TinyHAR: A Lightweight Deep Learning Model Designed for Human Activity Recognition

Yexu Zhou*
yexu.zhou@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Haibin Zhao*
haibin.zhao@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Yiran Huang
yiran.huang@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Hefenbrock
michael.hefenbrock@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Till Riedel
till.riedel@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

Michael Beigl
michael.beigl@kit.edu
Karlsruhe Institute of Technology
Karlsruhe, Germany

ABSTRACT

Deep learning models have shown excellent performance in human activity recognition tasks. However, these models typically require large amounts of computational resources, which makes them inefficient to deploy on edge devices. Furthermore, the superior performance of deep learning models relies heavily on the availability of large datasets to avoid over-fitting. However, the expensive efforts for labeling limits the amount of datasets. We address both challenges by designing a more lightweight model, called TinyHAR. TinyHAR is designed specifically for human activity recognition employing different saliency of multi modalities, multimodal collaboration, and temporal information extraction. Initial experimental results show that TinyHAR is several times smaller and often meets or even surpasses the performance of DeepConvLSTM, a state-of-the-art human activity recognition model.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; **Human computer interaction (HCI)**; • **Computing methodologies** → *Supervised learning by classification.*

KEYWORDS

human activity recognition, time series processing, lightweight deep learning model

ACM Reference Format:
Yexu Zhou, Haibin Zhao, Yiran Huang, Michael Hefenbrock, Till Riedel, and Michael Beigl. 2022. TinyHAR: A Lightweight Deep Learning Model Designed for Human Activity Recognition. In *The 2022 International Symposium on Wearable Computers (ISWC '22), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544794.3558467

*Authors contributed to the work equaly and should be regarded as co-first authors.

ISWC '22, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9424-6/22/09…$15.00
https://doi.org/10.1145/3544794.3558467

1 INTRODUCTION

Sensor streams can only be represented in abstract ways and the recorded data typically cannot be interpreted easily by humans [25]. This problem leads to difficulties in post-hoc annotation, which limits the availability and size of annotated human activity recognition (HAR) datasets. Given the complexity of sensor-based HAR tasks, such large datasets are typically necessary to apply state-of-the-art (SOTA) machine learning. Although deep learning (DL) models have shown extraordinary performance on HAR tasks, most DL models for HAR have large sizes (numerous trainable network parameters). When available data is limited, overly large network parameters make the model prone to overfitting, limiting or even jeopardizing its generalization performance. The second challenge arises from the fact that wearable devices that are intended to use the HAR model typically have limited resources. As a result, an excessive number of network parameters complicates the deployment of such models on end devices.

To address these challenges, it is desirable to design an efficient and lightweight DL model. By reviewing related work, we found only few works that considered designing a lightweight HAR model. To this end, we propose an efficient and lightweight DL model which has small model size and low inference latency. In summary, the contributions of this work are:

- We introduce a set of designing guidelines to design lightweight DL models for HAR.
- Based on the summarized design guidelines, we propose an efficient HAR model named TinyHAR.
- We conduct extensive experiments on six benchmark HAR datasets and the result shows great improvement both in performance and model size.

The proposed TinyHAR model meets or even surpass the optimized DeepConvLSTM [4] with reduced model size by more than 93% on five datasets. This reduction in model size facilitates the deployment of the model on common off the shelf low-cost devices such as the Seed Studio's RISC-V based Sipeed MAix BiT platform, which we are currently using for developments. The implementation of TinyHAR can be currently found in GitHub[1] and may be freely used for wearable computing applications.

[1] https://github.com/teco-kit/ISWC22-HAR

2 RELATED WORK

Deep learning for HAR. DCNN [27] first demonstrated the potential of convolutional neural networks (CNNs) for HAR tasks, which has excellent local dependency extraction capability. To investigate how to optimally extract and fuse features from multimodal sensor data, MCNN [18] proposed multibranch CNNs. However, CNNs usually need to be stacked very deep to obtain a larger receptive field to capture the long-term temporal dependencies. To efficiently capture the long-term dependencies, different recurrent neural networks (RNNs) variants were introduced to HAR tasks. Additionally, various CNN-RNN hybrid models were proposed [20, 21, 30]. DeepConvLSTM [21], a popular benchmark HAR model, first used a CNN subnet to extract local features from different sensor modalities and feed the extracted features into a long-short-term-memory (LSTM). However, RNNs still have difficulties capturing temporal dependencies over long ranges, as they suffer from the "forgetting" defect [11]. In recent years, self-attention based models [26] have shown superior performance over RNN-based models in capturing long-term temporal dependencies. The utilization of attention mechanism for HAR has also been explored [17, 19]. Almost all work has focused on adding more innovative structures to the DL models for improving the ability of models to extract features. However, most previous works have typically not taken the size of the model into account.

Lightweight DL models for HAR. Only a few works considered to design lightweight and efficient DL models for HAR. Based on the benchmark DeepConvLSTM model, the works in [4, 22] attempt to optimize the number of channels or LSTM layers of DeepConvLSTM model. In [7], only shallow CNN model is considered. Since CNN models need to be stacked multiple times to obtain global information, the work in [7] concentrates on improving the performance of CNNs by incorporating the classical machine learning model. The work in [15] attempts to use grouped convolution in CNN-based models, but the optimized models still have more than half million parameters. We thus try to investigate smaller model structures designed from scratch while considering the unique characteristics of the HAR tasks.

3 METHODOLOGY

Figure 1 gives an overview of the proposed TinyHAR model. In this section, we first introduce the design guidelines, that we gathered from previous experimentation with HAR models. Then, we propose our TinyHAR that follows those guidelines.

3.1 Practical Guidelines for Efficient HAR Model Design

Designing an optimal, lightweight DL model requires careful consideration of the characteristics of target tasks and the factors which could reduce the inference time and operations number. Based on these two considerations, we developed the following guidelines to design lightweight HAR models:

- **G1: The Extraction of local temporal context should be enhanced.** Unlike the NLP task, where each word has meaning in the sequence, the values at a single point in time series bring limited information [1, 21].
- **G2: Different sensor modalities should be treated unequally.** These modalities include different sensor types and wearing positions. Only some modalities are informative for the recognition of certain activities, while other modalities may also contain patterns that are irrelevant to the activity [16, 28]. Irrelevant modalities may influence the recognition and undermine the performance [6].
- **G3: Multi-modal fusion.** The activity is carried out collectively through the movement of the various body parts. Extracting features without considering interaction between different modalities may limit the models' performance [14].
- **G4: Global temporal information extraction.** Human activities are sequentially embedded as transient information in the sensor readings. Information on some time steps may show more salient patterns [19] than their temporal surrounding.
- **G5: The temporal dimension should be reduced appropriately.** Compared to image data, HAR data usually has a much larger temporal than spatial dimension. Excessively long temporal dimension poses a problem to the effective extraction of global temporal dependencies [11]. Reducing the temporal dimension can alleviate this problem and also reduce the computational cost.

3.2 TinyHAR

Following the guidelines above, TinyHAR consists of five parts. The input data of the model $\boldsymbol{X} \in \mathbb{R}^{T \times C \times F}$, where T denotes the temporal sliding window size, C is the number of sensor channels, and F indicates the number of filters ($F = 1$ for the raw data input, which has not been processed).

3.2.1 ***Individual Convolutional Subnet.*** To enhance the local context, we applied a convolutional subnet to extract and fuse local initial features from the raw data (**G1**). Considering the varying contribution of different modalities, each channel is separately processed through four individual convolutional layers (**G2**). For each convolutional layer, ReLU nonlinearities and batch normalization [10] are used. Individual convolution means that the kernels have only 1D structure along the temporal axis (the kernel size is 5×1). To reduce the temporal dimension (**G5**), the stride in each layer is set to 2. All four convolutional layers have the same number of filters F. The output shape of this convolutional subnet is thus $\mathbb{R}^{T^{*} \times C \times F}$, where T^{*} denotes the reduced temporal length.

3.2.2 ***Transformer encoder: Cross-Channel Info Interaction.*** Work [1] successfully adopted self-attention mechanism to learn the collaboration between sensor channels. Inspired by this, we utilized one transformer encoder block [26] to learn the interaction, which is performed across the sensor channel dimension (**G2**) at each time step. The transformer encoder block consists of a scaled dot-product self-attention layer and a two-layers Fully Connected (FC) feed-forward network. The scaled dot-product self-attention is used to determine relative importance for each sensor channel by considering its similarity to all the other sensor channels. Subsequently, each sensor channel utilized these relative weights to aggregate the features from all the other sensor channels. Then the feed-forward layer is applied to each sensor channel, which further

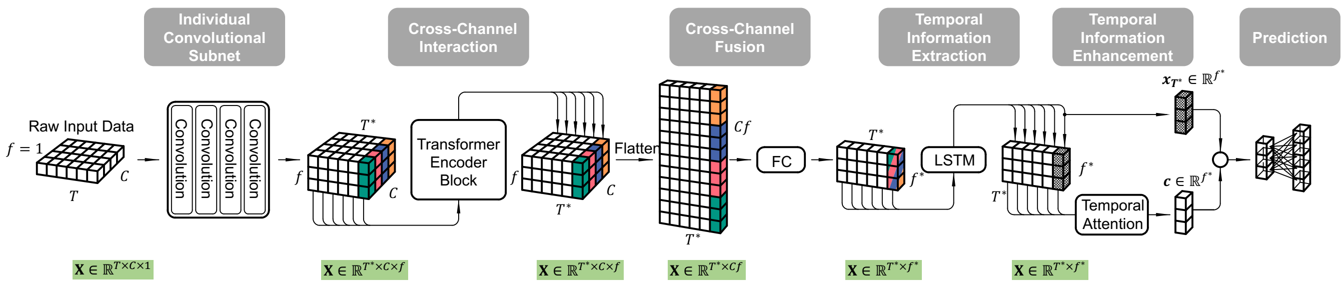

Figure 1: Overview of the proposed Algorithm.

fused the aggregated feature of each sensor channel. Until now, the features of each channel are contextualized with the underlying cross-channel interactions.

3.2.3 *Fully Connected Layer: Cross-Channel Info Fusion.*

In order to fuse the learned features from all sensor channels (**G3**), we first vectorize these representations at each time step, $X \in \mathbb{R}^{T^* \times C \times F}$ *to* $X \in \mathbb{R}^{T^* \times CF}$. Then one FC layer is applied to weighted summation of all the features. Compared to the attention mechanism used in [16], in which the features of same sensor channel share the same weights, FC layer allows different features of same sensor channel to have different weights. Such flexibility of the FC layer leads to more sufficient feature fusion. This FC layer works also as a bottleneck layer in the proposed TinyHAR, which reduce the feature dimension to F^*. In our work we set $F^* = 2F$.

3.2.4 *One-Layer LSTM: Global Temporal Info Extraction.*

After the features are fused across sensor and filter dimension, we obtain a sequence of refined feature vectors $\in \mathbb{R}^{T^* \times F^*}$ ready for sequence modeling. We then apply one LSTM layer to learn the global temporal dependencies.

3.2.5 *Temporal Attention: Global Temporal Info Enhancement.*

Given that not all time steps equally contribute to recognition of the undergoing activities, it is crucial to learn the relevance of features at each time step in the sequence. Following the work in [16], we generate a global contextual representation $\boldsymbol{c} \in \mathbb{R}^{F^*}$ by taking a weighted average sum of the hidden states (features) at each time step. The weights are calculated through a temporal self-attention layer. Because the feature at the last time step $\boldsymbol{x}_{T^*} \in \mathbb{R}^{F^*}$ has the representation for the whole sequence, the generated global representation $\boldsymbol{c}$ is then added to the $\boldsymbol{x}_{T^*}$. Here, we introduce a trainable multiplier parameter γ to $\boldsymbol{c}$, which allows the model has the ability to flexibly decide, whether to use or discard the generated global representation $\boldsymbol{c}$.

4 EXPERIMENT

4.1 Experiment Setup

In this section, we describe our evaluation methodology, including benchmark HAR datasets, compared baselines and model training.

Benchmark HAR datasets. In order to validate the effectiveness of the proposed TinyHAR, we evaluate it on six widely used HAR benchmark datasets [1, 8, 14, 28]. The datasets were selected to exhibit a great diversity in terms of the sensing modalities used, installation locations, sampling frequency, data collection scenarios and activities to be recognized. Important information is summarized in Table 1. Defined window sizes are consistent with the settings used in [8, 14, 24].

Dataset	Freq (Hz)	*#Subject*	*#Class*	*#Sensors*	*#SW*
PAMAP2 [23]	33	9	12	18 A,G	5.12 s
Opportunity [5]	30	4	18	79 A,G,M	1 s
Skoda [29]	33	1	10	30 A	2.56 s
DSADS [3]	25	30	12	6 A,G	5 s
Daphnet [2]	64	10	2	9 A	1 s
WISDM [13]	20	36	6	3 A	5 s

Table 1: For all datasets used in experiments, ***#Sensors*** **represents the total number and type of sensor channels used (A = accelerometer, G = gyroscope and M = magnetometer).** ***#SW*** **denotes the sliding window length. On dataset Opportunity, the 18 classes contain 17 sporadic gestures and a Null class.**

Evaluation rules and metric. On all datasets (except the Skoda), we performed Leave-One-Subject-Out (LOSO) Cross-Validation (CV) to assess the performance of the model with focus on inter-subject generalization. As there is only one subject in Skoda dataset, a 5-fold CV was performed on this data. During training, the data was split by a sliding window with 50% overlap between adjacent windows. For the test data, the window was slid forward by one time step [1]. We used the macro average F1-score $F1_M$ as the evaluation metric, which reflects the ability of the model to identify each activity, regardless of commonly unbalanced class distributions. The CV experiments are performed for five runs with pseudo-random number seeds (1,2,3,4,5). For each run, the mean $F1_M$ score of all subjects are calculated. Then, the scores of five runs are averaged and reported as the final performance.

Baselines. We compare the TinyHAR with the optimized DeepConvLSTM from [4], which is widely used as a benchmark model in various works. Specifications of all layers of the model are consistent with [4]. Furthermore, we introduce a model shrinking hyperparameter width multiplier α [9] to thin DeepConvLSTM uniformly at each layer. By setting $\alpha = \{0.25, 0.5, 0.75\}$, the model size will be reduced to approximately 1/2, 1/4 and 1/16 respectively (referred as DeepConvLSTM_0_75, DeepConvLSTM_0_50 and DeepConvLSTM_0_25). We adjust the filter number F of TinyHAR, so that has a comparable number of parameters to DeepConvLSTM_0_25.

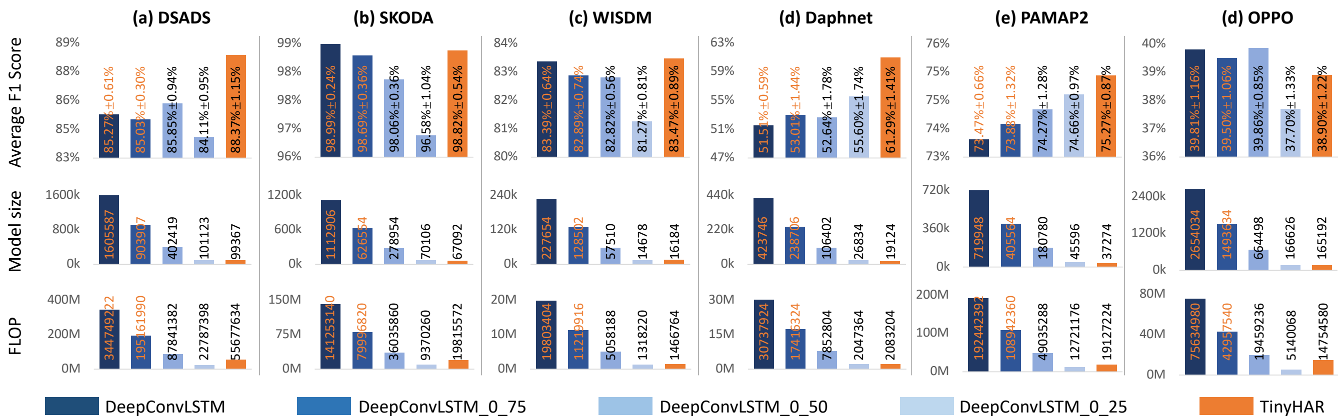

Figure 2: LOSO CV Performance comparison. For each dataset, the averaged $F1_M$, the corresponding model size (number of trainable parameters) and number of FLOPs are shown separately.

Training. The training is performed using the Adam optimizer [12] with an initial learning rate 10^{-4}. The learning rate decays to 10% with a patience equal to 5 epochs. We train models for a maximum of 150 epochs with early stopping, and the batch size is 256. All models were trained on a single NVIDIA A100 40G GPU.

4.2 Result

As summarized in Figure 2, TinyHAR by design has the smaller size than the baseline. However, it achieves a better performance on the three datasets (PAMAP2, DSADS and Daphnet) compared to its competitors. On these three datasets, TinyHAR outperforms the baseline in original size in terms of $F1_M$ by 1.8%, $3.1\% and 9.78\%$. We observed an overfitting trend of DeepConvLSTM variants for the PAMAP2 and Daphnet dataset in comparison with the reduced DeepConvLSTM model variants. Therefore, for Daphnet data, we additionally compared TinyHAR to a DeepConvLSTM with α = 0.21 (model size 17786), which achieved an $F1_M$ of 55.93% ± 1.75%. Although the performance has gotten better, it is still 5.36% lower than TinyHAR. For PAMAP2 dataset, we additionally reduced α to 0.23 (model size 36254). In this case, we observe a degradation of the performance (74.36% ± 1.02%)

Compared to the baseline, TinyHAR obtains slightly lower $F1_M$ on Skoda dataset and slightly higher $F1_M$ on WISDM dataset. Significant tests show that there is no statistically significant difference in the performance between both models. TinyHAR is, however, much smaller than the comparison model and performs significantly better than DeepConvLSTM_0_25 with a similar size. We speculate that the reason for the positive correlation of performance with model size for the baseline is that, Skoda data was collected from only one subject, so it does not penalize overfitting as the statistical independence of training and test samples within the CV is smaller. Similarly, the WISDM dataset was collected from 36 subjects through controlled, laboratory conditions, which reduces the effect of generalizability. Even under this condition, the TinyHAR still reaches the same performance with around 6% model size. The Performance on Dahpnet, PAMAP2, SKODA, WISDM, and DSADS suggests that TinyHAR has a great capability to capture temporal-spatial patterns in multimodal sensing.

However, TinyHAR obtains lower $F1_M$ score on Opportunity. Compared to other datasets, dataset Opportunity has much more sensor channels. We assume that the poor performance is owing to the fact that the model is too small to effectively extract and fuse information from so many channels. Thus, we increased the filter number F from 28 to 42 (model size 370082). The model's $F1_M$ performance of 41.22% ± 1.19% in this case (as predicted) exceeds the performance of all DeepConvLSTM variants.

As shown in Figure 2 (third row), we also utilized the computational cost (number of FLOPs) as an efficiency metric. The TinyHAR model has a much smaller number of FLOPs compared to the DeepConvLSTM model. Compared to the DeepConvLSTM_0_25 model, the TinyHAR model achieves better results on all datasets, although it has a higher number of FLOPs.

5 CONCLUSION

Through a HAR-specific design, we were able to develop a lightweight but highly competitive DL model. Particularly, when taking the model size into account, which is of great importance for wearable computing, the model clearly outperforms the DeepConvLSTM as an example of a state of the art HAR model in our experiments.

Instead of only adapting an architecture from another domain and letting the optimizer do its magic, different saliency of multi modalities, multimodal collaboration and temporal information extraction were specifically translated into a network architecture to achieve this performance. We believe that this shows that there is still a great potential for improvement in HAR models with the focus on deployment on wearable computing devices.

ACKNOWLEDGMENTS

This work has been partially supported by the Carl-Zeiss-Foundation as part of "stay young with robots" (Jubot) project, the German Ministry of Research and Education as part of the SDIL (01IS19030A), and the Ministry of Economic Affairs, Labour and Tourism Baden-Württemberg as part of CC-KING.

REFERENCES

[1] Alireza Abedin, Mahsa Ehsanpour, Qinfeng Shi, Hamid Rezatofighi, and Damith C Ranasinghe. 2021. Attend and Discriminate: Beyond the State-of-the-art for Human Activity Recognition Using Wearable Sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 1 (2021), 1–22.

[2] Marc Bachlin, Daniel Roggen, Gerhard Troster, Meir Plotnik, Noit Inbar, Inbal Meidan, Talia Herman, Marina Brozgol, Eliya Shaviv, Nir Giladi, et al. 2009. Potentials of Enhanced Context Awareness in Wearable Assistants for Parkinson's Disease Patients with the Freezing of Gait Syndrome. In *2009 International Symposium on Wearable Computers*. IEEE, 123–130.

[3] Billur Barshan and Murat Cihan Yüksek. 2014. Recognizing Daily and Sports Activities in Two Open Source Machine Learning Environments Using Body-worn Sensor Units. *Comput. J.* 57, 11 (2014), 1649–1667.

[4] Marius Bock, Alexander Hölzemann, Michael Moeller, and Kristof Van Laerhoven. 2021. Improving Deep Learning for HAR with Shallow LSTMs. In *2021 International Symposium on Wearable Computers*. 7–12.

[5] Ricardo Chavarriaga, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R Millán, and Daniel Roggen. 2013. The Opportunity Challenge: A Benchmark Database for On-body Sensor-based Activity Recognition. *Pattern Recognition Letters* 34, 15 (2013), 2033–2042.

[6] Kaixuan Chen, Lina Yao, Dalin Zhang, Bin Guo, and Zhiwen Yu. 2019. Multi-agent Attentional Activity Recognition. *arXiv preprint arXiv:1905.08948* (2019).

[7] Yves Luduvico Coelho, Francisco de Assis Souza dos Santos, Anselmo Frizera-Neto, and Teodiano Freire Bastos-Filho. 2021. A Lightweight Framework for Human Activity Recognition on Wearable Devices. *IEEE Sensors Journal* 21, 21 (2021), 24471–24481.

[8] Nils Y Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. *arXiv preprint arXiv:1604.08880* (2016).

[9] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861* (2017).

[10] Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *International conference on machine learning*. PMLR, 448–456.

[11] Urvashi Khandelwal, He He, Peng Qi, and Dan Jurafsky. 2018. Sharp Nearby, Fuzzy Far Away: How Neural Language Models Use Context. *arXiv preprint arXiv:1805.04623* (2018).

[12] Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* (2014).

[13] Jennifer R Kwapisz, Gary M Weiss, and Samuel A Moore. 2011. Activity Recognition Using Cell Phone Accelerometers. *ACM SigKDD Explorations Newsletter* 12, 2 (2011), 74–82.

[14] Shengzhong Liu, Shuochao Yao, Jinyang Li, Dongxin Liu, Tianshi Wang, Huajie Shao, and Tarek Abdelzaher. 2020. GlobalFusion: A Global Attentional Deep Learning Framework for Multisensor Information Fusion. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–27.

[15] Tianyi Liu, Shuoyuan Wang, Yue Liu, Weiming Quan, and Lei Zhang. 2022. A Lightweight Neural Network Framework Using Linear Grouped Convolution for Human Activity Recognition on Mobile Devices. *The Journal of Supercomputing* 78, 5 (2022), 6696–6716.

[16] Haojie Ma, Wenzhong Li, Xiao Zhang, Songcheng Gao, and Sanglu Lu. 2019. Attnsense: Multi-level Attention Mechanism for Multimodal Human Activity Recognition.. In *IJCAI*. 3109–3115.

[17] Saif Mahmud, M Tonmoy, Kishor Kumar Bhaumik, AK Mahbubur Rahman, M Ashraful Amin, Mohammad Shoyaib, Muhammad Asif Hossain Khan, and Amin Ahsan Ali. 2020. Human Activity Recognition from Wearable Sensor Data Using Self-attention. *arXiv preprint arXiv:2003.09018* (2020).

[18] Sebastian Münzner, Philip Schmidt, Attila Reiss, Michael Hanselmann, Rainer Stiefelhagen, and Robert Dürichen. 2017. CNN-based Sensor Fusion Techniques for Multimodal Human Activity Recognition. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 158–165.

[19] Vishvak S Murahari and Thomas Plötz. 2018. On Attention Models for Human Activity Recognition. In *Proceedings of the 2018 ACM international symposium on wearable computers*. 100–103.

[20] Ronald Mutegeki and Dong Seog Han. 2020. A CNN-LSTM Approach to Human Activity Recognition. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE, 362–366.

[21] Francisco Javier Ordóñez and Daniel Roggen. 2016. Deep Convolutional and Lstm Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* 16, 1 (2016), 115.

[22] Shalli Rani, Himanshi Babbar, Sonya Coleman, Aman Singh, and Hani Moaiteq Aljahdali. 2021. An Efficient and Lightweight Deep Learning Model for Human Activity Recognition Using Smartphones. *Sensors* 21, 11 (2021), 3845.

[23] Attila Reiss and Didier Stricker. 2012. Introducing A New Benchmarked Dataset for Activity Monitoring. In *2012 16th international symposium on wearable computers*. IEEE, 108–109.

[24] Charissa Ann Ronao and Sung-Bae Cho. 2016. Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks. *Expert systems with applications* 59 (2016), 235–244.

[25] Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, Soren Brage, Nick Wareham, and Cecilia Mascolo. 2021. SelfHAR: Improving Human Activity Recognition Through Self-training with Unlabeled Data. *arXiv preprint arXiv:2102.06073* (2021).

[26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *Advances in neural information processing systems* 30 (2017).

[27] Jianbo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. 2015. Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. In *Twenty-fourth international joint conference on artificial intelligence*.

[28] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. 2017. DeepSense: A Unified Deep Learning Framework for Time-series Mobile Sensing Data Processing. In *Proceedings of the 26th international conference on world wide web*. 351–360.

[29] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Tröster. 2008. Activity Recognition from On-body Sensors: Accuracy-power Trade-off by Dynamic Sensor Selection. In *European Conference on Wireless Sensor Networks*. Springer, 17–33.

[30] Yexu Zhou, Michael Hefenbrock, Yiran Huang, Till Riedel, and Michael Beigl. 2020. Automatic Remaining Useful Life Estimation Framework with Embedded Convolutional LSTM as the Backbone. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 461–477.