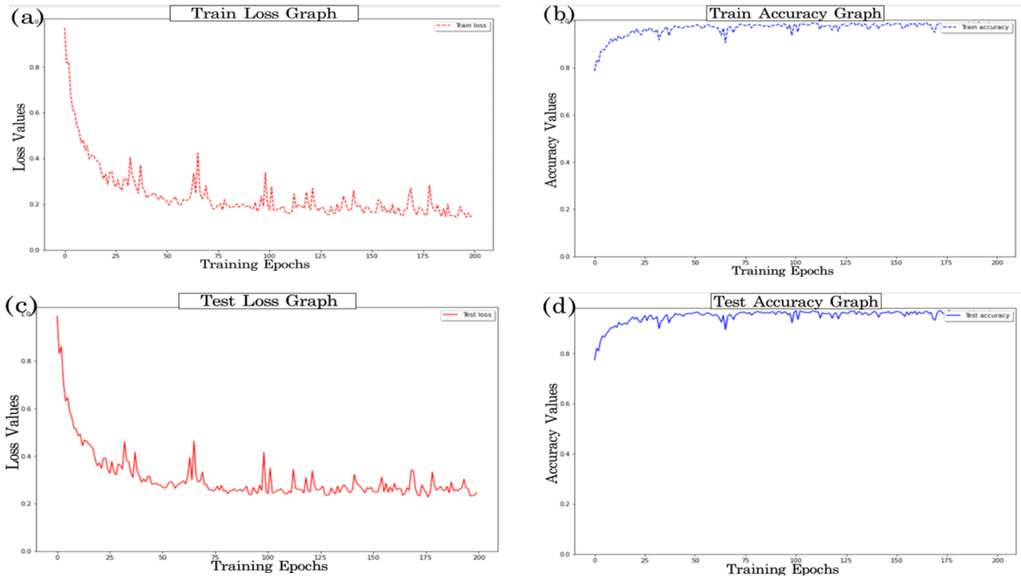

Fig 3. Train loss, train accuracy, test loss and test accuracy graphs of the model with batch size 32.

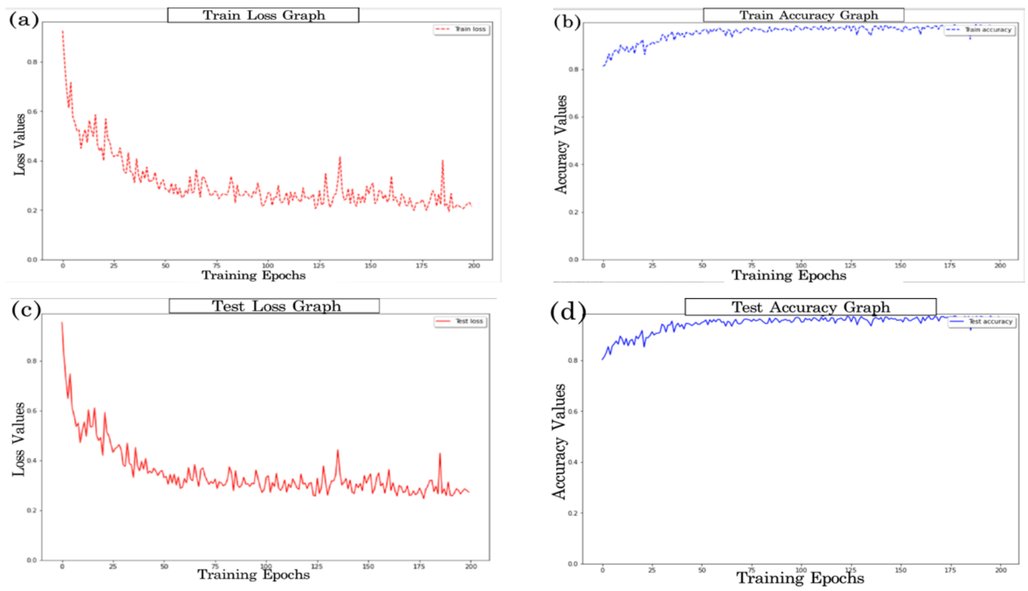

Fig 4. Train loss, train accuracy, test loss and test accuracy graphs of the model with optimum hyperparameters (number of LSTM layers=4, number of epochs=400, batch size=32)

VI. CONCLUSIONS

This study investigated the application of LSTM for HAR. Moreover, the hyperparameters were fine-tuned to obtain optimal results for the model as these parameters affect the accuracy and loss. The hyperparameters which were evaluated include: number of epochs, number of LSTM layers and batch size. LSTM layers were varied between 2-6, batch size in the range 32 – 128 with increments of 32 and the epoch values were tested for 300, 350 and 400. As a result of all hyperparameter tests, the optimum hyperparameter values were determined as LSTM layer number 4, epoch number 400 and batch size 128. And the highest accuracy achieved with these hyperparameters was 97.18%. In future, we will integrate LSTM with Parameter-Less Self-Organizing Map for hyperparameter fine-tuning.

REFERENCES

[1] Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access*, *8*, 56855-56866.

[2] Mukherjee, D., Mondal, R., Singh, P. K., Sarkar, R., & Bhattacharjee, D. (2020). EnsemConvNet: a deep learning approach for human activity recognition using smartphone sensors for healthcare applications. *Multimedia Tools and Applications*, *79*(41), 31663-31690.

[3] Agarwal, P., & Alam, M. (2020). A lightweight deep learning model for human activity recognition on edge devices. *Procedia Computer Science*, *167*, 2364-2373.

[4] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, "Deep residual bidir-lstm for human activity recognition using wearable sensors," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[5] Zhu, W., Lan, C., Xing, J., Zeng, W., Li, Y., Shen, L., & Xie, X. (2016). Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks. *arXiv preprint arXiv:1603.07772*.

[6] Nweke, H. F., Teh, Y. W., Al-Garadi, M. A., & Alo, U. R. (2018). Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Systems with Applications*, *105*, 233-261.

[7] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep learning for sensor-based activity recognition: A survey. *Pattern Recognition Letters*, *119*, 3-11.

[8] Ronao, C. A., & Cho, S. B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. *Expert systems with applications*, *59*, 235-244.

[9] Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing*, *62*, 915-922.

[10] Pienaar, S. W., & Malekian, R. (2019, August). Human activity recognition using LSTM-RNN deep neural network architecture. In *2019 IEEE 2nd Wireless Africa Conference (WAC)* (pp. 1-5). IEEE.

[11] Wang, L., & Liu, R. (2020). Human activity recognition based on wearable sensor using hierarchical deep LSTM networks. *Circuits, Systems, and Signal Processing*, *39*(2), 837-856.

[12] Ullah, M., Ullah, H., Khan, S. D., & Cheikh, F. A. (2019, October). Stacked Lstm Network for Human Activity Recognition Using Smartphone Data. In *2019 8th European Workshop on Visual Information Processing (EUVIP)* (pp. 175-180). IEEE.

[13] Barut, O., Zhou, L., & Luo, Y. (2020). Multi-task LSTM Model For Human Activity Recognition and Intensity Estimation using Wearable Sensor Data. *IEEE Internet of Things Journal*.

[14] Sansano, E., Montoliu, R., & Belmonte Fernández, Ó. (2020). A study of deep neural networks for human activity recognition. *Computational Intelligence*.

[15] Hernández, F., Suárez, L. F., Villamizar, J., & Altuve, M. (2019, April). 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.

[16] Time series prediction using lstm deep neural networks." [Online]. Available: https://www.altumintelligence.com/articles/a/Time-Series-Prediction-Using-LSTM-Deep-Neural-Networks

[17] "Softmax Activation Function with Python." [Online]. Available: https://machinelearningmastery.com/softmax-activation-function-with-python/