Task-Adaptive Wavelet Decomposition with Quadrature Mirror Filters for Time Series Classification

Felipe G. S. Teodoro[1], **Clodoaldo A. M. Lima**[1]

[1]Programa de Pós-Graduação em Sistemas de Informação (PPgSI)
Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (EACH-USP)
Av. Arlindo Béttio, 1000 – 03828-000 – São Paulo – SP – Brazil

fteodoro@usp.br, c.lima@usp.br

Abstract. *Deep models for time series classification (TSC) are usually fed raw signal windows, and fixed wavelet families such as Daubechies-4 (Db4) target generic smoothness rather than the discriminative frequency content of a given task; it is unclear* which *architectures benefit from replacing fixed filters with trainable ones. We investigate a* Learnable Wavelet Transform *(LWT): a differentiable front-end that learns per-channel low-pass filter coefficients end-to-end with the classifier, derives the high-pass filter structurally through the Quadrature Mirror Filter (QMF) relationship, and can be warm-started from the Db4 basis by a closed-form least-squares fit. Crossing four input conditions (Raw, fixed Db4, LWT, LWT-Warm) with five backbones (MLP, CNN, LSTM, CNN-LSTM, Transformer) on three benchmarks (Ford-A, SCP1, UWaveGesture), we test the hypothesis that the value of a wavelet front-end is backbone-dependent. The results confirm it: globally expressive backbones are largely insensitive to the front-end, whereas a fixed Db4 decomposition repeatedly collapses convolutional and recurrent backbones to the majority class, and making the filter learnable reverses every collapse, peaking at* $+56.4$ *accuracy points over Db4 on the low-data UWaveGesture.*

1. Introduction

Time series classification (TSC) underpins applications from industrial fault detection and brain–computer interfaces to gesture recognition, and deep neural networks have become the dominant modelling tool for the task [Ismail Fawaz et al. 2019]. A recurring and often under-examined design decision is the *input representation*: many pipelines feed raw windows directly to the model, even though real-world signals superimpose fast oscillations on slow trends in an inherently multi-scale structure.

The Discrete Wavelet Transform (DWT) offers a natural multi-resolution decomposition that separates these scales through an iterated two-channel filter bank [Mallat 2008], but fixed wavelet families such as Daubechies-4 (Db4) are designed for *generic* smoothness [Daubechies 1992], and the frequency band that discriminates between classes is dataset- and even channel-dependent. Trainable wavelet filter banks have therefore been studied in several domains – unsupervised condition monitoring [Michau et al. 2022], denoising [Frusque and Fink 2024], image classification [Le et al. 2024] and financial forecasting [Jia et al. 2026] – but almost always on a single downstream architecture, and rarely in supervised TSC. Two questions thus remain

open: *which* backbone families actually benefit from a learnable multi-scale front-end, and whether a fixed wavelet is at least a harmless default when they do not.

In this paper we investigate a **Learnable Wavelet Transform (LWT)**,[1] a modular, fully differentiable front-end that learns one low-pass filter per input channel end-to-end with the classifier and derives the complementary high-pass filter through the Quadrature Mirror Filter (QMF) conjugate relationship [Smith and Barnwell 1984], kept structural at every training step as in [Michau et al. 2022]. A closed-form least-squares (Moore–Penrose) *warm-start* initialises the filter-generating network so that it reproduces the Db4 basis before joint training, after which every parameter is free. The layer is architecture-agnostic and adds only a few thousand parameters. Its ingredients – QMF-structured trainable filter banks and Daubechies warm-starts – exist separately in prior work (Section 2); the focus here is their joint evaluation in supervised TSC.

Our working hypothesis is that the value of an explicit multi-scale decomposition is *backbone-dependent*: architectures with a global receptive field can compensate internally for a mismatched representation, whereas convolutional and recurrent ones cannot. To test it, we run a controlled protocol on three complementary public benchmarks – Ford-A (univariate, binary [Dau et al. 2019]), SelfRegulationSCP1 (6-channel EEG, binary [Bagnall et al. 2018]) and UWaveGesture (3-axis accelerometer, 8 classes, low-data [Liu et al. 2009]) – crossing four input conditions (Raw, fixed Db4, LWT, LWT-Warm) with five backbones (MLP, CNN, LSTM, CNN-LSTM, Transformer) under a shared hyperparameter grid and equal tuning budget.

This paper makes three contributions: (i) a controlled study across five backbone families and three regimes that isolates when a learnable decomposition is beneficial and documents a concrete failure mode – a fixed Db4 front-end repeatedly collapses convolutional and recurrent backbones to the majority-class solution, and learning the filter reverses every such collapse; (ii) the LWT front-end configuration, combining a structurally enforced QMF parameterisation with a closed-form Moore–Penrose Db4 warm-start, a combination prior work has explored only separately and outside supervised TSC; and (iii) an audit of the learned filters establishing that the QMF relationship is satisfied exactly while orthogonality and perfect reconstruction are relaxed, and that the filters specialise per channel away from the Db4 initialisation.

Section 2 reviews related work; Sections 3 and 4 describe the method and the experimental setup; Section 5 presents the results; Section 6 concludes.

2. Related Work

Models for TSC form a wide and actively growing landscape, from convolutional ensembles such as InceptionTime [Ismail Fawaz et al. 2020] and random-kernel methods such as ROCKET [Dempster et al. 2020] to meta-ensembles such as HIVE-COTE 2.0 [Middlehurst et al. 2021] and attention-based models [Vaswani et al. 2017]. Rather than attempting to cover this landscape, our study selects five widely used architectural families – fully connected, convolutional, recurrent, hybrid convolutional-recurrent and Transformer [Ismail Fawaz et al. 2019] – as *representatives* of qualitatively distinct processing styles. Comparative surveys report that no single family dominates, and that

[1] Source code and experimental scripts: https://github.com/felipesteodoro/LearnableWaveletLayer.

performance is often as sensitive to how the signal is presented to the network as to the network itself [Ismail Fawaz et al. 2019]; still, most pipelines operate directly on the raw signal, leaving the separation of frequency scales entirely to the model, even though many real signals – vibration traces, EEG, accelerometer readings – superimpose fast transients on slow trends.

The DWT provides exactly such a separation through an iterated two-channel filter bank [Mallat 2008], whose behaviour is entirely determined by the mother wavelet and the number of levels; classical families such as Daubechies are designed for generic smoothness rather than for any particular discriminative task [Daubechies 1992], and the QMF construction [Smith and Barnwell 1984] is the standard device for deriving the high-pass branch from the low-pass one (Section 3).

A growing body of work therefore replaces fixed filters with trainable ones: wavelet convolutional networks [Liu et al. 2019], scattering transforms [Bruna and Mallat 2013] and parameterised band-pass front-ends such as SincNet [Ravanelli and Bengio 2018]. [Recoskie and Mann 2018] first framed the 1D DWT as a constrained network trained by autoencoding, with the QMF and orthogonality conditions imposed as soft penalties; DeSpaWN [Michau et al. 2022] instead embeds the quadrature-mirror relationship *structurally* in the architecture, for unsupervised monitoring of high-frequency signals; L-WPT [Frusque and Fink 2024] learns a wavelet *packet* transform initialised from a Daubechies basis for denoising, the QMF property holding only at initialisation; and lattice-parameterised units enforce orthogonality by construction for image classification, also initialised from Daubechies filters [Le et al. 2024]. Closer to time series, DB2-TransF [Gupta and Tripathi 2025] replaces attention with Daubechies-initialised learnable filters for forecasting, GPT4FTS [Jia et al. 2026] attaches a Db4-initialised learnable stationary wavelet transform to a GPT backbone for financial prediction, and WaveLSFormer [Li et al. 2026] trains low- and high-pass kernels with spectral regularisers for equity trading – in these three, the two filters are trained independently, without a structural QMF coupling.

Our front-end claims no new class of transform: it combines the structural QMF parameterisation of [Michau et al. 2022] with a Daubechies warm-start in the spirit of [Frusque and Fink 2024, Le et al. 2024], obtained here in closed form by least squares. What distinguishes this study is the evaluation: prior learnable-wavelet works validate the front-end on a single downstream architecture, mostly outside supervised classification, whereas we ask – under one controlled protocol spanning five backbone families and three regimes – *which* backbones actually benefit from a learnable multi-scale decomposition, and document a failure mode in which a fixed wavelet, far from being neutral, collapses the downstream model. A direct empirical comparison against these neighbouring front-ends under the same protocol is a natural next step, which we leave as future work.

3. Method

The system under study places a learnable wavelet front-end before an otherwise standard classification backbone, so the multi-resolution decomposition adapts to each dataset's spectral characteristics jointly with the classifier. Figure 1 summarises the pipeline: an input window is presented to one of four interchangeable front-ends; a bias-free linear projection equalises the channel dimensionality; and a shared backbone produces the pre-

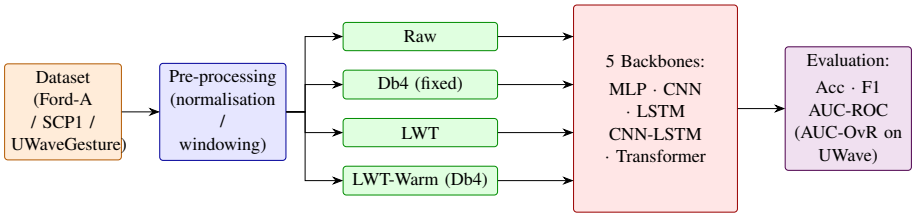


Figure 1. Pipeline overview: four interchangeable front-ends, a bias-free linear projection, and a shared backbone scored with task-appropriate metrics.

diction. Because the projection and the backbone are identical across conditions, any performance difference is attributable to the front-end rather than to model capacity.

3.1. Discrete Wavelet Transform

The DWT decomposes a discrete signal $\mathbf{x} \in \mathbb{R}^L$ through iterative convolution with a two-channel filter bank consisting of a low-pass filter $\mathbf{h}$ and a high-pass filter $\mathbf{g}$, each followed by downsampling by a factor of two (Eq. (2)). Iterating on the approximation branch over $\mathcal{L}$ levels produces a multi-resolution representation that captures signal energy at progressively coarser temporal scales [Mallat 2008]. In classical families such as Daubechies, the coefficients are fixed design choices optimised for generic smoothness and perfect reconstruction [Daubechies 1992]; when applied to a specific classification task they may not capture the most discriminative frequency content, which motivates the learnable extension below. The *Db4* condition applies the exact eight-coefficient Daubechies-4 filter pair, with no modification, so the fixed baseline is a valid orthogonal wavelet; *LWT* replaces this fixed bank with the learnable QMF-structured layer of Section 3.2, whose filter length K is a searched hyperparameter.

3.2. Learnable Wavelet Transform with QMF

Rather than treating the filter bank as a fixed pre-processing step independent of the learning objective, the LWT layer casts the low-pass filter $\mathbf{h} \in \mathbb{R}^K$ of each channel as trainable parameters embedded *inside* the network and updated by the same optimiser that fits the backbone. The decomposition is therefore shaped end-to-end so that the multi-scale representation $\left[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(\mathcal{L})}, \mathbf{A}^{(\mathcal{L})}\right]$ retains the frequency content most discriminative for the task, while the QMF relationship keeps the two branches structurally band-complementary.

Filter bank and decomposition. The high-pass filter is not a free parameter; it is derived deterministically from the low-pass one through the Quadrature Mirror Filter conjugate relationship,

$$g[n] = (-1)^n \, h[K-1-n], \quad n = 0, \ldots, K-1, \tag{1}$$

which makes $H(e^{j\omega})$ and $G(e^{j\omega})$ mirror images about $\omega = \pi/2$, so the two branches are structurally band-complementary. This is all the parameterisation guarantees: the mirror structure of Eq. (1) is enforced exactly at every training step, but, when $\mathbf{h}$ is otherwise free, it does not by itself ensure orthogonality, complete alias cancellation or perfect reconstruction [Smith and Barnwell 1984, Daubechies 1992]. The LWT is therefore best

described as a *discriminative filter bank structured by QMF*: the soft penalties below keep it near the admissible region, and Section 5.5 audits which properties actually survive training. Writing $\mathbf{A}^{(0)} = \mathbf{x}$, each decomposition level convolves the current approximation with the two filters and downsamples the result by two,

$$\mathbf{A}^{(\ell)} = \big(\mathbf{A}^{(\ell-1)} * \mathbf{h}\big)_{\downarrow 2}, \qquad \mathbf{D}^{(\ell)} = \big(\mathbf{A}^{(\ell-1)} * \mathbf{g}\big)_{\downarrow 2}, \qquad \ell = 1, \ldots, \mathcal{L}, \tag{2}$$

so that level ℓ halves the temporal resolution and isolates the octave band $[\,\pi/2^{\ell},\, \pi/2^{\ell-1}]$ in $\mathbf{D}^{(\ell)}$. After $\mathcal{L}$ levels the detail bands and the final approximation are zero-padded to a common length and concatenated channel-wise into the representation $\mathbf{Z} = \big[\mathbf{D}^{(1)}, \ldots, \mathbf{D}^{(\mathcal{L})}, \mathbf{A}^{(\mathcal{L})}\big]$, which the bias-free projection maps to the backbone's input width. The layer learns one filter *per input channel and per level*: at each level, a filter-generating head outputs $K \cdot C$ coefficients for C channels, each channel is decomposed independently by its own QMF-derived pair, and the warm-start below replicates the Db4 target for every channel. The decomposition can therefore specialise to the heterogeneous spectral content of each channel, a property the audit of Section 5.5 confirms empirically.

Training objective. Because $\mathbf{g}$ is a deterministic function of $\mathbf{h}$ through Eq. (1), the layer exposes only the low-pass coefficients (K per channel), and all gradients flow back to them. To keep the learned filter within a neighbourhood of admissible orthogonal wavelets, three soft penalties are added to the downstream classification loss $\mathcal{L}_{\text{cls}}$ (class-weighted cross-entropy), giving the joint objective

$$\mathcal{L} \;=\; \mathcal{L}_{\text{cls}} \;+\; \lambda_e\big(\|\mathbf{h}\|_2 - 1\big)^2 \;+\; \lambda_g\Big(\textstyle\sum_n g[n]\Big)^2 \;+\; \lambda_s \textstyle\sum_{n=1}^{K-2}\big(h[n+1] - 2h[n] + h[n-1]\big)^2. \tag{3}$$

The energy term anchors the filter norm to unity, preventing the trivial collapse $\mathbf{h} \to \mathbf{0}$ and unbounded growth; the DC-rejection term drives the high-pass response to zero mean, a necessary condition for a wavelet detail branch; and the smoothness term discourages spurious ripple in the coefficients. Each penalty is applied to every per-channel filter, and the three weights are kept fixed across all datasets, so the decomposition adapts to the data without per-dataset retuning of the regulariser. The entire front-end – the small filter-generating network introduced next – adds between 1.6×10^3 and 8.5×10^3 parameters depending on K and C, at least an order of magnitude fewer than any backbone used here, so performance differences cannot be attributed to added capacity.

Db4 warm-start (LWT-Warm). To avoid placing the filter far from valid wavelet bases at initialisation, a one-shot warm-start brings the LWT close to the Db4 prior *before* joint training, while leaving every parameter free afterwards. The low-pass filter is produced by a small internal network: a feature extractor (two Dense+ReLU layers) maps a fixed time-index vector $\mathbf{t} \in \mathbb{R}^K$ to a latent embedding $\mathbf{z} = \phi(\mathbf{t})$, and a linear head produces $\mathbf{h} = \mathbf{z}\,\mathbf{W}_{\text{head}}$. The warm-start fits *only* the head by closed-form least squares,

$$\mathbf{W}^{\star}_{\text{head}} = \arg\min_{\mathbf{W}} \big\| \mathbf{z}\,\mathbf{W} - \mathbf{h}_{\text{Db4}} \big\|_2^2 = \mathbf{z}^{+}\,\mathbf{h}_{\text{Db4}}, \tag{4}$$

where $\mathbf{h}_{\text{Db4}}$ are the Daubechies-4 low-pass coefficients, replicated for every input channel, and $\mathbf{z}^{+}$ is the Moore–Penrose pseudo-inverse. When $K \neq 8$ the target is a length-K

Table 1. Benchmark datasets and their regimes (Ch.: channels; Len.: length; Cl.: classes).

Dataset	Domain	Ch.	Len.	Train/Val/Test	Cl.	Regime
Ford-A	engine sensor	1	500	3060/541/1320	2	univ., binary
SCP1	EEG (BCI)	6	896	227/41/293	2	multiv., binary
UWaveGesture	accelerometer	3	315	102/18/320	8	multiv., low-data

linear interpolation of the Db4 coefficients – not itself a valid orthogonal wavelet (its power-complementarity deviation is 12%, 129% and 356% for $K{=}4, 16, 32$, vs. zero for the exact Db4), but serving purely as an *initialisation*; the fixed Db4 baseline always uses the exact eight-coefficient filter. Once joint training starts, head and base net are updated freely with the backbone. The variant *without* this step is denoted *LWT*; *with* it, *LWT-Warm*.

4. Experimental Setup

4.1. Datasets

We use three public benchmarks spanning complementary regimes of supervised TSC (Table 1): *Ford-A* [Dau et al. 2019], univariate engine-sensor signals for binary fault detection with strong periodic noise; *SelfRegulationSCP1* [Bagnall et al. 2018], 6-channel EEG from the slow-cortical-potentials paradigm, a canonical non-stationary, low-SNR setting; and *UWaveGesture* [Liu et al. 2009], 3-axis accelerometer readings for 8-gesture recognition with only 102 training examples, a low-data regime in which the axes differ markedly in spectral content.

4.2. Input Conditions, Backbones and Protocol

Each dataset is run through an identical pipeline crossing **four input conditions** – *Raw* (no decomposition), *Db4* (fixed exact Db4 DWT), *LWT* (learnable filter, random initialisation) and *LWT-Warm* (learnable filter, Db4 warm-start) – with the **five backbone families** whose default structural hyperparameters are listed in the top panel of Table 2, yielding 20 models per dataset. All backbones share a bias-free linear projection before their first layer, so the channel dimensionality is identical across conditions and any performance difference is attributable to the front-end rather than to model capacity.

The configuration of the LWT front-end and the shared training settings are reported in the bottom panel of Table 2. The number of decomposition levels $\mathcal{L}$ is matched to each series length, whereas the optimiser, learning-rate schedule, filter-network width and regulariser weights are kept fixed across all datasets and conditions; all wavelet conditions apply the same per-band layer normalisation before the backbone. We adopt the official train/validation/test splits and train, for each (backbone, condition) pair, one model per point of an exhaustive hyperparameter grid under a fixed global seed. The backbone axes – dropout rate, ℓ_2 weight, layer widths and, where applicable, convolutional kernel sizes, recurrent units, attention settings and a temporal subsampling factor for the LSTM – are shared by the four conditions, which therefore receive the same tuning budget on the axes they share; the learnable conditions add the filter length K as one further axis (the CNN alone fixes its convolutional kernel sizes at the default). This yields, per dataset

Table 2. Default backbone architectures (top) and shared front-end/training configuration (bottom).

Backbone structural hyperparameters	
MLP	dense $[256, 128, 64]$, ReLU
CNN	Conv1D $[64, 128, 256]$, k= $[7, 5, 3]$, BN, dense $[128, 64]$
LSTM	units $[128, 64]$, dense $[64]$
CNN-LSTM	Conv1D $[64, 128]$, k= $[5, 3]$, LSTM $[100, 50]$, dense $[64]$
Transformer	4 heads, d_k=64, d_{ff}=128, 2 blocks, sinusoidal pos. enc.
LWT front-end and training	
Filter K (searched) / levels $\mathcal{L}$	$K \in \{4, 8, 16\}$ ($\{8, 16, 32\}$ on SCP1) / $\mathcal{L}$=2 (Ford-A), 3 (SCP1, UWave)
Units / regularisers	32 / λ_e, λ_g=10^{-2}, λ_s=10^{-3}
Optimiser / batch	Adam, lr 10^{-3} / 256 (64 for SCP1)
LR schedule	ReduceLROnPlateau ($\times 0.5$, patience 7, min 10^{-6})
Epochs / patience	100 / 15 (early stopping on validation loss)
Search axes (backbones)	dropout 0.15–0.4; ℓ_2 10^{-4}–10^{-2}; widths, kernels, heads per family
Model selection	best validation loss; fixed global seed

and per fixed condition (Raw, Db4), 18 runs for the MLP, 36 for the CNN and the CNN-LSTM, 48 for the Transformer and 96 for the LSTM; each learnable condition trains 54 (MLP), 54 (CNN), 108 (CNN-LSTM), 144 (Transformer) and 288 (LSTM) runs, for a total of 1,764 runs per dataset and 5,292 overall. Tables report the mean and standard deviation of test performance across the grid, so the spread measures robustness to the hyperparameter choice; the best configuration per pair, used for the statistical tests of Section 5.4, is selected on validation loss.

Metrics are task-appropriate: Accuracy, macro-F_1 and AUC-ROC for the binary tasks (Ford-A, SCP1), and Accuracy and macro-F_1 for the 8-class UWaveGesture. Pairwise differences between conditions are assessed with the paired Wilcoxon signed-rank test [Wilcoxon 1945] over the five backbones, with Holm–Bonferroni correction [Holm 1979] for the six simultaneous comparisons; with n=5 paired backbones the smallest attainable exact p-value is 0.0625.

5. Results and Discussion

Table 3 reports test accuracy (mean $\pm$ standard deviation across the hyperparameter grid of Section 4.2) for every (backbone, input condition) pair on the three benchmarks, revealing three regimes that recur with notable consistency across datasets.

5.1. Three Regimes by Backbone

The first regime is one of *representational saturation* in globally expressive backbones: the MLP and the Transformer are largely insensitive to the input representation (on Ford-A the four conditions span only 0.842–0.856 for the MLP, whose differences fall within one standard deviation on every dataset) – with a global receptive field, these architectures appear to reconstruct an adequate multi-scale representation internally. Ford-A's Transformer is the instructive exception: on this strongly periodic signal accuracy rises monotonically from Raw to Db4 to LWT to LWT-Warm ($0.747 \rightarrow 0.889 \rightarrow 0.894 \rightarrow 0.901$), and LWT-Warm delivers the global best of the dataset.

Table 3. Test accuracy per backbone and input condition (mean ± std across the grid). Best condition per backbone in bold; overall best per dataset in *bold italic*. Uniform chance is 0.5 on Ford-A and SCP1 and 0.125 on UWaveGesture.

Dataset	Backbone	Raw	Db4 (fixed)	LWT	LWT-Warm
Ford-A (univ., binary)	MLP	0.842 ± 0.017	0.846 ± 0.016	0.852 ± 0.066	$\mathbf{0.856 \pm 0.079}$
	CNN	0.639 ± 0.150	$\mathbf{0.742 \pm 0.181}$	0.713 ± 0.178	0.717 ± 0.171
	LSTM	$\mathbf{0.599 \pm 0.119}$	0.516 ± 0.000	0.535 ± 0.052	0.538 ± 0.060
	CNN-LSTM	0.706 ± 0.191	0.530 ± 0.051	$\mathbf{0.750 \pm 0.155}$	0.728 ± 0.164
	Transformer	0.747 ± 0.138	0.889 ± 0.097	0.894 ± 0.093	$\boldsymbol{0.901 \pm 0.081}$
SCP1 (EEG, binary)	MLP	$\mathbf{0.839 \pm 0.073}$	0.816 ± 0.038	0.817 ± 0.060	0.813 ± 0.073
	CNN	$\mathbf{0.775 \pm 0.069}$	0.598 ± 0.135	0.670 ± 0.157	0.638 ± 0.140
	LSTM	0.773 ± 0.067	0.721 ± 0.156	0.823 ± 0.052	$\boldsymbol{0.830 \pm 0.036}$
	CNN-LSTM	0.732 ± 0.055	0.509 ± 0.017	$\mathbf{0.811 \pm 0.056}$	0.807 ± 0.057
	Transformer	0.812 ± 0.033	$\mathbf{0.816 \pm 0.031}$	0.800 ± 0.050	0.807 ± 0.044
UWaveGest. (8 cl., low-data)	MLP	0.822 ± 0.020	0.811 ± 0.046	$\boldsymbol{0.823 \pm 0.025}$	0.822 ± 0.029
	CNN	0.198 ± 0.066	0.309 ± 0.129	$\mathbf{0.516 \pm 0.206}$	0.510 ± 0.194
	LSTM	0.671 ± 0.107	0.384 ± 0.171	$\mathbf{0.740 \pm 0.069}$	0.737 ± 0.068
	CNN-LSTM	0.446 ± 0.104	0.138 ± 0.030	$\mathbf{0.702 \pm 0.094}$	0.696 ± 0.103
	Transformer	0.743 ± 0.097	0.660 ± 0.165	$\mathbf{0.760 \pm 0.072}$	0.753 ± 0.121

A second, sharply different regime appears in the convolutional and recurrent backbones, which are highly sensitive to the front-end: a *fixed* Db4 transform repeatedly collapses training to a degenerate, majority-class solution – LSTM+Db4 attains 0.516 ± 0.000 on Ford-A (chance on a balanced binary task), CNN-LSTM+Db4 attains 0.509 ± 0.017 on SCP1 and 0.138 ± 0.030 on UWaveGesture, barely above uniform chance for eight classes. Making the filter learnable removes the collapse in every such cell, with CNN-LSTM recovering to 0.811 on SCP1 and 0.702 on UWaveGesture. The CNN-LSTM cell on Ford-A is telling: the randomly-initialised LWT (0.750) edges out LWT-Warm (0.728), because when Db4 itself collapses the architecture a warm-start from it can land in the same degraded basin while random initialisation escapes – why both variants are retained.

The third regime, where the LWT advantage is largest, emerges as channel dimensionality and data scarcity increase, peaking on UWaveGesture (102 training examples, eight classes, three heterogeneous channels): on the CNN, Raw collapses to 0.198, fixed Db4 reaches 0.309 and LWT jumps to 0.516, while on CNN-LSTM the gain of LWT over Db4 reaches $+56.4$ pp ($0.138 \rightarrow 0.702$). Here the learnable decomposition acts as a strong inductive prior, shifting representational load from the classifier to the front-end – exactly the help a low-data setting needs.

5.2. Aggregate View Across Metrics

Table 4 averages Accuracy, macro-F_1 and AUC over the five backbones, confirming that the conclusions are not metric-specific: the learnable conditions improve on fixed Db4 on every dataset and metric, by margins widening towards the low-data UWaveGesture ($+0.246$ accuracy, $+0.284$ macro-F_1). Macro-F_1 tracks accuracy closely, so the gains are not an artefact of class imbalance, and on both binary datasets the highest AUC is obtained by a learnable variant; Raw's fractionally highest mean accuracy on SCP1 merely reflects

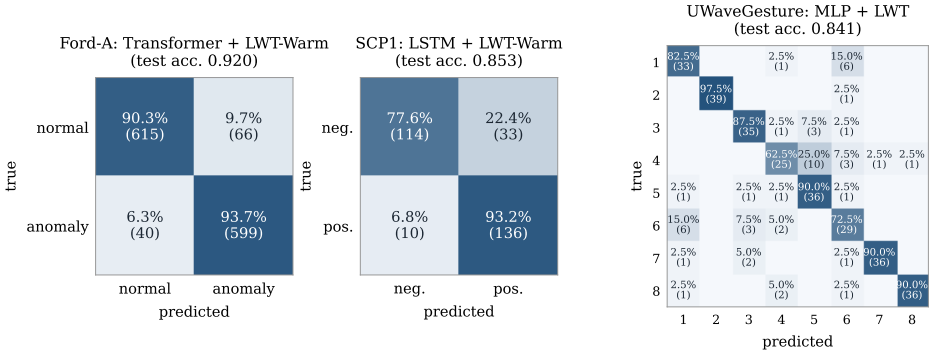


Figure 2. Test confusion matrices of the best experiment per dataset, retrained under its selected configuration (cells: row percentage over count).

Table 4. Across-backbone mean of Accuracy (Acc), macro-F_1 (F1) and AUC per condition. Best per dataset and metric in bold; AUC-OvR not reported for UWaveGesture.

	Ford-A			SCP1			UWaveGesture	
Condition	Acc	F1	AUC	Acc	F1	AUC	Acc	F1
Raw	0.706	0.645	0.807	**0.786**	**0.778**	0.893	0.576	0.544
Db4	0.705	0.623	0.795	0.692	0.643	0.817	0.460	0.407
LWT	**0.749**	**0.692**	0.854	0.784	0.765	**0.905**	**0.706**	**0.691**
LWT-Warm	0.748	0.691	**0.860**	0.779	0.759	0.900	0.705	0.690

the saturated MLP and Transformer dominating the average. Figure 2 shows the test confusion matrices of each dataset's best experiment: errors are roughly class-balanced on Ford-A, concentrated in the negativity class on SCP1, and confined to a few similar-gesture pairs on UWaveGesture ($1\leftrightarrow6$, $4\rightarrow5$) – the learnable front-end's gains do not come from biasing predictions towards a dominant class.

5.3. Anatomy of the Fixed-Db4 Collapse

The collapse is not random under-performance but a reproducible *degenerate solution*: the near-zero standard deviation across the entire hyperparameter grid (e.g. 0.516 ± 0.000 over all 96 LSTM+Db4 configurations on Ford-A) is the signature of a model converging to the majority class regardless of the hyperparameters chosen. Two factors compound: a fixed Db4 filter imposes a spectral profile misaligned with the discriminative band – and since its coefficients cannot move, the easiest descent direction for the classifier is the constant predictor – while dyadic downsampling shrinks a 500-point window to ≈ 125 samples in the coarsest sub-band, removing the temporal granularity a unidirectional LSTM needs to localise the periodic oscillation. That both LWT variants escape this basin admits a geometric reading: Db4 is a point of the para-unitary manifold that is uninformative for the task, and learning the low-pass coefficients lets the layer move to a signal-adapted filter – an option a fixed transform structurally lacks.

5.4. Statistical Significance

Paired Wilcoxon signed-rank tests over the five backbones – each represented by its best-configuration accuracy selected on validation loss, hence differing from the across-grid

means of Table 3 – with Holm–Bonferroni correction do not reach the conventional threshold on any dataset, as expected: with $n{=}5$ the smallest attainable exact p-value is 0.0625, and six simultaneous comparisons are corrected. The *direction*, however, is consistent: the median across-backbone improvement of LWT-Warm over fixed Db4 is $+4.2$ pp on Ford-A, $+5.1$ pp on SCP1 and $+12.5$ pp on UWaveGesture; on SCP1, four of the six comparisons reach the exact minimum $p_{\min} = 0.0625$, the maximal agreement attainable with five paired samples. The gains are also systematic rather than lucky best-of-N outcomes: on SCP1 the LSTM moving from Raw to LWT-Warm both raises the mean ($0.773 \rightarrow 0.830$) and tightens the spread (σ: $0.067 \rightarrow 0.036$), and on UWaveGesture the recovered cells shift the whole distribution upward at a spread comparable to Raw – what matters when a single trained model, not an oracle-selected run, is deployed.

5.5. Audit of the Learned Filters

To verify which filter-bank properties survive training, we audited the best LWT-Warm checkpoints with three metrics: the *QMF residual* $\max_n |g[n] - (-1)^n h[K{-}1{-}n]|$; the *high-pass DC sum* $\sum_n g[n]$, zero for an admissible detail branch; and the *perfect-reconstruction (PR) deviation* $\max_\omega |P(\omega){-}\bar{P}|/\bar{P}$ of the power-complementarity function $P(\omega) = |H(e^{j\omega})|^2 + |H(e^{j(\omega+\pi)})|^2$, which vanishes for a para-unitary bank ($P \equiv 2$). The QMF residual is *exactly* zero at every level and channel, confirming that the parameterisation of Eq. (1) is preserved throughout optimisation, and the low-pass ℓ_2 norm converges to 0.97–1.00. Conversely, perfect reconstruction and admissibility are *not* maintained: on UWaveGesture the PR deviation ranges from 79% to 169% and the DC sum departs markedly from zero, with strong per-channel heterogeneity (from -0.902 on the x-axis to $+1.651$ on the y-axis). This is the empirical signature of the design: the LWT operates inside the QMF-structural space but outside the strict para-unitary manifold, trading admissibility for discriminative power, and it learns *specialised* decompositions per accelerometer axis rather than copies of the Db4 initialisation. A worked example makes this concrete: retraining the best CNN-LSTM LWT-Warm configuration ($K{=}4$) and resynthesising a test signal from its decomposition yields a relative ℓ_2 error of 90%, against machine precision for the exact Db4 pair – the learned front-end is a discriminative transform, no longer an invertible one.

5.6. Discussion and Practical Guidance

Three lessons emerge: the value of an explicit multi-scale front-end is *backbone-dependent* – largely redundant for globally expressive models but decisive for convolutional and recurrent ones, where a fixed Db4 transform is actively harmful; learning the filter is strictly safer than committing to Db4 *a priori*; and the QMF parameterisation trains stably across datasets without per-dataset tuning, the warm-start being useful but not universally optimal. The apparent instability across datasets and architectures is thus not noise but the phenomenon itself: the effect of the representation is conditional on the backbone, which is why any comparison of input representations must span multiple families – the Transformer alone suggests Db4 is excellent, the CNN-LSTM alone that it is catastrophic. For practitioners: with a convolutional or recurrent backbone and scarce data, a learnable wavelet front-end is the higher-expected-value choice; and since the warm-start can inherit a bad basin when Db4 itself collapses the architecture, deploying both LWT variants and selecting on validation is a cheap hedge.

6. Conclusion

We investigated the Learnable Wavelet Transform (LWT), a differentiable front-end for time series classification that learns per-channel low-pass coefficients end-to-end with any backbone, deriving the high-pass branch structurally through a QMF parameterisation and initialising from a closed-form Db4 warm-start. The controlled study across three benchmarks and five backbones isolates *when* a learnable decomposition helps: globally expressive backbones are largely insensitive to the front-end, whereas on convolutional and recurrent ones a fixed Db4 transform repeatedly collapses training to the majority class while learning the filter reverses every collapse (largest gain: $+56.4$ pp over Db4 on the low-data UWaveGesture). An audit confirms that the QMF relationship holds exactly while orthogonality and perfect reconstruction are relaxed, and that the filters adapt away from Db4 per channel.

The study has four main limitations: statistical power (with $n{=}5$ paired backbones the smallest attainable exact Wilcoxon p-value is 0.0625, so the evidence is directional and consistent rather than conventionally significant); a single official train/validation/test split per dataset – on UWaveGesture only 18 validation examples select among dozens of configurations, a genuine risk of validation overfitting; variability measured across hyperparameter configurations under a fixed seed, not across random seeds; and scope – three benchmarks, with no direct empirical comparison against neighbouring learnable-wavelet front-ends such as DeSpaWN, L-WPT or lattice-based units. Natural extensions follow directly: repeated holdout or cross-validation; a direct comparison of the neighbouring front-ends under this same multi-backbone protocol; ablations of initialisation (random vs. interpolated Db4 vs. closed-form warm-start), filter length, depth and regularisers; and the learned frequency responses as a band descriptor.

References

Bagnall, A., Dau, H. A., Lines, J., Flynn, M., Large, J., Bostrom, A., Southam, P., and Keogh, E. (2018). The UEA multivariate time series classification archive, 2018. *arXiv preprint arXiv:1811.00075*.

Bruna, J. and Mallat, S. (2013). Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886.

Dau, H. A., Bagnall, A., Kamgar, K., Yeh, C.-C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., and Keogh, E. (2019). The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia, PA.

Dempster, A., Petitjean, F., and Webb, G. I. (2020). ROCKET: Exceptionally fast and accurate time series classification using random convolutional kernels. *Data Mining and Knowledge Discovery*, 34(5):1454–1495.

Frusque, G. and Fink, O. (2024). Robust time series denoising with learnable wavelet packet transform. *Advanced Engineering Informatics*, 62:102669.

Gupta, M. and Tripathi, A. M. (2025). DB2-TransF: All you need is learnable daubechies wavelets for time series forecasting. *arXiv preprint arXiv:2512.10051*.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2019). Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963.

Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D. F., Weber, J., Webb, G. I., Idoumghar, L., Muller, P.-A., and Petitjean, F. (2020). InceptionTime: Finding AlexNet for time series classification. *Data Mining and Knowledge Discovery*, 34(6):1936–1962.

Jia, R., Liu, Z., Zhu, P., Cheng, D., and Liang, Y. (2026). Beyond fixed patches: Enhancing GPTs for financial prediction with adaptive segmentation and learnable wavelets. *arXiv preprint arXiv:2505.02880*.

Le, A. D., Jin, S., Bae, Y.-S., and Nguyen, T. Q. (2024). A lattice-structure-based trainable orthogonal wavelet unit for image classification. *IEEE Access*, 12:88715–88727.

Li, S., Cheng, D., Zhang, A., and Liu, L. (2026). A learnable wavelet transformer for long-short equity trading and risk-adjusted return optimization. *arXiv preprint arXiv:2601.13435*.

Liu, J., Zhong, L., Wickramasuriya, J., and Vasudevan, V. (2009). uWave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6):657–675.

Liu, P., Zhang, H., Lian, W., and Zuo, W. (2019). Multi-level wavelet convolutional neural networks. *IEEE Access*, 7:74973–74985.

Mallat, S. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, Burlington, MA, 3 edition.

Michau, G., Frusque, G., and Fink, O. (2022). Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proceedings of the National Academy of Sciences*, 119(8):e2106598119.

Middlehurst, M., Large, J., Flynn, M., Lines, J., Bostrom, A., and Bagnall, A. (2021). HIVE-COTE 2.0: A new meta ensemble for time series classification. *Machine Learning*, 110(11):3211–3243.

Ravanelli, M. and Bengio, Y. (2018). Speaker recognition from raw waveform with SincNet. In *IEEE Spoken Language Technology Workshop (SLT)*, pages 1021–1028.

Recoskie, D. and Mann, R. (2018). Learning sparse wavelet representations. *arXiv preprint arXiv:1802.02961*.

Smith, M. J. T. and Barnwell, T. P. (1984). A procedure for designing exact reconstruction filter banks for tree-structured subband coders. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 9:421–424.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.