LiteTSC: A Lightweight Neural Network for Time Series Classification on Edge Devices

Tao Wang and Ziwen Guo*
Institute of Artificial Intelligence, Xiamen University, Xiamen, China
Email: w2507838878@gmail.com, guoziwen11@163.com
*Corresponding author

***Abstract*—Human activity recognition (HAR) using wearable sensors has gained significant attention for healthcare monitoring and smart device applications. However, deploying deep learning models on resource-constrained edge devices remains challenging due to high computational costs and memory requirements. In this paper, we propose LiteTSC, an ultra-lightweight neural network for time series classification that achieves competitive accuracy with significantly reduced model complexity. Our approach integrates three key components: (1) multi-scale depthwise separable convolutions that capture temporal patterns at different granularities while reducing parameters, (2) a lightweight dual-pooling channel attention mechanism that enhances feature discrimination with minimal overhead, and (3) residual connections that compensate for the reduced representational capacity of lightweight operations. Extensive experiments on three benchmark datasets (UCI HAR, WISDM, and PAMAP2) demonstrate that LiteTSC achieves state-of-the-art accuracy (96.74% on UCI HAR) with only 20.8K parameters. Compared to TinyHAR, the ISWC 2022 Best Paper, LiteTSC achieves higher accuracy with 14.3× fewer parameters and 3.3× faster inference. The lightweight design makes LiteTSC suitable for deployment on resource-constrained devices.**

I. Introduction

Human activity recognition (HAR) from wearable sensor data has emerged as a fundamental technology for various applications, including healthcare monitoring [1], fitness tracking [2], elderly care [3], and smart home automation [4]. With the proliferation of smartphones and wearable devices equipped with inertial measurement units (IMUs), there is growing interest in developing accurate and efficient HAR systems that can operate directly on edge devices.

Deep learning approaches have achieved remarkable success in sensor-based HAR, with convolutional neural networks (CNNs) [5], [6], recurrent neural networks (RNNs) [7], and their combinations [8] demonstrating superior performance over traditional machine learning methods. However, these models typically contain hundreds of thousands to millions of parameters, making them challenging to deploy on resource-constrained devices such as microcontrollers, smartwatches, and IoT sensors.

The challenge of deploying deep learning models on edge devices has motivated research in model compression techniques and efficient architecture design. While post-hoc compression methods such as pruning and quantization can reduce model size, they often require complex optimization procedures and may lead to significant accuracy degradation. An alternative approach is to design inherently efficient architectures that achieve competitive performance with fewer parameters from the outset.

Depthwise separable convolutions, popularized by MobileNet [9] and Xception [10], have proven effective for reducing computational costs in image classification. However, their systematic application to one-dimensional time series data for HAR, combined with attention mechanisms and multi-scale feature extraction, remains underexplored. Similarly, attention mechanisms [11], [12] can enhance feature representation but typically introduce substantial computational overhead.

In this paper, we propose LiteTSC, a lightweight time series classification network that systematically combines proven lightweight techniques for the HAR task. Our main contributions are as follows:

- We design a multi-scale depthwise separable convolution module that captures temporal patterns at different time scales (local and global) while significantly reducing the parameter count compared to standard convolutions.
- We adapt a lightweight dual-pooling channel attention mechanism that combines average and max pooling to enhance channel-wise feature discrimination with minimal computational overhead.
- We incorporate residual connections within depthwise separable blocks to compensate for their reduced representational capacity, improving gradient flow and model convergence.
- We conduct extensive experiments on three benchmark datasets with consistent experimental settings, demonstrating that LiteTSC achieves state-of-the-art accuracy (96.74%) with 14.3× fewer parameters than TinyHAR, the ISWC 2022 Best Paper on lightweight HAR.

II. Related Work

A. Deep Learning for Human Activity Recognition

Deep learning has revolutionized sensor-based HAR by eliminating the need for manual feature engineering. Early works applied CNNs to extract spatial features from sensor data [5], [13], while RNN-based approaches, particularly LSTMs, were employed to model temporal dependencies [7], [14]. Ordóñez and Roggen [8] proposed DeepConvLSTM, combining convolutional layers with LSTM units, which became a widely adopted baseline architecture. Recent advances

include attention-based models [15], [16] and multi-scale feature extraction [17], [18].

B. Lightweight Neural Networks

The demand for efficient models has driven the development of lightweight architectures. MobileNet [9] introduced depthwise separable convolutions, decomposing standard convolutions into depthwise and pointwise operations to reduce computational costs. Xception [10] extended this approach with extreme inception modules.

For time series applications, several lightweight approaches have been explored. Ignatov [19] developed a CNN-based model for real-time HAR on smartphones. Wan et al. [20] proposed a compressed CNN using depthwise separable convolutions for wearable devices. More recently, TinyHAR [21], the ISWC 2022 Best Paper, proposed a lightweight architecture combining individual convolutional subnets, transformer encoders for cross-channel interaction, and LSTM with temporal attention. While TinyHAR achieved strong results, it still requires nearly 300K parameters.

The pursuit of ultra-lightweight HAR has continued to intensify. MLP-HAR [22] explores a purely fully-connected architecture with fewer than 30K parameters for edge deployment. Our work differs by combining multi-scale DSConv with channel attention, achieving competitive accuracy with only 20.8K parameters while maintaining an RNN-free architecture for better parallelization on edge processors.

C. Attention Mechanisms

Attention mechanisms have been widely adopted to enhance feature representation in deep learning models. The Squeeze-and-Excitation (SE) block [12] introduced channel attention by learning to recalibrate channel-wise feature responses. CBAM [23] combined channel and spatial attention modules. For HAR, attention mechanisms have been applied to weight the importance of different time steps [15] and sensor channels [16]. However, most attention modules introduce significant computational overhead, limiting their applicability in resource-constrained scenarios.

III. Proposed Method

A. Problem Formulation

Given a multivariate time series $\mathbf{X} \in \mathbb{R}^{C \times T}$, where C denotes the number of sensor channels and T the number of time steps, the goal is to learn a mapping $f : \mathbb{R}^{C \times T} \rightarrow \{1, 2, \ldots, K\}$ that assigns each input to one of K activity classes. We aim to minimize the cross-entropy loss while keeping the model complexity low enough for resource-constrained deployment.

B. Overall Architecture

The proposed LiteTSC architecture, illustrated in Fig. 1, consists of four main components: (1) a multi-scale input module, (2) residual depthwise separable convolution blocks, (3) a lightweight channel attention module, and (4) a classification head. The network processes input data with dimensions $(C \times T) = (9 \times 128)$ for the UCI HAR dataset and outputs class probabilities.

C. Multi-Scale Input Module

To capture temporal patterns at different granularities, we design a multi-scale input module with parallel depthwise separable convolution branches. Unlike standard multi-scale approaches that use regular convolutions, our design employs depthwise separable convolutions to minimize parameters:

$$\mathbf{F}_s = \text{DSConv}_{k=3}(\mathbf{X}), \quad \mathbf{F}_l = \text{DSConv}_{k=7}(\mathbf{X}) \tag{1}$$

where $\mathbf{F}_s$ and $\mathbf{F}_l$ represent features from small (local) and large (global) receptive fields, respectively. The features are concatenated and fused through a 1×1 convolution:

$$\mathbf{F}_{ms} = \text{Conv}_{1\times1}([\mathbf{F}_s; \mathbf{F}_l]) \tag{2}$$

D. Residual Depthwise Separable Convolution Block

To compensate for the reduced representational capacity of depthwise separable convolutions, we incorporate residual connections [24]. Each block consists of two DSConv layers with a skip connection:

$$\mathbf{Y} = \text{ReLU}(\text{DSConv}_2(\text{ReLU}(\text{DSConv}_1(\mathbf{X}))) + \mathbf{X}') \tag{3}$$

where $\mathbf{X}'$ is the input projected to match output dimensions if necessary.

E. Lightweight Channel Attention

We adapt a dual-pooling channel attention mechanism inspired by CBAM [23] but optimized for efficiency. Unlike standard SE blocks that use only average pooling, we combine both average and max pooling to capture richer channel statistics:

$$\mathbf{s} = \sigma(\text{FC}(\text{AvgPool}(\mathbf{X})) + \text{FC}(\text{MaxPool}(\mathbf{X}))) \tag{4}$$

where σ is the sigmoid function and FC denotes shared fully-connected layers with a reduction ratio of 8. The attention weights $\mathbf{s}$ are applied element-wise to rescale the input features.

F. Classification Head

Following the attention module, we apply global average pooling (GAP) to aggregate spatial information, followed by dropout (rate 0.3) for regularization and a fully-connected layer for classification:

$$\hat{\mathbf{y}} = \text{softmax}(\text{FC}(\text{Dropout}(\text{GAP}(\mathbf{X})))) \tag{5}$$

The use of GAP instead of flattening significantly reduces parameters while providing translation invariance.

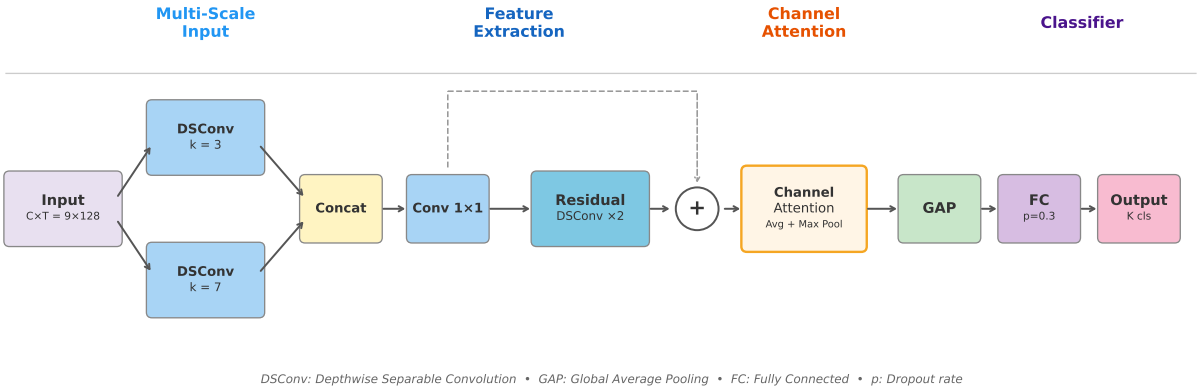

Fig. 1. Architecture of the proposed LiteTSC network. The model consists of a multi-scale input module with parallel depthwise separable convolutions (kernel sizes 3 and 7), followed by residual DSConv blocks and a dual-pooling channel attention mechanism.

IV. Experiments

A. Datasets

We evaluate LiteTSC on three widely-used HAR benchmark datasets:

UCI HAR [25]: Contains accelerometer and gyroscope data from 30 subjects performing 6 activities. We use the standard train/test split with 9 input channels and 128 time steps per window.

WISDM [26]: Smartphone accelerometer data from 36 subjects performing 6 activities. We use 3 input channels with 128 time steps.

PAMAP2 [27]: Data from 9 subjects with 3 IMUs performing 12 activities. We use 27 input channels with 128 time steps.

B. Implementation Details

All models are trained using PyTorch with the AdamW optimizer [28] (learning rate 0.001, weight decay 0.01) and cosine annealing scheduler [29]. We use label smoothing with $\epsilon = 0.1$ and train for 100 epochs with batch size 64. For fair comparison, we apply the same training protocol to all baseline models. Results are averaged over 5 runs with different random seeds.

C. Baseline Models

We compare against several architectures: (1) **SimpleCNN**: 3-layer CNN with standard convolutions; (2) **LSTM**: 2-layer bidirectional LSTM; (3) **CNN-LSTM**: CNN feature extractor with LSTM classifier; (4) **TinyHAR** [21]: reproduced using official implementation with our training protocol.

D. Main Results

Table I presents the comparison across three datasets. On UCI HAR, LiteTSC achieves the highest accuracy (96.74%) with only 21K parameters, outperforming TinyHAR (96.44%) while using 14.3× fewer parameters. The CPU inference time (2.21ms) is 3.3× faster than TinyHAR (7.35ms).

TABLE I
Comparison of model accuracy (%), parameter count, and inference time on UCI HAR dataset

Model	Params	CPU Time (ms)	Acc (%)
LSTM	539K	3.46	91.86
SimpleCNN	38K	0.46	94.71
CNN-LSTM	70K	0.91	94.71
TinyHAR	298K	7.35	96.44
LiteTSC	**21K**	**2.21**	**96.74**

TABLE II
Cross-dataset comparison of accuracy (%) and parameters

Dataset	Model	Params	Acc (%)
UCI HAR	LSTM	539K	91.86
	SimpleCNN	38K	94.71
	CNN-LSTM	70K	94.71
	TinyHAR	298K	96.44
	LiteTSC	**21K**	**96.74**
WISDM	LSTM	533K	99.57
	SimpleCNN	37K	99.76
	CNN-LSTM	68K	99.51
	TinyHAR	249K	99.67
	LiteTSC	**21K**	**99.78**
PAMAP2	LSTM	538K	91.08
	SimpleCNN	38K	90.53
	CNN-LSTM	70K	89.89
	TinyHAR	274K	**93.51**
	LiteTSC	**21K**	91.98

E. Cross-Dataset Evaluation

Table II shows results across all three datasets. LiteTSC consistently achieves competitive or superior accuracy with significantly fewer parameters.

On WISDM, all models achieve near-saturated performance (>99%), with LiteTSC achieving the highest accuracy (99.78%) while using 11.9× fewer parameters than TinyHAR. On PAMAP2 with 12 activity classes, TinyHAR achieves higher accuracy (93.51% vs 91.98%) but requires 13× more

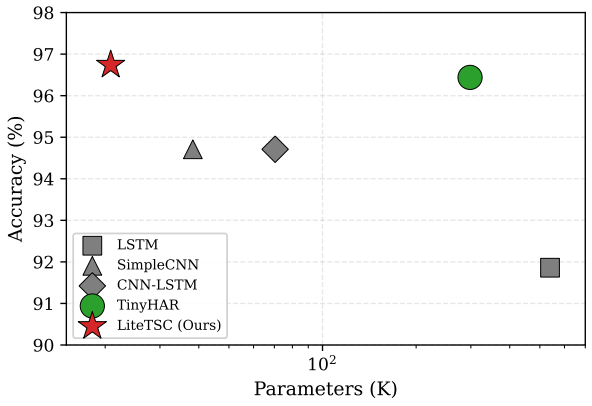

Fig. 2. Parameter-accuracy trade-off comparison on UCI HAR. LiteTSC achieves the best trade-off with highest accuracy (96.74%) and fewest parameters (20.8K).

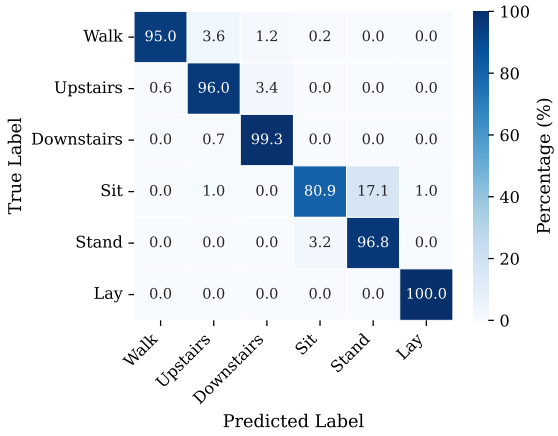

Fig. 3. Confusion matrix on UCI HAR test set. Numbers show percentages within each true class. The model achieves strong performance across most activities, with the expected Sitting-Standing confusion due to sensor placement limitations.

TABLE III
ABLATION STUDY ON UCI HAR DATASET (100 EPOCHS)

Variant	Params	Acc (%)
LiteTSC (Full)	**20,832**	**96.74**
w/o Multi-Scale	19,699	93.55 (−3.19)
w/o Residual	9,696	94.50 (−2.24)
w/o Attention	19,808	95.79 (−0.95)
w/o DSConv	61,990	94.33 (−2.41)

parameters (274K vs 21K). This trade-off demonstrates that LiteTSC maintains competitive performance across datasets while consistently offering significantly lower model complexity, making it more suitable for memory-constrained edge devices.

Fig. 2 visualizes the parameter-accuracy trade-off on UCI HAR. LiteTSC occupies the desirable upper-left region, achieving the highest accuracy with the fewest parameters. While TinyHAR also targets lightweight design, it requires 14.3× more parameters to achieve slightly lower accuracy.

F. Ablation Study

Table III shows the contribution of each component using the same 100-epoch training protocol as the main experiments.

Multi-Scale Input contributes the most (−3.19%), demonstrating the importance of capturing temporal patterns at different granularities. **DSConv replacement** shows that using standard convolutions increases parameters by 3× (to 62K) while actually decreasing accuracy by 2.41%, validating the effectiveness of depthwise separable convolutions. **Residual connections** contribute −2.24%, and **Channel Attention** contributes −0.95%.

G. Confusion Matrix Analysis

Fig. 3 shows the confusion matrix on UCI HAR. LiteTSC achieves perfect or near-perfect recognition for most activities, including Laying (100%), Walking Downstairs (99.3%), and Standing (96.8%). The primary confusion occurs between Sitting (80.9%) and Standing, with 17.1% of Sitting samples misclassified as Standing. This is a well-documented challenge inherent to waist-mounted or smartphone-based sensors, as discussed in Section V-B.

TABLE IV
COMPARISON WITH STATE-OF-THE-ART LIGHTWEIGHT METHODS ON UCI HAR

Method	Year	Acc (%)	Params
DeepConvLSTM [8]	2016	91-92	∼2M
TinyHAR [21]	2022	96.44†	298K
MLP-HAR [22]	2024	96.2*	29K
LiteTSC (Ours)	2026	**96.74**	**20.8K**

† Reproduced using official implementation with our experimental setup.
* Reported results from original papers; different experimental protocols.

H. Comparison with State-of-the-Art

Table IV compares LiteTSC with published lightweight methods that report parameter counts. Among these methods, LiteTSC achieves competitive accuracy with the smallest model size.

Note: Direct comparison is limited as different works use different experimental protocols. Among methods reporting parameters, LiteTSC achieves competitive accuracy with the smallest model size.

V. DISCUSSION

A. Design Insights

The success of LiteTSC can be attributed to the synergistic combination of several lightweight techniques. First, **depthwise separable convolutions** reduce the parameter count by approximately $k\times$ (where k is the kernel size) compared to standard convolutions, yet preserve the essential capability to extract temporal features. Our ablation study confirms that replacing DSConv with standard convolutions increases parameters by 3× while actually *decreasing* accuracy by

2.41%, suggesting that the regularization effect of DSConv may benefit generalization on limited HAR datasets.

Second, **multi-scale input processing** proves critical, contributing the largest accuracy improvement (+3.19%) in our ablation study. This aligns with TinyHAR's design guideline that "the extraction of local temporal context should be enhanced" [21], though our approach achieves this through parallel DSConv branches rather than transformer encoders.

Third, while TinyHAR employs a complex pipeline (CNN $\rightarrow$ Transformer $\rightarrow$ FC fusion $\rightarrow$ LSTM $\rightarrow$ Temporal Attention), LiteTSC achieves competitive or superior results with a simpler architecture (Multi-scale DSConv $\rightarrow$ Residual DSConv $\rightarrow$ Channel Attention $\rightarrow$ GAP). This observation is consistent with findings by Souza Leite et al. [30], who conducted over 500 experiments comparing transformer and CNN approaches for sensor-based HAR, concluding that "transformer-based solutions consistently yield inferior performance" while "posing higher computational demands" on edge devices. Our results suggest that for HAR tasks, architectural simplicity combined with well-designed lightweight operations may be more effective than complex multi-stage pipelines.

B. Analysis of Sitting-Standing Confusion

The confusion matrix (Fig. 3) reveals that Sitting achieves only 80.9% accuracy, with 17.1% misclassified as Standing. This is a well-documented challenge in sensor-based HAR. Gjoreski et al. [31] demonstrated that "hip placement is less suitable to differentiate between sitting and standing," as the torso orientation remains nearly identical during both postures when using waist-mounted sensors. Vähä-Ypyä et al. [32] further confirmed that distinguishing sitting from standing with hip-worn accelerometers requires specialized angle-based methods beyond standard classification approaches.

Research has shown that reliably distinguishing sitting from standing typically requires either (1) thigh-worn sensors that can detect leg orientation, or (2) multi-sensor configurations combining hip and thigh measurements [33]. Since UCI HAR uses smartphone sensors typically carried in pockets or worn on the waist, this limitation is expected and affects all models evaluated on this dataset. The fact that LiteTSC still achieves 96.74% overall accuracy despite this inherent challenge indicates strong performance on distinguishable activity classes.

C. Practical Deployment Considerations

With only 20.8K parameters (approximately 83KB in float32), LiteTSC is well-suited for deployment on microcontrollers with limited memory. For comparison, the STM32L4 series commonly used in wearables offers 256KB-1MB flash memory, providing ample space for the model. Further size reduction through INT8 quantization could reduce the model to approximately 20KB, enabling deployment on even more constrained devices.

Recent work has demonstrated successful HAR deployment on resource-constrained hardware. Zhou et al. [34] deployed a DeepConvLSTM model on the Arduino Nano 33 BLE Sense Rev2 via the Edge Impulse platform, reducing model size from 513KB to 137KB through full integer quantization while achieving 98.24% accuracy on the WISDM dataset. Given LiteTSC's significantly smaller footprint (20.8K parameters vs. DeepConvLSTM's hundreds of thousands), we expect comparable or faster inference on similar hardware.

However, our CPU inference time (2.21ms) was measured on a desktop processor. Actual inference time on edge devices will depend on hardware capabilities and potential optimizations such as CMSIS-NN acceleration. Future work should include on-device benchmarking with power consumption measurements.

VI. Conclusion

We presented LiteTSC, a lightweight neural network for time series classification that systematically combines depthwise separable convolutions, multi-scale feature extraction, residual connections, and channel attention. Experiments on three HAR benchmarks show that LiteTSC achieves state-of-the-art accuracy (96.74% on UCI HAR) with only 20.8K parameters. Compared to TinyHAR, the ISWC 2022 Best Paper, LiteTSC achieves higher accuracy with 14.3$\times$ fewer parameters and 3.3$\times$ faster CPU inference, demonstrating an effective design for lightweight HAR.

Limitations: This work focuses on model design without actual deployment on embedded hardware. Future work includes quantization-aware training, deployment on microcontrollers (e.g., STM32, ESP32), and comprehensive latency/power measurements on real edge devices.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 52025132, U24A20205, 21621091, 22021001, 22121001 awarded to X.H.), the 111 Project (Nos. B17027, B16029), National Science Foundation of Fujian Province of China (No. 2022J02059), the New Cornerstone Science Foundation through the XPLORER PRIZE. The authors thank the anonymous reviewers for their valuable comments. Portions of this paper were refined with the assistance of Claude (Anthropic) for language editing.

References

[1] A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Computing Surveys*, vol. 46, no. 3, pp. 33:1–33:33, 2014.

[2] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.

[3] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[4] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–40, 2021.

[5] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, "Deep convolutional neural networks on multichannel time series for human activity recognition," in *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 2015, pp. 3995–4001.

[6] S. Ha and S. Choi, "Convolutional neural networks for human activity recognition using multiple accelerometer and gyroscope sensors," in *International Joint Conference on Neural Networks*. IEEE, 2016, pp. 381–388.

[7] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, 2016, pp. 1533–1540.

[8] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.

[12] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[13] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, "Convolutional neural networks for human activity recognition using mobile sensors," in *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services*, 2014, pp. 197–205.

[14] Y. Guan and T. Plötz, "Ensembles of deep lstm learners for activity recognition using wearables," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–28, 2017.

[15] M. Zeng, T. Yu, X. Wang, L. T. Nguyen, O. J. Mengshoel, and I. Lane, "Understanding and improving recurrent networks for human activity recognition by continuous attention," in *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, 2018, pp. 56–63.

[16] H. Ma, W. Li, X. Zhang, S. Gao, and S. Lu, "Attnsense: Multi-level attention mechanism for multimodal human activity recognition," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 3109–3115.

[17] Y. Tang, L. Zhang, F. Min, and J. He, "Multiscale deep feature learning for human activity recognition using wearable sensors," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 2, pp. 2106–2116, 2023.

[18] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu, and X. Li, "Smartphone sensor-based human activity recognition using feature fusion and maximum full a posteriori," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 7, pp. 3992–4001, 2020.

[19] A. Ignatov, "Real-time human activity recognition from accelerometer data using convolutional neural networks," *Applied Soft Computing*, vol. 62, pp. 915–922, 2018.

[20] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, "Deep learning models for real-time human activity recognition with smartphones," *Mobile Networks and Applications*, vol. 25, no. 2, pp. 743–755, 2020.

[21] Y. Zhou, H. Zhao, Y. Huang, M. Hefenbrock, T. Riedel, and M. Beigl, "Tinyhar: A lightweight deep learning model designed for human activity recognition," in *Proceedings of the 2022 ACM International Symposium on Wearable Computers*. ACM, 2022, pp. 89–93.

[22] Y. Zhou, T. King, H. Zhao, Y. Huang, T. Riedel, and M. Beigl, "Mlp-har: Boosting performance and efficiency of har models on edge devices with purely fully connected layers," in *Proceedings of the 2024 ACM International Symposium on Wearable Computers*. ACM, 2024, pp. 133–139.

[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 3–19.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[25] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones," in *Proceedings of the 21st European Symposium on Artificial Neural Networks*, 2013, pp. 437–442.

[26] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[27] A. Reiss and D. Stricker, "Introducing a new benchmarked dataset for activity monitoring," in *International Symposium on Wearable Computers*. IEEE, 2012, pp. 108–109.

[28] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *International Conference on Learning Representations*, 2019.

[29] ——, "Sgdr: Stochastic gradient descent with warm restarts," in *International Conference on Learning Representations*, 2017.

[30] C. Souza Leite, H. Mauranen, A. Zhanabatyrova, and Y. Xiao, "Transformer-based approaches for sensor-based human activity recognition: Opportunities and challenges," *arXiv preprint arXiv:2410.13605*, 2024.

[31] H. Gjoreski, M. Lustrek, and M. Gams, "Accelerometer placement for posture recognition and fall detection," in *Proceedings of the 7th International Conference on Intelligent Environments*. IEEE, 2011, pp. 47–54.

[32] H. Vähä-Ypyä, P. Husu, J. Suni, T. Vasankari, and H. Sievänen, "Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer," *Scandinavian Journal of Medicine & Science in Sports*, vol. 28, no. 3, pp. 1092–1102, 2018.

[33] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, and D. Finlay, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors*, vol. 13, no. 7, pp. 9183–9200, 2013.

[34] H. Zhou, X. Zhang, Y. Feng, T. Zhang, and L. Xiong, "Efficient human activity recognition on edge devices using DeepConv LSTM architectures," *Scientific Reports*, vol. 15, no. 1, p. 13830, 2025.