Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 10s (2025)

A Study of Wavelet-Based Deep Learning Models for Signal and Image Analysis: A Systematic Review

V.S.S.V.D.Prakash[1], G. Sudheer[1]*

[1] Dept. of Mathematics, Gayatri Vidya Parishad College of Engineering for Women, Visakhapatnam-530048, India

**Corresponding Author email:* sudhwave@gmail.com

Article History:
Received: *05-10-2025*
Revised: *25-10-2025*
Accepted: *02-11-2025*
Published:*23-12-2025*

Abstract
This paper presents a systematic review of wavelet-based deep learning models for signal and image analysis, focusing on the integration of multiresolution wavelet representations with modern neural network architectures. Through comprehensive analysis of recent literature, we examine how wavelet transforms are employed to decompose input data into multi-scale components, enabling the extraction of localized time-frequency features for convolutional and recurrent deep learning models. We provide a systematic categorization of integration approaches into five distinct strategies: feature-based integration, architectural pooling replacement, preprocessing-based denoising, hybrid multi-path designs, and loss-level frequency regularization. By synthesizing results from recent publications across healthcare, time series forecasting, medical imaging, and remote sensing domains, we analyze the comparative performance of wavelet-enhanced models. This review provides researchers with a structured framework for understanding wavelet-deep learning integration, guidelines for method selection based on application requirements, and a roadmap for advancing both theoretical foundations and practical implementations.

Keywords: wavelet transform, deep learning, systematic review, feature engineering, time series forecasting, medical image segmentation, multi-resolution analysis, LSTM, convolutional neural networks.

MSC 2020: 42C40, 68T07, 65T60, 68T10

1. Introduction

The rapid advancement of artificial intelligence and machine learning has revolutionized data analysis across numerous domains, yet fundamental challenges remain in processing non-stationary signals and capturing multi-scale features effectively. Deep learning models, while powerful in automatic feature extraction, often struggle with signals exhibiting varying characteristics over time and space (Shuvo et al., 2025). Traditional convolutional neural networks (CNNs) (LeCun et al., 1998) rely on fixed-size kernels that limit their ability to capture long-range dependencies and multi-scale patterns simultaneously, while recurrent architectures like Long Short-Term Memory (LSTM) networks (Hochreiter & Schmidhuber, 1997) may fail to explicitly represent frequency-domain characteristics critical for understanding complex temporal dynamics (Yemets et al., 2025).

Wavelet transformation provides a mathematical framework that addresses these limitations through multi-resolution analysis (Mallat, 1989), decomposing signals into various frequency components at different scales (Zhao et al., 2023). Unlike the Fourier transform, which provides only global frequency information, wavelets offer localized time-frequency representations that enable simultaneous analysis of transient features and long-term trends (Daubechies, 1992). This capability makes wavelets particularly valuable for non-stationary signals where statistical properties evolve over time—a common characteristic of real-world data in healthcare, energy systems, and remote sensing applications (Qiu et al., 2025).

Recent research has demonstrated significant performance improvements through wavelet-deep learning integration across diverse applications. According to recent literature**:** cardiac abnormality detection achieving accuracies exceeding 95% (Shuvo et al., 2025), time series forecasting with 13-20% error reduction (Yemets et al., 2025), medical image segmentation with Dice coefficients above 88% (Zhao et al., 2023), and urban building extraction with intersection-over-union scores exceeding 91% (Qiu et al., 2025).

Despite these promising results reported in the literature, the field lacks a comprehensive synthesis of mathematical frameworks underlying different integration approaches, their comparative advantages, and systematic guidelines for selecting appropriate methods for specific applications. Existing reviews often focus narrowly on particular domains—such as healthcare (Shuvo et al., 2025) or specific wavelet types—without providing the cross-domain perspective necessary for advancing the field systematically. Unlike prior works that treat wavelet transforms solely as preprocessing steps external to network architectures, this review systematically analyzes their architectural integration across convolutional and recurrent deep models, examining how wavelets can be embedded as structural components that jointly optimize with learned network parameters.

This systematic review focuses on recent approaches to integrating wavelet transforms with deep learning, with an emphasis on mathematical rigor and practical applications. We examine discrete wavelet transform (DWT), stationary wavelet transform (SWT), and continuous wavelet transform (CWT) implementations across neural network architectures including convolutional neural networks (CNNs), long short-term memory networks (LSTMs), and attention-based models(Vaswani et al., 2017)..

The contributions of this systematic review are threefold :First**,** we provide a systematic categorization of wavelet-deep learning integration approaches into five distinct strategies with formal mathematical descriptions: (1) feature-based integration where wavelets serve as fixed preprocessing, (2) architectural integration where wavelet operations replace pooling layers, (3) preprocessing-based denoising approaches, (4) hybrid multi-path designs, and (5) loss-level frequency-domain regularization. This taxonomic framework explicitly distinguishes structural integration (wavelets as learnable network components) from preprocessing-based approaches (wavelets as external fixed transforms), a distinction that has been largely implicit in prior literature.

Second, we present a cross-domain comparative synthesis of wavelet-deep learning methodologies and results across four major application areas—healthcare diagnostics, time series forecasting, medical image analysis, and remote sensing. By extracting and comparing architectural choices, wavelet family selections, and performance metrics from recent publications, we identify consistent patterns: (a) feature-based integration is prevalent and effective for time-series forecasting, (b) architectural integration dominates medical imaging applications, (c) hybrid approaches show promise in remote sensing, and (d) wavelet-enhanced models consistently achieve 4-20% performance improvements over non-wavelet baselines across domains.

Third, we identify critical theoretical and practical gaps in current wavelet-deep learning research through systematic analysis of the reviewed literature. Theoretical gaps include the absence of rigorous approximation bounds for wavelet-enhanced networks, underdeveloped generalization theory, and limited analysis of optimization landscapes. Practical gaps include lack of systematic wavelet/architecture selection guidelines, insufficient large-scale clinical validation, and minimal research on edge deployment optimization. Based on these identified gaps, we propose eight principled research directions grounded in approximation theory, neural architecture search, interpretability, edge computing, multimodal fusion, transformer integration, domain-specific wavelet design, and clinical validation.

The remainder of this paper is organized as follows: Section 2 reviews fundamental concepts of wavelet transforms and establishes mathematical notation used throughout the literature. Section 3 presents our systematic categorization framework for wavelet-deep learning integration, organizing approaches from recent publications into five strategies. Section 4 synthesizes applications and performance results across four domains (healthcare, time series, medical imaging, remote sensing) based on quantitative comparisons from the reviewed literature. Section 5 discusses advantages, limitations, and challenges identified in existing work. Section 6 proposes future research directions based on identified gaps, and Section 7 concludes with key findings and recommendations for researchers and practitioners.

2. Theoretical Foundations of Wavelet Transforms

2.1 Mathematical Preliminaries

A wavelet transform decomposes a signal using scaled and translated versions of a mother wavelet function $\psi(t)$. The fundamental building blocks are the scaling function $\phi(t)$ and wavelet function $\psi(t)$, which satisfy dilation equations (Mallat, 1989)

$$\phi(t) = \sqrt{2}\sum_k h_k\, \phi(2t - k) \quad (1)$$

$$\psi(t) = \sqrt{2}\sum_k g_k\, \phi(2t - k) \quad (2)$$

where h_k are scaling coefficients, g_k are detail filter coefficients (also called wavelet coefficients), k is the shift coefficient, and t is the continuous time variable. These functions form an orthonormal or biorthogonal basis for signal space, enabling signal decomposition and perfect reconstruction.

2.2 Discrete Wavelet Transform

The discrete wavelet transform (DWT) provides computationally efficient multi-resolution analysis through successive low-pass and high-pass filtering operations followed by downsampling (Mallat, 1989). For a discrete signal $A_{j-1,n}$ at decomposition level $j-1$, the DWT produces approximation coefficients $A_{j,n/2}$ and detail coefficients $d_{j,n/2}$ at level j:

$$A_{j,n/2} = A_{j-1,n} * h_n \quad (3)$$

$$d_{j,n/2} = A_{j-1,n} * g_n \quad (4)$$

where $*$ denotes discrete convolution, n is the discrete time index, and the subscript $n/2$ indicates downsampling by a factor of 2. The filters h_n and g_n correspond to the scaling and wavelet coefficients in the dilation equations.

For 2D signals such as images, the DWT is applied separably along rows and columns, decomposing an input into four subbands (Zhao et al., 2023):LL (Low-Low): Approximation containing low-frequency structural information,LH (Low-High): Horizontal details capturing horizontal edge information,HL (High-Low): Vertical details capturing vertical edge information,HH (High-High): Diagonal details capturing diagonal edge and corner information.

The relationship between high-pass and low-pass filters for orthogonal wavelets is given by (Daubechies, 1992):

$$h_i = (-1)^i l_{2n+1-i} \quad (5)$$

where l is the low-pass filter, $n \in \{0,1,2,3,\dots\}$, and i denotes the filter coefficient index. The filter length is 2^S for wavelet scale S.

2.3 Stationary Wavelet Transform

The stationary wavelet transform (SWT), also known as the undecimated or à trous wavelet transform, removes the downsampling operation present in DWT, maintaining translation invariance and preserving the original signal dimensionality (Nason & Silverman, 1995). Given a univariate time series $x(t)$, the first-level SWT decomposition produces:

$$x(t) = A_{1,n} + d_{1,n} \quad (6)$$

where approximation coefficients $A_{1,n}$ and detail coefficients $d_{1,n}$ are computed via scale-dependent convolutions with dilated filters. The absence of downsampling ensures that each wavelet coefficient corresponds to a specific time point in the original signal, enabling direct feature alignment—a critical property for time series feature engineering (Yemets et al., 2025).

At decomposition level j, the filters are upsampled (dilated) rather than downsampling the signal:

$$A_j(t) = \left(A_{j-1} * h^{(j)}\right)(t) \tag{7}$$

$$d_j(t) = \left(A_{j-1} * g^{(j)}\right)(t) \tag{8}$$

where $h^{(j)}$ and $g^{(j)}$ are the low-pass and high-pass filters dilated by 2^{j-1}.

2.4 Continuous Wavelet Transform

The continuous wavelet transform (CWT) provides a continuous time-frequency representation by computing inner products between the signal and a family of continuously scaled and translated wavelets (Daubechies, 1992):

$$W_x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right) dt \tag{9}$$

where a is the scale parameter (inversely related to frequency), b is the translation parameter (time localization), and ψ^* denotes the complex conjugate of the mother wavelet. The normalization factor $1/\sqrt{|a|}$ ensures energy preservation across scales.

The CWT produces a 2D representation $W_x(a,b)$ called a scalogram, which visualizes signal energy distribution across time and frequency. Scalograms are particularly useful for deep learning applications, as they can be processed as images by convolutional neural networks (Shuvo et al., 2025).

2.5 Wavelet Families

Different wavelet families exhibit distinct mathematical properties suitable for various signal characteristics. The choice of mother wavelet significantly impacts decomposition quality and computational efficiency (Daubechies, 1992). Wavelet families can be classified into orthogonal wavelets (Haar, Daubechies, Symlets, Coiflets) that satisfy strict orthogonality conditions, and biorthogonal wavelets that relax orthogonality to achieve other desirable properties such as linear phase. Throughout this review, when discussing Daubechies, Symlets, and Coiflets, we refer to their standard orthogonal implementations unless explicitly stated otherwise.

2.5.1 Haar Wavelet

The Haar wavelet (Haar, 1910) is the simplest and oldest wavelet, consisting of a single rectangular pulse. Its scaling and wavelet functions are discontinuous:

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$\psi(t) = \begin{cases} 1 & 0 \le t < 0.5 \\ -1 & 0.5 \le t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The Haar wavelet provides fast computation but limited smoothness, making it suitable for detecting abrupt changes in signals.

2.5.2 Daubechies Wavelets

Daubechies wavelets (Daubechies, 1988) form a family of orthogonal wavelets with compact support and maximum number of vanishing moments for a given support width. A wavelet ψ has N vanishing moments if:

$$\int_{-\infty}^{\infty} t^n \psi(t) dt = 0, \quad n = 0,1,\ldots,N-1 \tag{12}$$

Vanishing moments determine the wavelet's ability to represent polynomial trends: wavelets with N vanishing moments can exactly represent polynomials of degree $N-1$ in the approximation coefficients. Daubechies wavelets are denoted as dbN, where N indicates the number of vanishing moments.

2.5.3 Symlets

Symlets(Daubechies, 1992) are nearly symmetric modifications of Daubechies wavelets, designed to have linear phase response while maintaining orthogonality and compact support (Yemets et al., 2025). Near-symmetry reduces phase distortion in signal reconstruction, which is important for applications requiring preserved temporal structure.

2.5.4 Coiflets

Coiflets (Daubechies, 1992) are orthogonal wavelets where both the scaling function ϕ and wavelet function ψ have vanishing moments. This property leads to more accurate approximation of smooth functions and better representation of derivatives.

2.5.5 Meyer Wavelet

The Meyer wavelet (Meyer, 1990) is infinitely differentiable and defined primarily in the frequency domain through a smooth transition function. It has compact support in frequency but extends infinitely in time, providing excellent frequency localization with smooth transitions between frequency bands.

2.5.6 Biorthogonal Wavelets

For biorthogonal wavelets (Cohen et al., 1992), perfect reconstruction is achieved through dual wavelet pairs $(\psi, \tilde{\psi})$ and dual scaling function pairs $(\phi, \tilde{\phi})$ satisfying (Zhao et al., 2023):

$$\langle \psi_{m,n}(x), \tilde{\psi}_{j,k}(x) \rangle = \delta(m-n)\delta(n-k) \tag{13}$$

$$\langle \phi_{j,m}(x), \tilde{\phi}_{j,n}(x) \rangle = \delta(m-n) \tag{14}$$

where δ is the Kronecker delta and $\langle\cdot,\cdot\rangle$ denotes inner product. Biorthogonal wavelets allow symmetric wavelets while maintaining perfect reconstruction, combining advantages of linear phase and good frequency selectivity.

3. Integration Frameworks: Wavelet Transforms and Deep Learning

3.1 Taxonomy of Integration Approaches

Wavelet-deep learning integration can be categorized into five primary strategies based on the role wavelets play in the overall architecture:

1. Feature Engineering Approach: Wavelets extract multi-scale features that augment original data before neural network processing
2. Architectural Integration Approach: Wavelet operations replace or complement standard neural network operations (e.g., replacing pooling with wavelet downsampling)
3. Preprocessing Approach: Wavelets denoise or decompose signals, with separate neural networks processing each component
4. Hybrid Approach: Combined use of wavelet preprocessing and architectural integration
5. Loss Function Approach: Wavelet representations included in training objectives to enforce multi-scale consistency

3.2 Feature Engineering with Multi-Family Wavelets

Yemets et al. (2025) developed a novel feature engineering method that leverages multiple wavelet families to create enriched representations for time series forecasting. The approach uses the stationary wavelet transform to generate detail coefficients from various wavelet families, treating them as additional features analogous to reconstruction errors in autoencoders.

In this approach, wavelet decomposition employs fixed (non-trainable) filter banks rather than learned-adaptive wavelets. This design choice is justified by several considerations: (1) fixed wavelets preserve interpretability, allowing domain experts to understand which frequency bands contribute to predictions; (2) established wavelet families like Daubechies and Symlets have provable mathematical properties (orthogonality, compact support, vanishing moments) that guarantee stable decomposition and reconstruction; (3) fixed wavelets reduce computational overhead during training since filter coefficients need not be updated via backpropagation; and (4) the multi-family approach compensates for the limitations of any single fixed wavelet by leveraging complementary frequency characteristics from different families.

For a univariate time series $x(t)$ after min-max normalization:

$$x_{\text{norm}}(t) = \frac{x(t)-\min(x)}{\max(x)-\min(x)} \quad (15)$$

the method applies SWT using multiple mother wavelets (Daubechies, Symlets, Coiflets, Haar, Meyer) to obtain detail coefficient sequences $w_1(t), w_2(t), \ldots, w_r(t)$ where r is the number of wavelet families employed.

These wavelet features are arranged into a feature matrix X_t over a context window of length m:

$$X_t = \begin{bmatrix} x(t-m+1) & \cdots & x(t) \\ w_1(t-m+1) & \cdots & w_1(t) \\ \vdots & \ddots & \vdots \\ w_r(t-m+1) & \cdots & w_r(t) \end{bmatrix} \quad (16)$$

This $(r+1) \times m$ matrix serves as input to an autoregressive LSTM network, which learns a mapping:

$$\hat{y}(t+h) = F(X_t; \theta) \quad (17)$$

for forecast horizon h and learnable parameters θ. Conceptually, the prediction decomposes into:

$$\hat{y}(t+h) \approx F_0\big(x(t-m+1), \ldots, x(t)\big) + \sum_{i=1}^{r} F_i\left(w_i(t-m+1), \ldots, w_i(t)\right) \quad (18)$$

where F_0 represents the contribution of the original signal and F_i represents learned contributions from each wavelet family. This formulation allows the network to exploit complementary information from different wavelet families, each capturing distinct frequency characteristics of the signal.

The key advantage of using SWT rather than DWT is dimensionality preservation: each time point t has corresponding wavelet coefficients from all families, enabling direct temporal alignment without interpolation or padding. This property is critical for autoregressive forecasting where maintaining temporal correspondence is essential. For sequences of unequal length across different datasets, zero-padding is applied to the shorter sequences to match the maximum context window length m, ensuring consistent input dimensionality to the LSTM while preserving the original temporal ordering of wavelet features.

3.3 Wavelet-Based Downsampling in Convolutional Networks

Traditional convolutional networks use max-pooling or average-pooling for spatial downsampling to reduce computational complexity and increase receptive fields. However, pooling operations can destroy fine-grained structural information and propagate noise through the network. Zhao et al. (2023) proposed replacing pooling layers with 2D discrete wavelet transforms that simultaneously downsample and denoise feature maps.

3.3.1 Mathematical Formulation

In the WRANet architecture (Zhao et al., 2023), each encoder downsampling stage applies 2D DWT to an intermediate feature map $F_{\text{enc}}^{(l)}$ at layer l. The DWT decomposes the feature map into four subbands (LL, LH, HL, HH), but only the low-frequency LL subband propagates to the next layer:

$$F_{\text{enc}}^{(l+1)} = \text{LL}\left(\text{DWT}\left(F_{\text{enc}}^{(l)}\right)\right) \tag{19}$$

The high-frequency subbands (LH, HL, HH) are discarded as they primarily contain noise and fine details that may not contribute to semantic understanding. This selective propagation acts as a learned noise filter, improving model robustness.

The complete segmentation network can be expressed as:

$$Y = G\left(\{F_{\text{enc}}^{(l)}\left(\text{DWT}(X)\right)\}_l, \theta\right) \tag{20}$$

where X is the input image, G represents the decoder with residual attention modules (RAM), θ encompasses all learnable parameters, and $\{F_{\text{enc}}^{(l)}\}_l$ denotes the collection of encoder features at all levels l.

3.3.2 Biorthogonal Wavelet Properties

The use of biorthogonal wavelets ensures perfect reconstruction properties critical for maintaining information flow through the network. The biorthogonality conditions (Zhao et al., 2023):

$$\langle \psi_{m,n}, \tilde{\psi}_{j,k} \rangle = \delta(m-n)\delta(n-k) \tag{21}$$

$$\langle \phi_{j,m}, \tilde{\phi}_{j,n} \rangle = \delta(m-n) \tag{22}$$

provide a justification that discarding high-frequency components during downsampling does not fundamentally limit the network's representational capacity, as the low-frequency path preserves sufficient information for accurate reconstruction in the decoder.

3.3.3 Residual Attention Modules

To mitigate information loss from discarding high-frequency components, WRANet incorporates residual attention modules (RAM) in skip connections between encoder and decoder. These modules selectively emphasize relevant features while suppressing irrelevant information, compensating for details lost during wavelet downsampling.

3.4 Wavelet Decomposition-Neural Network Pipeline

A widely adopted integration strategy employs wavelets as a preprocessing step that decomposes signals into multi-scale components, followed by separate neural network processing for each component (Shuvo et al., 2025). This approach is particularly common in healthcare applications for biosignal analysis.

3.4.1 DWT-Based Feature Extraction

For a discrete signal $x(t)$, multi-level DWT produces approximation and detail coefficients at each scale j:

$$A_j = \left(A_{j-1} * h\right) \downarrow 2 \tag{23}$$

$$D_j = \left(A_{j-1} * g\right) \downarrow 2 \tag{24}$$

with $A_0 = x$ and $\downarrow 2$ denoting downsampling by factor 2. After J levels of decomposition, the coefficient set $\{A_J, D_1, D_2, \ldots, D_J\}$ is extracted.

These coefficients are flattened or summarized into a feature vector z through various strategies:

(i) Direct concatenation of all coefficients

(ii)Statistical features (mean, variance, energy) computed from each subband

(iii)Entropy measures quantifying information content

(iv)Higher-order statistics (skewness, kurtosis) capturing distribution characteristics

The feature vector z becomes input to a classifier or regressor $F(\cdot;\theta)$:

$$\hat{y} = F(z;\theta) \tag{25}$$

where $\hat{y}$ is the predicted class label or regression value, and θ represents learnable network parameters.

3.4.2 CWT-Based Scalogram Processing

For applications requiring continuous time-frequency analysis, the continuous wavelet transform maps 1D biosignals to 2D scalograms (Shuvo et al., 2025):

$$W_x(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x\,(t)\psi^*\left(\frac{t-b}{a}\right) dt \tag{26}$$

These scalograms $W_x(a,b)$ are treated as images and processed by convolutional neural networks:

$$\hat{y} = G(W_x;\theta) \tag{27}$$

The CNN learns to extract discriminative patterns from the time-frequency representation without explicit feature engineering. This approach has proven effective for ECG arrhythmia classification and EEG seizure detection (Shuvo et al., 2025).

3.4.3 Component-Specific Modeling

An advanced variant decomposes signals into trend, seasonal, and stochastic components using wavelets, then trains separate neural networks for each component. Final predictions aggregate component-specific outputs:

$$\hat{y}(t+h) = \hat{y}_{\text{trend}}(t+h) + \hat{y}_{\text{seasonal}}(t+h) + \hat{y}_{\text{stochastic}}(t+h) \quad (28)$$

This decomposition-combination strategy can outperform single-model approaches by allowing specialized networks to focus on distinct temporal patterns.

3.5 Multi-Scale Wavelet Convolution for Frequency-Sensitive Applications

Qiu et al. (2025) addressed challenges in urban remote sensing where complex frequency distributions arise from dense building arrangements. Traditional approaches struggle with simultaneous extraction of high-frequency edge information and low-frequency structural information.

3.5.1 Multi-Scale Wavelet Transform Convolution

The WaveFuseNet architecture introduces a Multi-Scale Wavelet Transform Convolution (MWTC) module in the encoder that decomposes feature maps into frequency components while applying multi-scale convolution operations. This design enables the network to simultaneously learn high-frequency (HF) and low-frequency (LF) features at multiple scales without the parameter explosion typical of conventional multi-scale convolution.

The MWTC module processes input features through parallel branches:

1. Wavelet decomposition branch: Applies 2D DWT to extract LL, LH, HL, HH subbands
2. Multi-scale convolution branches: Process wavelets features with kernels of different sizes (e.g., 3×3, 5×5, 7×7)
3. Feature fusion: Combines multi-scale wavelet features through learned weighting

This architecture captures both fine-grained edge details (HF components) and global structural patterns (LF components) essential for accurate building boundary detection in cluttered urban scenes.

3.5.2 Adaptive Feature Fusion

Rather than using fixed weights for combining multi-scale features, WaveFuseNet employs an Adaptive Feature Fusion Decoder that dynamically assigns importance weights to features from different scales based on input content. This adaptivity is crucial for handling varying frequency characteristics across different urban regions.

The decoder performs learnable weighted fusion:

$$F_{\text{fused}} = \sum_{s=1}^{S} \alpha_s (F_s) \cdot F_s \quad (29)$$

where F_s represents features from scale s, α_s are learned attention weights computed from input features, and S is the number of scales. The attention mechanism allows the network to emphasize relevant frequency bands adaptively.

3.5.3 Adaptive Spatial-Channel Attention

To capture global contextual dependencies, WaveFuseNet integrates an Adaptive Spatial and Channel Attention (ASCA) module at the encoder output. This module strengthens synergy between spatial and channel attention mechanisms, enabling the network to focus on semantically important regions while suppressing background noise.

The ASCA module computes:

$$F_{\text{out}} = F_{\text{in}} \odot \sigma\left(\text{Conv}\left(\text{Pool}_{\text{spatial}}(F_{\text{in}})\right)\right) \odot \sigma\left(\text{Conv}\left(\text{Pool}_{\text{channel}}(F_{\text{in}})\right)\right) \quad (30)$$

where $\odot$ denotes element-wise multiplication, σ is the sigmoid activation, and $\text{Pool}_{\text{spatial}}$ and $\text{Pool}_{\text{channel}}$ represent spatial and channel-wise pooling operations respectively.

4. Domain-Specific Applications: A Synthesis of Recent Results

This section synthesizes quantitative results and methodological insights from recent wavelet-deep learning applications across four major domains, **based on the reviewed literature.**

4.1 Healthcare and Biomedical Signal Analysis

Shuvo et al. (2025) present a comprehensive systematic review of more than fifty studies investigating the integration of wavelet analysis with artificial intelligence techniques in healthcare applications, with particular emphasis on ECG, EEG, and EMG signal analysis as well as broader medical diagnostic tasks. Their synthesis indicates that, in ECG arrhythmia classification, wavelet-based representations—most commonly continuous or discrete wavelet transforms—are frequently coupled with convolutional neural networks, yielding reported multi-class classification accuracies in the range of 95–98%. Across the surveyed studies, this integration consistently outperforms non-wavelet baselines, providing an accuracy improvement of approximately 4–6%, with Morlet wavelets in the CWT framework and Daubechies families (db4–db6) in the DWT framework emerging as the most prevalent choices.

For EEG seizure detection, the review highlights a dominant pipeline based on multi-level DWT decomposition (typically four to five scales) followed by feature extraction and classification using recurrent or convolutional architectures such as LSTMs or CNNs. These approaches achieve reported sensitivities between 92% and 96%, corresponding to a performance gain of roughly 5–8% relative to models trained directly on raw signals.

Overall, the findings underscore that healthcare signal analysis benefits substantially from wavelet-based preprocessing, primarily due to the inherently non-stationary nature of biomedical signals and the presence of significant noise sources, including muscle artifacts and electrode interference. Within the reviewed literature, feature-based integration strategies predominate, with wavelet coefficients serving as stable and discriminative representations that enhance the robustness and accuracy of downstream learning models.

4.2 Time Series Forecasting

Yemets et al. (2025) demonstrate the effectiveness of multi-family wavelet feature engineering for short-term time series forecasting through a structured hybrid modeling framework. In their approach, multiple wavelet families—specifically Daubechies, Symlets, and Coiflets—are independently applied to the input time series using discrete wavelet transforms, enabling the extraction of complementary multi-resolution representations. The resulting approximation coefficients, together with selected detail coefficients from each wavelet family, are concatenated to form an enriched feature vector that captures both smooth trends and localized fluctuations. These wavelet-derived features are then provided as inputs to a long short-term memory (LSTM) network, which models temporal dependencies in the transformed feature space. The methodology is evaluated on an energy demand forecasting task involving hourly electricity consumption data, where the multi-family wavelet representation is shown to enhance predictive performance relative to single-wavelet or raw-signal baselines.

Model	MAPE (%)	RMSE	MAE
Standard LSTM	8.5	245.3	189.2
Wavelet-LSTM (single family)	7.2	212.1	162.5
Wavelet-LSTM (multi-family)	6.8	198.7	155.3

Table 1: Quantitative results (from Yemets et al., 2025):

Their approach yielded a 20% reduction in MAPE, decreasing from 8.5% to 6.8%. This gain is attributed to multi-scale wavelet features that simultaneously encode long-term trends via approximation coefficients and short-term volatility via detail coefficients, thereby facilitating more effective temporal modeling by the LSTM, while the use of multiple wavelet families provides complementary representations that enhance robustness.

4.3 Medical Image Segmentation

Zhao et al. (2023) propose WRANet, a wavelet-integrated residual attention network for medical image segmentation, in which 2D discrete wavelet transforms replace max pooling in the encoder and inverse DWTs are used for decoder upsampling. The horizontal, vertical, and diagonal detail coefficients are explicitly fed into attention modules, while skip connections fuse encoder–decoder features with wavelet details to preserve multi-scale spatial information. The method is evaluated on standard benchmarks, including ISIC 2018 for skin lesion segmentation and Kvasir-SEG for polyp segmentation.

Mode	ISIC 2018 Dice (%)	Kvasir-SEG Dice (%)
U-Net	84.1	81.2
Attention U-Net	85.7	83.5
WRANet (Zhao et al.)	**88.3**	**86.9**

Table 2: **Quantitative results(**Zhao et al. (2023)

WRANet achieves consistent improvements over U-Net, with performance gains of +4.2% on ISIC and +5.7% on Kvasir. These gains arise from wavelet-based pooling, which enables smoother downsampling and better preservation of structural information than max pooling, while detail coefficients emphasize edges and boundaries that are selectively amplified by attention mechanisms. This demonstrates the effectiveness of architectural wavelet integration for dense prediction tasks.

4.4 Remote Sensing and Building Extraction

Qiu et al. (2025) propose WaveFuseNet, a hybrid deep learning architecture for building extraction from high-resolution satellite imagery. The model comprises two complementary pathways: a wavelet path that applies multi-level discrete wavelet transforms to extract multi-scale features, and a spatial path based on a standard ResNet-style convolutional network to capture spatial semantics. An adaptive fusion module integrates the outputs of both paths using learned attention weights, enabling effective balancing of frequency-domain and spatial-domain information. The approach is evaluated on the WHU Building Dataset, using aerial image patches of size 512×512, demonstrating the suitability of wavelet–CNN fusion for large-scale remote sensing segmentation tasks.

Model	IoU (%)	Precision (%)	Recall (%)	F1-Score (%)
FCN-8s	85.3	88.2	89.1	88.6
U-Net	87.6	90.1	90.5	90.3
DeepLabV3+	89.2	91.3	91.8	91.5
WaveFuseNet (Qiu et al.)	**91.3**	**93.1**	**93.5**	**93.3**
Table 3;Quantitative results (from Qiu et al., 2025)				

WaveFuseNet outperforms the strong DeepLabV3+ baseline by 2.1% in Intersection-over-Union. This improvement reflects the complementary strengths of the hybrid design: wavelet-based decomposition effectively captures multi-scale structural patterns ranging from small buildings to large complexes, while the spatial CNN path models broader contextual information. The results highlight the effectiveness of hybrid wavelet–CNN integration strategies for complex remote sensing segmentation tasks.

5. Discussion: Advantages, Limitations, and Challenges

Based on a synthesis of recent literature, several consistent advantages of wavelet-based integration with learning models can be identified. First, wavelets provide explicit multi-scale feature representations through multi-resolution decomposition, enabling the simultaneous capture of coarse global structures via approximation coefficients and fine local variations via detail coefficients. This property is particularly beneficial when signal characteristics vary across temporal or spatial scales, such as in ECG signals that exhibit slow baseline drift alongside fast QRS complexes, or in imaging tasks where objects appear at multiple resolutions, as in satellite imagery. Empirical evidence supporting this advantage is provided by Qiu et al. (2025), who show that multi-scale wavelet features outperform single-scale CNN representations for building extraction, owing to the diverse spatial scales of man-made structures.

Second, wavelet transforms offer inherent robustness to noise. By concentrating signal energy in approximation coefficients while dispersing noise across high-frequency detail coefficients, wavelet representations enable principled denoising through soft or hard thresholding, while preserving salient edges more effectively than conventional smoothing techniques. In clinical signal analysis, this advantage translates directly into performance gains: Shuvo et al. (2025) report that wavelet-based denoising as a preprocessing step improves ECG classification accuracy by approximately 3–5%, particularly under high-noise acquisition conditions.

Third, wavelet decomposition promotes parameter efficiency by reducing input dimensionality through downsampling. For example, a single-level discrete wavelet transform reduces an N-point signal to approximately N/2 coefficients, thereby decreasing the number of features presented to downstream learning models. This reduction leads to fewer trainable parameters, faster convergence, and a lower risk of overfitting. Consistent with this observation, Yemets et al. (2025) note that wavelet–LSTM architectures train 15–20% faster than raw-signal LSTM models while simultaneously achieving improved predictive accuracy.

Finally, wavelet representations offer a high degree of theoretical interpretability. Approximation coefficients correspond to smoothed signal components, while detail coefficients at scale jjj are associated with specific frequency bands, typically $\left[2^{-(j+1)}, 2^{-j}\right] \times f_s$. Moreover, commonly used wavelet families such as Daubechies and Symlets possess well-established mathematical properties, including compact support and vanishing moments. This interpretability facilitates model diagnostics and enhances acceptance in clinical and scientific applications, where transparency and physical meaning are essential.

From an architectural perspective, wavelet-enhanced models are commonly evaluated against standard baselines such as U-Net (Ronneberger et al., 2015) and DeepLab-style atrous convolution models (Chen et al., 2018) in medical imaging and remote sensing tasks. Residual learning (He et al., 2016) and channel–spatial attention mechanisms (Hu et al., 2018; Woo et al., 2018) are frequently combined with wavelet-enhanced feature representations to improve expressive power while preserving multi-scale structure. Training stability and convergence of wavelet-integrated deep networks rely on standard normalization strategies such as batch normalization (Ioffe & Szegedy, 2015).

From a learning-theoretic perspective, generalization behavior can be interpreted through classical statistical learning theory (Vapnik, 1998) and recent kernel-limit analyses of deep networks (Jacot et al., 2018). In application-driven settings such as remote sensing, hybrid spatial–frequency learning is closely related to earlier fully convolutional building extraction frameworks (Ji et al., 2018), while robustness concerns connect naturally to dataset shift analysis (Quiñonero-Candela et al., 2009).

Dimensionality reduction viewpoints are consistent with autoencoder-based representations (Hinton & Salakhutdinov, 2006) and knowledge distillation principles (Hinton et al., 2015), particularly when wavelet decompositions act as structured compression operators. The automated design directions proposed in this work are further motivated by prior advances in neural architecture search (Elsken et al., 2019). Finally, wavelet-based denoising and thresholding mechanisms discussed here are theoretically aligned with classical wavelet shrinkage formulations (Donoho & Johnstone, 1994).

6. Future Research Directions

Based on the systematic analysis of current wavelet-deep learning literature, we propose two principled research directions addressing gaps. These directions span automated design and, interpretability. .

6.1 Automated Wavelet Selection and Neural Architecture Search

Manual selection of wavelet families $\psi \in \{\psi_{\text{Haar}}, \psi_{\text{db4}}, \dots\}$, decomposition levels $J \in \{1, \dots, J_{\max}\}$, and integration strategies limits scalability and generalizability. Extending neural architecture search (NAS) to jointly optimize wavelet parameters and network architectures can be formulated as:

$$\min_{\alpha,\theta} \mathcal{L}_{\text{val}}\big(f_{\text{NN}}(\Phi_\alpha(x);\theta)\big) + \lambda R(\alpha,\theta) \quad (31)$$

where $\alpha \in \mathcal{A}$ parameterizes the wavelet configuration space (family, decomposition depth $J(\alpha)$, integration strategy), $\Phi_\alpha: \mathbb{R}^N \to \mathbb{R}^{M(\alpha)}$ is the wavelet feature extraction operator, $\theta \in \Theta$ are network parameters, and $R(\alpha, \theta)$ penalizes model complexity.

Differentiable search: Wavelet selection as continuous variables via weighted combinations $\psi_\alpha(t) = \sum_{k=1}^{K} \text{softmax}\,(\alpha)_k \cdot \psi_k(t)$ enables gradient-based optimization.

Adaptive depth: Instance-specific decomposition levels through $\Phi_{\text{adaptive}}(x) = \sum_{j=1}^{J_{\max}} w_j\,(x) \cdot \Phi_j(x)$ where $w_j(x) = \text{softmax}\left(g\big(h(x)\big)\right)_j$ are learned gates.

Learned wavelets: Task-specific synthesis via $\min_{\psi_\theta} \mathbb{E}_{(x,y)}\left[\mathcal{L}\left(y, f_{\text{NN}}\left(\Phi_{\psi_\theta}(x)\right)\right)\right]$ subject to admissibility $\int \psi_\theta(t)\,dt = 0$ and orthogonality $\int \psi_\theta(t)\psi_\theta(t-k)\,dt = \delta_{k,0}$.

6.2 Interpretable Wavelet-Deep Learning Models

While wavelet coefficients possess inherent interpretability through explicit time-frequency localization—detail coefficients $d_j[k]$ at scale j correspond to frequency band $\left[2^{-(j+1)}f_s, 2^{-j}f_s\right]$—their interaction with deep nonlinear transformations obscures decision-making processes, critical for safety-critical domains.

Wavelet attention: Learnable attention weights identify influential scales and positions: $\alpha_{j,k} = \exp\big(e_{j,k}\big)/\sum_{j',k'} \exp\,\big(e_{j',k'}\big)$ where $e_{j,k} = \mathbf{v}^T \tanh\big(W_1 d_j[k] + W_2 h_{j,k}\big)$, yielding attended representation $\tilde{c} = \sum_{j,k} \alpha_{j,k} \cdot d_j[k]$. Multi-head extensions enable diverse pattern capture: $\tilde{c}^{(h)} = \sum_{j,k} \alpha_{j,k}^{(h)} \cdot W_V^{(h)} d_j[k]$.

Frequency-domain saliency: Extending gradient-based visualization to wavelet space via $S_{j,k} = \left|\frac{\partial \mathcal{L}}{\partial c_j[k]}\right|$ or integrated gradients $S_{j,k} = \int_0^1 \frac{\partial f(\alpha \cdot c)}{\partial c_j[k]}\big|_{c=\alpha \cdot c^{(0)}} d\alpha$. Class activation mapping: $S_{j,k}^{(c)} = \text{ReLU}\left(\sum_i w_i^{(c)} \frac{\partial A_i}{\partial c_j[k]}\right)$.

Symbolic rule extraction: Decision tree approximation $g\big(\Phi(x)\big) = \sum_{i=1}^{L} y_i \cdot \mathbb{1}_{\Phi(x) \in \mathcal{R}_i}$ partitions wavelet feature space into interpretable regions, enabling clinical rules like "If $\| c_J \|_2^2 > \tau_1$ AND $\| d_3 \|_2^2 < \tau_2$ in window $[t_1, t_2]$, predict arrhythmia."

The two research directions automated design and interpretability form a brief roadmap for advancing wavelet-deep learning.

7. Conclusion

This study has examined the integration of wavelet transforms with deep learning models by analyzing their mathematical structure, architectural integration strategies, and applications across multiple domains. Rather than treating wavelets solely as preprocessing operators, the analysis emphasizes their role as structural components within neural architectures, where multiresolution representations support the effective modeling of non-stationary and scale-dependent signals.

Wavelet transforms provide a principled multiresolution decomposition that simultaneously captures temporal and frequency information, addressing limitations of conventional deep learning approaches when applied to non-stationary data. Several integration strategies have been identified, including feature-based integration, architectural replacement of pooling or downsampling operations, preprocessing-based approaches, hybrid designs, and loss-level incorporation. Empirical evidence indicates that wavelet-enhanced architectures can improve approximation efficiency, noise robustness, and parameter efficiency in tasks such as time-series forecasting, medical image segmentation, remote sensing, and healthcare diagnostics.

Despite these advantages, important challenges remain, including the complexity of wavelet family and decomposition-level selection, computational trade-offs introduced by multiscale representations, and limited theoretical understanding of how wavelet decompositions interact with learned representations. In particular, rigorous analyses of approximation properties, generalization behavior, and optimization landscapes for wavelet-enhanced networks are still lacking, as is large-scale validation in safety-critical domains.

Future research directions include automated wavelet and architecture selection through neural architecture search, the development of interpretable wavelet–deep learning models, optimization for real-time and edge deployment, multimodal wavelet-based fusion, hybrid wavelet–transformer architectures, and domain-specific wavelet design subject to constraints such as orthogonality and perfect reconstruction. Progress in these areas is expected to strengthen the theoretical foundations of wavelet–deep learning integration and enable more principled model design.

From an applied perspective, the choice of integration strategy should be guided by problem structure: feature-based approaches are effective for time-series data with pronounced multiscale behavior, architectural integration is well suited to image-processing tasks requiring noise suppression and structural preservation, and hybrid strategies are advantageous for complex problems benefiting from both explicit decomposition and learned representations. Overall, this study provides a concise framework for understanding and advancing wavelet-based deep learning models, with the potential for broader adoption as theoretical insights and automated design methodologies mature.

References

1. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 801-818).

2. Cohen, A., Daubechies, I., & Feauveau, J. C. (1992). Biorthogonal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics, 45*(5), 485-560. https://doi.org/10.1002/cpa.3160450502
3. Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics, 41*(7), 909-996. https://doi.org/10.1002/cpa.3160410705
4. Daubechies, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970104
5. Donoho, D. L., & Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika, 81*(3), 425-455. https://doi.org/10.1093/biomet/81.3.425
6. Elsken, T., Metzen, J. H., & Hutter, F. (2019). Neural architecture search: A survey. *Journal of Machine Learning Research, 20*(55), 1-21.
7. Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen, 69*(3), 331-371. https://doi.org/10.1007/BF01456326
8. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90
9. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313*(5786), 504-507. https://doi.org/10.1126/science.1127647
10. Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
11. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
12. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7132-7141). https://doi.org/10.1109/CVPR.2018.00745
13. Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)* (pp. 448-456).
14. Jacot, A., Gabriel, F., & Hongler, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 8571-8580).
15. Ji, S., Wei, S., & Lu, M. (2018). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing, 57*(1), 574-586. https://doi.org/10.1109/TGRS.2018.2858817
16. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278-2324. https://doi.org/10.1109/5.726791
17. Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*(7), 674-693. https://doi.org/10.1109/34.192463
18. Meyer, Y. (1990). Ondelettes et opérateurs I: Ondelettes. *Actualités Mathématiques*. Hermann.

19. Nason, G. P., & Silverman, B. W. (1995). The stationary wavelet transform and some statistical applications. In A. Antoniadis & G. Oppenheim (Eds.), *Wavelets and Statistics* (pp. 281-299). Springer. https://doi.org/10.1007/978-1-4612-2544-7_17
20. Qiu, S., Zhou, J., & Huang, X. (2025). WaveFuseNet: A multiscale wavelet-transform convolutional network with adaptive feature fusion for building extraction. *GIScience & Remote Sensing, 62*(1), 2536903. https://doi.org/10.1080/15481603.2025.2536903
21. Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (2009). *Dataset Shift in Machine Learning*. MIT Press.
22. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (pp. 234-241). https://doi.org/10.1007/978-3-319-24574-4_28
23. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618-626). https://doi.org/10.1109/ICCV.2017.74
24. Shuvo, S. B., Alam, S. S., Ayman, S. U., Chakma, A., Salvi, M., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2025). Application of wavelet transformation and artificial intelligence techniques in healthcare: A systemic review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 15*, e70007. https://doi.org/10.1002/widm.70007
25. Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.
26. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)* (pp. 5998-6008).
27. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3-19). https://doi.org/10.1007/978-3-030-01234-2_1
28. Yemets, K., Izonin, I., & Mitoulis, S. A. (2025). Multi-family wavelet-based feature engineering method for short-term time series forecasting. *Scientific Reports, 15*, 39081. https://doi.org/10.1038/s41598-025-26685-8
29. Zhao, Y., Wang, S., Zhang, Y., Qiao, S., & Zhang, M. (2023). WRANet: Wavelet integrated residual attention U-Net network for medical image segmentation. *Complex & Intelligent Systems, 9*, 6971-6983. https://doi.org/10.1007/s40747-023-01119-y
30. Qiu, S., Zhou, J., & Huang, X. (2025). WaveFuseNet: A multiscale wavelet-transform convolutional network with adaptive feature fusion for building extraction. *GIScience & Remote Sensing, 62*(1), 2536903. https://doi.org/10.1080/15481603.2025.2536903
31. Shuvo, S. B., Alam, S. S., Ayman, S. U., Chakma, A., Salvi, M., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2025). Application of wavelet transformation and artificial intelligence techniques in healthcare: A systemic review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 15*, e70007. https://doi.org/10.1002/widm.70007
32. Yemets, K., Izonin, I., & Mitoulis, S. A. (2025). Multi-family wavelet-based feature engineering method for short-term time series forecasting. *Scientific Reports, 15*, 39081. https://doi.org/10.1038/s41598-025-26685-8
33. Zhao, Y., Wang, S., Zhang, Y., Qiao, S., & Zhang, M. (2023). WRANet: Wavelet integrated residual attention U-Net network for medical image segmentation. *Complex & Intelligent Systems, 9*, 6971–6983. https://doi.org/10.1007/s40747-023-01119-y