PAMUKKALE ÜNİVERSİTESİ
Mühendislik Bilimleri Dergisi

PAMUKKALE UNIVERSITY
Journal of Engineering Sciences

Cilt | Vol: XX • Sayı | No: X • ss | pp: XXXX-XXXX • Yıl | Year: 20XX

Wavelet-Based multi-frequency image transformation for time series classification

Mehmet Kurnaz, Celal Alagöz

ARTICLE INFO

Dates:
Received: 28.01.2026
Revision: 03.03.2026
Accepted: 17.03.2026

doi:
10.65206/pajes.1873789

Corresponding author:
Mehmet Kurnaz
(mkurnaz@sivas.edu.tr)

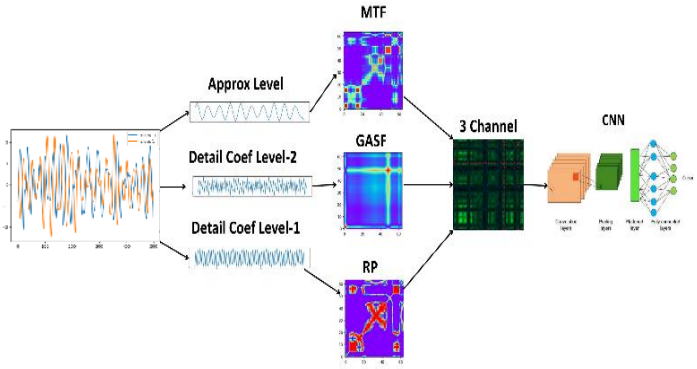
Author addresses:
Department of Computer Engineering, Sivas Science and Technology University, Sivas, 58000, Turkey
(mkurnaz@sivas.edu.tr; alagozc@sivas.edu.tr)

ABSTRACT

Context—Time series classification (TSC) plays a critical role in many application domains such as healthcare, fault diagnosis, and signal analysis, where signals are often non-stationary and noisy. Although convolutional neural networks (CNNs) have achieved strong performance, their effectiveness is limited when operating directly on raw signals, particularly in capturing multi-frequency characteristics. Image-based representations such as Gramian Angular Field (GAF), Markov Transition Field (MTF), and Recurrence Plot (RP) have enabled CNNs to exploit temporal structures more effectively. However, most existing approaches generate these images directly from raw signals, leaving the potential of frequency-aware representations underexplored.

Objective—The objective of this study is to develop a hybrid time series classification framework tailored for sensor-based signals with rich frequency content. In particular, the study investigates whether generating GAF, MTF, and RP images from different Discrete Wavelet Transform (DWT) components can enhance the representation quality of frequency-dependent time series. By assigning distinct image encodings to approximation and detail coefficients, the proposed approach aims to improve dataset representations by capturing complementary information across multiple frequency bands commonly observed in sensor data. The effectiveness of the framework is evaluated on benchmark datasets from the UCR archive.

Method—In the proposed method, each time series is decomposed using a two-level Discrete Wavelet Transform with the db4 wavelet to obtain approximation and detail coefficients. These coefficients are then independently transformed into image representations: MTF from approximation coefficients, GAF from medium-frequency components, and RP from high-frequency components. The resulting images are combined into multi-channel inputs and classified using lightweight CNN architecture. Additional experiments using direct image transformation and multi-branch and multi-channel CNN designs are conducted to assess the contribution of wavelet-based representations.



Results—Experimental results on multiple UCR benchmark datasets demonstrate that the proposed DWT-based image representations improve classification accuracy, particularly for long and noisy time series. The wavelet-based approach shows clear advantages in datasets characterized by complex temporal dynamics and low signal-to-noise ratios. For short and relatively clean signals, CNNs trained on raw image representations achieve comparable or slightly better performance. These findings indicate that frequency-aware image encoding enhances robustness without increasing model complexity.

Conclusion—This study shows that integrating Discrete Wavelet Transform with image-based time series representations provides a more balanced and informative feature space for CNN-based classification. By mapping different DWT components to GAF, MTF, and RP, the proposed framework captures temporal patterns across multiple frequency scales. The results highlight the importance of frequency-specific encoding, especially for challenging real-world signals. Future work may explore deeper architectures and adaptive wavelet-image assignments to further improve classification performance.

Key Words—Time series classification; CNN; Discrete wavelet transform; Image transformation; Gramian angular field

I. INTRODUCTION

The TSC has become an essential task in various real-world applications. Deep learning models, particularly CNNs, have shown remarkable success in extracting spatial and temporal features from time series data [1]–[4]. However, CNNs often struggle to capture information at multiple frequency levels, which limits their ability to represent non-stationary signals effectively. The DWT provides a powerful mechanism to analyze time series data simultaneously in both time and frequency domains [5],[6]. By decomposing signals into approximation and detail coefficients, DWT enables more robust feature extraction and noise reduction, improving the representation of complex temporal dynamics [7],[8].

Image-based representations such as the GAF [9], MTF [9], and RP [10] have enabled the adaptation of powerful computer vision architectures to time series data. Wang and Oates [10] were among the first to encode time-series data into GAF and MTF images and classify them using a tiled CNN.

Another study [11] introduced RP as an image-based representation for time series. Their method visualizes the recurrence of system states over time, effectively creating texture-like matrices from one-dimensional signals. By applying a simple CNN with two convolutional layers and one fully connected layer, they achieved strong results on several UCR datasets, showing that RP-based images can enhance classification performance especially in small-scale settings. The most similar study [7] to this work extended this line of work by combining multiple image representations—such as grayscale, RP, MTF, and Gramian Angular Difference Field (GADF)—and integrating DWT to enhance texture clarity and suppress noise. The DWT was used as a preprocessing step before image encoding, which improved feature quality. Their experiments on six UCR datasets showed that wavelet-enhanced image features substantially improve clustering and classification outcomes. Compared to this study [7], our study focuses on using the three-level DWT decomposition to separate low- and high-frequency components before applying image transformations. Instead of using the GASF, we adopt the Gramian Angular Summation Field (GASF) for approximated (low-frequency) components, while the detail coefficients (CD1, CD2) are utilized for MTF and RP generation. In general, image encodings allow CNNs to exploit spatial patterns and textures in the data. For instance, CNNs applied to RP images have improved human activity and EEG classification [12]. These methods leverage the fact that neighboring pixels in the image can represent time-domain relationships (for GAF/MTF) or phase-space recurrences (for RPs), enabling CNN to learn meaningful features [13].

One of the specialized studies [14] focused on multivariate physiological signals, converting them into GAF and MTF representations and training a CNN with attention pooling to classify cardiopulmonary exercise test data. Their approach demonstrated that image-based encodings can generalize well to complex biomedical time series. Mariani et al. [15] fused RP and Gramian Angular Fields (both GASF and GADF) to build three-channel image representations of time-series data. They employed Bayesian optimization to tune image resolution and CNN hyperparameters automatically. Their GAF-RP-CNN-BO framework achieved high accuracy on both univariate and multivariate datasets from UCR and UCI repositories. Other recent works also contribute to this domain. Wu et al. [16] employ RP, GASF, GADF, and MTF together for clustering financial time series, showing that 2D image structures can reveal inter-period correlations that are hidden in raw sequence data. Another financial time series work [17] is to use ensembles of CNNs on GAF-based images for financial time series forecasting, leveraging multi-resolution inputs. Jin et al. [18] propose a multi-channel fusion method combining GAF, RP, and MTF representations for classification, reinforcing the idea that ensemble or fused image encodings yield stronger models. Lee and Lee [19] in the domain of indoor localization convert RSSI time series into MTF (and RP) images and apply CNNs to recognize location classes. In contrast to these fusion-based approaches, ViSemble [20] explicitly treats each image transformation (such as GAF, MTF, and RP) as an independent visual view, trains separate classifiers for each representation, and combines their predictions through an ensemble strategy, enabling a systematic comparison of visual encodings for time series classification.

II. METHOD

In this study, we developed a deep learning framework for classifying time series data. Our approach combines wavelet decomposition with multiple image-based transformations and evaluates performance using repeated resampling. Our approach combines wavelet decomposition with multiple image-based transformations and evaluates performance using repeated resampling. The methodology is summarized in Fig. 1. The components are DWT, image-based transformation, and CNN architecture. Detailed information is provided in the following.

A. Transformations

1. Discrete wavelet transform (DWT)

DWT decomposes a discrete-time signal using a two-channel filter bank composed of a low-pass analysis filter $h[t]$ and a high-pass analysis filter $g[t]$, followed by dyadic downsampling

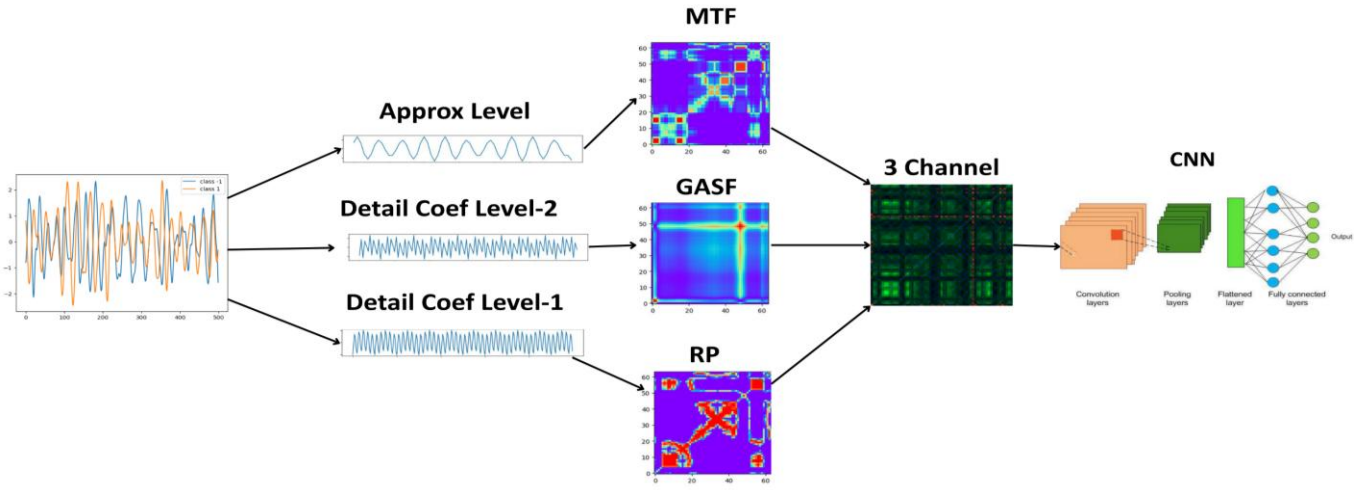


Figure 1. Illustrates the overall workflow of the proposed method, showing how raw time series are first decomposed using wavelet transforms, then converted into multiple 2D image representations, and finally classified using a CNN-based model.

(decimation by a factor of two). At each decomposition level, the signal is convolved with these filters and subsequently down sampled, separating low- and high-frequency components while reducing temporal resolution. The approximation part captures the low-frequency trend, while the detail part represents high-frequency variations.

For a discrete signal $x[t]$ the approximation coefficients at level j are defined as:

$$A_j[k] = \sum_n x[t]h[2k - t], \quad (1)$$

and the detail coefficients are

$$D_j[k] = \sum_n x[t]g[2k - t], \quad (2)$$

where the term $2k$ explicitly represents dyadic down sampling (decimation by 2).

The low-pass branch produces approximation coefficients capturing coarse-scale (low-frequency) components of the signal, whereas the high-pass branch generates detail coefficients representing fine-scale (high-frequency) variations.

At decomposition level J, the original signal can be reconstructed using the multiresolution representation defined by

$$x(t) = A_j(t) + \sum_{j=1}^{J} D_i(t), \quad (3)$$

where $A_j(t)$ denotes the approximation (low-frequency) component at the coarsest scale and $D_i(t)$ represents the detail component at scale j.

This process enables efficient feature extraction by separating trend information from short-term fluctuations, which can significantly improve model robustness and noise resistance. In this study, a two-level DWT decomposition ($J = 2$) is employed using the Daubechies-4 (db4) wavelet, implemented via the Mallat multiresolution algorithm. This results in one approximation component (cA) and two detail components (cD_1) and (cD_2), corresponding to low-, mid-, and high-frequency information, respectively. The choice of the db4 wavelet is motivated by its favorable time–frequency localization properties and its widespread use in time-series analysis. A two-level decomposition provides a balanced trade-off between frequency resolution and signal length preservation, allowing effective separation of global trends and localized fluctuations without excessive signal fragmentation. This multiresolution decomposition facilitates robust feature extraction and improves noise resistance in subsequent image-based encoding and classification stages.

2. Gramian angular summation field (GASF)

The GAF converts a normalized time series into a 2D image by using polar coordinates. Each element in the image represents the angular relationship between two time points.

$$X = \{Ri, j = 1\}, \quad (4)$$

where N denotes the length of the time series and X_i represents the signal value at time index i. Prior to transformation, the time series is normalized to the interval $[-1, 1]$ to ensure numerical stability and to satisfy the requirements of the polar encoding.

$$\phi_i = \arccos(r_i),\ r_i = \frac{t_i}{N} \quad (5)$$

where ϕ_i represents the angular value corresponding to the normalized time-series sample X_i. The GASF is then constructed as:

$$G_{i,j} = cos(\phi_i + \phi_j);\ i, j = 1, 2, \dots, N \quad (6)$$

The resulting $N \times N$ matrix encodes the global temporal correlations between all pairs of time points in the series. This representation preserves temporal dependency information in a structured visual form, making it well suited for convolutional neural network (CNN)-based feature learning and classification.

3. Recurrence plot (RP)

The Recurrence Plot (RP) is a visual tool that reveals repeating patterns within a time series by representing the pairwise similarity between time points. The RP is defined as

$$R_{i,j} = \Theta\left(\varepsilon - \left|\left|x_i - x_j\right|\right|\right), \quad (7)$$

where $\Theta(.)$ is the Heaviside step function and ε is a predefined threshold. A value of $R_{i,j} = 1$ that the system states x_i and x_j are sufficiently close in phase space, signifying recurrence. The resulting binary image encodes the dynamical behavior of the signal, making it suitable for texture-based feature extraction through convolutional models.

4. Markov transition field (MTF)

The Markov Transition Field (MTF) represents the transition dynamics of a time series as a two-dimensional matrix by encoding the probability of state transitions over time. First, the normalized series x is discretized into Q quantile bins to compute a first-order Markov transition matrix $P \in \mathrm{R}^{QxQ}$, where

$$P_{u,v} = P_r(x_{t+1} \in bin\ \ v | x_t \in bin\ \ u). \quad (8)$$

Then, the MTF is constructed as

$$M_{i,j} = P_{q_i,q_j}, \quad (9)$$

where q_i and q_j denote the quantile bins corresponding to time steps i and j. The resulting image representation encodes transition probabilities across the entire time span, preserving temporal order information in a format suitable for CNN-based classification.

B. Image-based representation of time series

Before feeding the data into the neural network, each one-dimensional signal is transformed into a three-channel image. This transformation is performed in two different approaches.

1. Wavelet-based GAF-MTF-RP approach

Each signal is first decomposed using the DWT with db4 wavelets up to level 2. The resulting coefficients correspond to different frequency bands. The approximation coefficients (CA) capture the overall trend and low-frequency variations of the signal, providing a compact representation of its long-term behavior. These coefficients represent the overall trend of the time series and are particularly informative for understanding gradual transitions and long-term dependencies. The MTF transformation is applied to the CA coefficients because it encodes the probabilistic transition between signal states as a structured matrix. This mapping allows the visualization of gradual temporal changes and provides a clear view of how the system evolves over time. The GASF was generated from the CD2 coefficients, corresponding to the medium-frequency band. These coefficients balance both global and local temporal information. Finally, the RP was derived from CD1 coefficients, which represent high-frequency information. RP visualizes the recurrence of states in the reconstructed phase space, allowing the detection of local variations and rapid fluctuations that are often masked in lower-frequency domains. This mapping enhances the representation of short-term dynamics critical for discriminating against similar temporal patterns. By combining MTF, GAF, and RP derived from different frequency bands, the proposed approach tries to achieve a comprehensive representation that encapsulates both long-term trends and high-frequency variations, thus improving the discriminative power of CNN-based models for time series classification tasks.

As shown in Fig. 2, the ECG5000 time series was decomposed into three levels using the DWT. This hierarchical representation

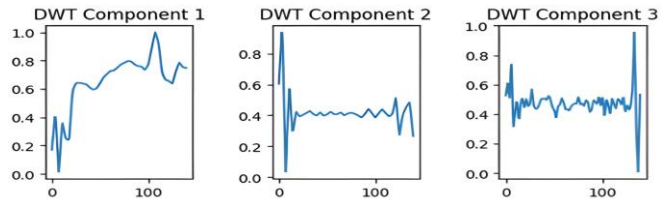


Figure 2. Three-level DWT decomposition of an ECG5000 time series using db4 wavelet. The first component represents the final approximation coefficients (A3), capturing the low-frequency trend of the signal. The remaining components correspond to detail coefficients (D3 and D2), representing high-frequency variations at different scales.

allows the extraction of frequency-specific information, where CD1 captures sharp variations, CD2 encodes medium-scale transitions, and CA3 retains the global signal trend. These coefficients were then mapped into MTF, GAF, and RP images to form a three-channel representation.

As shown in Fig. 3, after the DWT decomposition, the coefficients were transformed into image representations—GAF, MTF, and RP. Each of these methods encodes different aspects of the temporal structure: GAF focuses on correlations and global trends, MTF models transition probabilities, and RP emphasizes recurrence behaviors. Combining these images enables a comprehensive understanding of both short-term dynamics and long-term dependencies within time series. Finally, these three representations are stacked to form a three-channel image, each of which conveys complementary information from different frequency bands. This fusion enhances the discriminative power of the representation for classification tasks.

2. Direct GAF-MTF-RP transformation

To evaluate the effect of wavelet decomposition, we also preformed experiments without DWT. In this version the original signals were directly converted into GAF, RP, and MTF images, again combined into a three-channel structure.

Unlike the original study by Costa et al. [10], which employs GoogLeNet-based architecture, we adopted the CNN architecture proposed in the same study to ensure a fair comparison with our framework. All experiments were therefore performed using the same network structure, and differences in performance can be attributed to the input representations rather than the model design.

3. Multi-branch design approach

In the multibranch configuration, two independent CNN branches are constructed, each receiving a different type of time-series image representation as input. The first branch processes images generated directly from the raw time-series signals, namely the GASF, MTF, and RP representations. The second branch processes images from transformed signals obtained via DWT. Specifically, the raw signals are decomposed into approximation and detail components, and the resulting wavelet-based signals are subsequently converted into GASF, MTF, and RP images. Each branch independently extracts high-level features through its own convolutional and pooling layers. The learned feature vectors from both branches are then concatenated at the fully connected layer level, enabling the model to jointly exploit complementary information captured from raw-domain and wavelet-domain representations. This design allows the network to learn specialized filters for each representation while preserving their structural differences before fusion. Figure 4 shows the overall multibranch CNN architecture and the feature-level fusion strategy used in this study.

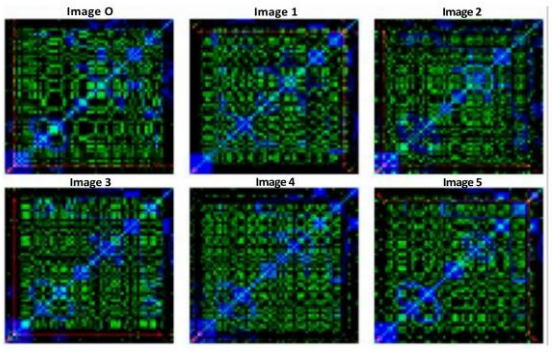


Figure 3. Example of MTF, GAF and RP representations generated from DWT coefficients of an ECG5000 sample.

1. Multi-channel CNN approach

In contrast, the multichannel configuration adopts a single CNN architecture where multiple representations are stacked along the channel dimension of the input. In this setup, six image channels are jointly provided to the network: raw GASF, raw MTF, raw RP, wavelet-based GASF, wavelet-based MTF, and wavelet-based RP representations. These six channels form a unified multi-channel input tensor that is processed by a single CNN. Unlike the multibranch approach, all representations in the multichannel model share the same convolutional filters from the earliest layers. This design encourages the network to learn cross-representation correlations at low and intermediate feature levels, treating different time-series encodings as complementary channels of a single input rather than as independent feature extractors. Figure 5 illustrates the proposed multichannel CNN architecture and the stacking of raw and wavelet-based image representations.

C. CNN architecture

For classification, we employ a lightweight convolutional neural network composed of two convolutional blocks followed by fully connected layers. The network consists of 3×3 convolutional kernels with gradually increasing feature maps (16 and 32 filters), interleaved with max-pooling layers for spatial down sampling. This compact architecture is intentionally chosen to avoid excessive model complexity and overfitting, particularly given the relatively limited size of benchmark time-series datasets. Since the proposed DWT-based multiresolution image encoding already captures rich temporal and frequency-specific information, a shallow CNN is sufficient to learn discriminative spatial patterns for effective classification. This design ensures that performance gains can be attributed to the representation learning strategy rather than to an overly complex classifier. Finally, the proposed model's performance is evaluated using standard benchmark datasets such as UCR and compared with existing state-of-the-art approaches.

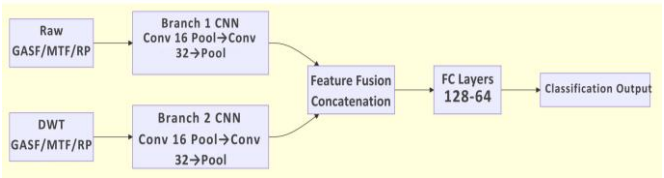


Figure 4. Example of multibranch CNN architecture.

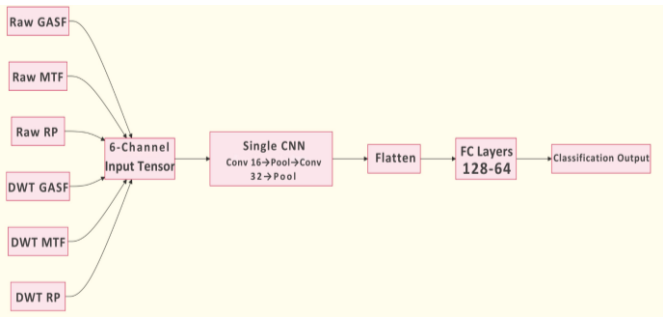


Figure 5. Example of multibranch CNN architecture.

D. Datasets

The following datasets from the UCR Time Series Classification Archive [21] are used in this study, representing a comprehensive suite of benchmarks across diverse domains, signal properties, and classification challenges.

Car: This sensor-based automotive diagnostics dataset contains 1D time series (e.g., from accelerometers) capturing a car's operational state. Its two classes differentiate between normal and faulty conditions, presenting moderate-length series with discriminative vibration or state-related temporal patterns.

CinCECGTorso (CinC-ECG-Torso): A biomedical dataset of ECG recordings from the torso surface, designed to localize the origin of ventricular activation. Its four classes correspond to anatomical locations. The signals are long, noisy, and contain complex morphological patterns, posing a significant challenge in separating physiological noise from class-discriminative features.

ECG5000: A canonical biomedical benchmark with 5,000 synthetic ECG heartbeats across five classes (one normal, four arrhythmias). The series are clean, of fixed length, and discrimination hinges on the shape, amplitude, and timing of characteristic P, QRS, and T waves.

Earthquakes: A seismology dataset for distinguishing between seismic recordings that precede major earthquakes (positive class) and those that do not. The time series are long and exhibit low signal-to-noise ratios, requiring detection of subtle foreshock patterns within background noise.

FordA & FordB: Paired automotive sensor datasets for engine fault diagnosis via audio or vibration sensors. Both involve binary classification of normal vs. faulty engine operation. FordA was recorded in a standard acoustic environment, while FordB presents a more challenging, noisy setting. Both feature long series where discriminative features are often spectral.

Mallat: A synthetic dataset fundamental to signal processing and wavelet analysis. It contains eight classes of complex, computer-generated waveforms. Discrimination relies on identifying distinct spectral and multi-scale properties, making it an ideal test for methods capturing frequency-domain features.

Wafer: A semiconductor manufacturing dataset comprising sensor traces from silicon wafer etching chambers. The binary task identifies "normal" vs. "abnormal" processes leading to defective chips. The long series often contain critical, subtle local deviations (e.g. a minor pressure drop) within an otherwise stable trajectory.

TwoPatterns: A controlled synthetic benchmark where each of the four classes is defined by the presence and order of two fundamental base patterns (e.g. spikes, steps). It tests a model's ability to recognize and combine localized shape features irrespective of their position.

ShapeletSim: A synthetic dataset created explicitly to validate shapelet-based classification methods. One class contains a defining, short discriminative shapelet embedded randomly within noise, while the other class lacks it. It isolates the challenge of discovering phase-invariant, local subsequences.

MixedShapeRegularTrain: A composite synthetic benchmark formed by merging instances from several simpler shape datasets into a five-class problem. It provides a more comprehensive test for general shape-discrimination algorithms against a variety of fundamental temporal shapes (e.g. cylinder, bell) with noise and misalignment.

InlineSkate: A motion capture dataset of an inline skater performing seven different maneuvers. The univariate series are very long and derived from multi-variate sensors, capturing complex, periodic, and semi-periodic motion patterns characteristic of specific skating styles.

NonInvasiveFetalECGThorax (NIFECG): A highly challenging biomedical dataset of thoracic ECG recordings from pregnant women. The 42-class task involves identifying the fetal heart rate from the extracted fetal QRS complex, which is obscured by the much stronger maternal ECG and biological noise, resulting in an extremely low signal-to-noise ratio.

SmallKitchen: A smart environment dataset from a kitchen, where time series represent state-changes of contact switches on objects (e.g. drawers, doors) during meal preparation. Classes correspond to high-level activities. The signals are binary or step-like, with discriminative information residing in the precise timing, duration, and sequence of events.

SmoothSubspace: A synthetic dataset designed to test a classifier's ability to exploit global, smooth trajectories. Classes are generated from different smooth paths in a latent space, meaning successful classification requires recognizing the holistic evolution of the series rather than local shapes or alignments.

FaceAll: An image-derived dataset created by converting 2D facial contours into 1D time series (e.g. radial distance from centroid). The multi-class task is person identification. The series represents holistic shape profiles, with challenges including intra-class variation and the cyclic nature of the contour data.

This collection ensures evaluation across critical dimensions including signal length, noise level, discriminative feature type (local shape vs. global trajectory), and domain complexity, providing a robust foundation for methodological comparison.

E. Data preparation and resampling

For each dataset, we split the time series into training and testing sets using predefined indices from the UCR repository. We repeated this procedure for 28 resamples (folds) to ensure statistical reliability. Each resample provides different training

Table 1. Summary of selected UCR time series datasets.

UCR Datasets	Number of classes	Size of training set	Size of testing set	Time series length	Type
Car	4	60	60	577	Automotive (sensor)
CinCECGTorso	4	40	1380	1639	ECG (sensor)
ECG5000	5	500	4500	140	ECG (sensor)
Earthquakes	2	322	139	512	Seismic (sensor)
FordA	2	3601	1320	500	Engine (sensor)
FordB	2	3636	810	500	Engine (sensor)
Mallat	8	55	2345	1024	Simulated
TwoPatterns	4	1000	4000	128	Simulated
ShapeletSim	2	20	180	500	Simulated
MixedShapesRegularTrain	5	500	2425	1024	IMAGE
InlıneSkate	7	100	550	1882	Motion (sensor)
NonInvasiveFetalECGThorax	42	1800	1965	750	ECG
SmallKitchen	3	375	375	720	Device (sensor)
SmoothSubspace	3	150	150	15	Simulated
Wafer	2	1000	6164	152	Manufacturing (sensor)

and testing subsets, allowing robust evaluation of the model. This procedure ensures that each dataset is evaluated multiple times under different splits, improving the generalization assessment. After transformation, the three resulting matrices (GAF, MTF, RP) were normalized and resized to a fixed image size depending on the signal length:

For signals shorter than 64 points: image size = timestamps

64–299 points: 64 × 64

300–599 points: 128 × 128

600+ points: 256 × 256

F. Training and evaluation metrics

For each dataset and each resample, the proposed model was trained for 10 epochs with a batch size of 16. The Adam optimizer was used due to its fast convergence and robustness across different datasets. Depending on the number of classes, either binary cross-entropy (for binary classification tasks) or categorical cross-entropy (for multi-class tasks) was employed as the loss function.

During training, a model checkpointing strategy was applied: the model weights corresponding to the highest validation accuracy were saved and later restored before the evaluation phase. This ensured that performance metrics were computed using the best-performing model for each resample, rather than the final training epoch.

All experiments were conducted using predefined UCR resample indices, guaranteeing fair and reproducible comparisons across datasets. After training, classification performance was evaluated on the test split using accuracy as the primary metric, along with balanced accuracy, F1-score, AUC, and negative log-likelihood for further analysis. Training and inference durations were also recorded to assess computational efficiency.

G. Implementation Details

Simulations were implemented using TensorFlow (Keras API) as the deep learning framework. The convolutional neural network architecture was constructed using tensorflow.keras.layers and trained with the Adam optimizer.

Wavelet decomposition was performed using the PyWavelets (pywt) library. Image-based time series transformations—including GAF, MTF, and RP—were generated using the pyts.image module. The code used in this study is available at https://github.com/mehmet-kurnaz/Wavelet-Based-image-transformation-for-TSC.

III. RESULTS and DISCUSSION

A. Accuracy comparison

The Critical Difference (CD) diagram provides a comprehensive performance comparison of the evaluated methods across 10 representative UCR datasets, focusing on sensor-based frequency data. Among these, eleven datasets correspond to real sensor-based measurements, namely Car, CinCECGTorso, ECG5000, Earthquakes, FordA, FordB, InlineSkate, NonInvasiveFetalECGThorax, SmallKitchen, Mallat and Wafer. These datasets include biomedical signals (ECG), seismic recordings, engine sensor measurements, industrial process data, and motion-based sensor signals. The remaining four datasets (TwoPatterns, ShapeletSim, MixedShapesRegularTrain, and SmoothSubspace) are synthetic benchmarks. When restricting the statistical analysis to the eleven sensor-origin datasets, the proposed method demonstrates improved mean performance compared to its overall average across all 15 datasets, highlighting its effectiveness for frequency-rich real-world signals. It is important to note that although the proposed wavelet-based approach is relatively simple and still in its experimental stages, it successfully joins the top-performing group. While this model currently requires further refinement and various experimental operations, its ability to compete with complex ensembles suggests that incorporating wavelet transformations effectively enhances feature extraction for frequency-dependent sensor datasets. Figure 6 presents the Critical Difference (CD) diagram obtained from the Friedman test followed by the Nemenyi post-hoc analysis across 11 representative UCR datasets. The diagram ranks the evaluated classifiers according to their average performance, where lower ranks indicate better overall accuracy.

TDE achieves the best overall ranking (1.6250), followed by RISE and BOSS. The proposed Wavelet-based GAF-MTF-RP method obtains an average rank of 3.8750, clearly outperforming its non-wavelet counterpart and the CNN-DTW baseline. Importantly, the proposed approach is positioned within the statistically competitive group, indicating that its performance differences with top ensemble methods are not statistically significant at the chosen confidence level.

These findings demonstrate that incorporating discrete wavelet decomposition prior to image-based transformations substantially improves discriminative capability for frequency-sensitive sensor signals. The results suggest that multi-resolution feature enhancement contributes meaningfully to classification robustness across heterogeneous datasets. Across selected UCR datasets, the performance of DWT-based representations varied depending on the dataset's signal characteristics, as summarized in Table 1. For long and noisy series such as FordA, FordB, Wafer, and CinCECGTorso, DWT-based representations consistently outperformed the raw time series, suggesting that noise reduction and trend preservation provided by approximation coefficients enhance model stability. In contrast, for trend-dominated or relatively clean frequency-based signals such as ECG200 and ECG500, the raw time series achieved slightly better performance than the DWT-enhanced versions. Although these datasets contain frequency components, their patterns are already stable and low in noise, which means the wavelet smoothing can remove subtle but informative variations. Consequently, while DWT helps suppress unwanted fluctuations, it may also blur fine temporal structures that contribute to discriminative accuracy. This shows that even for frequency-driven signals, the raw temporal dynamics can sometimes provide richer information for CNN-based models. The TwoPatterns dataset exhibited dataset-specific behavior, indicating that pattern segmentation may play a larger role than frequency decomposition. For datasets with repeating motifs, such as ShapeletSim and MixedShapes, the DWT-based RP and MTF representations achieved higher accuracy, especially when mid-frequency detail bands were used. This supports the idea that recurrence and state-transition patterns are better captured in the middle-frequency range, aligning with Mallat's findings on wavelet-based multiscale structures.

However, datasets like Trace and Coffee showed reduced performance after DWT decomposition, possibly due to excessive smoothing of the approximation component. These results suggest that while DWT effectively suppresses noise, it can also remove meaningful small-scale variations when over-averaged.

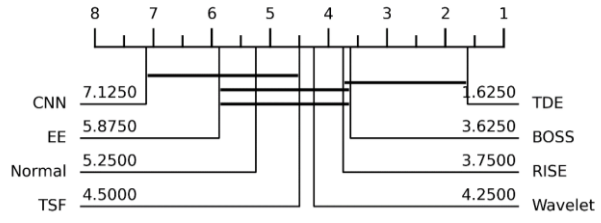


Figure 6. Critical difference (CD) diagram comparing the average ranks of the evaluated methods across the selected 10 UCR sensor datasets.

In high-frequency components MTF-based representations sometimes improved performance on datasets with abrupt transitions—particularly in biomedical and industrial datasets such as Wafer—though the improvement remained modest. InlineSkate performed poorly for both raw and transformed inputs, likely due to its irregular temporal dynamics and low inter-class separability. Similarly, Yoga and NonInvasiveFetalECGThorax favored raw signal representations, as their morphological differences are more prominent in the temporal rather than frequency domain.

Overall, the findings suggest that DWT-based image transformations enhance performance for long, noisy, or highly textured signals, but may not benefit datasets where fine temporal details are crucial. The complementary strengths of GAF, MTF, and RP representations provide a richer perspective on signal dynamics when stacked as multi-channel inputs, especially when frequency bands are chosen to match the dataset's intrinsic complexity. The analysis focuses on representative UCR datasets exhibiting the largest performance gaps between the two approaches. Among the selected datasets, the proposed DWT-based multi-channel representation (DWT + CNN) shows superior performance on 5 datasets—Mallat (+0.091), FordB (+0.077), SmallKitchen (+0.047), FordA (+0.043), and ShapeletSim (+0.034)—where frequency-domain characteristics, long temporal dependencies, or structured noise play a dominant role. In contrast, the standard CNN outperforms the DWT-based representation on 5 datasets, namely TwoPatterns (+0.414 CNN), ACSF1 (+0.356 CNN), CBF (+0.339 CNN), Trace (+0.316 CNN), and SmoothSubspace (+0.222 CNN). These datasets are typically characterized by shorter sequences, simpler structures, or highly discriminative local patterns, where explicit time–frequency decomposition offers limited benefit. In addition, 5 datasets—Earthquakes (+0.002 CNN), Car (+0.012 CNN), MixedShapes (+0.018 DWT), Wafer (+0.002 DWT), and FaceAll (±0.000)—exhibit only marginal performance differences, indicating close competition between the two representations. Figure 7 shows this dot-plot-based comparison, illustrating both the magnitude and direction of performance differences between DWT+CNN and the baseline CNN across the selected datasets.

Furthermore, as shown in Table 2, the proposed method demonstrates competitive performance on small CNN models when compared with several state-of-the-art time series classification methods, including 1-NN DTW [22], BOSS [23], TSF [24], RISE [25], CNN [26], FCN [26], ViSemble [20], TDE [27], and EE [28]. Notably, on the FordA, FordB, Wafer, and CinC ECG Torso datasets. For each dataset, the mean accuracy values reported in the literature are used to provide a representative baseline

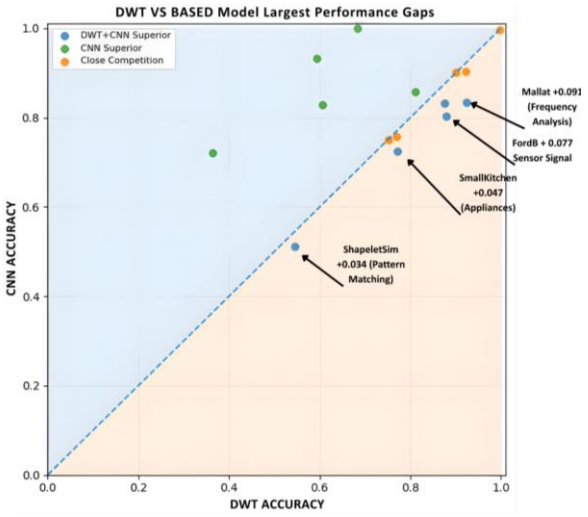


Figure 7. Dataset-level accuracy differences between DWT+CNN and the baseline CNN on selected UCR datasets.

comparison. These datasets share common characteristics: strong frequency components, high sensor noise, and non-stationary temporal dynamics. This indicates wavelet-integrated multi-channel representations can capture informative spectral–temporal patterns that purely temporal models tend to overlook. The proposed method attains a mean accuracy of 0.7476 across all datasets. However, when examining frequency-rich sensor datasets such as FordA, FordB, Wafer, and CinCECGTorso, the proposed approach demonstrates competitive performance relative to baseline CNN and classical methods.

Notably, the standard deviation analysis reveals that ensemble-based methods exhibit lower performance variance, whereas the proposed method shows higher variability due to reduced performance on synthetic datasets (e.g., TwoPatterns and SmoothSubspace). When restricting evaluation to real sensor-based datasets, the performance gap narrows considerably, highlighting the suitability of wavelet-integrated representations for non-stationary and frequency-dominant signals.

To further analyze the effect of representation fusion, we compare single image-based time series representations with fusion-based approaches constructed by combining multiple transforms. The evaluated methods include individual representations such as GADF, GASF, RP, MTF, Continuous Wavelet Transform (CWT), and First-Order Difference (FOD), as well as fusion strategies formed by merging GAF, MTF, and RP representations, and their combinations with DWT coefficients.

Table 2. Summary of selected UCR time series datasets.

UCR Datasets	Wavelet Based	Baseline-CNN	1—NN DTW	FCN	ViSemble	TDE	ER	BOSS	TSF	RISE	CNN
Car	0.7689	0.7562	0.6750	0.6667	0.7667	0.8622	0.7833	0.8411	0.7461	0.7700	0.7400
CinCECGTorso	0.9159	0.8882	0.6685	0.7648	0.9659	0.9839	0.9331	0.8614	0.9715	0.9542	0.6601
ECG5000	0.9344	0.9384	0.9251	0.9229	0.9389	0.9436	0.9371	0.9388	0.9431	0.9368	0.9304
Earthquakes	0.7520	0.7496	0.7021	0.6763	0.7482	0.7474	0.7429	0.7460	0.7472	0.7482	0.6866
FordA	0.8750	0.8321	0.5784	0.9273	0.8909	0.9255	0.7350	0.9205	0.8184	0.9416	0.8994
FordB	0.8797	0.8022	0.6461	0.7852	0.6914	0.9096	0.7728	0.9078	0.7852	0.9195	0.8686
Mallat	0.9243	0.8334	0.9371	0.8183	0.8960	0.8994	0.9591	0.9506	0.9248	0.9636	0.9422
TwoPatterns	0.4421	0.8562	1.00	0.9987	0.9815	0.9990	1.00	0.9909	0.9903	0.4382	0.9923
ShapeletSim	0.5446	0.5107	0.6562	0.7000	0.5389	0.9948	0.9379	1.000	0.5133	0.7857	0.4985
MixedShapesRegularTrain	0.9217	0.9032	0.8863	0.8219	0.9212	0.9745	0.8508	0.9273	0.9197	0.9394	0.8439
InlıneSkate	0.2893	0.3093	0.3944	0.3200	0.3145	0.5261	0.4667	0.4818	0.3657	0.3850	0.2836
NonInvasiveFetalECGThorax	0.5906	0.6675	0.8355	0.8870	0.8997	0.8344	0.8508	0.8298	0.8836	0.9116	0.8587
SmallKitchen	0.7717	0.7245	0.6542	0.7093	0.8107	0.8387	0.6896	0.8013	0.8068	0.7524	0.6603
SmoothSubspace	0.6055	0.8277	0.8600	0.9733	0.98	0.8466	0.9846	0.4024	0.9862	0.8711	0.8716
Wafer	0.9981	0.9965	0.9841	0.9959	0.9959	0.9998	0.9965	0.9988	0.9969	0.9954	0.9634
Mean	0.7476	0.7730	0.7610	0.7978	0.8227	0.8857	0.8427	0.8399	0.8266	0.8208	0.7800
Max. Resample	1.0	1.0	1.0	0.9986	1.0	1.0	1.0	1.0	1.0	0.9983	0.9972
Min. Resample	0.2450	0.2490	0.3545	0.2956	0.3056	0.4490	0.3345	0.34	0.3254	0.34	0.2290
Std.	0.2096	0.1739	0.1654	0.1900	0.2034	0.1219	0.1444	0.1735	0.1767	0.1803	0.1912

Figure 8. DWT with MTF GASF RP approach achieved the highest accuracy gain relative to the baseline models.

The results show that fusion-based representations do not consistently outperform single-transform methods. In several samples, individual representations achieve comparable or higher accuracy than their fused counterparts. This observation indicates that when a single transform already captures the dominant characteristics of the signal, combining it with additional representations may introduce redundancy rather than complementary information. In particular, fusion strategies involving CWT or FOD tend to exhibit less stable performance across samples. Although CWT provides detailed time–frequency information, it is sensitive to noise and scale selection. Similarly, FOD emphasizes local variations, which may negatively affect classification in noisy or short time series. When these representations are included in a fusion scheme, their limitations can influence the overall performance. In contrast, fusion methods that incorporate DWT coefficients with GAF, MTF, and RP representations show more stable behavior. The multi-resolution structure of DWT allows frequency-related information to be represented in a compact form, making it more compatible with image-based transforms. However, even in these cases, fusion does not guarantee improvement over the best-performing single representation.

Overall, these results suggest that the success of fusion-based approaches strongly depends on the choice and compatibility of the combined representations. Simply increasing representational diversity does not necessarily lead to better performance and may even degrade accuracy when irrelevant or redundant information is introduced. To further investigate this phenomenon, Fig. 8 provides a sample-level accuracy comparison. Specifically, the heatmap displays the top 7 datasets where the 'DWT with MTF GASF RP' fusion method achieved the highest accuracy gain compared to the mean performance of all other individual representations. These specific sample indices were selected to highlight the 'best-case' scenarios where DWT coefficients provide the most significant complementary spectral–temporal information, effectively overcoming the limitations of single-domain transforms.

The motivation behind the multi-branch architecture is to process each DWT component separately, allowing the network to learn more specialized spectral–temporal features from different frequency bands. Unlike the 6-channel approach, where all DWT components are merged at the input level, the multi-branch design preserves the unique characteristics of each sub-band and reduces feature interference during learning.

Figure 9 shows a consistent performance gain of the multi-branch architecture over the 6-channel model across several UCR datasets. In all evaluated cases, the multi-branch approach achieves equal or higher accuracy, indicating that independent branch processing provides a more effective feature representation than simple channel-wise fusion. For datasets with strong temporal or frequency-dependent patterns, such as CinCECGTorso and FordB, the multi-branch model yields moderate but stable improvements over the 6-channel design. This suggests that while both models benefit from wavelet-based representations, the multi-branch architecture offers greater flexibility in capturing discriminative patterns at different scales. The classical DWT+CNN model often performs between these two approaches, indicating that it captures useful information but lacks the capacity to adaptively emphasize different frequency components. In datasets with short or clearly defined signals, such as GunPoint, the performance gap between the two approaches is smaller. In these cases, simple temporal patterns are already well captured by basic convolutional structures, limiting the additional benefit of architectural complexity. Nevertheless, the multi-branch model still maintains a consistent advantage over the 6-channel approach.

Overall, the results demonstrate that separating DWT components into independent branches leads to more robust and consistent performance gains compared to channel-based fusion. Although both the 6-channel and multi-branch architectures aim to improve upon classical DWT+CNN, neither approach consistently surpasses raw CNN models on all challenging datasets. However, when wavelet-based representations are beneficial, the multi-branch architecture provides a more effective and stable solution.

B. Time performance

In our pipeline, the time-series signals are first converted into images and then fed into CNN. Crucially, the image-conversion stage is executed in parallel on the processor threads, which significantly reduces the preprocessing overhead. Because we parallelized the transformation of multiple samples at once, the

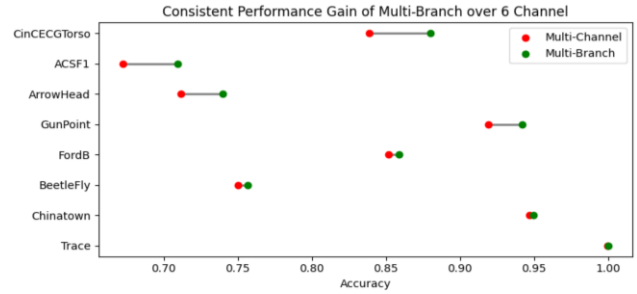


Figure 9. Accuracy comparison of single and fusion-based time-series image representations.

end-to-end time (image conversion + CNN training/testing) is much lower than if each sample were processed sequentially. This parallel image-conversion approach contributes to the very fast training durations we observed. Therefore, our decision to apply image-conversion in parallel allows us to leverage the computational capacity of our setup efficiently, and it helps explain why our training and test durations are so low. In this study, we evaluated our method on multiple time-series classification data sets. Training and testing durations were measured on a system equipped with an Intel i5 (11th generation) processor and an RTX 3060 GPU (12 GB VRAM). Image resizing and processing were carried out using scikit-image and PIL. Parallel preprocessing of image transformations was implemented using the joblib library to accelerate computation.

Across datasets, the mean transformation, training, and testing times were 3.81 s, 8.47 s, and 0.46 s, respectively. The dataset with the highest transformation time was NonInvasiveFetalECGThorax1 (13.26 s), while the lowest was SmoothSubspace (0.06 s). Individual datasets required a range of 1.43–41.48 s for the complete pipeline, with cumulative execution across all datasets totaling 191.05 s.

These results indicate that image-conversion imposes minimal overhead relative to CNN training, and that our parallel preprocessing approach efficiently leverages computational resources. The low durations suggest that the pipeline is suitable for real-time or near-real-time applications in IoT and time-series monitoring contexts.

C. Wavelet performance

To analyze the impact of wavelet selection on classification performance, four widely used wavelet families—Haar, db2, db4, and sym4—were evaluated across 15 datasets under identical experimental conditions. Each configuration was trained and tested using the same preprocessing pipeline, model architecture, and evaluation protocol, ensuring that any observed performance differences originate solely from the representation choice rather than from optimization or training variability.

These wavelets were selected to represent different levels of decomposition complexity and smoothness characteristics. The Haar wavelet was included as a simple and computationally efficient baseline with sharp discontinuity modeling capability. The Daubechies family members db2 and db4 were chosen due to their widespread use in time-series analysis and their ability to capture progressively smoother signal structures through increasing vanishing moments. In particular, db4 provides a balance between temporal localization and smooth approximation, making it a commonly preferred option in practical signal processing applications. The sym4 wavelet was additionally considered as a near-symmetric alternative designed to reduce phase distortion while preserving similar regularity properties.

The statistical analysis confirms that the performance differences among the evaluated wavelet families are negligible. The Friedman test indicates no statistically significant difference across methods ($p = 0.713$), suggesting that classification performance remains consistent regardless of the selected wavelet representation.

Pairwise Wilcoxon signed-rank tests further support this observation, as all comparisons yield p-values greater than 0.05, indicating the absence of pairwise superiority between any wavelet family.

Although minor variations in mean accuracy are observed (Haar: 0.754, db2: 0.753, sym4: 0.751, db4: 0.748), these differences remain practically insignificant. The average rank analysis also shows closely grouped rankings, reinforcing the stability of performance across wavelets.

For completeness, a one-way ANOVA analysis was additionally performed and produced a very high p-value ($p = 0.999$), further confirming the lack of statistically meaningful differences.

Figure 10 presents the mean accuracy results obtained from four different wavelet families (Haar, db2, db4, and sym4) across 15 datasets. Overall, the figure shows that the performance of the wavelet methods follows a very similar trend, indicating stable behavior across datasets. The accuracy values increase gradually from more challenging datasets to easier ones, and the differences between methods remain small.

However, some dataset-specific variations can still be observed. For example, in the NFECG1 dataset, the Haar wavelet achieves noticeably higher accuracy compared to the other wavelet families. This suggests that the Haar representation may capture important signal characteristics more effectively for this particular dataset. In contrast, on the SmoothSubspace dataset, the Haar wavelet shows a relatively lower accuracy compared to the other methods. While db2, db4, and sym4 maintain similar performance levels, Haar experiences a small performance drop. This difference indicates that certain datasets may favor smoother or more complex wavelet bases instead of simpler ones like Haar.

Despite these local variations, the overall differences across datasets remain minor. The figure clearly illustrates that no single wavelet family consistently outperforms the others. Therefore, the results support the conclusion that the overall success of the approach comes from the multi-representation learning strategy rather than from a specific wavelet type.

IV. CONCLUSION

The results of this study demonstrate that the Discrete Wavelet Transform (DWT) is effective in extracting frequency-based features from sensors, biomedical like time series datasets. By decomposing signals into approximation and detail coefficients, DWT helps to reduce noise and highlight relevant patterns across low-, medium-, and high-frequency bands. Our experiments on the UCR benchmark datasets indicate that DWT-based multi-channel representations can improve classification performance, particularly for noisy and non-stationary signals. This highlights the importance of frequency-domain analysis for sensor data and shows that careful selection of wavelet components can significantly enhance CNN-based time series classification.

Experimental evaluation on 15 representative UCR datasets revealed that the proposed DWT+CNN framework improved classification accuracy on 9 datasets, particularly those characterized by long duration, strong frequency components, or high noise levels. The most notable performance gains were observed on Mallat, FordB, SmallKitchen and FordA where frequency-domain decomposition enhanced discriminative feature extraction.

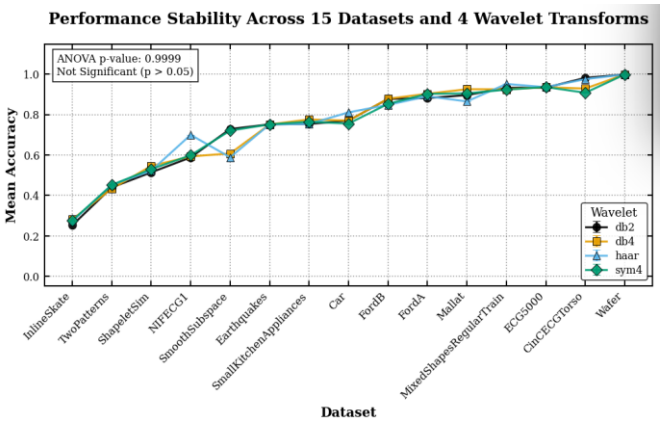


Figure 10. Accuracy plot of 4 wavelet across 15 selected datasets.

On datasets dominated by short sequences or strongly localized temporal patterns, such as TwoPatterns and SmoothSubspace, the raw CNN model achieved higher accuracy, indicating that explicit frequency decomposition may not benefit signals where local temporal structures are already highly separable. Importantly, for frequency-sensitive and noisy datasets, the proposed approach consistently outperformed the raw CNN and demonstrated consistently competitive performance compared to several state-of-the-art methods reported in literature, despite relying on a relatively small CNN architecture. This observation suggests that the performance gains primarily originate from the proposed multi-representation learning strategy rather than from increased model complexity or network scale.

These findings indicate that wavelet-based multi-resolution image encoding provides measurable advantages for complex, non-stationary sensor signals, while its benefits are dataset-dependent. The results confirm that frequency-aware representation learning enhances robustness without increasing model complexity.

AUTHOR STATEMENT

Plagiarism Check—The article has been scanned with iThenticate and found to be compliant with the journal's plagiarism policy.
Conflict of Interest—There is no conflict of interest with any person/organization.
Ethics Committee Approval—Ethics committee approval is not required for this article.
Use of Artificial Intelligence Tools—No artificial intelligence tools were used for content generation, literature review, data analysis, or evaluation in the preparation of this article.
Funding— No institutional/financial support was received for this study.
Data availability—No new data was generated or analyzed in this study.
CRediT Author Contribution—Conceptualization, methodology development, implementation, and manuscript drafting, (**Mehmet Kurnaz**); supervised the research, software, provided guidance, visualization (**Celal Alagöz**)
Acknowledgment—The authors thank the anonymous reviewers for their constructive comments and improvements to the manuscript.

REFERENCES

[1] H. Sharabiani, S. Darabi, S. Harford, E. Douzali, F. Karim, H. Johnson, S. Chen, "Asymptotic dynamic time warping calculation with utilizing value repetition", *Knowledge and Information Systems*, 57(2), (2018), 359–388.
[2] W. Chen, K. Shi, "Multi-scale attention convolutional neural network for time series classification", *Neural Networks*, 136, (2021), 126–140.
[3] H. Kang, T. H. Lee, J. Lee, "A graph convolutional network for time series classification using recurrence plots", *Applied Intelligence*, 55(15), (2025), 972.
[4] H. I. Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. Webb, L. Idoumghar, P. A. Muller, F. Petitjean, "InceptionTime: Finding AlexNet for time series classification", *Data Mining and Knowledge Discovery*, 34(6), (2020), 1936–1962.
[5] C. Alagöz, "Crossfire: Cross-domain feature integration for robust time series classification", *PeerJ Computer Science*, 11, (2025), e3328.
[6] X. T. Li, T. Y. Li, Y. Wang, "GW-DC: A deep clustering model leveraging two-dimensional image transformation and enhancement", *Algorithms*, 14(12), (2021), 349.
[7] X. T. Li, K. Zhou, F. Xue, Z. B. Chen, Z. Q. Ge, X. Chen, K. Song, "A wavelet transform-assisted convolutional neural network multi-model framework for monitoring large-scale fluorochemical engineering processes", *Processes*, 8(11), (2020), 1480.
[8] X. Y. Lu, Y. Li, X. Chen, Y. Q. Li, Y. B. Liu, "Discrete wavelet transform assisted convolutional neural network equalizer for PAM VLC system", *Optics Express*, 32(6), (2024), 10429–10443.
[9] Z. G. Wang, T. Oates, "Imaging time-series to improve classification and imputation", *24th International Joint Conference on Artificial Intelligence (IJCAI)*, Buenos Aires, Argentina, 25–31 July 2015.
[10] H. V. Costa, A. G. R. Ribeiro, V. M. A. Souza, "Fusion of image representations for time series classification with deep learning", *33rd International Conference on Artificial Neural Networks and Machine Learning,* Lugano, Switzerland, 17-20 September 2024.
[11] N. Hatami, Y. Gavet, J. Debayle, "Classification of time-series images using deep convolutional neural networks", *10th International Conference on Machine Vision (ICMV),* Vienna, Austria, 13–15 November 2017.
[12] N. Filimonova, M. Specovius-Neugebauer, E. Friedmann, "Determination of the time-frequency features for impulse components in EEG signals", *Neuroinformatics*, 23(2), (2025), 17.
[13] Y. Molina-Tenorio, A. Prieto-Guerrero, E. Rodriguez-Colina, L. A. Vásquez-Toledo, O. A. Olvera-Guerrero, "Gramian angular field and convolutional neural networks for real-time multiband spectrum sensing in cognitive radio networks", *Sensors*, 25(12), (2025), 3580.
[14] Y. Sharmal, N. Coronato, D. E. Browne, "Encoding cardiopulmonary exercise testing time series as images for classification using convolutional neural network", *44th Annual International Conference of the IEEE-Engineering-in-Medicine-and-Biology-Society (EMBC),* Glasgow, Scotland, 11-15 July 2022.
[15] M. Mariani, P. Appiah, O. Tweneboah, "Fusion of recurrence plots and gramian angular fields with Bayesian optimization for enhanced time-series classification", *Axioms*, 14(7), (2025), 528.
[16] J. Wu, Z. L. Zhang, R. Tong, Y. Zhou, Z. F. Hu, K. T. Liu, "Imaging feature-based clustering of financial time series", *PLoS ONE*, 18(7), (2023), doi: 10.1371/journal.pone.0288836.
[17] S. Barra, S. M. Carta, A. Corriga, A. S. Podda, D. R. Recupero, "Deep learning and time series-to-image encoding for financial forecasting", *IEEE-CAA Journal of Automatica Sinica*, 7(3), (2020) 683–692.
[18] X. B. Jin, A. Q. Yang, T. L. Su, J. L. Kong, Y. T. Bai, "Multi-channel fusion classification method based on time-series data", *Sensors*, 21(13), (2021), 4391.
[19] H. Lee, J. Lee, "Convolutional model with a time series feature based on RSSI analysis with the Markov transition field for enhancement of location recognition", *Sensors*, 23(7), (2023), 3453.
[20] V. M. A. Souza, P. S. Veiga, A. G. R. Ribeiro, "ViSemble: A fast ensemble approach for time series classification with multiple visual representations", *Knowledge-Based Systems*, 309, (2025), 112864.
[21] H. A. Dau, E. Keogh, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, Y. Chen, B. Hu, N. Begum, A. Bagnall, A. Mueen, G. Batista, "The UCR Time Series Classification Archive", https://www.cs.ucr.edu/~eamonn/time_series_data_2018, (23.2.2026).
[22] H. Ding, G. Trajcevski, P. Scheuermann, X. Y. Wang, E. Keogh, "Querying and mining of time series data: Experimental comparison of representations and distance measures", *Proceedings of the VLDB Endowment*, 1(2), (2008), 1542–1552.
[23] P. Schäfer, "The BOSS is concerned with time series classification in the presence of noise", *Data Mining and Knowledge Discovery*, 29(6), (2015), 1505–1530.
[24] H. T. Deng, G. Runger, E. Tuv, M. Vladimir, "A time series forest for classification and feature extraction", *Information Sciences*, 239, (2013), 142–153.
[25] M. Flynn, J. Large, A. Bagnall, "The contract random interval spectral ensemble (c-RISE): The effect of contracting a classifier on accuracy", *14th International Conference on Hybrid Artificial Intelligence Systems (HAIS),* Leon, Spain, 4-6 September 2019.
[26] B. D. Zhao, H. Z. Lu, S. F. Chen, J. L. Liu, D. Y. Wu, "Convolutional neural networks for time series classification", *Journal of Systems Engineering and Electronics*, 28(1), (2017), 162–169.
[27] M. Middlehurst, J. Large, G. Cawley, A. Bagnall, "The temporal dictionary ensemble (TDE) classifier for time series classification", *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2020)*, Ghent, Belgium, 14-18 September 2020.
[28] J. Lines, A. Bagnall, "Time series classification with ensembles of elastic distance measures", *Data Mining and Knowledge Discovery*, 29(3), (2015), 565–592.