Building robust models for Human Activity Recognition from raw accelerometers data using Gated Recurrent Units and Long Short Term Memory Neural Networks

Jeremiah Okai[1], Stylianos Paraschiakos[1,2], Marian Beekman[2], Arno Knobbe[1] and Cláudio Rebelo de Sá[3]

***Abstract*— Human Activity Recognition (HAR) is a growing field of research in biomedical engineering and it has many potential applications in the treatment and prevention of several diseases. Due to the recent advancement in technology, devices that collect position and orientation measurements (e.g. accelerometers and gyroscopes) are becoming ubiquitous. These measurements can then be used to train machine learning models for HAR. In this research, we propose one recurrent neural network architecture and a data augmentation approach for building robust and accurate models for HAR. We compared models with Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) layers. The proposed data augmentation approach was used to make the models robust to the cases where one or more sensors are missing. In this empirical study, we could also understand some relations between the ideal locations of the sensors in the participants and the types of activities performed. The proposed approaches were tested in the GOTOv dataset from a study which involved 35 participants performing 16 sedentary, ambulatory and lifestyle activities in a semi-structured environment. The results presented, clearly show that the models are able to detect these activities in a robust way.**

I. Introduction

Human Activity Recognition (HAR) is an active field of research in biomedical engineering. HAR can be defined as the capacity to interpret the human motions through sensors and determine its activity [1]. There has been much research in the field of HAR [2], [3], [4], mostly due to the recent advancement in sensing technologies. Sensors such as accelerometers can detect and respond to inputs from the physical environment [5]. Nowadays, more and more sensors are embedded in devices such as camera-based [6], [7], depth sensor-based [8], [9] and wearable-based sensors [10], [11] to collect measurements for HAR. One of the most prominent applications of HAR is in the health sector for treatment and prevention of several diseases [12]. For example, monitoring and treatment of chronic diseases in elderly people [13], [14], to encourage physical exercises for children with motor disabilities [15] or estimating energy expenditure to help in the treatment and prevention of obesity [16].

*This work was supported by Leiden University Medical Center

[1]Jeremiah Okai, Stylianos Paraschiakos and Arno Knobbe are with Leiden Institute of Advanced Computer Science, Universiteit Leiden, The Netherlands (`email: s.paraschiakos@lumc.nl`)

[2]Stylianos Paraschiakos and Marian Beekman are with Molecular Epidemiology, Dept. Biomedical Data Science, LUMC, Leiden, The Netherlands

[3]Cláudio Rebelo de Sá is with Data Science research group, University of Twente, Enschede, The Netherlands

In this work we empirically test two approaches for HAR using sensor data provided by the Leiden University Medical Center (LUMC). The GOTOv (Growing Old TOgether validation) study [17] data was colected from 35 participants performing 16 sedentary, ambulatory and lifestyle activities in a semi-structured environment. Accelerometers sensors were placed on multiple body locations, such as the wrist, ankle and chest for capturing the body motion. This information can then be used for predicting the activity that was performed. However, there are some challenges which arise when analyzing sensor information. For example, different activities with similar measurements (e.g. *walkingNormal* and *walkingSlow*. Besides, each own person has different pace for doing the same thing.

One common way to approach that is using machine learning for the detection of human activities [2], [3], [4]. Machine learning methods are able to extract relevant features and also learn complex relations in large datasets. Many different methods can be used for HAR, for example, Random Forest classifiers [4], Hidden Markov Models [2] and Conditional Random Field (CRF) [3]. In this work we propose and test two different neural network approaches, one with Gated Recurrent Units (GRU) and another with Long Short Term Memory (LSTM) layers.

Even though most of the participants in the GOTOv study were wearing sensors in the ankle, wrist and chest, some of them could not wear all the devices. This resulted in a dataset with missing information. For this reason, our neural network models for HAR needed also to be robust to different levels of missing information (e.g. no sensor in the wrist or only one sensor in the ankle). To train the models to deal with missing information, we proposed a data augmentation technique based on [18]. Empirical results showed that, the proposed data augmentation, improved the ability of the models to detect human activities in cases where measurements from one sensor, or a combination of sensors, were not provided.

This paper is structured as follows: in Section II, we discuss the materials and methods used to carry out the research; in Section III we present the experimental setup; in Section IV we discuss the results. Finally, in Section V we present the discussion and conclusion of our research.

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

II. MATERIALS AND METHODS

A. Recurrent Neural Networks

Recurrent neural networks (RNN) contain feedback loops within their hidden layers whose activation at each time depends on that of the previous layer [19]. Because of this, they are more suitable for dealing with sequential data. RNN have been used to solve a variety of problems, such as language modeling [20] or speech recognition [21].

Even though RNN are able to remember older information from sequences, they are not able to do this efficiently when the time difference in the sequences becomes too big (due to the vanishing gradient problem [22]). To avoid that, LSTM or GRU layers can be used.

LSTM and GRU are a type of RNN layer that contain memory cells [23]. Both LSTM and GRU contain memory cells that are used to store information. Gates control which information goes through the model. For that, they use a sigmoid function and pointwise multiplication operations. LSTM contain 3 gates, namely, *forget, input* and *output*. GRU have only 2 gates, the *reset* gate and the *update* gate. The reset gates controls which of the memory cell information needs to be forgotten. The update gates controls which information needs to be updated. This allows them to remember long sequences of information without loosing relevant information.

B. Data Augmentation

Data augmentation refers to a type of methods for the introduction of unobserved data or latent variables [24]. Different data augmentation approaches have been used in image classification [25], document analysis [26] or automotive industry [18]. These enhance the overall performance of machine learning models by preventing them from overfitting the data during the inductive learning phase [24].

C. Data

The dataset used for this experiment was provided by the Molecular Epidemiology department from LUMC and is referred as GOTOv study [17]. The GOTOv study was designed to serve multiple ageing studies in LUMC that collected data[1] from similar population and devices [27], [28], [29]. The study involved 35 participants (14 females and 21 males) with an average age of 65. Each participant performed 16 everyday activities while wearing 5 sensor devices on 6 different body locations. The devices were *GeneActiv accelerometers* and *Equivital*. We use the sensor measurements recorded by the GeneActiv sensors placed on the wrist, ankle and chest for this experiment. Each GeneActiv sensor records triaxial acceleration (+/- 8g) with a high sampling rate of 83Hz.

Every participant involved in the study had to follow a specific protocol for the activities performed. Before starting the activities, a calibrations step took place which lasted approximately 15 minutes. The activities were divided into 2 parts, *indoor* and *outdoor*. The indoor activities consisted of *lying down, sitting, standing* and several household chores, such as *dishwashing, stakingShelves* and *vacuumCleaning*. The outdoor activities included the different types of walking *walkingSlow, walkingNormal, walkingFast* and *cycling*. All participants were allowed a resting period of 1 minute standing between the different activities providing this way a clear demarcation to the signal data.

Even though each participant was supposed to complete a total of 16 activities, this was not always possible. This was because 7 of the participants could not perform the outdoor activities due to weather conditions.

[1]The data collection involving human subjects was approved by the Medical Ethical Committee of LUMC and was performed according to the Helsinki Declaration.

III. EXPERIMENTAL SETUP

A. Data pre-processing

In order to predict the human activities from the original data, several transformations had to be made. We started by standardizing the measurements to zero mean and a standard deviation of 1. Then, due to the choice of the methods, RNN, we had to build sequences of consecutive measurements associated with each activity and each participant. A sequence is generally defined as a finite or infinite list of terms arranged in a definite order [30].

We used a fixed time window of 2.5 seconds [31] which resulted in sequences of length 200, because of the frequency at which the sensor measurements were sampled (83Hz). Because activities were performed sequentially, measurements from two different activities could appear in the same sequence. To avoid that, sequences which only had part of the measurements from the respective activity were dropped.

Besides the previously mentioned transformations, we did not use any feature extraction or feature selection techniques. This was because deep neural networks are able to automatically learn feature representations during the training phase.

Finally, another important step was to deal with the class imbalance. In particular, *cycling* is the most represented class with 66858 instances and *walkingStairsUp* the minority class with only 1286 instances. For this reason, we under-sampled the data so that the models are trained with an equal number of instances per class.

B. Proposed approaches

We proposed one RNN architecture to tackle the problem of HAR. This architecture was implemented with two variants, one using LSTM layers and other one with GRU layers. As for selecting the number of hidden layers and neurons, since there is no rule of thumb to decide that [32], several other architectures were also previously tested. This final setting was obtained after a careful testing period which took approximately 2 weeks of experimental work.

The most promising architecture was in fact quite simple. It consisted of an input layer with 9 neurons, 3 hidden layers with 512 neurons and an output layer. The output layer contains 16 neurons, which corresponds to the number of targets (Section II-C). A dropout ratio of 50% between every two layers of the network was used to prevent overfitting.

During the training we used a *data Generator* to apply the data augmentation, which simulated the case where sensor measurements could be missing. To simulate the case of missing sensor measurements, at every batch, the measurements of a randomly selected sensor, or sensors, have their original values replaced with zeros. Once these values are set to zeros, the model is forced to train with the other input features [18]. This helps the model to be more robust to the cases where any sensor is missing.

Besides the data augmentation, the generator was also used to solve the problem of class imbalance that was present in our training data. For that, on every batch, the number of classes was automatically balanced. Moreover, different combinations of sequences were randomly selected at each batch. Thus, making it difficult for the model to memorize the training data.

C. Evaluation

We tested the performance of the different models using the Leave One Subject Out (LOSO) cross-validation. We trained the models with data from 25 participants and validated on data from 2 other participants. The data of the remaining participant, the test set, was then used to test the model after training was completed. To get the overall performance of the proposed approaches we averaged the accuracies of the different models.

Adding to that, to check for the robustness of the models to missing data, we tested how they performed on 7 different cases (or scenarios):

- *awc* sensor measurements from ankle, wrist and chest.
- *aw* sensor measurements from ankle and wrist.
- *ac* sensor measurements from ankle and chest.
- *wc* sensor measurements from wrist and chest.
- *a* sensor measurements from the ankle.
- *w* sensor measurements from the wrist.
- *c* sensor measurements from the chest.

We performed a t-test score for the 7 different cases. For that we used the *ttest_ind* python package.

IV. Results

In this section we present and discuss the results obtained from the two different approaches, the RNN with LSTM layers and the RNN with GRU layers. For simplicity, we refer the models as LSTM and GRU models. All models were trained in 200 epochs with a mini-batch size of 512.

The barplot in Figure 1 represents the average accuracy of LSTM and GRU models obtained in the different scenarios. Besides the better accuracy, the GRU models were faster to train when compared with the LSTM models. It took 22 seconds to complete an epoch of training using GRU models, while it took 222 seconds in the case of LSTM models.

For both models, there is an expected decline in the accuracy as sensor measurements are removed (see Figure 1). The relative difference (in p.p.) between GRU and LSTM models for the different scenarios (*awc, aw, ac, wc, a, w, c*) was 6, 7, 10, 7, 11, 12 and 15, respectively. This seems to indicate that, even though both models decrease in accuracy when there is less information, LSTM models have a bigger decrease in accuracy than GRU. The *t-test* indicated a significant difference between the LSTM and GRU models in all the different scenarios.

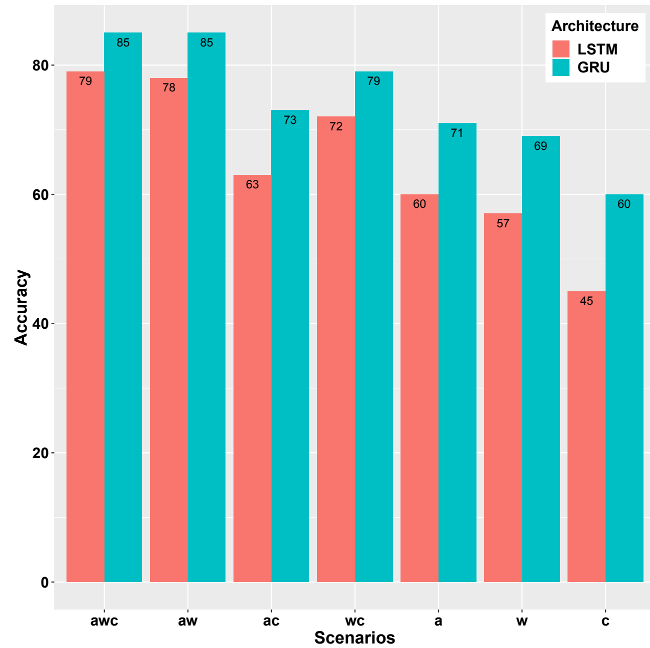

Fig. 1. Barplot showing the average accuracy of the LSTM and GRU models for the different scenarios. The x-axis represents the different test scenarios (*awc, ac, aw, wc, a, c, w*) and the y-axis the average accuracy.

The box and whisker plot in Figure 2 gives an overview of the accuracies obtained from the 28 the GRU models in the 7 scenarios. In the case of having measurements from all the 3 sensors, *awc*, the accuracy of the models ranged between 72% and 98%. These results are quite impressive, considering that HAR is a very challenging task, especially in the case where the number of activities is high.

In the scenarios where these models only had sensor measurements from two sensors, *aw*, *ac* and *wc*, the accuracy was only slightly affected. In particular for *ac* and *wc*, the range of accuracy was $58\% - 92\%$ and $57\% - 96\%$

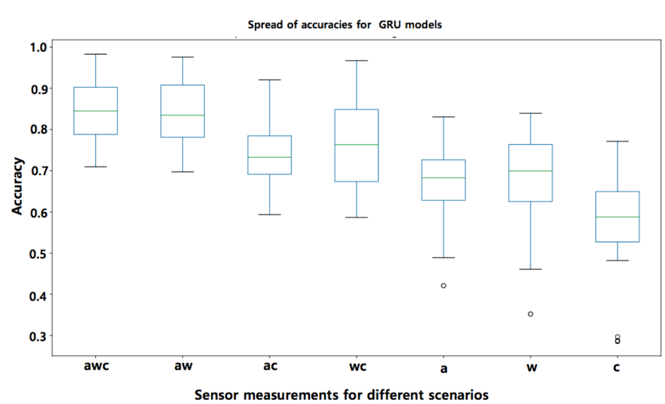

Fig. 2. Spread of accuracies for different scenarios where a sensor could be missing for the GRU models.

respectively. The exception, is the case where the models have no measurements from the chest sensor, *aw*, in which case the accuracy is almost not affected ($69\% - 97\%$). The latter can be explained by the fact that the studied activities, have more variation in the legs and arms than in the chest.

Finally, when sensor measurements from only one sensor are given to the models, *w*, *a* and *c*, we can observe a bigger drop in the accuracy. In particular, the loss of accuracy when using only the chest sensor measurements was the most striking which ranged between $52\% - 76\%$ if we ignore the outliers (Figure 2). This can be explained due to little signal changes recorded, as already mentioned. Nevertheless, it is important to note that with 16 classes, an accuracy within those ranges is quite satisfactory with only the chest sensor.

A. Using measurements from all sensors

Figure 3 shows the confusion matrix of the predictions of GRU models, in the scenario of having measurements from all the 3 sensors, *awc*. We observe that the models are good at distinguishing between most of the activities in this study. The highest misclassification is between the types of walking (*walkingSlow, walkingFast, walkingNormal*). This misclassification error might result from the different walking pace, which are known to differ for each individual. For example, a persons' fast pace might be slow for another.

Also, the models could not distinguish some *sittingChair* from *sittingSofa* activities. However, these two activities are also quite similar. *SittingChair* is the activity where participants are watching TV with the legs up on the couch, while *sittingSofa* is when the participants are reading a paper with their feet on the ground.

B. Using measurements from one sensor

In this section, for simpler comparison, the matrices are relative to the confusion matrix presented in Figure 3.

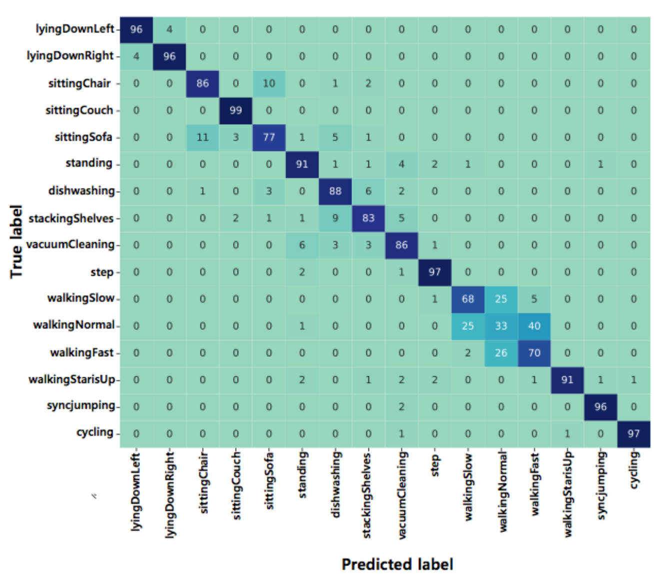

Fig. 3. Confusion matrix of the predictions from all models with ankle, wrist and chest sensor measurements.

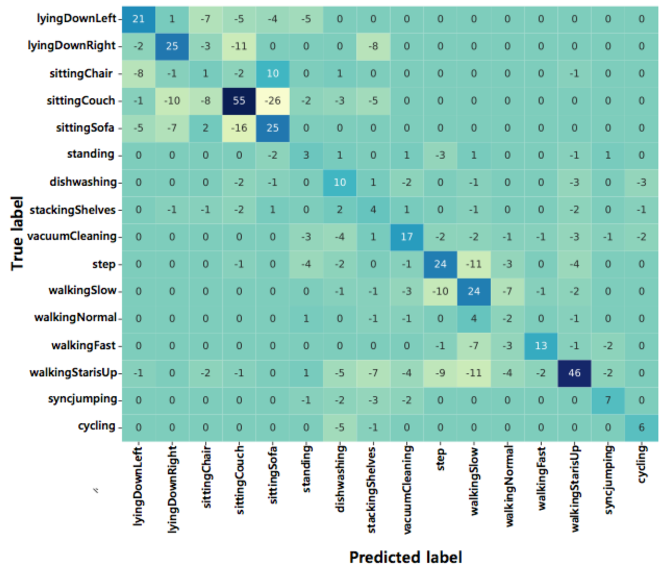

Fig. 4. Confusion matrix of the predictions from all models with wrist sensor measurements.

Therefore, the matrix in Figure 4, represents the difference (in p.p.) between two confusion matrices. This is obtained by subtracting the confusion matrix of the prediction in the *w* scenario from the *awc*. This means that, positive values indicate that the percentage is higher in the *awc* scenario.

In the scenario of having only the wrist measurements, *w*, the models performed worst than the *awc* in the detection of most of the activities. Specially, there was higher misclassification in the sitting and walking activities. In particular, the misclassification errors increased by 26 p.p. in the *sittingCouch* and *sittingSofa* activities. On the other hand, a few exceptions can be observed in activities without much movement of the legs (*standing* and *stackingShelves*).

We also analyse the predictions of the GRU models obtained from sensor measurements of the ankle only, i.e. scenario *a*. Figure 5 represents the matrix obtained by deducting the confusion matrix of the prediction in the *a* scenario from the *awc*. It can be observed a bigger decrease in accuracy of the models in the prediction of household activities, compared to the *awc* scenario. Also, without the wrist and chest measurements, the number of cases where the *sittingSofa* was misclassified as *sittingChair* increased 22 p.p.. Also, the other way around increased 18p.p.. This can be explained by the fact that, the participants in the sofa and chair have their hands in distinct positions.

Finally, we also look into the predictions of the GRU models obtained from sensor measurements of the chest, scenario *c*. Figure 6 represents the matrix obtained by deducting the confusion matrix of the prediction in the *c* scenario from the *awc*. From Figure 6, we observed that four activities can be predicted, from the measurements of the chest sensor, almost as good as with all the sensors (less or equal to 5 p.p. difference). The activities are *lyingDowLeft*, *lyingDowRight*, *walkingFast* and *syncjumping*.

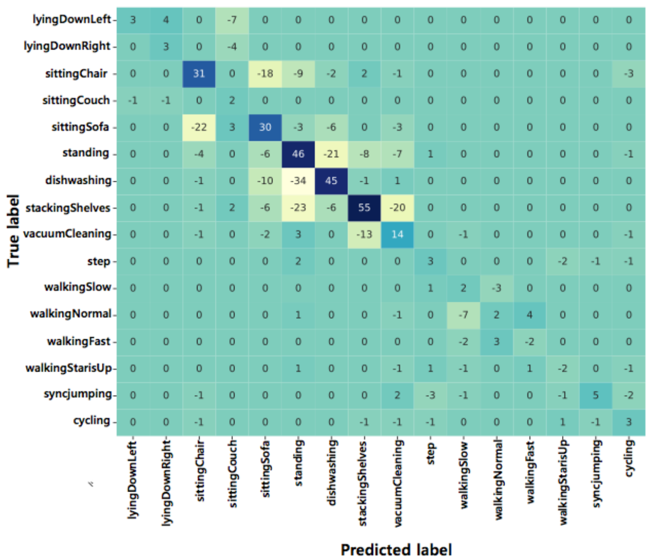

Fig. 5. Confusion matrix of the predictions from all models with ankle sensor measurements.

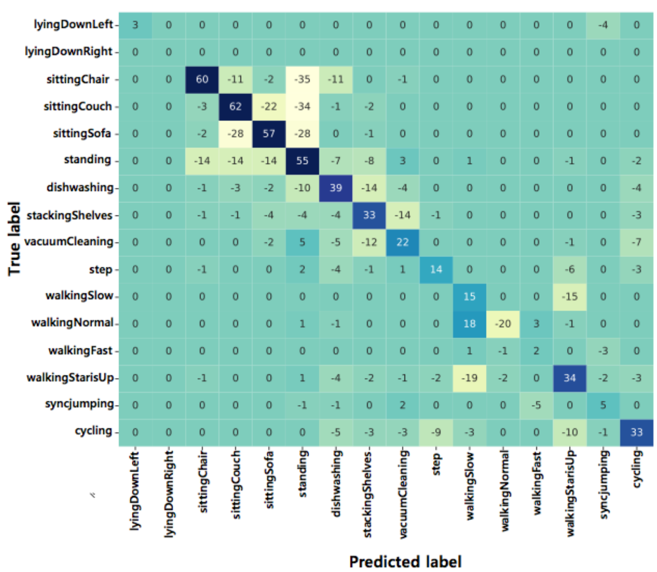

Fig. 6. Confusion matrix of the predictions from all models with chest sensor measurements.

V. Conclusion

In this work we proposed two variants of a simple neural network architecture for HAR. One variant of neural networks used Gated Recurrent Units (GRU) and the other Long Short Term Memory (LSTM) layers. We also proposed a data augmentation approach for making the models more robust to scenarios where the measurements from one or more sensors are missing. To test the models we used sensor data provided by the Leiden University Medical Center (LUMC), which involved 35 participants and 16 different activities (or classes).

We tested the two neural network approaches, one using LSTM layers and the other with GRU layers, and compared them. The empirical results clearly showed that models with GRU layers performed better when used in HAR. Considering the misclassification errors observed, when the model was given all the sensor measurements, we conclude that most of the mistakes are in distinguishing between the different types of walking (*walkingSlow, walkingFast, walkingNormal*). These can be due to the difference paces at which different participants go about walking. Also, some mistakes were observed between the *sittingChair* and the *sittingSofa* activities. Considering the similarity within these two types of activities, it makes sense that the models made some mistakes here.

We also tested the models in scenarios where we had missing data. From the results, we observed that the GRU models were more robust in the presence of missing data than LSTM models. Besides, we could also conclude that, using one sensor in the ankle and one in the wrist are enough for the detection of GOTOv study activities (see Figure 1).

Based on the results of missing sensors, we could observe which locations of the sensors were more relevant for the detection of some specific activities. As expected, for the detection of the activities which involved more movement in the legs, the models relied more on the measurements from the sensor in the ankle (see Figure 5). On the other hand, for the detection of the activities with more movement in the arms, the models relied more on the measurements from the sensor in the wrist (see Figure 4). Finally, because most of the activities involved the use of the upper part (arms) or lower part of the body (legs), the sensor in the chest was the least important.

As future work, we would like to improve the approach, possibly with the use of other data augmentation techniques. We would also like to test the proposed approach in more datasets.

References

[1] O. C. Ann and L. B. Theng, "Human activity recognition: a review," in *Control System, Computing and Engineering (ICCSCE), 2014 IEEE International Conference on*. IEEE, 2014, pp. 389–393.

[2] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010. [Online]. Available: https://doi.org/10.3390/s100201154

[3] E. Kim, S. Helal, and D. J. Cook, "Human activity recognition and pattern discovery," *IEEE Pervasive Computing*, vol. 9, no. 1, pp. 48–53, 2010. [Online]. Available: https://doi.org/10.1109/MPRV.2010.7

[4] P. Casale, O. Pujol, and P. Radeva, "Human activity recognition from accelerometer data using a wearable device," in *Pattern Recognition and Image Analysis - 5th Iberian Conference, IbPRIA 2011, Las Palmas de Gran Canaria, Spain, June 8-10, 2011. Proceedings*, 2011, pp. 289–296. [Online]. Available: https://doi.org/10.1007/978-3-642-21257-4_36

[5] H. Chan and A. Perrig, "Security and privacy in sensor networks," *computer*, vol. 36, no. 10, pp. 103–105, 2003.

[6] S. Wang, S. Zabir, and B. Leibe, "Lying pose recognition for elderly fall detection," in *Robotics: Science and Systems VII, University of Southern California, Los Angeles, CA, USA, June 27-30, 2011*, 2011. [Online]. Available: http://www.roboticsproceedings.org/rss07/p44.html

[7] M. Takahashi *et al.*, "Robust recognition of specific human behaviors in crowded surveillance video sequences," *EURASIP J. Adv. Sig. Proc.*, vol. 2010, 2010. [Online]. Available: https://doi.org/10.1155/2010/801252

[8] W.-H. Ong, L. Palafox, and T. Koseki, "Investigation of feature extraction for unsupervised learning in human activity detection," *Bulletin of Networking, Computing, Systems, and Software*, vol. 2, no. 1, pp. pp–30, 2013.

[9] A. A. Chaaraoui, J. R. Padilla-López, P. Climent-Pérez, and F. Flórez-Revuelta, "Evolutionary joint selection to improve human action recognition with RGB-D devices," *Expert Syst. Appl.*, vol. 41, no. 3, pp. 786–794, 2014. [Online]. Available: https://doi.org/10.1016/j.eswa.2013.08.009

[10] J. Yang, J. Lee, and J. Choi, "Activity recognition based on RFID object usage for smart mobile devices," *J. Comput. Sci. Technol.*, vol. 26, no. 2, pp. 239–246, 2011. [Online]. Available: https://doi.org/10.1007/s11390-011-9430-9

[11] N. Alshurafa, W. Xu, J. J. Liu, M. Huang, B. Mortazavi, C. K. Roberts, and M. Sarrafzadeh, "Designing a robust activity recognition framework for health and exergaming using wearable sensors," *IEEE J. Biomedical and Health Informatics*, vol. 18, no. 5, pp. 1636–1646, 2014. [Online]. Available: https://doi.org/10.1109/JBHI.2013.2287504

[12] D. E. Warburton, C. W. Nicol, and S. S. Bredin, "Health benefits of physical activity: the evidence," *Canadian medical association journal*, vol. 174, no. 6, pp. 801–809, 2006.

[13] L. Chen, C. D. Nugent, and H. Wang, "A knowledge-driven approach to activity recognition in smart homes," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 6, pp. 961–974, 2012. [Online]. Available: https://doi.org/10.1109/TKDE.2011.51

[14] O. Baños, M. Damas, H. Pomares, A. Prieto, and I. Rojas, "Daily living activity recognition based on statistical feature quality group selection," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 8013–8021, 2012. [Online]. Available: https://doi.org/10.1016/j.eswa.2012.01.164

[15] J. Huang, "Kinerehab: a kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities," in *The 13th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '11, Dundee, Scotland, UK, October 24-26, 2011*, 2011, pp. 319–320. [Online]. Available: https://doi.org/10.1145/2049536.2049627

[16] E. Sazonov *et al.*, "Monitoring of posture allocations and activities by a shoe-based wearable sensor," *IEEE Trans. Biomed. Engineering*, vol. 58, no. 4, pp. 983–990, 2011. [Online]. Available: https://doi.org/10.1109/TBME.2010.2046738

[17] S. Paraschiakos, R. Cachucho, M. Beekman, M. Moed, D. V. Heemst, S. Mooijaart, A. Knobbe, and E. Slagboom, "Activity recognition for tracking the elderly," *Under Review*, 2019.

[18] C. R. de Sá, A. K. Shekar, H. Ferreira, and C. Soares, "Building robust prediction models for defective sensor data using artificial neural networks," in *International Joint Conference SOCO'19-CISIS'19-ICEUTE'19 - Seville, Spain, May 13-15, 2019, Proceedings*, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-20055-8_14

[19] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, "Gated feedback recurrent neural networks," in *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, 2015, pp. 2067–2075. [Online]. Available: http://jmlr.org/proceedings/papers/v37/chung15.html

[20] S. Li and J. Xu, "A recurrent neural network language model based on word embedding," in *Web and Big Data - APWeb-WAIM 2018 International Workshops: MWDA, BAH, KGMA, DMMOOC, DS, Macau, China, July 23-25, 2018, Revised Selected Papers*, 2018, pp. 368–377. [Online]. Available: https://doi.org/10.1007/978-3-030-01298-4_30

[21] K. Lee, C. Park, N. Kim, and J. Lee, "Accelerating recurrent neural network language model based online speech recognition system," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2018, Calgary, AB, Canada, April 15-20, 2018*, 2018, pp. 5904–5908. [Online]. Available: https://doi.org/10.1109/ICASSP.2018.8462334

[22] Y. Hu, A. E. G. Huber, J. Anumula, and S. Liu, "Overcoming the vanishing gradient problem in plain recurrent networks," *CoRR*, vol. abs/1801.06105, 2018. [Online]. Available: http://arxiv.org/abs/1801.06105

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[24] D. A. Van Dyk and X.-L. Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States.*, 2012, pp. 1106–1114.

[26] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *7th International Conference on Document Analysis and Recognition (ICDAR 2003), 2-Volume Set, 3-6 August 2003, Edinburgh, Scotland, UK*, 2003, pp. 958–962. [Online]. Available: https://doi.org/10.1109/ICDAR.2003.1227801

[27] van de Rest *et al.*, "Metabolic effects of a 13-weeks lifestyle intervention in older adults: The growing old together study," *Aging*, vol. 8, no. 1, pp. 111–124, 2016. [Online]. Available: https://doi.org/10.18632/aging.100877

[28] S. Jansen *et al.*, "Human longevity is characterised by high thyroid stimulating hormone secretion without altered energy metabolism," in *Scientific reports*, 2015.

[29] E. M. F. van Sluijs *et al.*, "Effects of a web-based intervention on physical activity and metabolism in older adults: Randomized controlled trial," in *Journal of medical Internet research*, 2013.

[30] S. B. Volchan, "What is a random sequence?" *The American Mathematical Monthly*, vol. 109, no. 1, pp. 46–63, 2002. [Online]. Available: http://www.jstor.org/stable/2695767

[31] A. M. Khan *et al.*, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010. [Online]. Available: https://doi.org/10.1109/TITB.2010.2051955

[32] D. Stathakis, "How many hidden layers and nodes?" *International Journal of Remote Sensing*, vol. 30, no. 8, pp. 2133–2147, 2009.