International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 4, Number 2, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v4n2.30
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Wearable Activity Recognition for Advancing Elderly Care with Modified Transformer Model

Grant Li

Western Canada High School, Calgary, Alberta, Canada T2S 0B5

ABSTRACT

Technological devices such as smartphones can be utilized in tracking various human activities or movements through built-in accelerometers and gyroscopes. Data obtained from these inertial sensors can be utilized in various applications to assist people, including healthcare, human-computer interaction, and sports. As a result, further developments in the effective classification of time series data through machine and deep learning is highly valued and actively pursued. In this study, the transformer model, a deep learning architecture designed for sequential data such as natural language processing (NLP), has been utilized for analysis of time-series motion readings from wearable accelerometers. The transformer model in this study has been refined by incorporating a Long Short-Term Memory (LSTM) recurrent neural network (RNN) architecture. By leveraging the HAR70+ dataset with a wide range of activities, the modified transformer model in this study obtained a best accuracy of 95.85%, demonstrating that it can match the performance of state-of-the-art wearable activity recognition methods using Deep Neural Networks (DNN) and LSTM. Hence, the findings presented in this study suggest the future relevance of improved transformer or deep learning models to enhance the quality of life for seniors.

KEYWORDS

Transformer; Long Short-Term Memory (LSTM); Wearable activity recognition; Deep learning; Elderly care

1. INTRODUCTION

Wearable activity recognition has become increasingly more popular as a way of classifying motion data. At its core, it involves analyzing readings from motion sensors such as accelerometers and gyroscopes in order to accurately identify human behaviour through the use of machine or deep learning methods [7]. The utilization of wearable activity recognition has great potential in the advancement of healthcare and elderly care. As the well-being of senior citizens has become an increasingly important area in healthcare due to shifts towards an aging population, utilizing wearable activity recognition technology to monitor the daily activities of seniors is an area of ongoing exploration due to its potential in detecting sudden movements such as falls, allowing for increased independence for the elderly, and prompting early intervention [2]. Moreover, physical abilities and health differ amongst elderly subjects, making personalized monitoring and support over their daily activities a crucial aspect. However, relying solely on human surveillance over elderly subjects can be time consuming and inefficient. Wearable activity recognition offers the prospect of automating continuous monitoring and classification of the daily activities elderly subjects perform [6]. In addition to healthcare applications, this technology can also be used for sports in support of athletes [8] as well as human-computer interactions [9].

Currently, there have been significant focuses on the use of deep learning and neural networks as models for wearable activity recognition. Many deep learning models, including convolutional neural networks (CNNs), temporal convolutional networks, and RNNs such as LSTMs, have been explored and tested with motion sensor data to assess their performance in identifying human behaviour [10-12]. These models have made substantial progress and received significant results in accurately and precisely classifying human activities. However, tailoring these technologies to assessing senior behaviour remains a challenge [2]. This study addresses this challenge directly.

This study examines the potential of more recent deep learning models in their abilities to accurately and precisely classify daily activities performed by older adults. More specifically, although deep learning has been increasingly more common as a method for wearable activity recognition, existing studies have commonly focused on the usage of CNNs and LSTMs which, despite having achieved strong performance, are older models [13]. As a result, it is useful to assess the performance of more current deep learning models as they can offer improvements for wearable activity recognition in the context of elderly care. As such, this study focuses on examining the effectiveness of the transformer deep learning architecture for wearable activity recognition. Furthermore, this study also attempts to improve existing transformer architecture by incorporating LSTM to create a combined deep learning model.

Through an investigation of more up-to-date deep learning models, this study adds valuable information to preexisting advancements and knowledge about the role wearable activity recognition plays in terms of elderly healthcare. The assessment of an improved transformer architecture helps provide insight into the viability of current computer models for wearable activity recognition, allowing further modifications to be made based on areas of improvement for this study's model. Overall, this could lead to more personalized and efficient healthcare which would positively improve the well-being of both seniors and their families.

2. BACKGROUND

Among the extensive number of studies present regarding machine and deep learning for wearable activity recognition, some studies have been selected for further discussion in this section. The selected studies have been primarily referenced throughout this study and have offered state-of-the-art methods.

Ustad et al. [1] evaluated the performance of two extreme gradient boost (XGB) machine learning models. The first XGB model was trained with an updated version of the Human Activity Recognition Trondheim (HARTH) dataset [14-15]. Then, they trained a second XGB model using a combination of the HARTH and HAR70+ datasets [1] which they further trained on a leave-one-subject-out cross-validation (LOSO). The overall accuracy Ustad et al. achieved on the HAR70+ dataset which consists of motion data from elderly subjects is 94%, showcasing that their model has quite high performance but there still remains areas for future improvement.

Abbas et al. [2] conducted a study on the effectiveness of both machine and deep learning architectures in human activity recognition for elderly care by using the HAR70+ dataset. They assessed a diverse array of machine and deep learning algorithms including Extreme Gradient Boost (XGB) and Long Short-Term Memory networks (LSTM). In their study, active learning was implemented as a method for mitigating the burden of data labeling since it allows the algorithm to label interactively [2]. The study helped present the potential of active learning to enhance the performance of recognition systems and as a result, showcased its ability to allow for more personalized elderly healthcare that improves older adults' well-being.

Luptáková et al. [3] also assessed the effectiveness of deep learning on wearable sensor-based human activity recognition. However, rather than the DNN and LSTM utilized by Abbas et al., they investigated the performance of a transformer architecture that was adapted for time series analysis of motion signals. Although they used the KU-HAR dataset [16-17] which does not specifically focus on elderly care, it has more classes and a wider range of activities than the HAR70+ dataset utilized by Abbas et al. Moreover, the average accuracy obtained by the transformer model was 99.2% despite the larger dataset. As a result, this suggests that the adapted transformer architecture utilized by Luptáková et al. should also be able to achieve a high performance on wearable activity recognition in the context of elderly care.

Recent advancements in transformer models have focused on improving the implementation of attention mechanisms and layers for various applications. Reza et al. [4] presented a multi-head attention-based transformer model for use in the application of traffic forecasting to determine whether or not transformers are better at mitigating congestion compared to previous RNN state-of-the-art models. They conducted a comparative analysis of their attention mechanism transformer model with gated recurrent units (GRU) and LSTM on PeMS dataset [18], reaching a conclusion that transformers are more suitable in gaining long-range features. This demonstrates the promising performance of recent advancements in transformer models for use in traffic forecasting. More recently, Pareek et al. [5] also employed an attention mechanism in their transformer architecture for human activity recognition. They advanced the transformer model by leveraging both spatial and temporal attributes and combined a 3D CNN architecture with an attention layer. The model achieved accuracies of 98.09% and 99.09% on the Weizmann [19] and UCF101 [20] datasets respectively, suggesting the promising capabilities of recently advanced state-of-the-art transformers.

3. METHODS

3.1. Transformer Model

In the present study, the transformer model was developed using Python and built-in libraries including PyTorch, sci-kit learn, NumPy, and pandas. It initializes the dataset with ‘features’ and ‘labels’ and applies data augmentation in the form of adding Gaussian noise with mean of 0 and standard deviation of 0.01, random scaling with scale factor being chosen uniformly between 0.9 and 1.1 (scaled down or up by 10%), magnitude warping using a curve with a mean of 1.0 and a standard deviation of ‘sigma = 2’ that is applied to each sample independently, and random flipping of the input sample’s sign is set at a probability of 50%. The addition of data augmentation was utilized in order to enhance the diversity of the data during training which can help the model’s ability to generalize. Furthermore, since the HAR70+ dataset only contains seven classes, data augmentation was implemented so more training examples are available for the model and to help prevent overfitting. Through data augmentation, the model’s validation accuracy is projected to have a slight increase with it being closer to the training accuracy, as augmentation should reduce the model’s risk of overfitting on training subset noises. Data augmentation was only applied for the training subset in order to ensure a consistent validation of the model’s performance on original data.

The transformer model also consists of a data preprocessing function which reads in the HAR70+ dataset and then preprocesses it by forming sequences with length of 50. Further augmentation is applied by merging sequences. If the merge factor is greater than 1, merge_factor * sequence_length samples are merged into a new sequence. If the merge factor is equal to or lesser than 1, there is no merging and the features are reshaped into sequences with length of 50. These steps are taken in order to allow for longer-term dependencies and data trends to be captured by the model and therefore, improve the accuracy. Down sampling was also implemented and is controlled by downsample_factor. If it is greater than 1, every sample that is a multiple of downsample_factor is kept in each sequence. This is done to help prevent overfitting of the model since it skips some data points within sequences, allowing the model to better focus on general trends within data. The dataset is split into a training and temporary subset with 30% of data being allocated to the temporary set. This temporary subset is further divided into validation and testing subsets using a 50-50 split. The features are normalized to have similar magnitudes using Standard scaler which is fit to the training data. The same transformation is also applied to the validation and testing subsets to prevent data leakage during model training and testing.

The transformer model consists of a positional encoding class which allows it to gain a reference of where certain elements are in sequential data. This is necessary to include because the base transformer model lacks an inherent idea of sequential order since it processes inputs in parallel rather than in sequence. The transformer class itself consists of encoder, decoder, and dropout layers. Each encoder layer further consists of a multi-head attention mechanism and a feed forward neural network which are added to enhance the model's ability to capture complex relationships and details within the motion data. The encoder layers are stacked on top of each other and the number of layers is controlled in the hyperparameters. The decoder is a linear layer which maps the transformer model's output embeddings to the classes for classification. A dropout layer for regularization is utilized to randomly drop out neurons in the network for each forward pass in training in order to prevent overfitting and improve the model's generalization to unseen data.

3.2. LSTM Model

The LSTM component utilized in this study's model consists of LSTM layers, a fully connected layer, and a dropout layer. Each LSTM layer receives the number of features in the input sequence and produces a hidden dimension as the output which is the number of units in the LSTM's hidden state. Similar to the transformer layers, the LSTM layers are also stacked on top of each other. There is a fully connected layer which converts the output of the LSTM layers to the desired number of classes. A dropout layer, just like that of the transformer, is also incorporated in order to regularize the model by randomly dropping out neurons to prevent overfitting. A forward pass method is incorporated to define how the model receives inputs, processes them, and produces outputs for activity prediction.

3.3. Combined Model

The combined model class integrates both the transformer model and the LSTM model which were defined previously. The input passes through the transformer architecture where it is processed. A linear layer, transformer_projection, is used to project the transformer's output to match the hidden dimension size of the LSTM output. The LSTM class also processes the same input data and the output is captured. The combined model concatenates the outputs from both the transformer and LSTM architectures along the feature dimension, producing a combined feature vector that incorporates information from both models. Since the hidden dimension from both models is concatenated, the combined vector is of shape [batch_size, hidden_dim * 2]. The combined vector is passed through a fully connected layer which maps it to the number of classes and produces the model's predictions.

The transformer model utilized in this study was adopted from Luptáková et al. and was modified and converted from tensorflow and keras to PyTorch. A combined model with transformer and LSTM architectures was adopted for this study to improve the performance of the base transformer model. Since LSTM has exhibited strength for time series data and in the field of wearable activity recognition, this study aimed to explore how it could be utilized alongside the transformer model. The combined model was aimed to leverage strengths from both models, allowing for improved performance overall by merging complementary information from each architecture.

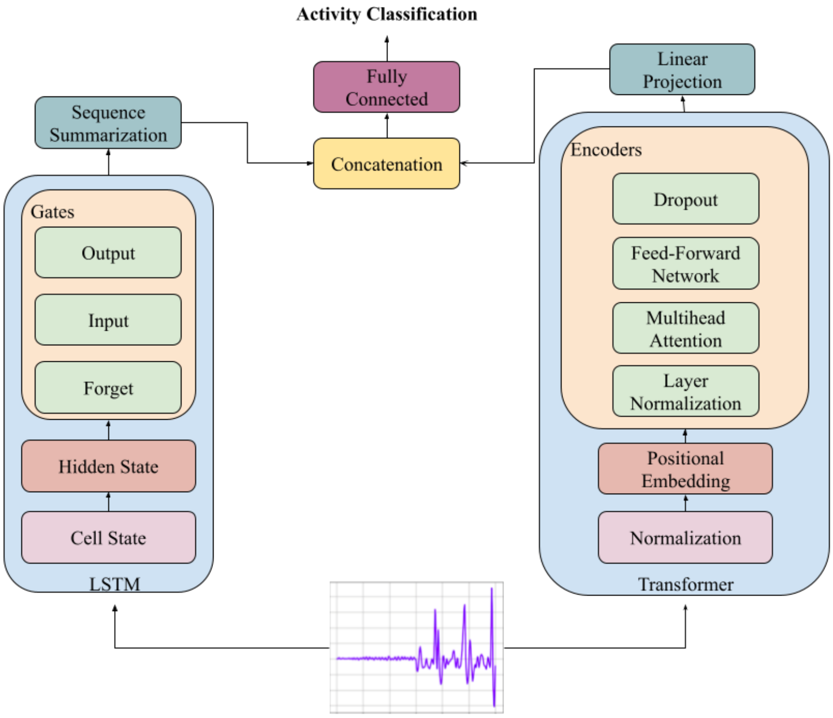

Figure 1. The combined model with LSTM and transformer architectures

3.4. Optimal Hyperparameters

The optimal hyperparameters were primarily adopted from Luptáková et al.'s study. The hyperparameters were slightly modified based on insight from the validation accuracy obtained during model training and final values were selected based on those that produced the highest accuracy when tested. Furthermore, modifications were made to hyperparameters, especially dropout rate, based on possible overfitting. Dropout rate was initially set at 0.1 but was tested with 0.2 and then 0.3 where it produced validation accuracies with small deviation from the training accuracy. Testing of multiple dropout rates alongside data augmentation allowed overfitting to be managed during model training and testing. The optimized hyperparameters used in this study are shown in table 1 below.

Table 1. Optimal hyperparameter values

Hyperparameter	Description	Value
Epochs	Number of training episodes	50
Num Layers	For LSTM, this refers to the number of LSTM layers	3
Embed Layer Size	For LSTM, this refers to the hidden layer size	128
Hidden Dimension	LSTM hidden dimension	128
Combined Dimension	Combined model dimension	128
Num Heads	# of heads in the multi-head attention mechanism	8
Dropout	Dropout rate applied between layers	0.2
Optimizer	Optimizer for the training model	Adam
AMSGrad	Variant of the Adam optimizer to address convergence issues	False
Label Smoothing	Smoothing of the traditional one-hot encoded labels	0.1
Learning Rate	Maximum learning rate value	1e-3
Warmup Steps	First iterations where learning rate is gradually increased	10
Batch Size	Number of training samples that are processed together	64
Global Clipnorm	Clipping of the gradients to prevent them from exploding	3.0

3.5. HAR70+ Dataset

Since this study focuses on wearable activity recognition in the context of elderly care, the HAR70+ dataset was selected from UCI Machine Learning Repository for model training. This dataset consists of motion readings obtained from 18 fit-to-frail older adult subjects during an approximate 40-minute semi-structured free-living period. Subjects were aged 70-95 years old and each wore two 3-axial accelerometers attached to the right thigh and lower back. The dataset contains annotated activities of walking, shuffling, stairs (ascending), stairs (descending), standing, sitting, and lying which were each relabeled in this study from 0 to 6 respectively.

4. RESULTS

The combined model's accuracy of 95.85%, compared to the transformer-only model's 94.80%, indicates that the inclusion of LSTM layers improves the model's ability to capture temporal dependencies in the data, even if the difference is slight. The macro F1-score, averaging 0.75 and peaking at 0.82, highlights a balanced performance across classes, suggesting that the model manages precision and recall well, especially given the imbalanced nature of wearable activity recognition data. The higher peak F1-score indicates that some epochs performed significantly better than others, possibly due to fluctuations in the model's sensitivity to noise or the optimization process. Further analysis could focus on the variance in F1-scores for specific classes, particularly those that may be harder to distinguish, such as 'sitting' or 'lying,' where performance may lag behind more distinct activities. This class-specific performance could guide additional data augmentation strategies or model refinement. While the macro F1-score shows a reasonable balance, the gap between the average and the best score suggests that further tuning of hyperparameters like learning rate and dropout could improve overall performance and reduce the risk of overfitting or underfitting. In summary, while the combined model demonstrates strong performance, deeper analysis into class-level trends and hyperparameter adjustments could unlock additional gains in accuracy and F1-scores.

5. ANALYSIS AND DISCUSSION

The combined model adopted in this study was studied in order to explore the potential that updated deep learning models such as transformers have compared to older machine and deep learning architectures. This goal was supported by findings presented in previous papers where a transformer architecture was able to outperform conventional machine learning methods such as in the case of Luptáková et al. where their transformer model achieved an average accuracy of 99.2% compared to the previous 89.67% using the same KU-HAR dataset.

The obtained average accuracy of 95.57%, the best accuracy of 95.85%, the average macro F1-score of 0.75, and the best macro F1-score of 0.82 showcase that the modified model performed strongly with the HAR70+ dataset. Before the implementation of the combined model and the LSTM class, the base transformer model itself obtained an average accuracy of 94.80%. Compared to the average accuracy of 95.57%, it can be noted that the merging of the LSTM and transformer architectures has improved the model's performance. Furthermore, the model used in this study can match or even surpass the performance of state-of-the-art methods for wearable activity recognition utilizing the HAR70+ dataset. For example, Abbas et al.'s study utilized the HAR70+ dataset and achieved a best accuracy of 95% using their LSTM architecture on 7 activities. Compared with this study's best accuracy, it can be seen that the combined model has slightly improved performance.

Since the transformer architecture is more current than the traditional machine and deep learning methods utilized in other studies, it would already offer improved performance for time series classification, primarily because of its ability to process inputs in parallel [21]. Additionally, the implementation of the LSTM architecture would also have led to improved results since it is designed to capture temporal patterns within sequential data, making it very effective for wearable activity recognition [2, 22-23]. The combination of transformer and LSTM models would have allowed for both efficient parallel processing and the capturing of temporal patterns which likely contributed to the improved performance seen in this study.

5.1. Limitations and Improvements

Although the accuracy and macro F1-score in this study can be considered quite strong, there is still area for improvement in both when compared to some other studies utilizing transformer architecture. For example, in Luptáková et al.'s study, despite using a different dataset which did not focus on elderly care, their transformer model was able to achieve an average accuracy of 99.2%. Furthermore, although the 0.75 macro F1-score indicates a good starting point for a robust model, there is lots of room for improvement in order to increase the value closer to 0.9 or 1.0. Particularly, the model appears to struggle more between activities which are similar in nature and may have overlapping features such as walking and running or sitting and standing. For instance, walking and running can share similar sensor signals due to potential overlapping acceleration patterns, making them difficult to distinguish from each other. This is further pronounced as there can be different speeds of walking and running such as speed walking or a light jog which may be difficult to distinguish between their sensor readings, resulting in misclassifications. This suggests the possibility of improvement for the combined model used in the current study.

Another possible limitation of this study is the risk of overfitting since the model has substantial capacity due to the merging of two deep learning models which both have many complex layers and dimensions. Moreover, the selected dataset itself poses some limitations which can lead to the risk of overfitting for the transformer model. The HAR70+ dataset has fewer classes and a limited set of activities compared to other datasets, such as KU-HAR, which can add to the model's risk of overfitting on the training data since it is more likely that the model relies on noise patterns rather than generalizable trends due to a lack of data diversity. Possible improvements to ensure better regularization for the combined model could be to incorporate weight decay or more advanced dropout techniques such as variational dropout where the same dropout mask is applied to each timestep. Additionally, an improvement would be to test the model against a variety of datasets, especially those that are larger, as this will provide greater insight into how the model generalizes to more diverse activities.

Lastly, a possible improvement could be the implementation of active learning. As seen with Abbas et al.'s study, the usage of active learning can offer improvements to the performance of both machine and deep learning models for wearable activity recognition. Active learning allows the model to query a user to label new data points based on which ones will offer greatest improvement and benefit to the model's performance [2]. The implementation of active learning can increase efficiency by reducing the burden of data labeling and has the potential to enhance model accuracy and adaptability, making it an area of future focus.

6. CONCLUSION

The results presented in this study demonstrate the capabilities and future potential of modifying transformer models by combining it with other deep learning neural networks. This study brings attention to the possible improvements such models can have on wearable activity recognition, especially when applied to elderly care. The dataset selected consists of motion readings obtained specifically from elderly subjects aged 70-95 years old which allows for model classification performance to be assessed for elderly care. Overall, the combined model proposed in the current study achieved a level of accuracy and precision that match some of the state-of-the-art methods.

In the future, the combined model should be tested on larger datasets which consist of more specific actions such as taking medicine. The results should then be applied to offer direct support for humans, especially elderly individuals. Moreover, future work should investigate if the model can be applied in real-time monitoring or be integrated into wearable devices, as well as what modifications to the model architecture would enhance its real-time performance and integration with Internet of Things devices.

REFERENCES

[1] Ustad, A., Logacjov, A., Trollebø, S.Ø., Thingstad, P., Vereijken, B., Bach, K., Maroni, N.S., 2023. Validation of an Activity Type Recognition Model Classifying Daily Physical Behavior in Older Adults: The HAR70+ Model. Sensors 23, 2368. https://doi.org/10.3390/s23052368

[2] S. Abbas et al., "Advancing Healthcare and Elderly Activity Recognition: Active Machine and Deep Learning for Fine- Grained Heterogeneity Activity Recognition," in IEEE Access, vol. 12, pp. 44949-44959, 2024, doi: 10.1109/ACCESS.2024.3380432. keywords: {Older adults; Legged locomotion; Medical services; Human activity recognition; Adaptation models; Feature extraction; Data models; Deep learning; Machine learning; Life testing; Active learning (AL); elderly activity recognition; human activity recognition (HAR); healthcare; deep learning (DL);machine learning (ML); lifestyle and technology}

[3] Dirgová Luptáková, I.; Kubovčík, M.; Pospíchal, J. Wearable Sensor-Based Human Activity Recognition with Transformer Model. Sensors 2022, 22, 1911. https://doi.org/10.3390/s22051911

[4] Reza, S., Ferreira, M. C., Machado, J. J. M., & Tavares, J. M. R. S. (2022). A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. Expert Systems with Applications, 202, 117275. https://doi.org/10.1016/j.eswa.2022.117275

[5] Pareek, G., Nigam, S. & Singh, R. Modeling transformer architecture with attention layer for human activity recognition. Neural Comput & Applic 36, 5515–5528 (2024). https://doi.org/10.1007/s00521-023-09362-7

[6] Teixeira, E., Fonseca, H., Diniz-Sousa, F., Veras, L., Boppre, G., Oliveira, J., Pinto, D., Alves, A. J., Barbosa, A., Mendes, R., & Marques-Aleixo, I. (2021). Wearable Devices for Physical Activity and Healthcare Monitoring in Elderly People: A Critical Review. Geriatrics, 6(2), 38. https://doi.org/10.3390/geriatrics6020038

[7] Yadav, S. K., Tiwari, K., Pandey, H. M., & Akbar, S. A. (2021). A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowledge-Based Systems, 223, 106970. https://doi.org/10.1016/j.knosys.2021.106970

[8] Taylor, K., Abdulla, U. A., Helmer, R. J. N., Lee, J., & Blanchonette, I. (2011). Activity classification with smart phones for sports activities. Procedia Engineering, 13, 428–433. https://doi.org/10.1016/j.proeng.2011.05.109

[9] Z. Song et al., "Attention-Oriented Action Recognition for Real- Time Human-Robot Interaction," 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021, pp. 7087-7094, doi: 10.1109/ICPR48806.2021.9412346. keywords: {Semantics;Pose estimation;Human-robot interaction;Real-time systems;Skeleton;Mobile robots;Data mining},

[10] M. A. Khatun et al., "Deep CNN-LSTM With Self-Attention Model for Human Activity Recognition Using Wearable Sensor," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 10, pp. 1-16, 2022, Art no. 2700316, doi: 10.1109/JTEHM.2022.3177710. keywords: {Sensors;Smart phones;Activity recognition;Featureextraction;Deeplearning;Wearablesensors;Gyroscopes;Sensors;smartphones;accelerometers;attention;gyroscopes;LSTM},

[11] Nair, N., Thomas, C., & Jayagopi, D. B. (2018). Human Activity Recognition Using Temporal Convolutional Network. Proceedings of the 5th International Workshop on Sensor-Based Activity Recognition and Interaction. https://doi.org/10.1145/3266157.3266221

[12] Liu L, He J, Ren K, Lungu J, Hou Y, Dong R. An Information Gain-Based Model and an Attention-Based RNN for Wearable Human Activity Recognition. Entropy. 2021; 23(12):1635. https://doi.org/10.3390/e23121635

[13] Yanhui, C. (2021, March 8). A Battle Against Amnesia: A Brief History and Introduction of Recurrent Neural Networks. Medium. https://towardsdatascience.com/a-battle-against-amnesia-a-brief-history-and-introduction-of-recurrent-neural-networks-50496aae6740

[14] A. Logacjov, K. Bach, A. Kongsvold, H. B. Bårdstu, and P. J. Mork, “HARTH: A Human Activity Recognition Dataset for Machine Learning,” Sensors, vol. 21, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/s21237853.

[15] K. Bach et al., “A Machine Learning Classifier for Detection of Physical Activity Types and Postures During Free-Living,” Journal for the Measurement of Physical Behaviour, vol. 1, no. aop, pp. 1–8, Dec. 2021, doi: 10.1123/jmpb.2021-0015.

[16] N. Sikder and A.-A. Nahid, “KU-HAR: An open dataset for heterogeneous human activity recognition,” Pattern Recognition Letters, vol. 146, pp. 46–54, Jun. 2021, doi: 10.1016/j.patrec.2021.02.024

[17] A.-A. Nahid, N. Sikder, and I. Rafi, “KU-HAR: An Open Dataset for Human Activity Recognition.” Mendeley, Feb. 16, 2021, doi: 10.17632/45F952Y38R.5

[18] PeMS Data Source | Caltrans. (n.d.). Dot.ca.gov. https://dot.ca.gov/programs/traffic-operations/mpr/pems-source

[19] Yu S., Tan D., Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on; 2006; Hong Kong, China. pp. 441–444.

[20] Khurram Soomro, Amir Roshan Zamir and Mubarak Shah, UCF101: A Dataset of 101 Human Action Classes From Videos in The Wild., CRCV-TR-12-01, November, 2012.

[21] Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. AI Open, 3. https://doi.org/10.1016/j.aiopen.2022.10.001

[22] Enes Koşar, & Billur Barshan. (2023). A new CNN-LSTM architecture for activity recognition employing wearable motion sensor data: Enabling diverse feature extraction. Engineering Applications of Artificial Intelligence, 124, 106529–106529. https://doi.org/10.1016/j.engappai.2023.106529

[23] Ankita, Rani S, Babbar H, Coleman S, Singh A, Aljahdali HM. An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones. Sensors. 2021; 21(11):3845. https://doi.org/10.3390/s21113845