ARZCORE RESEARCH / PREPRINT 2026

C3-DeltaMind: A 48K-Parameter Derivative-Augmented Network for WiFi CSI Human Activity Recognition

Osman Akkawi
Information Technology / Computer Science, City University, Tripoli, Lebanon
ORCID: 0009-0008-6151-1548

STATUS

Preprint - submitted to IEEE Sensors Journal

Peer review is in progress. This work has not been accepted or published by IEEE.

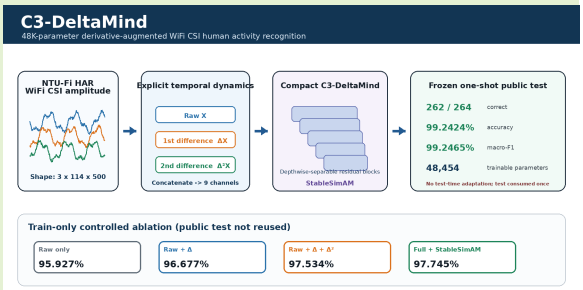
Public preprint landing page: https://arzcore.com/research/c3-deltamind

Public release: September 2026

C3-DeltaMind: A 48K-Parameter Derivative-Augmented Network for WiFi CSI Human Activity Recognition

Osman Akkawi

***Abstract*— WiFi channel state information (CSI) enables device-free human activity recognition (HAR), but strong benchmark accuracy is often obtained with models that are unnecessarily large for resource-constrained sensing. This paper presents C3-DeltaMind, a 48,454-parameter network that explicitly stacks normalized CSI amplitude, first temporal differences, and second temporal differences before compact depthwise-separable processing with parameter-free StableSimAM attention. On the predefined NTU-Fi HAR split, the frozen model was retrained on all 936 training samples and evaluated once on the 264-sample public test after a train-integrity gate and a permanent consumption lock were established. The model correctly classified 262/264 samples, achieving 99.2424% accuracy, 99.2465% macro-F1, and 97.7778% minimum class F1. A post-final train-only controlled ablation using five acquisition-order blocked folds and equal 48,454-parameter variants increased mean accuracy from 95.927% for raw CSI alone to 96.677% with the first temporal difference, 97.534% after adding the second difference, and 97.745% for the complete model. The results support explicit temporal differencing as the main architectural contribution while preserving a small learned parameter budget. The paper also documents the frozen training recipe, test-set governance, hashes, and limitations needed for reproducible interpretation.**



***Index Terms*— channel state information, compact neural networks, depthwise-separable convolution, edge AI, human activity recognition, reproducibility, SimAM, WiFi sensing.**

I. Introduction

WIFI sensing infers human motion from perturbations of wireless propagation and can operate without optical line of sight or wearable devices [1]. Channel state information provides fine-grained amplitude and phase measurements that support learning-based device-free sensing. SenseFi systematized public WiFi sensing datasets and baseline models and showed that compact models can be competitive with much larger networks [2].

NTU-Fi HAR is a useful benchmark for studying parameter-efficient models. In the SenseFi release, each processed example has shape $3 \times 114 \times 500$; the predefined split contains 936 training and 264 test examples over six balanced activities: box, circle, clean, fall, run, and walk [2], [3]. SenseFi reports 99.69% accuracy for MLP and BiLSTM, but the MLP uses 175.24M parameters and BiLSTM uses 0.209M. More recent models include RGANet at 99.24% [4], HMS-RWKV at 99.62% with about 0.076M parameters [5], and CNN-ECBAM at 100% using a pseudo-color image pipeline [6]. Because preprocessing and checkpoint policies differ, these point estimates are descriptive rather than a universal rank.

This research received no external funding.

O. Akkawi is with Information Technology / Computer Science, City University, Tripoli, Lebanon (e-mail: osmanjihadakkawi@outlook.com; ORCID: 0009-0008-6151-1548).

This work asks whether near-ceiling NTU-Fi HAR accuracy can be retained with a much smaller learned representation and stricter test-set governance. C3-DeltaMind explicitly encodes temporal derivatives before compact spatiotemporal filtering, combines depthwise-separable convolution [7] with a stabilized parameter-free SimAM module [8], and contains 48,454 trainable parameters. Development additionally exposed a strong acquisition-order effect, motivating a freeze-before-test protocol rather than repeated public-test monitoring.

The main contributions are: 1) a nine-channel representation of raw normalized CSI, first temporal difference, and second temporal difference; 2) a 48,454-parameter depthwise-separable residual network with StableSimAM; 3) a leakage-aware development and one-shot public-test protocol; and 4) a frozen public result of 99.2424% accuracy and 99.2465% macro-F1, accompanied by controlled train-only component ablation and cryptographic evidence.

II. Method

A. Input and derivative representation

Let X_{raw} denote the amplitude matrix from the `CSIamp` field. Following the processed NTU-Fi convention, each example is normalized with fixed constants and downsampled

along packet order:

$$X = \mathrm{reshape}\left(\frac{X_{\mathrm{raw}} - 42.3199}{4.9802}[:, :: 4], 3, 114, 500\right). \quad (1)$$

The first and second temporal differences are

$$D_t^{(1)} = X_t - X_{t-1}, \quad (2)$$

$$D_t^{(2)} = D_t^{(1)} - D_{t-1}^{(1)}, \quad (3)$$

with the first temporal position zero-filled. The model input is

$$Z_0 = \mathrm{concat}[X, D^{(1)}, D^{(2)}] \in \mathbb{R}^{9\times114\times500}. \quad (4)$$

The deterministic derivative operators introduce no learned parameters and expose velocity-like and acceleration-like changes before feature extraction.

B. Human-subjects and informed-consent statement

This study is a secondary computational analysis of the publicly released, processed NTU-Fi HAR benchmark distributed through SenseFi [2]. The author did not recruit participants, interact with participants, or access direct identifiers. Accordingly, no new informed consent was obtained for this secondary analysis. The present study relies on the dataset as released by the original investigators; ethics approval and consent procedures for the original data collection remain those of the source study and are not re-asserted here.

C. Compact backbone and StableSimAM

A 5×9 stem with stride $(2, 4)$ maps nine channels to 32 feature maps. Five depthwise-separable blocks use depthwise 5×7 filtering, 1×1 pointwise mixing, batch normalization, SiLU activation, dropout, and StableSimAM. Downsampling occurs in blocks 2 and 4. Global mean and maximum pooling are concatenated and passed through LayerNorm, 10% dropout, and a six-class linear head. Fig. 1 summarizes the network.

For a feature element x, StableSimAM computes $d = (x - \mu)^2$, $v = \sum d / \max(HW - 1, 1)$, and

$$y = x\,\sigma\left(\mathrm{clamp}\left(\frac{d}{4\max(v + 10^{-4}, 10^{-4})} + 0.5, -30, 30\right)\right). \quad (5)$$

All attention operations use FP32. The variance floor and clamp prevent non-finite behavior observed in earlier mixed-precision development. The complete model contains 48,454 trainable parameters, equivalent to 193,816 raw FP32 weight bytes (189.3 KiB). A direct convolution/linear count is approximately 135.04M MACs per sample; therefore parameter compactness should not be confused with minimum arithmetic cost or embedded latency.

III. Experimental Protocol

A. Dataset and development audit

The processed SenseFi NTU-Fi HAR split contains 156 training and 44 test examples per class, for 936 and 264 samples respectively [2]. The public test was not used for model selection. A random train-only development split initially reached 100% validation accuracy. Because this was suspiciously high, five contiguous filename-sorted blocked folds were used as an acquisition-order stress test, yielding 97.315% mean accuracy and 93.011% worst-fold accuracy. A train-only structure audit found no explicit subject/session identifier in filenames or MAT metadata, but found strong order structure, with a median adjacent-to-random summary-feature distance ratio of 0.570017. We therefore did not invent a subject-disjoint protocol; blocked folds were treated only as a stress diagnostic and the predefined train/test split was retained for the benchmark endpoint.

TABLE I

Architecture summary.

Stage	Output	Params
Derivative stack	$9 \times 114 \times 500$	0
Stem 5×9	$32 \times 57 \times 125$	13,024
DS block 1	$32 \times 57 \times 125$	2,272
DS block 2	$64 \times 29 \times 63$	3,360
DS block 3	$64 \times 29 \times 63$	6,592
DS block 4	$96 \times 15 \times 32$	8,704
DS block 5	$96 \times 15 \times 32$	12,960
LayerNorm + classifier	6	1,542
Total	–	**48,454**

B. Frozen training recipe

The final configuration was frozen before public-test access: seed 2026, width 32, five depthwise-separable blocks, batch size 24, AdamW [9], learning rate 10^{-3}, weight decay 10^{-4}, label smoothing 0.05, dropout 0.10, gradient clipping at 1.0, and 20 epochs. A cosine scheduler used $T_{\max} = 80$ and $\eta_{\min} = 10^{-5}$ while stopping at epoch 20 to preserve the development trajectory. FP32 was used throughout; AMP and TF32 were disabled and deterministic algorithms were enabled.

Frozen augmentation consisted of per-channel gain in $[0.96, 1.04]$ ($p = 0.75$), Gaussian noise with standard deviation 0.015 ($p = 0.70$), temporal roll of up to ±12 packets ($p = 0.45$), temporal masking of width 4–30 ($p = 0.30$), and subcarrier masking of width 2–9 ($p = 0.25$). No mixup or test-time adaptation was used.

C. One-shot public-test governance

The final runner retrained on all 936 training examples, then required finite states/logits, training accuracy $\geq 95\%$, every class predicted, minimum class recall $\geq 80\%$, and dominant prediction fraction $\leq 70\%$. After the gate passed, the checkpoint was saved and hashed, a permanent consumption lock was written, and only then was `test_amp` enumerated and evaluated once (Fig. 2). A failed integrity gate exits before test access.

IV. Results

A. Integrity gate and frozen public result

The final retraining passed all integrity checks. At epoch 20, training accuracy, macro-F1, and minimum class recall were 100%; all six prediction counts were 156; logits were finite from -4.0415 to 5.4590; and the dominant fraction was $1/6$. These checks detect collapse rather than estimate generalization.

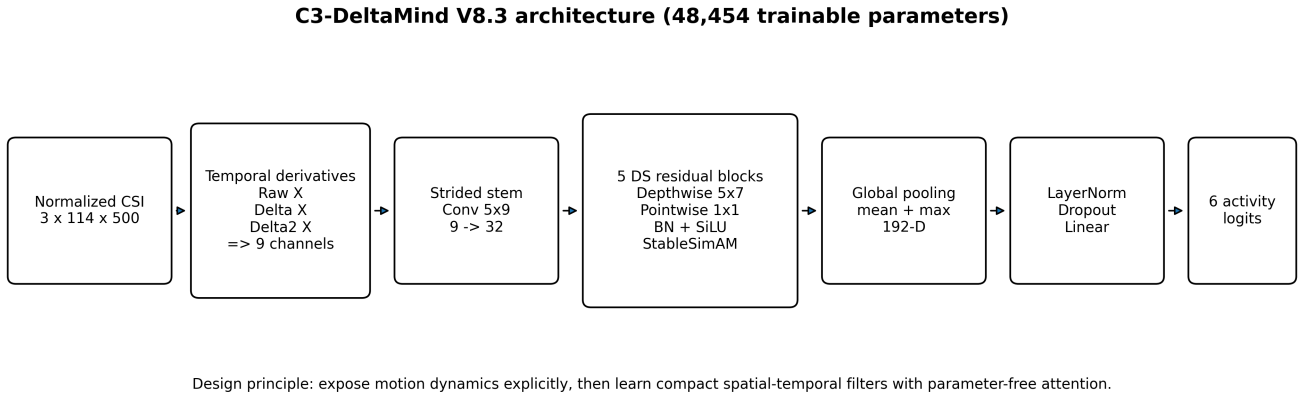


Fig. 1. C3-DeltaMind architecture. Explicit temporal derivatives are formed before compact depthwise-separable spatiotemporal filtering.

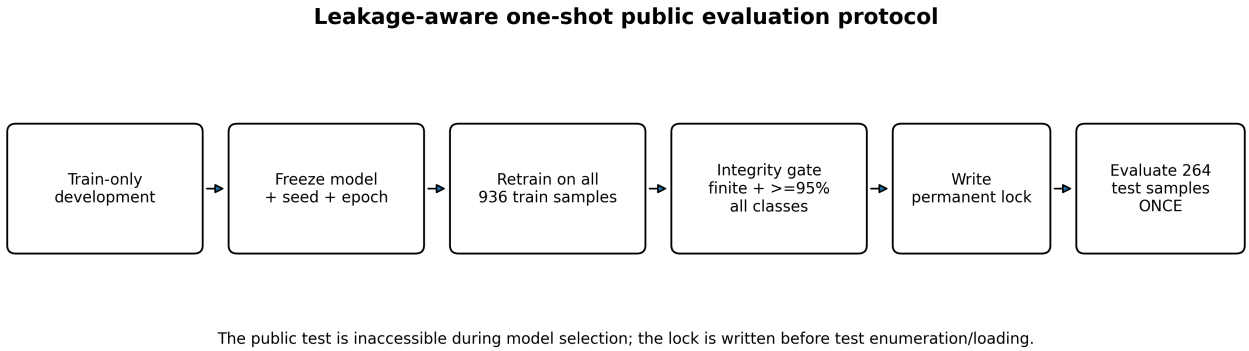


Fig. 2. Leakage-aware one-shot public evaluation. The public test is inaccessible during model selection; the consumption lock is written before test enumeration/loading.

TABLE II

FROZEN PUBLIC-TEST METRICS.

Metric	Result
Correct / total	262 / 264
Accuracy	99.2424%
Macro-F1	99.2465%
Minimum class F1	97.7778%
95% Wilson interval	97.28–99.79%
Prediction counts	[44,43,46,43,44,44]
Test-time adaptation	None

TABLE III

TRAIN-ONLY CONTROLLED ABLATION.

Variant	Mean Acc.	Worst Acc.	Macro-F1
Raw only	95.927	90.323	95.824
Raw + Δ	96.677	90.860	96.566
Raw + Δ + Δ^2, no SimAM	97.534	93.548	97.499
Full C3-DeltaMind	**97.745**	**93.548**	**97.696**

On the one-shot public test, C3-DeltaMind classified 262/264 examples correctly: 99.2424% accuracy, 99.2465% macro-F1, and 97.7778% minimum class F1. The 95% Wilson interval for accuracy is 97.28–99.79%. The two errors were circle→clean and fall→clean (Fig. 3). No test-time adaptation was performed.

B. Train-only controlled component ablation

After the public final was consumed, a component ablation was performed using `train_amp` only. Four equal-parameter variants kept the same nine-channel stem and 48,454 trainable parameters; absent derivative groups were zero-filled, and SimAM is parameter-free. Each variant used the same five contiguous filename-sorted blocked folds and the frozen training recipe. The public `test_amp` was neither enumerated nor loaded.

Adding the first temporal difference improved mean accuracy by 0.750 percentage points; the second difference added another 0.857 points; StableSimAM added 0.212 points. The fold-level effect of attention was not uniformly positive, so the defensible interpretation is that derivative augmentation provides the dominant gain and StableSimAM provides a smaller average benefit under this stress test.

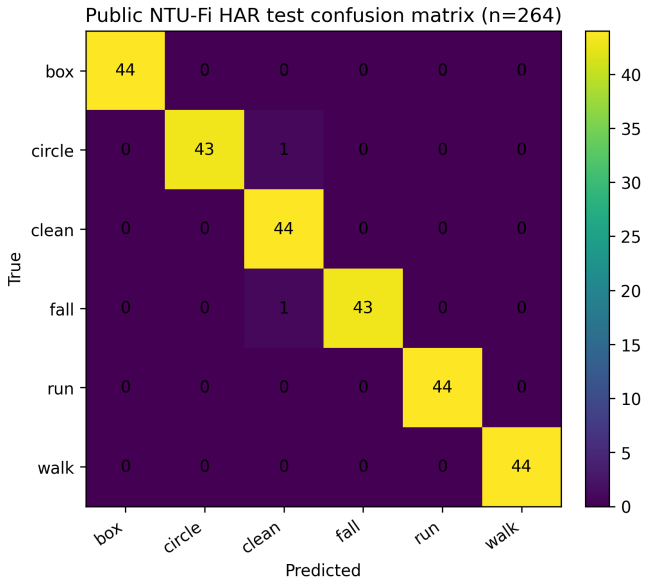


Fig. 3. Frozen one-shot public-test confusion matrix ($n = 264$).

TABLE IV

SELECTED PUBLISHED NTU-FI HAR POINT ESTIMATES.

Method	Acc.	Params	Note
SenseFi CNN-5 [2]	98.70	0.477M	benchmark
SenseFi BiLSTM [2]	99.69	0.209M	benchmark
SenseFi ResNet50 [2]	99.38	23.55M	benchmark
RGANet [4]	99.24	∼0.44M	different training
HMS-RWKV [5]	99.62	0.076M	different protocol
CNN-ECBAM [6]	100.00	–	pseudo-color RGB
C3-DeltaMind	**99.24**	**0.048M**	one-shot frozen

C. Descriptive comparison with published NTU-Fi HAR results

Table IV places the frozen point estimate beside selected published results. SenseFi values use its processed benchmark pipeline [2]. HMS-RWKV and RGANet use different training configurations [4], [5], while CNN-ECBAM converts CSI to pseudo-color RGB images [6]. These protocol differences preclude a strict universal ranking.

C3-DeltaMind is substantially smaller than the SenseFi CNN-5, BiLSTM, and ResNet50 baselines and is also smaller than the reported HMS-RWKV parameter count, while retaining a near-ceiling public-test point estimate. The result supports a parameter-efficiency claim, not a global accuracy-SOTA claim.

V. DISCUSSION

The ablation supports the central design hypothesis: making temporal velocity- and acceleration-like changes explicit reduces the burden on the learned network. The increase from 95.927% raw-only mean accuracy to 97.534% with both derivatives is larger than the additional StableSimAM gain. This is useful for compact sensing because deterministic feature formation costs no learned storage.

The large gap between random-split development (100%) and blocked stress validation demonstrates why CSI benchmarks should be interpreted with acquisition structure in mind. In the processed training set we found no explicit metadata that would justify a subject-disjoint split, so the paper reports blocked cross-validation only as a conservative stress test rather than a replacement official protocol.

Parameter efficiency and compute efficiency must also be separated. The 48,454 learned weights occupy only 189.3 KiB in FP32, but the current high-resolution input and stem require about 135.04M convolution/linear MACs per sample. Embedded latency, peak RAM, energy, and quantized accuracy were not measured for this exact NTU-Fi model. Earlier ESP32-C3 experiments on smaller DeltaMind prototypes are therefore not reused as evidence here.

Finally, the public test contains only 264 examples. One error changes accuracy by approximately 0.379 percentage points, and the 95% Wilson interval overlaps several published point estimates. Without matched repeated trials and paired predictions, differences of a few tenths of a point should not be interpreted as statistically decisive cross-paper rank differences.

VI. REPRODUCIBILITY, LIMITATIONS, AND ETHICS

Full SHA-256 values for the frozen model, final report, dataset-structure audit, and experiment packages are provided in the supplementary material together with the exact final runner and public-test policy. Exact frozen source and the train-only ablation runner are prepared for archival release. The dataset is not redistributed and should be obtained from SenseFi [2].

The main limitations are: one public HAR benchmark; no explicit subject/session labels in the processed training split; train-only rather than public-test component ablation; no exact-model embedded measurements; descriptive cross-paper comparison under nonidentical protocols; and a small public test set. These restrictions limit the claim to the frozen NTU-Fi HAR result rather than universal generalization.

Human-subjects and informed-consent context is stated explicitly in the Method section. This secondary analysis used only the publicly released, processed benchmark and did not involve new participant recruitment, interaction, or access to direct identifiers. Device-free WiFi sensing avoids image capture but can still reveal behavioral patterns, so real deployments should use purpose limitation, access control, data minimization, and consent where appropriate.

VII. CONCLUSION

C3-DeltaMind shows that near-ceiling NTU-Fi HAR performance can be obtained with a small learned parameter budget. The 48,454-parameter network combines explicit first- and second-order temporal differences, depthwise-separable residual filtering, StableSimAM, and mean/max pooling. Under a frozen one-shot public-test protocol it classified 262/264 examples correctly, reaching 99.2424% accuracy and 99.2465% macro-F1. Controlled train-only ablation identifies temporal derivatives as the dominant architectural contribution. The model should be described as parameter-compact rather than universally SOTA or fully microcontroller-ready; cross-environment generalization and exact-hardware efficiency remain future work.

Data and Code Availability

NTU-Fi HAR is available through the SenseFi benchmark. The C3-DeltaMind repository is currently private pending publication and IP review. Exact final-evaluation source, hashes, and train-only ablation code are prepared for archival release; no Zenodo DOI is claimed in this submission draft.

Declarations

Author contribution: Osman Akkawi: conceptualization, methodology, software, investigation, validation, formal analysis, visualization, derived benchmark artifact curation, and writing. *Funding:* This research received no external funding. *Conflict of interest:* The author declares no known conflict of interest directly affecting the reported work. Any future commercial or intellectual-property interest related to C3-DeltaMind will be disclosed to the journal if applicable.

Acknowledgment

OpenAI ChatGPT was used during method development for architecture brainstorming and code-drafting assistance, and during manuscript preparation for language and organization in the Abstract, Introduction, Discussion, declarations, supplementary text, and graphical-abstract layout. The author executed the experiments, reviewed and revised all generated code and text, verified every reported numerical result and hash, interpreted the findings, and accepts full responsibility for the scientific content. No AI-generated experimental data or benchmark measurements are reported.

References

[1] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[2] J. Yang, X. Chen, D. Wang, H. Zou, C. X. Lu, S. Sun, and L. Xie, "Sensefi: A library and benchmark on deep-learning-empowered wifi human sensing," *Patterns*, vol. 4, no. 3, p. 100703, 2023.

[3] J. Yang, X. Chen, H. Zou, D. Wang, Q. Xu, and L. Xie, "Efficientfi: Towards large-scale lightweight wifi sensing via csi compression," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13 086–13 095, 2022.

[4] J. Hu, F. Ge, X. Cao, and Z. Yang, "Rganet: A human activity recognition model for extracting temporal and spatial features from wifi channel state information," *Sensors*, vol. 25, no. 3, p. 918, 2025.

[5] Y. Liu and Z. Sheng, "Hms-rwkv: A hybrid multi-scale model with spatial-temporal adaptive fusion for efficient wifi sensing," in *Proc. IEEE Int. Conf. Communications (ICC)*, 2026, pp. 1–6.

[6] B. Hu and Y. Tong, "Human activity recognition using a cnn with an enhanced convolutional block attention module," *Wuhan University Journal of Natural Sciences*, vol. 31, no. 1, pp. 10–24, 2026.

[7] 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.

[8] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Machine Learning (ICML)*, vol. 139, 2021, pp. 11 863–11 874.

[9] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learning Representations (ICLR)*, 2019.