An Intrinsic Human Physical Activity Recognition from Fused Motion Sensor Data using Bidirectional Gated Recurrent Neural Network in Healthcare

Okeke Stephen[a], Samaneh Madanian[b] and Minh Nguyen[c]
Department of Computer Science and Software Engineering, Auckland University of Technology (AUT), 6 St. Paul Street, Auckland, New Zealand

Keywords: Human Physical Activity Recognition, Gated Recurrent Unit, Machine Leaning, Rectified Adam Optimiser, Deep Learning, Movement Recognition.

Abstract: An intrinsic bi-directional gated recurrent neural network for recognising human physical activities from intelligent sensors is presented in this work. In-depth exploration of human activity data is significant for assisting different groups of people, including healthy, sick, and elderly populations in tracking and monitoring their level of healthcare status and general fitness. The major contributions of this work are the introduction of a bidirectional gated recurrent unit and a state-of-the-art nonlinearity function called rectified adaptive optimiser that boosts the performance accuracy of the proposed model for the classification of human activity signals. The bidirectional gated recurrent unit (Bi-GRU) eliminates the short-term memory problem when training the model with fewer tensor operations, and the nonlinear function, a variant of the classical Adam optimiser provides an instant dynamic adjustment to the adaptive models' learning rate based on the keen observation of the impact of variance and momentum during training. A detailed comparative analysis of the proposed model performance was conducted with long-short-term-memory (LSTM), gated recurrent unit (GRU), and bi-directional LSTM. The proposed method achieved a remarkable landmark result of 99% accuracy on the test samples, outperforming the earlier architectures.

1 INTRODUCTION

With the rapid advancement of sensors and wearable devices, recently, detecting and classifying human physical activities from diverse sensor data has attracted enormous interest in computer vision and digital health. From the healthcare perspective, the current and well-known pressure on healthcare coupled with technological advancements shifted the focus from on-hospital services to home-based services at patients' homes.

This field has drastically grown with the ever-demanding societal need for elderly care, telehealth, and telerehabilitation (Tun, Madanian, & Mirza, 2021; Tun, Madanian, & Parry, 2020) preventive medicine, human-computer interface, sport and fitness, and intelligent surveillance. This makes the integration of sensors and computer vision suitable for everyday lives' workout monitoring and general healthcare directly related to physical activity recognitions driven by wearable sensors (Casale, Pujol, & Radeva, 2011; Ordóñez & Roggen, 2016).

However, despite all the advancements in the sensors, challenges in Human Physical Activity Recognition (HPAR) abound, and information representation is a prominent issue that inhibits sensor-driven HPAR. Traditional machine learning classification techniques rely on feature engineering and traditional information extraction from kinetic signals (Bevilacqua et al., 2018). Heuristic methods are employed to pick these features regarding tasks under process. However, they have created several issues in developing and deploying HPAR systems.

Therefore, a profound understanding of the application domain or expert guide is necessary for

[a] https://orcid.org/0000-0002-7495-6174
[b] https://orcid.org/0000-0001-6430-9611
[c] https://orcid.org/0000-0002-2757-8350

Stephen, O., Madanian, S. and Nguyen, M.
An Intrinsic Human Physical Activity Recognition from Fused Motion Sensor Data using Bidirectional Gated Recurrent Neural Network in Healthcare.
DOI: 10.5220/0011296400003280
In *Proceedings of the 19th International Conference on Smart Business Technologies (ICSBT 2022)*, pages 26-32
ISBN: 978-989-758-587-6; ISSN: 2184-772X

the feature extraction process (Bengio, 2013). Also, motions with complex patterns are not scalable with HPAR and, in most instances, yield abysmal results in dynamic data obtained from continuous streams of activities. Achieving an intrinsic high recognition accuracy with low computational demand is another crucial issue and obstacle in the wide deployment of such systems for healthcare. These challenges, recently, have given significant growth to deep learning methods for the HPAR.

The adoption of deep learning models for human signal detection and other classification tasks has become widespread thanks to the development and availability of smart and wearable devices, along with their collected data (Stephen, Maduh, & Sain, 2021). Deep learning models are capable of detecting and recognising spatial and temporal dependencies between signals and model scale-invariant features in them (Moya Rueda, Grzeszick, Fink, Feldhorst, & Ten Hompel, 2018; Zeng et al., 2014).

In this work, we deploy a bidirectional Gated Recurrent Unit (Bi-GRU), an advanced Recurrent Neural Network algorithm (RNN) with a rectified RAdam optimiser for the HPAR classification tasks. The RAdam stabilises the model training, improving the model convergence and generalisation using learning rate warmup heuristics (Liu et al., 2019). The GRU uses fewer Tensors to speed up its training and learning process, maximising resources, and reducing computational requirements. All these aim to make this model more suitable for real-life scenario applications and implementations.

2 RELATED WORKS

Human activity recognition (Figure 1) is an area of artificial intelligence (AI) application with continuous research interests and various studies focusing on recognising daily human activities, including sports, workouts, and sleep monitoring.

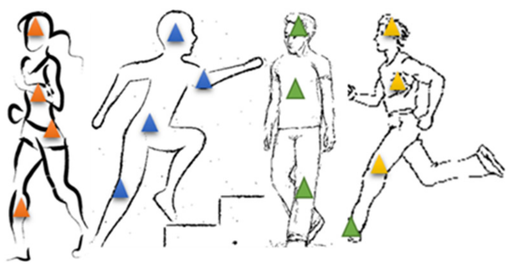


Figure 1: A cross-section of human activity points.

In the human activity recognition task, extracting discriminative features (Hernández, Suárez, Villamizar, & Altuve, 2019) to recognise the activity type is a vital yet challenging task. Different classifications and deep learning approaches have been used for human activity recognition most of which increased the models' complexity and the computational cost.

In a study that dealt with the adaptation of triaxial accelerometer data features, kernels of Convolutional Neural Networks (CNN) were altered to build a model to learn how to recognise human activities (Chen & Xue, 2015). Unlike digital image data that possess spatial connections in their pixels, sensor data are time series, and thus time series models are broadly used in human activity recognition. In another research, a long-short-term memory (LSTM) based deep learning neural network model was built to predict human activities on data collected from mobile sensors (Inoue, Inoue, & Nishida, 2018). In the same area, a Bi-directional LSTM model was used for human activity recognition (Edel & Köppe, 2016). In a healthcare monitoring research, wearable sensors were attached to people to collect data on their speed, heart rate, blood pressure level, and walking gait (Hammerla et al., 2015). These data were collected to detect Parkinson's disease in participants. Among these studies, a study involved space and time characteristics extraction (Ordóñez & Roggen, 2016) combing four layers of CNN and two layers of LSTM achieved a superior result compared to CNN only.

Recognition and monitoring of activity for sports is also an important area. Multiplayer confrontations (Subetha & Chitrakala, 2016) or individual movements data were collected from athletes using wearable devices (Ermes, Pärkkä, Mäntyjärvi, & Korhonen, 2008; Nguyen et al., 2015) data to explore and predict their shooting capability (Nguyen et al., 2015). Activity detection from videos is not left out as combined CNN and RNN approaches (Srivastava, Mansimov, & Salakhudinov, 2015) were deployed to recognise video-based activity tasks, and a landmark result was recorded, although due to high computational demands, the model training was extremely difficult. The proposed model will usher in a more compact and rapid method of recognising vital human daily activities.

3 THEORETICAL BACKGROUND AND METHOD

RNN, LSTM, GRU, and Bi-GRU are a family of deep learning models mainly used for sequential data training and inferencing due to their ability to handle recurrent patterns. The LSTM model is an advanced, RNN architecture built with a set of memory blocks or repeatedly connected subnets. It aims to solve the long-term dependency problems prevalent in the conventional RNNs caused by exploding or vanishing gradients resulting from the backpropagation process. Every memory block in the architecture comprises one or more self-connected memory cells and the input (three multiplicative units) (Graves, 2012). The forget gets f_t in LSTM architecture, determines what information to be kept or eliminated from memory, the input gate i_t is a channel where new information flows to the cell state. The memory update is a cell state vector C_t, which sums the previous memory through the forget gate and the new memory through the input gate. Finally, the output gate O_t conditionally decides which information from the memory should be released.

GRU is a compacted form of an LSTM in the RNN family (Cho et al., 2014). GRU operates with lesser parameters, unlike LSTM, because of the absence of output gates in its architecture. It has been confirmed that GRU can produce superior accuracy in some smaller datasets, such as the one we used in this work. Initialising t and the vector of the output at 0 in equation 1, the GRU operations are expressed as:

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z) \quad (1)$$

$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r)$$

$$h_t = z_t * h_{t-1} + (1 - z_t) * \emptyset_h(W_h x_t + U_h(r_t * h_{t-1}) + b_h) \quad (2)$$

The update gate z_t (1) performs a similar function, on like the input gate and the forget gate found in the LSTM architecture by determining which information to be ignored or to be added. The eset gate r_t, is responsible for determining the amount of information to discard. h_t in equation (2), is an output vector that releases the recurrent operations' outcome. The GRU is faster to train and inference because it uses fewer tensor operations than a typical RNN or LSTM model.

The Bidirectional GRU (Bi-GRU) works on the assumption that the outcome at the time t may or may not rely on the previous information and the subsequent information (Yang, Ng, Mooney, & Dong, 2017). At the beginning of its operation, it combines the cell state, the $q-th$ hidden unit and creates the reset gate τ_q (3) computed as shown below in (3):

$$\tau_q = \sigma([W_r x]_q + [U_r h(t-1)]_q) \quad (3)$$

Where σ denotes the sigmoid function, $[.]_q$ is the $q-th$ element of a vector, x and $h(t-1)$ are the input vector and pre hidden state, W_r *and* U_r representing the weight matrices, respectively. Subsequently, it merges the forget gate and input gate into a single update gate. The update gate μ_q is shown in the formula:

$$\mu_q = \sigma([W_\mu x]_q + [U_\mu h(t-1)]_q)$$

Then, the actual computation of the generated activation unit h_q is shown in equation (4) below:

$$h_q(t) = \mu_q h_q(t-1) + (1-\mu_q)(\tilde{h}_q)(t) \quad (4)$$

where,

$$\tilde{h}_q(t) = \tanh([Wx]_q + [U(p \odot \mathrm{h}(\mathrm{t}-1))]_q$$

$\odot$ represents the element-wise multiplication. We then use the element-wise sum to summate the forward and backward states generated by the Bi-GRU as the product of the qth signal (5).

$$h_q(t) = [\overrightarrow{(h_q(t))} \oplus \overleftarrow{(h_q(t))}] \quad (5)$$

4 EXPERIMENT AND RESULTS

This section presents the details of our data pre-processing and experiments. The experiments' results are also explained in this section.

4.1 Data Pre-processing and Experiment

For our research, we obtained the dataset from (Shoaib, Bosch, Incel, Scholten, & Havinga, 2014). The dataset comprises seven human physical activities, including biking, walking, sitting, running, jogging, standing, and walking upstairs/downstairs. All these activities are the everyday rudimentary daily human motion activities. Ten males, in the 25-30 age range, participated in the data collection exercise, and each participant performed three to four minutes of each of the activities. The data collection was performed in a university indoor building, excluding biking which was done outdoor. Five smartphones were fixed on each participant on five body parts (right arm, right wrist, right hips, left and right legs – check Shoaib et al. (2014)).

During the data pre-processing stage, the collected data was split into small segments solely for extracting critical features with the sliding window method.

In the data pre-processing phase, it was of utmost importance to select the appropriate sliding window size so that an exact value could be affixed to it. Since it has been proven from previous works that a window size of two seconds was appropriate for obtaining a meaningful performance in activity recognition (Hernández et al., 2019), only a window size of two seconds was used. Also, fifty sliding window steps and one-hundred-time steps for the sliding window length were used. For the feature extraction process, only twelve feature extractors were used to extract features from the experimental data frame.

From each participant, one thousand eight hundred activity segments were extracted for a single activity performed at a position. In some cases, data from three positions were fused together, resulting in obtaining 5400 segments of each activity for the total of the three positions.

We concatenated the pre-processed data and split them into train, validation, and test sets in this work. We assigned 80% to the train set and 10% each to the validation and test sets, respectively. During the model building process, a total of 32 hidden units, 0.000001 L2 regularizer (for both the kernel and bias parameters), 0.0001 learning rate and RMSprop optimiser were set across the compared models for consistency. Also, a SoftMax classifier was used in the dense classification layer with categorical cross-entropy as the loss function. Finally, 1024 was set as the batch size with 50 epochs.

4.2 Result Analysis

We compare our experimental result with models such as LSTM, GRU and Bi-directional LSTM. The result of this comparison is presented in Table 1.

Table 1: The results of different models.

Model	MSE	MSLE	Accuracy
LSTM	0.2392	0.0164	0.98
GRU	0.2432	0.0169	0.98
Bi-LSTM	0.2396	0.0165	0.98
Bi-GRU	0.1804	0.0113	0.99

From the experimental results, it can be concluded that our proposed model produced an overall 0.99 accuracy on the test samples, while the rest of the models yielded approximately 0.98 accuracies each.

In our calculations, MSE and MSLE estimate Mean Square Error (MSE) and MSLE (Mean Square Log Error) respectively, of the results obtained in the proposed model. MSEs and MSLEs in LSTM, GRU and Bi-LST are relatively negligible; however, MSE between the LSTM and Bi-GRU is 0.0588, GRU, and Bi-GRU are 0.0628, then between Bi-LSTM and Bi-GRU is 0.0592. With a total of 106,983 LSTM trainable parameters, learning rate (lr) of $1e-4$, $1e-6$ kernel regularizer, 1024 batch size, 40 epochs and a hidden unit of 32, we achieved a test accuracy of 98.1% and loss of 0.0697 as shown in Table 2 and Figure 2.

In Table 2, classes 0, 1, 2, 3, 4, 5, and 6 represent biking, walking downstairs, jogging, sitting, standing, walking upstairs and conventional walking, respectively.

Table 2: Confusion matrix with LSTM.

Class	Precision	Recall	F1-score	Support
0	0.99	0.99	0.99	724
1	0.99	0.97	0.98	722
2	1.00	0.97	0.98	724
3	1.00	1.00	1.00	716
4	1.00	1.00	1.00	720
5	0.97	0.91	0.94	716
6	0.89	0.99	0.94	716
Accuracy			0.981	5038
Macro Avg	0.98	0.98	0.98	5038
Weighted Avg	0.98	0.98	0.98	5038

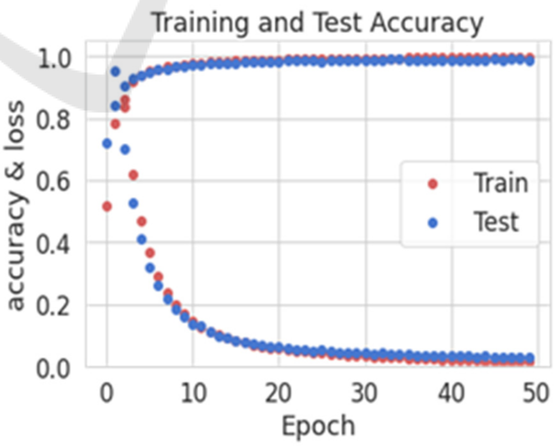


Figure 2: Model accuracy & loss with LSTM.

As shown in Table 2, we achieved 0.99 precision, recall and F1-score from 724 test samples of biking (0) activity; 0.99 precision, 0.97 recall, 0.98 F1-score from 722 test samples on the walking-downstairs (1) activity, respectively; 100% precision, 0.97 recall, 0.98 F1-score from 724 test samples on the jogging (2) activity, respectively. We achieved 100% precision, recall, and F1-score on the sitting (3) and standing (4) activities. For the sitting (3) activity, we used 716 test samples while we had 720 test samples for standing (4).

We obtained 0.97 precision, 0.91 recall and 0.94 F1-score from 716 test samples of walking upstairs (5) activity; 0.89 precision, 0.99 recall and 0.94 F1-score from 716 test samples of basic walking (5) activity and 0.98 macro and weighted average each on the 5038 test samples.

To increase the training and inference speed and as well as the overall performance of the model, we swapped the LSTM with a Bi-directional LSTM and GRU architectures separately, and we observed no meaningful change in the overall performance of the models on the test data, as shown in Figures 3 and 4, respectively.

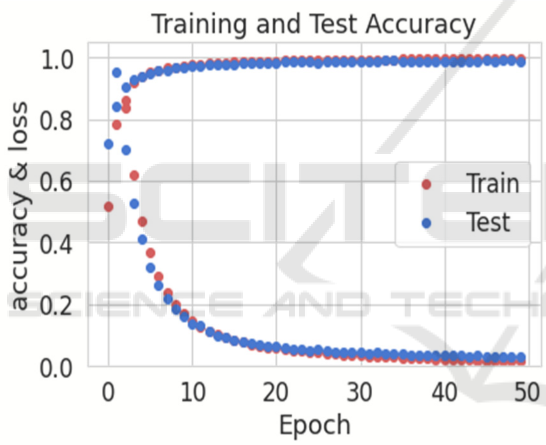


Figure 3: Model accuracy & loss with Bi-LSTM.

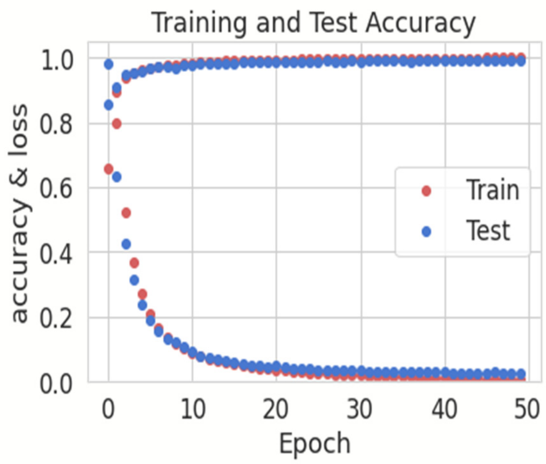


Figure 4: Model accuracy & loss with GRU.

Consequently, we introduced the Bi-GRU model into the set-up with all parameters remaining constant. We observed a remarkable difference in both the MSE, MSLE and the oval accuracy, as shown in Figures 5 and 6, respectively.

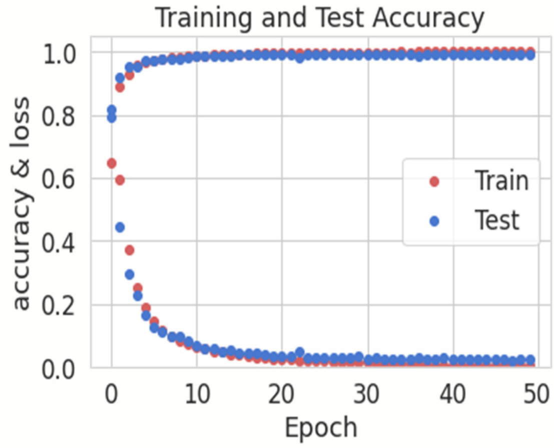


Figure 5: Model accuracy & loss with Bi-GRU.

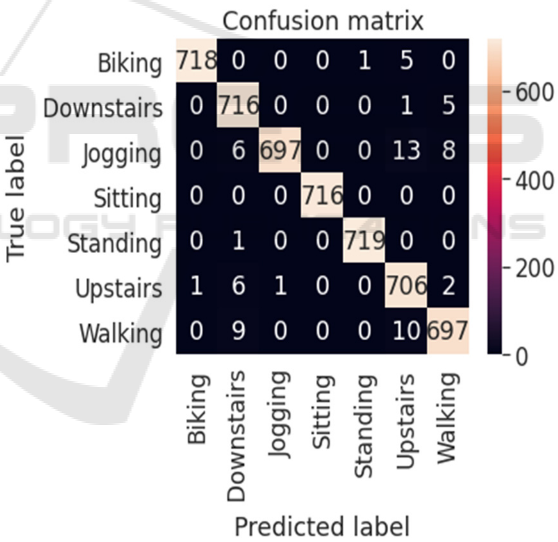


Figure 6: Matrix plot of the Bi-GRU model.

Further analysis of the result from the proposed model, as shown in the confusion matrix plot of Figure 6, the model accurately classified 718 biking data samples from the test sample and misclassified four as walking upstairs. Also, 716 test data points belonging to the downstairs activity were accurately classified, and five were misclassified. In addition, the model accurately classified 697 samples as actual jogging activities and misclassified 13 upstairs activities and eight as walking. Furthermore, the model got all 716 test data points right as sitting and

719 as standing while misclassifying only one point as downstairs activity, respectively.

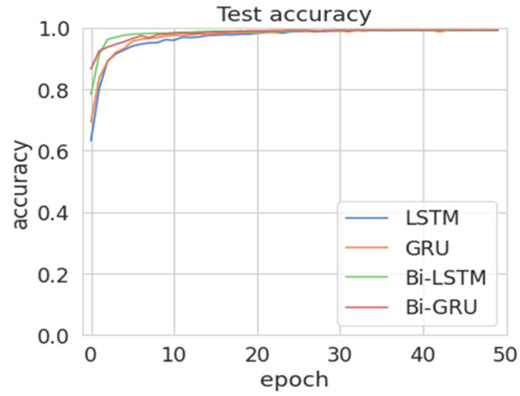
The proposed model correctly recognised 706 upstairs activity test data points as true and ten false. For the test data points belonging to the upstairs activity, the model appropriately recognised 706 while misclassify two as walking, six as downstairs activity, and one each for biking and jogging. Finally, the model accurately classified 697 walking test data correctly while mistakenly recognising ten as upstairs activity and nine as downstairs activity; this is because of the close similarity between these activities.

5 CONCLUSIONS

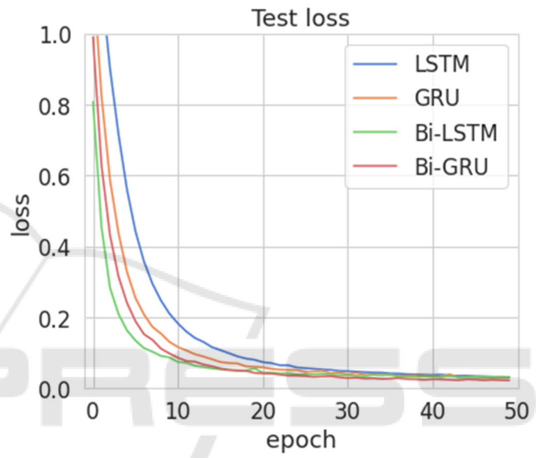
Human activity recognition and classification have become a demanding field in different domains, especially for healthcare and wellbeing. Alo, technology integration is an approach to get technologies into their full potential and use them to address business or health challenge (Madanian & Parry, 2019). Our proposed model could promote cost-effective, rapid and efficient telehealth monitoring and telerehabilitation, for different population groups such as the elderly and athletes.

To deploy such human activity recognition systems for real-life scenarios and applications, the system should be able to process tasks in almost real-time with high accuracy and low computational cost.

In this work, we presented an advanced RNN algorithm for human physical activity recognition tasks. We focused on seven distinct activities extracted from basic human daily activities using multi-sensor data point collection sources. We trained the proposed Bi-directional GRU architecture and evaluated the accuracy using separate test data samples. From the result of the extensive experiments we conducted, we discovered that the Bi-Directional GRU model is a good fit for solving human activity recognition problems when compared with conventional LSTM and GRU, as shown in the combined plots in Figures 7a & 7b. In future work, we plan to include more activities and scenarios and implement human activity detection, classification, and recognition in real-time.



(a)



(b)

Figure 7: (a) & (b) are combined performance accuracy and loss of the studied models.

REFERENCES

Bengio, Y. (2013). *Deep learning of representations: Looking forward.* Paper presented at the International conference on statistical language and speech processing.

Bevilacqua, A., MacDonald, K., Rangarej, A., Widjaya, V., Caulfield, B., & Kechadi, T. (2018). *Human activity recognition with convolutional neural networks.* Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases.

Casale, P., Pujol, O., & Radeva, P. (2011). *Human activity recognition from accelerometer data using a wearable device.* Paper presented at the Iberian conference on pattern recognition and image analysis.

Chen, Y., & Xue, Y. (2015). *A deep learning approach to human activity recognition based on single accelerometer.* Paper presented at the 2015 IEEE international conference on systems, man, and cybernetics.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Edel, M., & Köppe, E. (2016). *Binarized-blstm-rnn based human activity recognition.* Paper presented at the 2016 International conference on indoor positioning and indoor navigation (IPIN).

Ermes, M., Pärkkä, J., Mäntyjärvi, J., & Korhonen, I. (2008). Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *IEEE transactions on information technology in biomedicine, 12*(1), 20-26.

Graves, A. (2012). Supervised sequence labelling. In *Supervised sequence labelling with recurrent neural networks* (pp. 5-13): Springer.

Hammerla, N. Y., Fisher, J., Andras, P., Rochester, L., Walker, R., & Plötz, T. (2015). *PD disease state assessment in naturalistic environments using deep learning.* Paper presented at the Twenty-Ninth AAAI conference on artificial intelligence.

Hernández, F., Suárez, L. F., Villamizar, J., & Altuve, M. (2019, 24-26 April 2019). *Human Activity Recognition on Smartphones Using a Bidirectional LSTM Network.* Paper presented at the 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA).

Inoue, M., Inoue, S., & Nishida, T. (2018). Deep recurrent neural network for mobile human activity recognition with high throughput. *Artificial Life and Robotics, 23*(2), 173-185.

Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2019). On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*.

Madanian, S., & Parry, D. (2019). IoT, Cloud Computing and Big Data: Integrated Framework for Healthcare in Disasters. *Stud Health Technol Inform, 264*, 998-1002. doi:10.3233/shti190374

Moya Rueda, F., Grzeszick, R., Fink, G. A., Feldhorst, S., & Ten Hompel, M. (2018). *Convolutional neural networks for human activity recognition using body-worn sensors.* Paper presented at the Informatics.

Nguyen, L. N. N., Rodríguez-Martín, D., Català, A., Pérez-López, C., Samà, A., & Cavallaro, A. (2015). *Basketball activity recognition using wearable inertial measurement units.* Paper presented at the Proceedings of the XVI international conference on Human Computer Interaction.

Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors, 16*(1), 115.

Shoaib, M., Bosch, S., Incel, O. D., Scholten, H., & Havinga, P. J. (2014). Fusion of smartphone motion sensors for physical activity recognition. *Sensors, 14*(6), 10146-10176.

Srivastava, N., Mansimov, E., & Salakhudinov, R. (2015). *Unsupervised learning of video representations using lstms.* Paper presented at the International conference on machine learning.

Stephen, O., Maduh, U. J., & Sain, M. (2021). A Machine Learning Method for Detection of Surface Defects on Ceramic Tiles Using Convolutional Neural Networks. *Electronics, 11*(1), 55.

Subetha, T., & Chitrakala, S. (2016). *A survey on human activity recognition from videos.* Paper presented at the 2016 international conference on information communication and embedded systems (ICICES).

Tun, S. Y. Y., Madanian, S., & Mirza, F. (2021). Internet of things (IoT) applications for elderly care: a reflective review. *Aging Clinical and Experimental Research, 33*(4), 855-867. doi:10.1007/s40520-020-01545-9

Tun, S. Y. Y., Madanian, S., & Parry, D. (2020). Clinical Perspective on Internet of Things Applications for Care of the Elderly. *Electronics, 9*(11), 1925. Retrieved from https://www.mdpi.com/2079-9292/9/11/1925

Yang, L., Ng, T. L. J., Mooney, C., & Dong, R. (2017). *Multi-level attention-based neural networks for distant supervised relation extraction.* Paper presented at the AICS.

Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). *Convolutional neural networks for human activity recognition using mobile sensors.* Paper presented at the 6th international conference on mobile computing, applications and services.