

# STACKED LSTM NETWORK FOR HUMAN ACTIVITY RECOGNITION USING SMARTPHONE DATA

*Mohib Ullah<sup>1</sup>, Habib Ullah<sup>2</sup>, Sultan Daud Khan<sup>2</sup>, Faouzi Alaya Cheikh<sup>1</sup>*

<sup>1</sup>Department of Computer Science, Norwegian Univeristy of Science and Technology, Norway

<sup>2</sup> College of Computer Science and Engineering, University of Ha'il, Saudi Arabia

## ABSTRACT

Sensor-based human activity recognition is an essential task for automatic behavior analysis for sports player, senior citizens, and IoT applications. The traditional approaches are based on hand-crafted features which use fixed mathematical rules to extract the features from the input data and are not capable of incremental learning. In this paper, we proposed a stacked long Short-term memory (LSTM) network for recognizing six human behaviors from the smartphone data. The network consists of a five LSTM cell that is trained end-to-end on the sensor data. The network is preceded by a single layer neural network that pre-processes the data for the stacked LSTM network. An  $L_2$  regularizer is used in the cost function which helps the network in generalization. The network is evaluated on public domain UCI dataset and quantitative results are compared against six state-of-the-art methods. The performance is calculated in terms of precision-recall and the average accuracy. The proposed network improves the average accuracy by 0.93% as compared to the closest state-of-the-art method without any manual feature engineering.

**Index Terms**— human activity recognition, Sports player,  $L_2$  regularizer, sensor-data, stacked LSTM.

## 1. INTRODUCTION

With the exponential growth of computing technology, wearable electronics are widespread in human communities for daily usage. Especially, the use of smartphones is persuasive irrespective of the economic status of an individual. Thanks to miniaturization technology which makes it possible to integrate many sensors in an averaged priced smartphone. With the daily usage of the smartphone, the embedded sensors like accelerometer and gyroscope produce a large amount of useful data that can be used to automatically predict and classify human activities. Potentially, human activity recognition can be used in elderly house [1] especially in the countries where the average old population is on the rise e.g. western world. Similarly, it can help in analyzing the moves of a sports player [2] and consequently help to improve player performance. Likewise, sensor-based human activity recognition plays a vital rule in smart homes and IoT based technologies

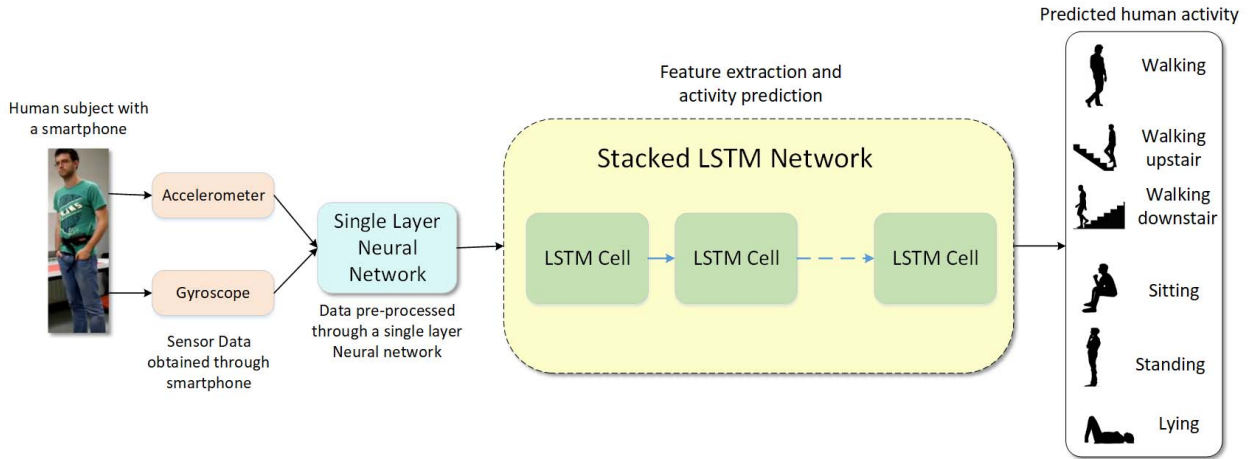
[3]. Due to its potential utilization in a wide spectrum of applications, it is an active topic of research in the computer vision and many breakthroughs have been achieved in the last few decades. In essence, human activity algorithm can be divided in the following two broad categories.

- Vision Based activity recondition
- Sensor based activity recognition

The aim of algorithms in both categories is the same. However, data acquisition and processing mode is very different. In the following, a brief overview of both categories is given.

### 1.1. Vision Based activity recondition

Vision-based activity recognition uses RGB or RGBD data to analyze human actions. Many techniques [4, 5, 6, 7, 8, 9, 10, 11, 12, 13] rely on holistic approaches and model human activity recognition as a scene understanding problem. For example, in [7, 6, 12, 13], first a dynamic system is defined from an optical flow field [14] of the video. Later, the stability analysis of the dynamic system is used to classify human behavior. Moreover, a social interaction model [7] is used to characterize the interaction between human in crowded scenes. Varol et al. [15] come up with a convolutional neural network (CNN) architecture named LTC-CNN (Long-term Temporal Convolution CNN) where they incorporate the extended temporal video information in the learning framework. They used raw pixel values and the optical flow fields as the representative features for modeling human actions. Different from visual data, Ullah et al. [16] used the information in Electroencephalography signal (EEG) to classify internal human emotions. They introduced an ensemble based learning algorithm which is trained on sparse coded discriminative subset of EEG channels. Similarly, Kong et al. [17] proposed a deep learning model where a CNN extract deep features from the video frames and the corresponding optical flow field and in the later stage fed to LSTM network for generating latent features which are consequently used for the action recognition. Tong et al. [18] proposed a two-stage approach wherein the first stage, a clustering technique is used for obtaining motion



**Fig. 1:** Proposed Approach consists of two networks. First, a single layer neural network pre-processes the data and then forwarded to the stacked LSTM network. The output of stacked LSTM is given to a six-way softmax that gives the probability of the six human activities under observation.

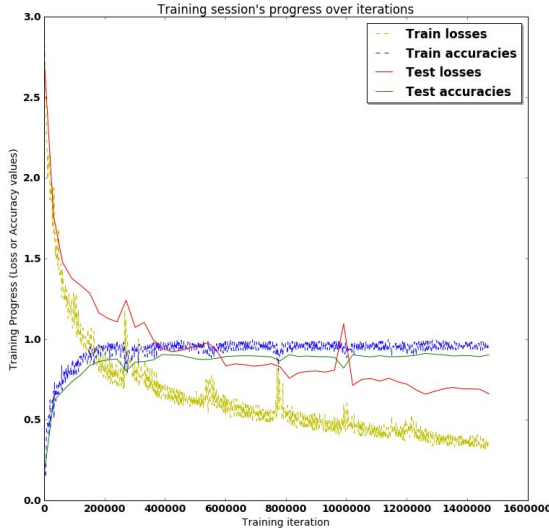
regions in the frame. In the second stage, matrix factorization with temporal dependencies constraints (TD-NMF) is proposed that exploit the spatiotemporal relationship of different subjects in the video for action recognition. A Siamese neural network has been proposed in [19] which can be used for tracking and activity classification. A comprehensive analysis of the latest vision-based action recognition can be viewed in the survey papers of Bux et al. [20] and Kong et al. [21].

## 1.2. Sensors Based activity recondition

Compared to vision-based action recognition, where the visual features (raw pixel, gradients, edges, orientation, etc.) and motion features (optical flow) are extracted from the scene, sensor-based recognition obtained the 1D data from the sensors. Kim et al. [22] used two sensor based hand-crafted features (mean, standard) in an ensemble of hidden Markov model for the activity recognition. The mean and standard deviation are obtained from the temporal window of raw sensor data. Compared to [22], Ronao et al. [23] proposed a two-stage hidden Markov model (CHMM). The two-stage model follows the modality of coarse to fine classification. Rather than using hand-crafted features, random forests variable importance measures is used to define the feature vector. Similarly, Anguita et al. [24] comes up with a computationally efficient model and incorporate fixed-point arithmetic in the feed forward phase of multi-class SVM model. Additionally, the One-Vs-All (OVA) strategy is adopted for the multi-class classification. From each temporal window, 17 value feature vector is calculated (mean, standard deviation, signal magnitude area, entropy, etc.) for training and testing the multi-class SVM. Seto et al. [25] proposed a clustering-based approach. Initially, the training data is segmented into the different cluster where each

cluster corresponds to an action. Based on the cluster, a reference template is generated for each action. The training and test samples are first evaluated against the templates and the distance between the sample and template is used as the feature vector. Consequently, this feature vector is used to train an SVM for action recognition. Hassan et al. [26] used means, median, and autoregression coefficients and processed it through kernel principal component analysis (KPCA) and linear discriminant analysis (LDA) for robustness. For activity recognition, a Deep Belief Network (DBN) is trained on the processed features. Hevesi et al. [27] combined gaze invariant features from the eye-tracker and features from a wrist-worn accelerometer to train a Naive Bayes classifier for classifying different physical activities of the construction worker (drilling, carrying, hand-tool activities, etc.). Li et al. [28] come up with an unsupervised learning technique for the automatic feature extraction. For the feature extraction, they used sparse auto-encoder while for the action classification, a radial Gaussian kernel SVM is trained. Similar to [24], the OVA strategy is adopted for the multi-class classification. Ignatov [29] proposed an ensemble of features approach where features extracted from a CNN are complemented by statistical features for better recognition.

Compared to these techniques, our approach doesn't rely on any feature engineering. Moreover, any post-processing is not required, and only raw data is used to train the network. The rest of the paper is organized in the following order. In section 2, a general overview of the proposed model is given. Different components of the proposed algorithm like the data normalization, stacked LSTM, the cost function, and the optimization strategy are given in section 3. Experimental setup, dataset, performance metric and the results are discussed in section 4. Section 5 concludes the paper with final remarks and future directions.



**Fig. 2:** The learning and testing curve of the underlying network is depicted. As the network trains, the training loss gradually reduces but after a certain point, the test accuracy stabilized which shows the optimal operating point of the network.

## 2. PROPOSED APPROACH

The block diagram of the proposed approach is given in Fig. 1. It mainly consists of two parts i.e. a single layer neural network and a network of stacked LSTM cells [30]. Initially, sensor data is obtained from the smartphone that is worn by a human subject. We used two types of sensor data i.e. accelerometer and the gyroscope. The raw sensor data is passed through a single layer neural network which acts like a pre-processing and normalized the input data for the succeeding network. The normalization is achieved through a linear discriminant function and ReLU activation. After that, the data is fed to the stacked LSTM network. The network consists of five LSTM cells which have learned the temporal dependencies of the sensor sequential data. The output of the stacked LSTM network is given to a six-way softmax which gives the individual probability of the six human behavior i.e. (walking, walking upstairs, walking downstairs, sitting, standing, lying). In the following section 3, network architecture and the cost function is briefly explained.

## 3. NETWORK ARCHITECTURE

In this section, different components of the proposed network is analyzed individually.

### 3.1. Normalization

The data obtained through the accelerometer and gyroscope have a different distribution. The range of values is also different. If the data is not normalized, the network can't

equally distribute the importance to each input and consequently larger values dominate the network training and a natural bias is learned by the network. To cope with such a problem, data is normalized that helps in removing the geometrical biases towards some of the dimensions of the input vector. There are different normalization strategies [31] like statistical normalization, Euclidean norm normalization, etc. In this work, we used a single layer network with ReLU activation to smooth out the discrepancies in the input data. In this way, the data coming from two sensors are treated in a fair manner. The normalized data is given as input to the 5 layer stacked LSTM network.

### 3.2. Cost function

In order to optimize the parameter of the underlying network, we used a standard cross entropy loss with a  $L_2$  regularizer. The  $L_2$  regularizer improves the generalization of the network and help the network in overfitting. Mathematically, the cost function is written as:

$$\mathcal{L} = \underbrace{-\frac{1}{B} \sum_{i=1}^B (Y_{(true)}^i \log(Y_{(predicted)}^i))}_{\text{Cross entropy loss}} + \underbrace{\lambda \sum_{j=1}^P \theta_j^2}_{L_2 \text{ regularizer}}$$

$B$  is the size of mini-batch,  $Y_{true}$  is the true and  $Y_{predicted}$  is the predicted class label of the activity,  $\lambda$  is the loss coefficient and  $\theta$  is the set of learnable parameters (weights and biases) of the network.

### 3.3. Regularizer

The classical problem of any deep learning model is overfitting i.e. when the model learns the training data well but don't give good performance on the testing data. Many techniques have been proposed to cope up with the problem like model pruning, skip connections, dropout, and regularization, etc. We have added a  $L_2$  regularizer in the cost function to mitigate the overfitting problem. The regularizer term can be written as:

$$\lambda \sum_{j=1}^P \theta_j^2$$

where  $\lambda$  is the regularizer coefficient or the loss coefficient and  $\theta$  is the learnable parameters. The  $L_2$  regularizer can be seen as the trade-off parameter between the model complexity and the model generalization. Hence, the trained network performs well on the testing data.

### 3.4. Optimizer

Standard stochastic gradient descent often gets stuck in the local minima and has an oscillatory behavior in the training

stage. To address this issue, Adaptive moment estimation (Adam optimizer) [32] is used that combines the momentum and the RMSprop in the gradient descent which essentially tune the network learning towards faster convergence at each iteration. It has a number of hyperparameters ( $\beta_1, \beta_2, \epsilon, \eta$ ) that are chosen empirically. In addition of the hyperparameters, in order to update the weights and bias, a few intermediate terms are calculated in each mini-batch that calculate the running average of the squared gradients ( $d_w, d_b, V_{dw}, V_{db}, S_{dw}, S_{db}$ ). Mathematically, it can be expressed as:

$$V_{dw} = \beta_1 V_{dw} + (1 - \beta_1)dw \quad V_{db} = \beta_1 V_{db} + (1 - \beta_1)db \quad (1)$$

$V_{dw}, V_{db}$  is seen as the momentum like update.

$$S_{dw} = \beta_2 S_{dw} + (1 - \beta_2)dw^2 \quad S_{db} = \beta_2 S_{db} + (1 - \beta_2)db^2 \quad (2)$$

Similarly,  $S_{dw}, S_{db}$  is seen as the RMSprop like update. At the  $t$  iteration, the intermediate parameters are updated as:

$$V_{dw}^{update} = V_{dw}/(1 - \beta_1^t) \quad V_{db}^{update} = V_{db}/(1 - \beta_1^t) \quad (3)$$

$$S_{dw}^{update} = S_{dw}/(1 - \beta_2^t) \quad S_{db}^{update} = S_{db}/(1 - \beta_2^t) \quad (4)$$

And finally, based on the updated intermediate parameters, the network weights and bias are updated as:

$$w_t = w_{t-1} - \eta \frac{V_{dw}^{update}}{\sqrt{S_{dw}^{update} + \epsilon}} \quad (5)$$

$$b_t = b_{t-1} - \eta \frac{V_{db}^{update}}{\sqrt{S_{db}^{update} + \epsilon}} \quad (6)$$

where  $\eta$  is the learning rate that is chosen empirically. The iterative learning continues until converge is achieved. The network training and test progress is shown in Fig. 2.

#### 4. EXPERIMENT

To evaluate the proposed method, we used publicly available UCI dataset [24]. The dataset is collected from 30 test subjects with age ranges from 19 to 48 years. Each subject carries out the activities (walking, walking up stairs, walking down stairs, sitting, standing and lying) while wearing a Samsung Galaxy S2 smartphone on the waist. To introduce ambidexterity in the network, each test subject performs the experiment twice, first wearing the smartphone on the right-hand side and second to the left-hand side. At 50 Hz, the embedded accelerometer and gyroscope of the smartphone records 3 axial linear accelerations and 3 axial angular velocities which are

manually labeled for each activity. Given the sequential nature of the data, a sliding window of size 2.56 seconds ( $2.56 \times 50 = 128$  readings/window) with 50% overlap is used for processing the data. The whole dataset is partitioned into training and testing set with 70% and 30% distribution of samples respectively. The detailed description of the dataset is given in [33, 34].

We implemented our model in python 3.5 using Tensorflow 1.3.0 and run on a single 12 GB NVIDIA TitanX GPU. Before inserting the data to the stacked LSTM, it passes through a single layer neural network with a ReLU activation which pre-processed the data. We used Adam optimizer with the following hyperparameters.

Hyperparameters	Value
Learning rate $\eta$	0.0025
Loss co-efficient $\lambda$	0.0015
$\beta_1$	0.9
$\beta_2$	0.999
$\epsilon$	$10^{-8}$

**Table 1:** Hyperparameters values

Techniques	Average Accuracy
Kim et al. [22]	83.51
Anguita et al. [24]	89.00
Ronao et al. [23]	91.76
Ronao et al. [35]	90.89
Seto et al. [25]	89.00
Li et al. [28]	92.16
Proposed	93.13

**Table 2:** Proposed method against six state-of-the-art methods. It can be seen that the proposed method improves the average accuracy by 0.93% as compared to the closest method.

##### 4.1. Performance Metric

We evaluate the performance of our network using the recall/precision rate and the average overall accuracy. Technically, we calculated a multi-class confusion matrix for the test data and based on its values, calculated the aforementioned quantities using the following relations:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

Where **TP** is the true positive and **FP** (false positive) shows the sum of the corresponding column in confusion matrix except TP. Similarly, **FN** (false negative) shows the sum of corresponding row except TP. For the multi-class confusion matrix, the average accuracy is calculated as:

$$AverageAccuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

where **TP**, **TN**, **FP**, and **FN** are the overall true positive, true negative, false positive and false negative of all the classes in the confusion matrix. In other words, the overall accuracy is the sum of off-diagonal elements divided by all the elements in the multi-class confusion matrix. The proposed approach is compared to six state-of-the-art methods. The average accuracy of our proposed method outperforms all the considered methods. The quantitative results are listed in Table 2.

## 5. CONCLUSIONS

We proposed a stacked LSTM network for human activity recognition using smartphone data. The network consists of five LSTM cell that is connected back to back. The network is preceded by a single layer neural network that takes the input data and normalized it for the stacked LSTM network. Similarly, a cross-entropy cost function is used with a  $L_2$  regularizer for training the network. The network trainable parameters are trained and optimized through Adam optimizer. The network is evaluated on public domain UCI dataset and quantitative results are calculated in terms of precision-recall and average accuracy. The results are compared to 6 state-of-the-art methods. The quantitative results show that the proposed network outperforms all the state-of-the-art methods. In the future, we will incorporate the temporal attention in the LSTM network. Additionally, we are planning to evaluate the network on more challenging datasets and extend the network to recognized more comprehensive human activities.

## 6. REFERENCES

- [1] Ya-Hung Chen, Ming-Je Tsai, Li-Chen Fu, Chia-Hui Chen, Chao-Lin Wu, and Yi-Chong Zeng, "Monitoring elder's living activity using ambient and body sensor network in smart home," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, pp. 2962–2967.
- [2] Jörg Spörri, Josef Kröll, Benedikt Fasel, Kamiar Aminian, and Erich Müller, "The use of body worn sensors for detecting the vibrations acting on the lower back in alpine ski racing," *Frontiers in physiology*, vol. 8, pp. 522, 2017.
- [3] Wonsik Lee, Seoungjae Cho, Phuong Chu, Hoang Vu, Sumi Helal, Wei Song, Young-Sik Jeong, and Kyungeun Cho, "Automatic agent generation for iot-based smart house simulator," *Neurocomputing*, vol. 209, pp. 14–24, 2016.
- [4] Paolo Rota, Habib Ullah, Nicola Conci, Nicu Sebe, and Francesco GB De Natale, "Particles cross-influence for entity grouping," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.
- [5] Habib Ullah, Muhammad Uzair, Mohib Ullah, Asif Khan, Ayaz Ahmad, and Wilayat Khan, "Density independent hydrodynamics model for crowd coherency detection," *Neurocomputing*, vol. 242, pp. 28–39, 2017.
- [6] Berkan Solmaz, Brian E Moore, and Mubarak Shah, "Identifying behaviors in crowd scenes using stability analysis for dynamical systems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 10, pp. 2064–2070, 2012.
- [7] Mohib Ullah, Habib Ullah, Nicola Conci, and Francesco GB De Natale, "Crowd behavior identification," in *2016 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 1195–1199.
- [8] Muhammad Saqib, Sultan Daud Khan, Nabin Sharma, and Michael Blumenstein, "Extracting descriptive motion information from crowd scenes," in *2017 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 2017, pp. 1–6.
- [9] Habib Ullah, Ahmed B Altamimi, Muhammad Uzair, and Mohib Ullah, "Anomalous entities detection and localization in pedestrian flows," *Neurocomputing*, vol. 290, pp. 74–86, 2018.
- [10] Khurram Soomro, Haroon Idrees, and Mubarak Shah, "Online localization and prediction of actions and interactions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 459–472, 2019.
- [11] Sultan D Khan, Stefania Bandini, Saleh Basalamah, and Giuseppe Vizzari, "Analyzing crowd behavior in naturalistic conditions: Identifying sources and sinks and characterizing main flows," *Neurocomputing*, vol. 177, pp. 543–563, 2016.
- [12] Habib Ullah and Nicola Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in *ICPR workshop on pattern recognition and crowd analysis*, 2012.
- [13] Habib Ullah, Mohib Ullah, and Muhammad Uzair, "A hybrid social influence model for pedestrian motion segmentation," *Neural Computing and Applications*, pp. 1–17, 2018.
- [14] Tal Nir, Alfred M Bruckstein, and Ron Kimmel, "Over-parameterized variational optical flow," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 205–216, 2008.
- [15] Gül Varol, Ivan Laptev, and Cordelia Schmid, "Long-term temporal convolutions for action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1510–1517, 2018.
- [16] Habib Ullah, Muhammad Uzair, Arif Mahmood, Mohib Ullah, Sultan Daud Khan, and Faouzi Alaya Cheikh, "Internal emotion classification using eeg signal with sparse discriminative ensemble," *IEEE Access*, vol. 7, pp. 40144–40153, 2019.

		Predicted behavior						Recall
		walking	walking upstairs	walking downstairs	sitting	standing	laying	
Actual behavior	walking	467	3	26	0	0	0	94.15%
	walking upstairs	7	438	26	0	0	0	92.99%
	walking downstairs	10	3	407	0	0	0	96.90%
	sitting	0	24	2	596	29	0	91.50%
	standing	0	3	0	47	402	0	88.93%
	laying	0	27	0	0	0	510	94.97%
	Precision	96.48%	87.95%	88.28%	92.69%	93.27%	100%	

**Table 3:** Confusion Matrix of the test data. The off-diagonal elements corresponds to the True positive. Other values in the column corresponds to the false positive while value along the row corresponds to the false negative.

- [17] Yu Kong, Shangqian Gao, Bin Sun, and Yun Fu, "Action prediction from videos via memorizing hard-to-predict samples," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [18] Ming Tong, Yiran Chen, Lei Ma, He Bai, and Xing Yue, "Nmf with local constraint and deep nmf with temporal dependencies constraint for action recognition," *Neural Computing and Applications*, pp. 1–25.
- [19] Mohib Ullah, Habib Ullah, and Faouzi Alaya Cheikh, "Single shot appearance model (ssam) for multi-target tracking," *Electronic Imaging*, vol. 2019, no. 7, pp. 466–1, 2019.
- [20] Allah Bux, Plamen Angelov, and Zulfiqar Habib, "Vision based human activity recognition: a review," in *Advances in Computational Intelligence Systems*, pp. 341–371. Springer, 2017.
- [21] Yu Kong and Yun Fu, "Human action recognition and prediction: A survey," *arXiv preprint arXiv:1806.11230*, 2018.
- [22] Yong-Joong Kim, Bong-Nam Kang, and Daijin Kim, "Hidden markov model ensemble for activity recognition using tri-axis accelerometer," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015, pp. 3036–3041.
- [23] Charissa Ann Ronao and Sung-Bae Cho, "Human activity recognition using smartphone sensors with two-stage continuous hidden markov models," in *2014 10th International Conference on Natural Computation (ICNC)*. IEEE, 2014, pp. 681–686.
- [24] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *International workshop on ambient assisted living*. Springer, 2012, pp. 216–223.
- [25] Skyler Seto, Wenyu Zhang, and Yichen Zhou, "Multivariate time series classification using dynamic time warping template selection for human activity recognition," in *2015 IEEE Symposium Series on Computational Intelligence*. IEEE, 2015, pp. 1399–1406.
- [26] Mohammed Mehedi Hassan, Md Zia Uddin, Amr Mohamed, and Ahmad Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Future Generation Computer Systems*, vol. 81, pp. 307–313, 2018.
- [27] P Hevesi, JA Ward, O Amiraslanov, G Pirkli, and P Lukowicz, "Wearable eye tracking for multisensor physical activity recognition," *International Journal On Advances in Intelligent Systems*, vol. 10, no. 1-2, pp. 103–116, 2018.
- [28] Yongmou Li, Dianxi Shi, Bo Ding, and Dongbo Liu, "Unsupervised feature learning for human activity recognition using smartphone sensors," in *Mining Intelligence and Knowledge Exploration*, pp. 99–107. Springer, 2014.
- [29] Andrey Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.
- [30] Kaisheng Yao, Trevor Cohn, Katerina Vylomova, Kevin Duh, and Chris Dyer, "Depth-gated lstm," *arXiv preprint arXiv:1508.03790*, 2015.
- [31] Tim Salimans and Durk P Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 901–909.
- [32] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [33] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones.," in *Esann*, 2013.
- [34] Machine Learning Repository, "Human activity recognition using smartphones data set," 2012, <https://archive.ics.uci.edu/ml/datasets/>.
- [35] Charissa Ann Ronao and Sung-Bae Cho, "Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors," *Proceedings of the Korean Information Science Society*, pp. 858–860, 2015.