JOURNAL OF RENEWABLE ENERGY AND SMART GRID TECHNOLOGY

Vol. 19, No. 1, January-June 2024

Recognition of Multiple Power Quality Disturbances Based on Discrete Wavelet Transform and Improved Long Short-Term Memory Networks

Supakan Janthong[1,2], Pornchai Phukpattaranont[1*]

[1]Department of Electrical & Biomedical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla 90110, Thailand
[2]Meter Department, Provincial Electricity Authority, Ranot District, Songkhla 90140, Thailand
***Corresponding author's email**: pornchai.p@psu.ac.th

Article info:

Received: 24 February 2024
Revised: 24 April 2024
Accepted: 20 May 2024
DOI:
10.69650/rast.2024.255814

Keywords:

Recognition
Multiple Power Quality Disturbances
Discrete Wavelet Transform
Improved Long Short-Term Memory Networks

ABSTRACT

In the past few years, distributed electricity generation from renewable sources, or microgrid systems, has been connected to the grid to increase power supply stability. This responds to government policy regarding commitment to using 100% renewable energy in operations (RE100) project efforts. This results in the entry of power electronic or non-linear equipment into the electrical system, making it more sensitive. Moreover, multiple power quality disturbances (PQDs) consist of a variety of single disturbances. Analysis of complex multi-label patterns is a challenging task. In this paper, we propose a methodology to address this challenge by leveraging Discrete Wavelet Transform (DWT) and Improved Long Short-Term Memory Networks (LSTM). Firstly, multiple PQDs are synthesized utilizing a mathematical model based on IEEE standards 1159-2019. Secondly, the obtained PQDs are decomposed into nine feature classes, yielding detailed (cDs) and approximation (cAs) coefficients through Five-Level DWT Decomposition. Furthermore, we conducted a comparative analysis of each component across five different wavelet functions: haar1, db4, bior1.3, coif2, and sym4. Thirdly, the cDs and cAs coefficients derived from each wavelet type undergo statistical analysis before being inputted into the LSTM model for classification of each feature class. Our results highlight that cD5 components obtained from the db4 wavelet exhibit the highest accuracy rate of 93.86%. This finding elucidates the significance of selecting appropriate wavelet types and compositions for the successful classification of multiple PQDs.

1. Introduction

1.1 Background

Presently, power quality inspections are conducted across various segments of the electrical system, encompassing the generation, transmission, and distribution systems, all of which hold significant importance. The proliferation of the modern energy industry, coupled with the widespread use of electronic equipment, has emerged as a common catalyst for disruptions within the system [1]. Furthermore, the integration of renewable energy sources and distributed generation in Thailand has introduced new challenges, as these sources often introduce fluctuations and intermittencies into the power grid [2]. This alongside the operation of industrial power equipment, exacerbates the occurrence of Power Quality Disturbances (PQDs). Without adequate monitoring and preventive measures, these disturbances can lead to unforeseen events such as equipment malfunctions, prolonged blackouts affecting widespread regions, or even pose risks to field mechanics [3]. In this context, the analysis and classification of PQDs become imperative, as they provide early warning signs and enable proactive interventions to mitigate potential risks. It's worth noting that PQ standards, such as those outlined in IEEE 1159-2019 [4], offer comprehensive guidelines for measuring electromagnetic phenomena associated with variations in voltage, current, and frequency resulting from changes in power supplies and loads. Adherence to these standards not only ensures the reliability and stability of the power grid but also facilitates the early detection and characterization of PQDs, thereby enhancing operational efficiency and minimizing downtime. Therefore, prioritizing the detection and classification of PQDs characteristics is paramount for maintaining a resilient and sustainable electrical infrastructure in Thailand [5].

Power quality encompasses the intricate interplay between electrical power and the performance of electrical equipment. When electrical equipment operates reliably without damage or stress, we deem the electrical power to be of good quality. Conversely, if equipment malfunctions, proves unreliable, or suffers damage during regular operation, we infer poor power quality.

Diving into the intricacies, power quality comprises several facets including voltage quality, current quality, supply quality, and

consumption quality [5]. It can be succinctly defined as any deviation in voltage, current, or frequency that leads to equipment failure or malfunction. Poor power quality often lurks as the primary culprit behind perplexing equipment trips, intermittent shutdowns, sporadic damage or component failures, erratic process performance, and inexplicable occurrences such as random lockups and data errors or overheating of power system components [5]. The spectrum of power quality problems causes a discernible deterioration in the performance of various sensitive electronic and electric equipment. An ideal scenario of good power quality entails several key attributes: Supply voltage staying within the guaranteed tolerance of its declared value; A waveform exhibiting a pure sine wave shape within permissible distortion limits; Voltage balance across all three phases; And continuous supply, ensuring uninterrupted availability. However, causes of poor power quality abound: Variations in voltage, magnitude, and frequency [5]. Frequency variations stem from system dynamics or harmonics injection. Consequently, the pristine sinusoidal nature of power system voltage or current waveforms dissipates, replaced by the presence of harmonics and other noise elements [5].

Fig. 1 (a) Suspension insulators damaged by arc burns, (b) Arrester explodes from overvoltage, (c) Voltage stress on the bushing, and (d) Damaged cable caused by flash overs.

PQDs manifest in various forms, each with distinct characteristics and impacts [6]: Transients, marked by instantaneous, nanosecond-range spikes in voltage; Interruptions, occurring when supply voltage or load current drops below 0.1 pu for a period not exceeding 1 minute; Voltage sags, where the RMS voltage dips between 10% and 90% of nominal voltage for half a cycle to a minute; Voltage swells, characterized by an increase in RMS voltage to 110%-180% of nominal voltage for up to 1 minute; Waveform distortion, signifying unexpected alterations in current and voltage waveforms as they traverse through a device. These PQDs events, when left unchecked, wreak havoc on electrical systems. For instance [6], voltage sags can disrupt manufacturing processes, causing equipment downtime and production losses (Fig.1 (a)). Transients may fry sensitive electronic components, leading to costly replacements (Fig.1 (b) and (c)). Voltage swells might damage equipment designed for specific voltage ranges, rendering them inoperable (Fig.1 (d)). Such disturbances not only incur financial losses but also compromise safety and reliability, underscoring the critical importance of mitigating PQD effects [6].

1.2 Literature Review

In the scope of power quality research, various methodologies have been proposed to detect and classify PQDs. These methodologies can be broadly categorized into three main approaches [7]: signal processing-based feature extraction, artificial intelligence-based classifiers, and optimization techniques for optimal feature selection.

Signal processing-based methods often utilize techniques such as wavelet transforms to extract relevant features from voltage or current signals. For example, Upadhyaya et al. introduced a novel approach using Second Generation Wavelet Transform (SGWT) to detect ten different PQ formats [8]. Their study demonstrated that SGWT outperformed traditional techniques like Discrete Wavelet Transform (DWT) in terms of both speed and effectiveness. However, the paper lacked detailed evaluation of the results and explanation of the detection process, which limits the reproducibility and understanding of their findings. Similarly, the study in [9] employed wavelet transform extraction, including Continuous Wavelet Transform (CWT), DWT, and Multi-Resolution Analysis (MRA), followed by detection using Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs). While the approach showed promise, there was insufficient information provided on how the PQDs model was generated or the parameters used in the classification model, hindering a comprehensive understanding of the methodology.

In other studies, such as [10] and [11], the focus was on analyzing harmonic components, yet there was a lack of clarity regarding the source of the data, or the criteria used for determining the results. Furthermore, Arvez et al. [12] utilized the Simulink toolbox in MATLAB and DWT with level 2 components, such as Low-Pass Filters and High-Pass Filters, to extract features like Root Mean Square (RMS), average, standard deviation, and approximation coefficients. They then applied for a one-class Support Vector Machine (OCSVM) for classification, achieving high accuracy. However, the authors did not provide an explanation of how data from smart meters was utilized, which is crucial for understanding the applicability of their approach in real-world scenarios. Moreover, studies in [13] and [14] employed Stockwell's Transform (ST) for data extraction, focusing specifically on PQDs related to wind energy according to IEEE standards. However, the lack of clarity regarding the classifier and its evaluation methodology poses limitations to their findings, making it challenging to assess the robustness and generalizability of their results.

From the above mentioned, several observations can be made:

(1) PQDs data generation lacks uniformity, with varying resolution values derived from mathematical equations and laboratory data. This inconsistency in data generation methods can lead to discrepancies in results and hinder comparison between different studies.
(2) Extraction techniques vary based on their intended objectives, necessitating a standardized approach. The lack of standardization makes it difficult to assess the efficacy and generalizability of different extraction techniques and hampers the reproducibility of results.
(3) While classifiers play a crucial role in the characterization process, their effectiveness is contingent upon adequate evaluation. However, many studies fail to provide

comprehensive evaluation methodologies, making it challenging to gauge the reliability and robustness of their proposed models.

(4) Evaluation methodologies are essential for substantiating the efficacy of proposed methods but are often overlooked or inadequately described. The absence of clear evaluation criteria and metrics makes it difficult to assess the performance of different approaches.

The present study's primary contributions are outlined below:

(1) Introduction of multiple PQDs generated through mathematical equations to enable the model to learn diverse interference patterns. This approach enhances the model's ability to accurately detect and classify a wider range of PQDs.

(2) Utilization of DWT to categorize elements into levels, facilitating focused analysis of essential components. By categorizing elements into distinct levels, the model can better isolate and prioritize relevant features, improving the efficiency and effectiveness of the classification process.

(3) Implementation of Long Short-Term Memory Networks (LSTM) to incorporate sequential information, enhancing accuracy in PQDs classification. By analyzing data recurring directions, the model can capture temporal dependencies and patterns of disturbance waveforms.

(4) Adoption of accuracy metrics and confusion matrix for comprehensive assessment of the model's performance, ensuring robust evaluation of results.

The remainder of this paper is structured as follows: Section 2 delineates the theoretical framework and methodologies, encompassing decomposition using DWT, Five-Level DWT decomposition, Wavelet Families, and LSTM. Section 3 elucidates the experimental procedures, while Section 4 offers a succinct overview of the results and subsequent discussion, including any identified limitations. Finally, Section 5 provides a summary of the research findings and outlines potential avenues for future investigations.

2. Theory and Methods

In the endeavor to classify PQDs effectively, it is imperative to delve into the intricate workings of signal processing techniques, particularly decomposition using DWT. DWT is a powerful tool for analyzing non-stationary signals, allowing for the extraction of valuable features that are crucial for accurate classification. Moreover, classifiers and evaluation tools are elaborated upon in detail within this section.

2.1 Decomposition using DWT

DWT serves as the cornerstone of feature extraction in PQD classification [15]. The process involves breaking down the input signal into different frequency bands or levels, each representing a specific scale of detail. This hierarchical decomposition enables the identification of relevant features at various resolutions, facilitating the characterization of PQD with different temporal and spectral characteristics.

DWT decomposes a signal into approximation and detail coefficients at different resolution levels. The DWT operation can be expressed as:

$$x(t) = \sum_k c_k(t) + \sum_j d_j(t) \tag{1}$$

where $x(t)$ represents the original signal, $c_k(t)$ represents the approximation coefficients level k, $d_j(t)$ represents the detail coefficients at level j. The number of decomposition levels determines the resolution and frequency bands obtained from the decomposition.

2.2 Five-Level DWT Decomposition

The input signal is decomposed into approximation and detail coefficients at a five-resolution level [16]. This decomposition provides a coarse representation of the signal's overall trend (approximation coefficient) and its high-frequency components (detail coefficients) as shown in Fig. 2. In five-level DWT decomposition, the input signal is decomposed into approximation (smooth) and detail (detail) coefficients at a single resolution level j. The decomposition can be represented as:

$$x(t) = c_1(t) + d_1(t) \tag{2}$$

Here, $c_1(t)$ represents the approximation coefficients, capturing the low-frequency components of the signal, while $d_1(t)$ represents the detail coefficients, highlighting the high-frequency components.

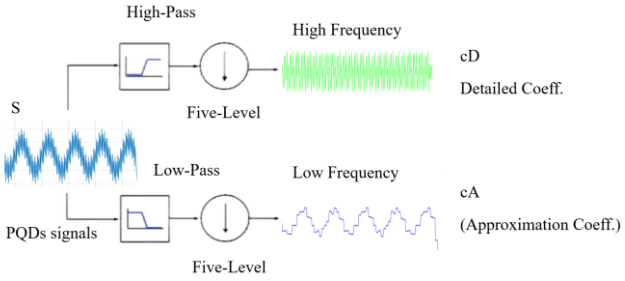


Fig. 2 Five-level DWT decomposition.

2.3 Introduction to Wavelet Families

Several wavelet families are commonly employed in PQD classification, each possessing unique properties and suitability for different types of signals [17]. Some of the prominent wavelet families include:

(1) Haar Wavelet:

The Haar wavelet is the simplest wavelet function, characterized by its step-like waveform [18]. It is particularly well-suited for detecting sudden changes or discontinuities in signals, making it an ideal choice for capturing transient PQD events.

The Haar wavelet is defined by its scaling function $\varphi(t)$ and wavelet function $\psi(t)$, given by:

$$\varphi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & otherwise \end{cases} \tag{3}$$

The wavelet function $\psi(t)$ is defined as follows:

$$\psi(t)=\begin{cases}1 & 0\le t<\frac{1}{2},\\ -1 & \frac{1}{2}\le t<1,\\ 0 & otherwise\end{cases} \quad (4)$$

The Haar wavelet effectively captures sudden changes or discontinuities in signals due to its step-like waveform.

(2) Daubechies Wavelet:

Daubechies wavelets, also known as db wavelets, are widely used due to their orthogonality and compact support properties. They offer excellent time-frequency localization, making them suitable for analyzing signals with both short and long-term variations, such as voltage sags and swells [19]. For example, the Daubechies-4 (db4) wavelet has 4 vanishing moments and is widely used in signal processing applications.

(3) Biorthogonal Wavelet:

Biorthogonal wavelets are characterized by their biorthogonality property, which allows for a more flexible decomposition of signals compared to orthogonal wavelets. They are particularly useful for analyzing signals with non-symmetric features or complex dynamics, such as harmonic distortions [19]. The biorthogonality property allows for a more flexible decomposition of signals compared to orthogonal wavelets.

(4) Coiflets Wavelet:

Coiflets, or coif wavelets, are designed to provide a smoother transition between approximation and detail coefficients, making them suitable for signals with gradual variations or smooth trends [20]. They offer a balance between time and frequency localization, making them versatile for analyzing various types of PQD.

(5) Symlets Wavelet:

Symlets, or symmlets, are similar to Daubechies wavelets but offer improved symmetry properties, making them better suited for analyzing signals with symmetric features or periodic components [20]. They provide efficient representation of signals with both smooth and oscillatory characteristics, making them a popular choice for PQD classification tasks.

The five types of wavelet shape features used in this experiment are shown in Table 1. Incorporating these wavelet families into the classification framework allows for the extraction of discriminative features from PQD signals, enabling accurate and robust classification of different disturbance events. Moreover, understanding the working principles, structures, and equations involved in each wavelet family empowers researchers to tailor their classification approach to the specific characteristics of the signals under analysis, thereby enhancing the effectiveness and reliability of the classification system.

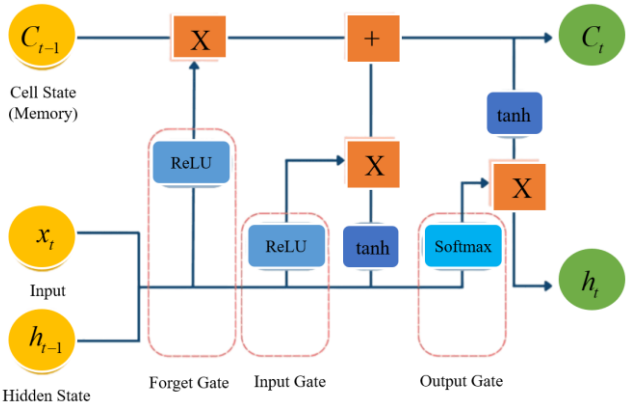


Fig. 3 LSTM Architecture used in this paper.

2.4 Long Short-Term Memory (LSTM)

Section Long Short-Term Memory (LSTM) networks have emerged as powerful tools for sequential data analysis, including the characterization and classification of PQDs [21]. LSTM belongs to the family of Recurrent Neural Networks (RNNs), specialized for handling sequential data with long-range dependencies. Unlike traditional feedforward neural networks, RNNs have feedback connections that allow them to process sequences of inputs by maintaining internal states. LSTM networks address the vanishing gradient problem encountered in standard RNNs, making them better suited for capturing long-term dependencies in sequential data as shown in Fig. 3.

The fundamental unit of an LSTM network is the LSTM cell, which contains multiple gates responsible for regulating the flow of information. These gates include:

1. Forget Gate: Controls what information from the previous cell state should be discarded.

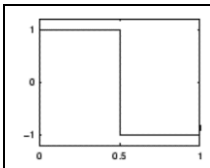
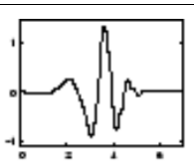
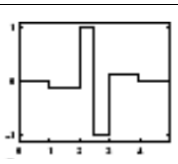
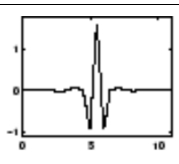
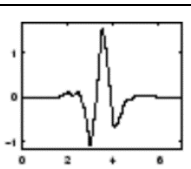
2. Input Gate: Determines what new information should be added to the cell state.

3. Output Gate: Regulates the information that will be output from the cell state.

The LSTM cell utilizes these gates to selectively update its internal state based on the input sequence, allowing it to retain relevant information over long time intervals while discarding irrelevant or redundant information.

The computations within an LSTM cell can be described by the following equations:

Table 1 Visualize the wavelets included in this work.

				
haar1	db4	bior1.3	coif2	sym4
Haar Wavelet	Daubechies Wavelet	Biorthogonal Wavelet	Coiflets Wavelet	Symlets Wavelet

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (7)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \quad (8)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (10)$$

Where:

f_t, i_t and o_t are the forget, input, and output gate activations, respectively.

$\widetilde{C}_t$ represents the candidate cell state.

C_t is the updated cell state.

h_t is the output of the LSTM cell.

2.5 Evaluating Performance

(1) Accuracy:

Accuracy is a commonly used metric for evaluating the performance of classification models, including those using LSTM for PQD classification. It measures the proportion of correctly classified instances among the total number of instances in the dataset [22]. Mathematically, accuracy is defined as:

$$\text{Accuracy} = \frac{\text{Number of Correctly Classified Instances}}{\text{Total Number of Instances}} \times 100\% \quad (11)$$

The context of LSTM-based PQD classification, accuracy indicates the model's ability to correctly classify different types of PQDs based on the input features extracted using DWT.

(2) Confusion Matrix:

A confusion matrix is a tabular representation of the actual and predicted classifications produced by a classification algorithm [22]. It provides valuable insights into the model's performance by summarizing the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). For example, these values are arranged in a 2x2 matrix as follows:

	Predicted Negative	Predicted Positive
Actual Negative	TN	FP
Actual Positive	FN	TP

Each cell in the confusion matrix represents a specific classification outcome, such as the TP value indicates the number of instances that is correctly classified (i.e., power quality disturbances).

In addition to the accuracy measure, to ensure a comprehensive assessment of the model's performance when applied to real-world scenarios, precision, recall, and F1-score are also employed to evaluate the model's effectiveness, as presented in Equations (12)-(15).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$F1 - score = \frac{2TP}{2TP + FP + FN} \quad (15)$$

3. Experimental detail

The experimental procedure encompasses four primary phases: data preparation, feature extraction, classification of PQDs, and model evaluation. These phases are elucidated in Fig. 4. The subsequent sections will explicate each stage in the following sequence.

3.1 Preprocessing

In this study, our primary focus lies in the identification of intricate PQDs patterns, with particular emphasis on multiple PQDs. These phenomena emerge as a result of the amalgamation or interaction of multiple patterns, a scenario often encountered due to the diverse nature of loads and the integration of alternative energy sources.

Table 2 illustrates the classification of each PQDs class model, meticulously adhering to the IEEE-1159-2019 standard guidelines [4]. The classification process is facilitated by an integral mathematical model implemented through MATLAB, a widely utilized computational tool in this experimental setup. Within the confines of this table, individual PQD events (designated as C-1 to C-9) are delineated based on distinct characteristics such as waveform magnitude, duration, and morphology [24]. Additionally, leveraging the aforementioned mathematical framework, we construct graphical representations, as elucidated in Table 3.

It is imperative to note that the data encapsulated within each PQD class undergoes a randomization process, employing specified parameter ranges to ensure robustness and comprehensiveness in our analysis. As part of our research methodology, we meticulously generated a dataset comprising 900 instances, each meticulously curated to encompass 100 samples, thus ensuring ample coverage and representation across various PQD scenarios.

3.2 Feature Extraction

After completing the data preparation phase, the distinctive attributes of each PQDs class were extracted utilizing DWT. For this purpose, we employed a MATLAB application named 'Wavelet Analyzer' [25]. The configuration parameters were set as follows: the dataset size was fixed at 500 samples, and the Wavelet function comprised several types including haar1, db4, bior1.3, coif2, and sym4. Additionally, a constant level value of 5 was established. During the experimentation, we iteratively adjusted the loop value for each Wavelet function format and subsequently recorded the resulting values from the decomposition process (in Algorithm 1). This decomposition encompassed approximations at levels 1 to 5 and coefficients of approximations at levels 1 to 5. A representative signal characteristic is illustrated in Fig. 5.

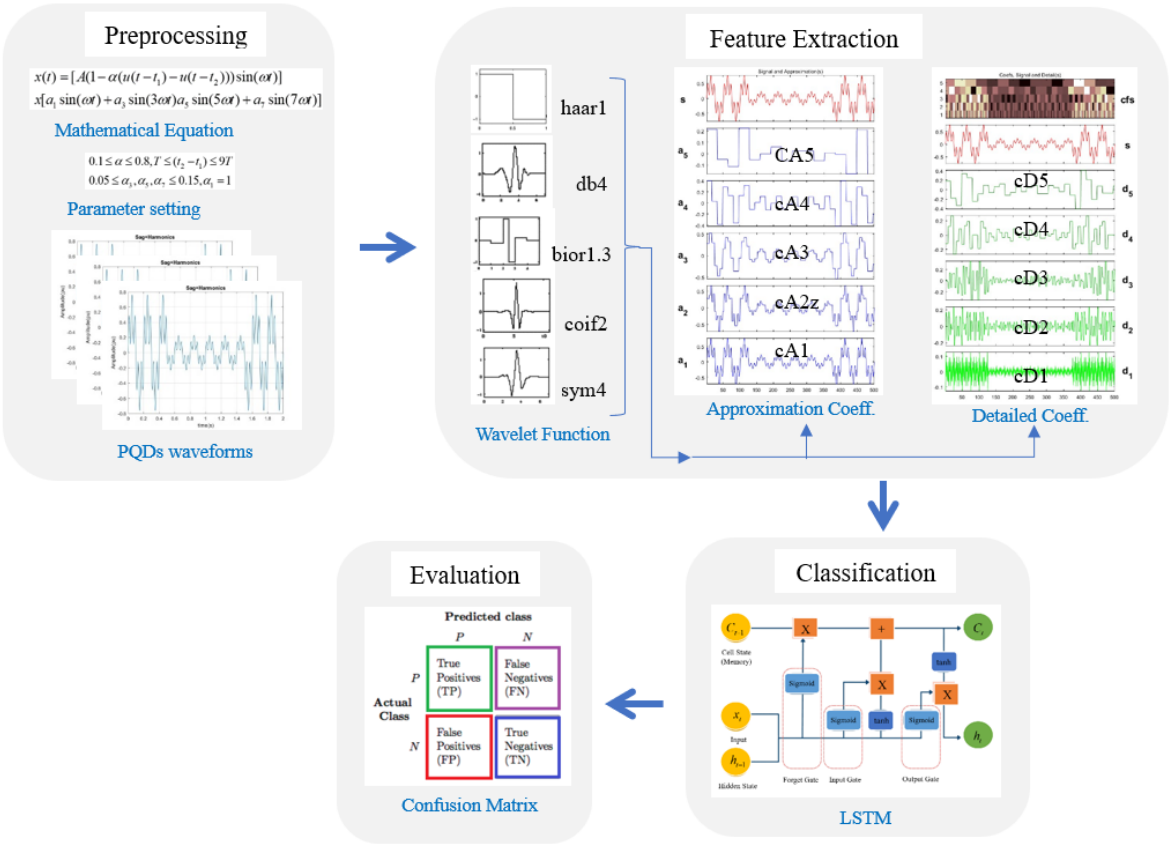


Fig. 4 Overview of the proposed methods.

Table 2 Categories of Multiple PQDs.

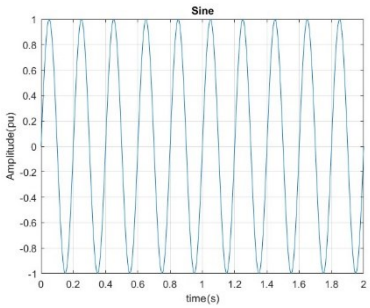
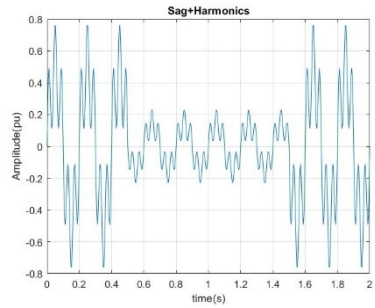
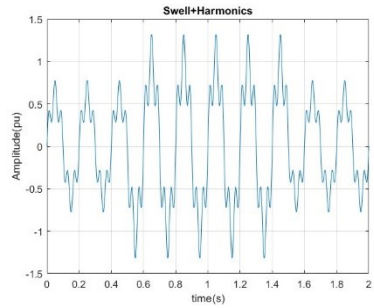
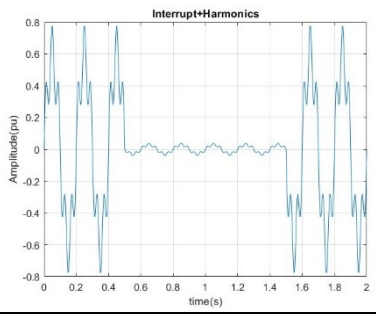
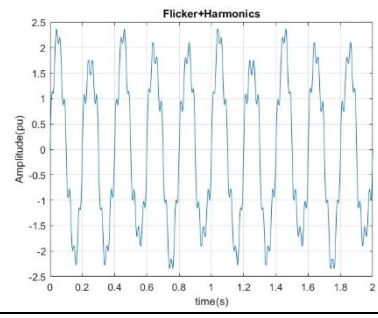
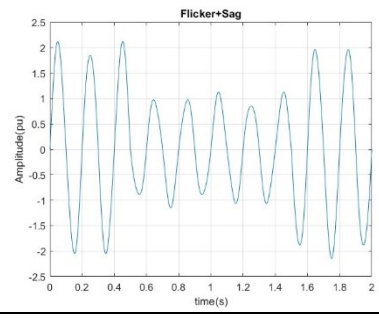
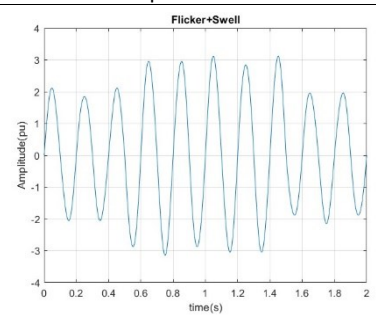
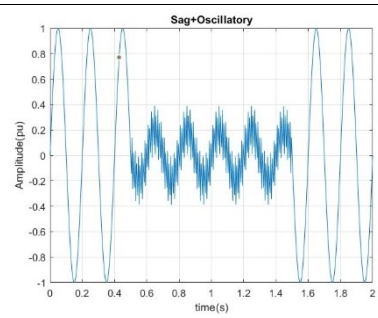
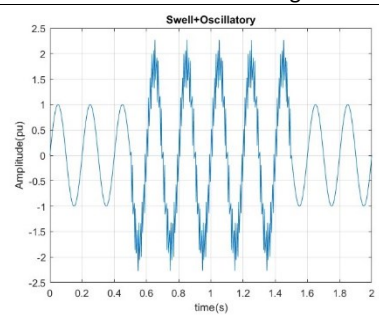
Class	PQD Event	Mathematical Model	Parameter setting
C-1	Pure Sinusoidal	$x(t)=A\sin(\omega t)$	$0.9<A<1.1$
C-2	Sag with Harmonics	$x(t)=[A(1-\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)]$ $x[a_1\sin(\omega t)+a_3\sin(3\omega t)a_5\sin(5\omega t)+a_7\sin(7\omega t)]$	$0.1\le\alpha\le0.8, T\le(t_2-t_1)\le9T$ $0.05\le\alpha_3,\alpha_5,\alpha_7\le0.15,\alpha_1=1$
C-3	Swell with Harmonics	$x(t)=[A(1+\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)]$ $x[a_1\sin(\omega t)+a_3\sin(3\omega t)a_5\sin(5\omega t)+a_7\sin(7\omega t)]$	$0.1\le\alpha\le0.9, T\le(t_2-t_1)\le9T$ $0.05\le\alpha_3,\alpha_5,\alpha_7\le0.15,\alpha_1=1$
C-4	Interruption with Harmonics	$x(t)=[A(1-\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)]$ $x[a_1\sin(\omega t)+a_3\sin(3\omega t)a_5\sin(5\omega t)+a_7\sin(7\omega t)]$	$0.9\le\alpha\le1, T\le t_2-t_1\le9T,$ $0.05\le\alpha_3\le0.15, 0.05\le\alpha_5\le0.15$
C-5	Flicker with Harmonics	$x(t)=[1+\lambda\sin(\beta\omega t)]\times[a_1\sin(\omega t)+$ $a_3\sin(3\omega t)+a_5\sin(5\omega t)+a_7\sin(7\omega t)]$	$0.1\le\lambda\le0.2, 5\le\beta\le20$ $0.05\le\alpha_3,\alpha_5,\alpha_7\le0.15,\alpha_1=1$
C-6	Flicker with Sag	$x(t)=[1+\lambda\sin(\beta\omega t)]\times[1-\alpha(u(t-t_1)-u(t-t_2))]\sin(\omega t)$	$0.1\le\lambda\le0.2, 5\le\beta\le20$ $0.1\le\alpha\le0.8, T\le(t_2-t_1)\le9T$
C-7	Flicker with Swell	$x(t)=[1+\lambda\sin(\beta\omega t)]\times[1+\alpha(u(t-t_1)-u(t-t_2))]\sin(\omega t)$	$0.1\le\alpha\le0.2, 5\le\beta\le20$ $0.1\le\alpha\le0.9, T\le(t_2-t_1)\le9T$
C-8	Sag with Oscillatory	$x(t)=[A(1-\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)]$ $+a_t\exp(-\frac{t-t_1}{\tau})(u(t-t_1)-u(t-t_2))$	$0.1\le\alpha\le0.8, 0.5T\le(t_2-t_1)\le9T$ $0.1\le\alpha_t\le4, 8ms\le\tau\le30ms$
C-9	Swell with Oscillatory	$x(t)=[A(1+\alpha(u(t-t_1)-u(t-t_2)))\sin(\omega t)]$ $+a_t\exp(-\frac{t-t_1}{\tau})(u(t-t_1)-u(t-t_2))$	$0.1\le\alpha\le0.8, 0.5T\le(t_2-t_1)\le9T$ $0.1\le\alpha_t\le4, 8ms\le\tau\le30ms$

Where: Fundamental frequency (f) =50Hz, $\omega=2\pi f$, Sampling frequency $(f_s)=3.2\,kHz$, Time duration of each event $(T)=0.6\sec$, parameters such as α,β control the variations, amplitudes of harmonic components $\alpha_3,\alpha_5,\alpha_7$ an exponential variation controlled by τ,

Algorithm 1: DWT Extraction	
1.	**Define Function:**
1.1	Extract_features(data, wavelet_family, wavelet_level)
2.	**Initialize Features List:**
2.1	Store the extracted features for all samples in the data.
3.	**Add Pre-generated Waveform (Modify as needed):**
3.1	Take a pre-generated waveform
4.	**Loop Through Data Samples:**
4.1	Iterate through each sample in the input data using a for loop
5.	**Perform DWT on Each Sample:**
5.1	Within the loop, perform DWT on the current sample using pywt.dwt function.
5.2	Specify the chosen wavelet family based on the input parameter (wavelet_family)
5.3	Store the resulting DWT coefficients in a variable named coeffs
6.	**Decompose Coefficients Level:**
6.1	Loop from level 1 to 5 wavelet_level (In each iteration)
6.1.2	Separate the approximation (approx) and detail (detail) coefficients from coeffs.
6.1.3	Update coeffs to hold the detail coefficients in the next level
6.1.4	Append both approx and detail coefficients to the decomposed_coeffs list
7.	**Combine Features**
7.1	Calculate standard deviation (np.std(sample)) and mean (np.mean(sample))
8.	**Return Features:**

Algorithm 2: LSTM Classification	
1.	**Data Loading and Preprocessing**
1.1	X_data: stores the preprocessed features (DWT_features)
1.2	y_data: stores the corresponding labels for the data
2.	**Define Number of Folds:**
2.1	Define a variable: num_rounds (set 9 classes)
3.	**K-Fold Cross-Validation:**
3.1	Define a KFold object (kf)
3.2	Set the number of splits (n_splits) to num_rounds
3.3	Set shuffle to randomly shuffle the data before splitting
4.	**Iterate Through Folds:**
4.1	Loop through each fold using a for loop (i, (train_index, test_index)) in kf.split(X)
4.2	The loop iterates num_rounds times
4.3	Print the current fold number (i+1) and total number of folds (num_rounds)
5.	**Split Data into Training and Testing Sets (for each fold):**
5.1	Use indexing with train_index and test_index for the current fold.
6.	**Reshape Features for LSTM:**
6.1	Reshape X_train and X_test to a 2D format suitable for 1D LSTM.
6.2	The new shape is (number_of_samples, number_of_features, 1)
7.	**Build LSTM Model:**
7.1	Define a Sequential model (model)
7.2	Add a LSTM layer with 32 units, ReLU activation, and input shape based on X_train
7.3	Add a Dense layer with 128 units and ReLU activation
7.4	Add a final Dense layer with num_classes and softmax activation for classification.
8.	**Compile the Model:**
8.1	Compile with Adam optimizer, sparse_categorical_crossentropy loss function
9.	**Train the Model:**
9.1	Train the model on X_train and y_train for 50 epochs with a batch size of 32
10.	**Evaluate the Model:**
10.1	Evaluate the model on X_test and y_test
10.2	Store the accuracy in a list named accuracies
11.	**Print performance metrics**
12.	**Calculate and print average accuracy**

Table 3 Multiple PQDs waveforms.

		
C-1: Pure Sine	C-2: Sag with Harmonics	C-3: Swell with Harmonics
		
C-4: Interruption with Harmonics	C-5: Flicker with Harmonics	C-6: Flicker with Sag
		
C-7: Flicker with Swell	C-8: Sag with Oscillatory	C-9: Swell with Oscillatory

3.3 Classification

Following the feature extraction phase, both cDs and cAs are obtained for each component, as outlined in Fig. 5 (C-9: Swell with Oscillatory). Of particular interest are the values of cD5 and cA5 (yellow frame), as they exhibit distinct waveforms for each class of PQDs. Details of the decomposition for the remaining classes (C-1 to C-8) are presented in the appendix (Table 8). In the experimental dataset, each wavelet function is utilized as input to facilitate result comparison. To enhance the efficiency of dataset partitioning for learning, k-fold cross-validation is implemented with k set to 9.

For the LSTM model, structured as depicted in Fig. 3, the following configurations are employed:

- A layer with 32 units utilizing the Rectified Linear Unit (ReLU) activation function.
- Input shape: defined in a 1D waveform.
- Dense1: Fully connected layer with 128 units and ReLU activation function.
- Dense2: Fully connected output layer with a number of units equal to the total output classes, employing the softmax activation function.

The comprehensive classification process is outlined in Algorithm 2. The experimentation was conducted on a Windows 11 platform with Jupyter Lab, utilizing an Intel Core i7-12700H 2.9 GHz CPU and 16 GB of RAM. Python served as the primary programming language, supplemented by libraries including Keras, Scikit-learn, NumPy, Pandas, and PyWavelets.

3.4 Evaluation

Another crucial aspect of the experiment involves assessing the accuracy of the model's performance. The evaluation entails scrutinizing the efficacy of the confusion matrix, which serves as a potent tool for gauging outcomes and facilitating comparisons across different classes.

4. Results and Discussion

In the experiment, the dataset obtained through the synthesis process using the mathematical equations outlined in Table 2 consisted of 900 instances spanning 9 distinct classes. Sample signals corresponding to each class are depicted in Table 3. Subsequently, the data was partitioned into individual classes to undergo the feature extraction process employing DWT. This process yielded both cDs and cAs features, with each wavelet type including haar1, db4, bior1.3, coif2, and sym4. Physical characteristics exemplifying each wavelet type are presented in Table 1. The component extraction procedure is outlined in

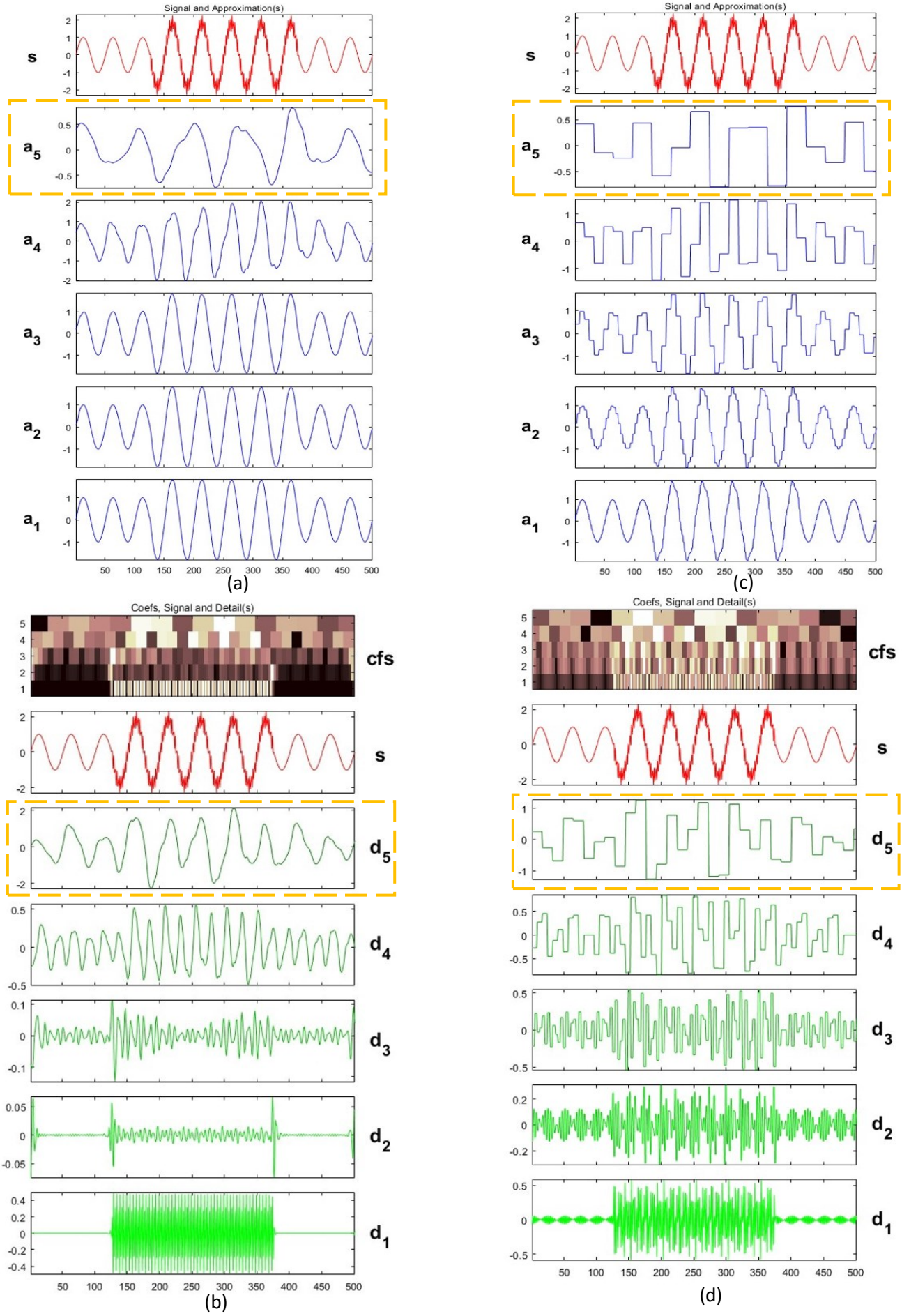


Fig. 5 (a)-(b) best (db4) and (c)-(d) worst (haar1) wavelet functions at level 5 for C-9: Swell with Oscillatory.

Algorithm 1. By the results of decomposing components, we show an example of the best (db4) and worst (haar1) wavelet functions in detail in Fig. 5. The outcomes derived from the decomposition process were then fed into the optimized LSTM model, as depicted in Algorithm 2. To delineate the characteristics of each form of PQDs, the following observations were made:

Firstly, the 1D signal undergoes decomposition using DWT via both high pass and low pass filters, with the parameter level set to 5. This results in the generation of Detailed Coefficients (cD1 to cD5) and Approximation Coefficients (cA1 to cA5), constituting a single waveform data. Consequently, each component can be segregated into 10 distinct sets of features, implying a total of 900x10=9000 sets of decomposed data per wavelet type. Therefore, in the experiment, we opt to utilize a variety of wavelet types, totaling 5. This leads to a cumulative count of 9000x5=45,000 sets of data employed in the analysis. cAs represent the coarsest scale of the signal. They capture the overall trends or low-frequency components of PQDs. These coefficients provide a compressed representation of the original signal by summarizing its general characteristics. Suitable for capturing large-scale patterns or trends in PQDs. Since they contain low-frequency information, they are useful for capturing global features of the signal, such as the overall shape or general trends. cDs represent the differences or high-frequency components between successive approximation levels. They capture the finer details or high-frequency components of the signal that are associated with specific levels of detail or resolution in PQDs. Suitable for capturing local variations or fine-scale features in the data. They provide information about abrupt changes or edges in PQDs, which can be useful for detecting sharp transitions or localized patterns.

Secondly, from the decomposed dataset, both cDs and cAs obtained from five distinct wavelet types were subjected to preliminary characterization using statistical data analysis. Specifically, we chose to utilize the mean (avg) and standard deviation (sd) to conduct a comparative analysis for each class. The outcomes of this analysis are presented in Table 4 for cDs and Table 5 for cAs.

Table 4 Comparison of individual cDs coefficients and wavelet types.

C.	cD	harr1		db4		bior1.3		coif2		sym4	
		avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
1	cD1	-1.30E-17	7.05E-01	1.04E-14	-1.98E-05	-9.31E-18	6.29E-02	1.02E-13	1.48E-06	1.02E-13	1.48E-06
	cD2	1.30E-16	6.99E-01	3.88E-05	6.94E-04	1.28E-04	-6.34E-02	1.84E-05	5.31E-05	1.84E-05	5.31E-05
	cD3	9.16E-04	6.77E-01	4.40E-04	4.69E-03	3.98E-05	-6.47E-02	8.52E-05	-1.48E-03	8.52E-05	-1.48E-03
	cD4	9.16E-04	5.96E-01	3.91E-04	-7.40E-03	2.94E-04	-6.13E-02	-7.01E-04	2.25E-02	-7.01E-04	2.25E-02
	cD5	**8.46E-03**	**3.34E-01**	**1.31E-02**	**7.73E-03**	**5.37E-03**	**7.10E-02**	**-8.97E-04**	**-2.91E-03**	**-8.97E-04**	**-2.91E-03**
2	cD1	8.13E-17	3.00E-01	2.44E-14	-7.61E-03	9.01E-17	3.94E-02	3.06E-14	3.21E-04	2.29E-15	9.63E-04
	cD2	-4.49E-18	2.86E-01	6.31E-06	5.74E-03	1.88E-04	-3.65E-02	3.53E-05	4.09E-03	2.50E-04	2.06E-02
	cD3	1.20E-03	2.49E-01	3.97E-04	7.07E-05	-1.75E-04	-1.29E-02	-7.84E-05	-9.44E-04	-1.18E-04	2.99E-03
	cD4	1.20E-03	2.21E-01	3.23E-04	3.72E-03	-3.65E-05	-7.67E-03	-7.71E-04	1.06E-02	-9.96E-04	6.47E-03
	cD5	**2.96E-03**	**1.28E-01**	**6.57E-03**	**-1.14E-02**	**8.48E-04**	**-9.17E-03**	**-4.54E-04**	**1.34E-02**	**-1.69E-03**	**7.48E-03**
3	cD1	-2.22E-17	6.19E-01	1.89E-14	-1.42E-02	5.28E-17	1.04E-01	4.75E-14	1.21E-03	2.05E-15	3.64E-03
	cD2	7.44E-17	6.05E-01	1.45E-05	1.29E-02	1.59E-04	-9.76E-02	2.84E-05	9.09E-03	2.15E-04	4.59E-02
	cD3	1.04E-03	5.69E-01	3.83E-04	-2.57E-04	-1.06E-04	-7.77E-02	-2.92E-05	-2.41E-03	-9.48E-05	2.24E-03
	cD4	1.04E-03	5.00E-01	3.17E-04	-4.11E-03	5.48E-05	-3.20E-02	-7.02E-04	1.19E-02	-9.06E-04	2.96E-02
	cD5	**4.24E-03**	**2.83E-01**	**7.86E-03**	**3.22E-03**	**2.00E-03**	**1.60E-02**	**-5.46E-04**	**-2.67E-04**	**-2.09E-03**	**1.86E-03**
4	cD1	8.42E-17	3.14E-01	1.89E-14	-5.59E-03	1.73E-17	5.20E-03	4.75E-14	3.56E-05	1.99E-15	1.07E-04
	cD2	-3.60E-18	3.07E-01	1.45E-05	6.43E-04	1.59E-04	-4.88E-03	2.84E-05	4.54E-04	2.15E-04	2.29E-03
	cD3	1.04E-03	2.86E-01	3.83E-04	-8.47E-04	-1.06E-04	-2.29E-03	-2.92E-05	-4.36E-04	-9.48E-05	6.91E-04
	cD4	1.04E-03	2.54E-01	3.17E-04	1.84E-03	5.48E-05	-1.70E-03	-7.02E-04	2.36E-03	-9.06E-04	4.37E-03
	cD5	**4.24E-03**	**1.49E-01**	**7.87E-03**	**3.63E-04**	**2.00E-03**	**2.08E-03**	**-5.49E-04**	**-1.15E-04**	**-2.09E-03**	**-2.73E-03**
5	cD1	2.40E-11	1.44E+00	-7.31E-07	-5.12E-02	2.40E-11	2.92E-01	2.86E-06	4.87E-03	2.56E-06	1.73E-02
	cD2	2.40E-11	1.42E+00	1.34E-04	-2.43E-03	4.23E-04	-2.37E-01	6.40E-05	1.43E-02	4.17E-04	1.29E-01
	cD3	2.56E-03	1.36E+00	1.62E-03	-2.60E-03	5.60E-05	-1.28E-01	6.45E-04	-1.80E-02	-1.74E-04	-7.07E-04
	cD4	2.56E-03	1.20E+00	1.32E-03	1.52E-02	5.44E-04	3.26E-02	-2.18E-03	5.63E-02	-3.15E-03	3.50E-02
	cD5	**1.37E-02**	**6.94E-01**	**2.26E-02**	**-2.71E-02**	**5.50E-03**	**1.85E-01**	**-3.77E-03**	**-2.12E-02**	**-8.40E-03**	**-6.28E-03**
6	cD1	1.20E-11	1.12E+00	-3.13E-07	-8.42E-05	1.20E-11	6.32E-02	1.23E-06	-4.97E-05	1.10E-06	5.26E-05
	cD2	1.20E-11	1.11E+00	7.79E-05	3.67E-04	2.66E-04	-5.64E-02	4.02E-05	-5.34E-04	3.77E-04	-3.68E-04
	cD3	1.76E-03	1.07E+00	1.09E-03	5.86E-03	2.12E-04	-5.49E-02	4.72E-04	-4.43E-03	-6.23E-05	1.31E-03
	cD4	1.76E-03	9.47E-01	9.45E-04	8.93E-03	7.22E-04	4.02E-03	-1.48E-03	4.24E-02	-2.19E-03	4.21E-02
	cD5	**1.61E-02**	**5.46E-01**	**2.49E-02**	**-4.45E-03**	**9.77E-03**	**3.21E-02**	**-2.31E-03**	**1.43E-02**	**-7.91E-03**	**1.33E-02**
7	cD1	2.40E-11	1.80E+00	-3.13E-07	-1.04E-04	2.40E-11	1.20E-01	1.23E-06	-4.82E-05	1.10E-06	5.57E-05
	cD2	2.40E-11	1.78E+00	7.79E-05	4.66E-04	2.66E-04	-1.31E-01	4.02E-05	-2.15E-04	3.77E-04	-3.68E-04
	cD3	1.76E-03	1.73E+00	1.09E-03	2.36E-03	2.12E-04	-1.84E-01	4.72E-04	-6.16E-03	-6.23E-05	1.67E-03
	cD4	1.76E-03	1.52E+00	9.45E-04	-1.16E-02	7.22E-04	-1.02E-01	-1.48E-03	-1.11E-02	-2.19E-03	4.53E-02
	cD5	**1.61E-02**	**8.59E-01**	**2.49E-02**	**1.64E-03**	**9.77E-03**	**9.84E-02**	**-2.31E-03**	**-1.58E-02**	**-7.91E-03**	**-6.91E-03**
8	cD1	-6.45E-07	5.09E-01	-6.45E-07	-1.98E-05	-6.45E-07	1.54E-02	-6.45E-07	1.48E-06	-6.45E-07	3.09E-06
	cD2	-6.45E-07	5.05E-01	3.82E-05	6.94E-04	1.27E-04	-2.67E-02	1.78E-05	-2.52E-04	1.85E-04	2.12E-04
	cD3	9.15E-04	4.89E-01	4.39E-04	2.45E-03	3.91E-05	-1.18E-02	8.46E-05	-3.02E-04	-5.15E-05	8.05E-04
	cD4	9.15E-04	4.31E-01	3.91E-04	5.03E-03	2.93E-04	-1.17E-02	-7.01E-04	1.94E-02	-9.14E-04	1.55E-02
	cD5	**8.46E-03**	**2.48E-01**	**1.31E-02**	**-1.38E-02**	**5.37E-03**	**-1.98E-02**	**-9.01E-04**	**1.43E-02**	**-3.60E-03**	**8.35E-03**
9	cD1	1.61E-06	1.03E+00	1.61E-06	-1.98E-05	1.61E-06	6.29E-02	1.61E-06	1.48E-06	1.61E-06	3.09E-06
	cD2	1.61E-06	1.02E+00	4.05E-05	-7.34E-04	1.29E-04	-6.34E-02	2.00E-05	2.55E-04	1.87E-04	-5.30E-05
	cD3	9.18E-04	9.84E-01	4.42E-04	7.47E-03	4.14E-05	-1.14E-01	8.68E-05	-1.48E-03	-4.93E-05	8.05E-04
	cD4	9.18E-04	8.66E-01	3.93E-04	-1.77E-02	2.96E-04	-6.13E-02	-6.99E-04	1.92E-02	-9.12E-04	4.48E-02
	cD5	**8.46E-03**	**4.87E-01**	**1.31E-02**	**4.98E-03**	**5.37E-03**	**2.21E-02**	**-8.93E-04**	**-5.48E-04**	**-3.60E-03**	**7.88E-03**

Table 5 Comparison of individual cAs coefficients and wavelet types.

C.	cA	harr1		db4		bior1.3		coif2		sym4	
		avg	sd	avg	sd	avg	sd	avg	sd	avg	sd
1	cA1	-0.0074	0.7121	-0.0074	0.7134	-0.0075	0.7149	-0.0074	0.7134	-0.0074	0.7134
	cA2	-0.0072	0.7059	-0.0072	0.7136	-0.0076	0.7202	-0.0073	0.7136	-0.0073	0.7136
	cA3	-0.0070	0.6839	-0.0080	0.7125	-0.0119	0.7326	-0.0087	0.7121	-0.0087	0.7121
	cA4	-0.0029	0.5859	-0.0068	0.6845	-0.0007	0.7139	-0.0125	0.6830	-0.0125	0.6830
	cA5	**0.0079**	**0.3269**	**0.0080**	**0.2558**	**0.0169**	**0.2706**	**-0.0063**	**0.2573**	**-0.0063**	**0.2573**
2	cA1	0.0179	0.2983	0.0180	0.3025	0.0178	0.3049	0.0179	0.3025	0.0179	0.3025
	cA2	0.0194	0.2848	0.0190	0.2953	0.0176	0.2979	0.0189	0.2955	0.0185	0.2948
	cA3	0.0202	0.2439	0.0192	0.2609	0.0193	0.2716	0.0194	0.2628	0.0204	0.2599
	cA4	0.0189	0.2086	0.0187	0.2492	0.0201	0.2621	0.0178	0.2510	0.0178	0.2489
	cA5	**0.0245**	**0.1195**	**0.0231**	**0.1001**	**0.0235**	**0.1147**	**0.0176**	**0.1070**	**0.0160**	**0.1027**
3	cA1	-0.0366	0.6243	-0.0365	0.6282	-0.0369	0.6310	-0.0366	0.6282	-0.0366	0.6282
	cA2	-0.0336	0.6094	-0.0350	0.6223	-0.0376	0.6290	-0.0351	0.6226	-0.0357	0.6225
	cA3	-0.0306	0.5858	-0.0340	0.5943	-0.0373	0.6155	-0.0345	0.5961	-0.0328	0.5981
	cA4	-0.0231	0.5046	-0.0304	0.5745	-0.0240	0.6028	-0.0348	0.5745	-0.0361	0.5753
	cA5	**-0.0125**	**0.2737**	**-0.0157**	**0.2116**	**-0.0065**	**0.2152**	**-0.0242**	**0.2270**	**-0.0256**	**0.2260**
4	cA1	0.0326	0.3126	0.0326	0.3146	0.0326	0.3159	0.0326	0.3146	0.0326	0.3146
	cA2	0.0331	0.3063	0.0333	0.3122	0.0326	0.3144	0.0332	0.3122	0.0329	0.3118
	cA3	0.0328	0.2808	0.0329	0.2993	0.0329	0.3093	0.0330	0.3004	0.0338	0.2985
	cA4	0.0294	0.2395	0.0311	0.2843	0.0317	0.2978	0.0301	0.2848	0.0303	0.2846
	cA5	**0.0342**	**0.1379**	**0.0339**	**0.1137**	**0.0332**	**0.1355**	**0.0270**	**0.1263**	**0.0250**	**0.1218**
5	cA1	-0.0203	1.4496	-0.0201	1.4551	-0.0206	1.4589	-0.0203	1.4551	-0.0203	1.4550
	cA2	-0.0174	1.4272	-0.0180	1.4436	-0.0208	1.4590	-0.0182	1.4436	-0.0190	1.4439
	cA3	-0.0144	1.3787	-0.0202	1.4315	-0.0274	1.4741	-0.0211	1.4320	-0.0170	1.4352
	cA4	-0.0094	1.1868	-0.0164	1.3756	-0.0044	1.4396	-0.0288	1.3713	-0.0320	1.3721
	cA5	**0.0155**	**0.6759**	**0.0136**	**0.5364**	**0.0316**	**0.5809**	**-0.0170**	**0.5146**	**-0.0219**	**0.5028**
6	cA1	0.0495	1.1312	0.0495	1.1334	0.0496	1.1357	0.0495	1.1334	0.0495	1.1334
	cA2	0.0488	1.1207	0.0499	1.1338	0.0500	1.1444	0.0498	1.1337	0.0498	1.1337
	cA3	0.0470	1.0841	0.0478	1.1304	0.0425	1.1600	0.0466	1.1293	0.0492	1.1327
	cA4	0.0427	0.9263	0.0450	1.0762	0.0509	1.1218	0.0358	1.0738	0.0350	1.0812
	cA5	**0.0610**	**0.5393**	**0.0621**	**0.4317**	**0.0706**	**0.4841**	**0.0339**	**0.4187**	**0.0279**	**0.4017**
7	cA1	-0.0805	1.8126	-0.0805	1.8161	-0.0807	1.8197	-0.0805	1.8161	-0.0805	1.8161
	cA2	-0.0792	1.7970	-0.0802	1.8164	-0.0813	1.8333	-0.0802	1.8163	-0.0803	1.8163
	cA3	-0.0771	1.7423	-0.0820	1.8139	-0.0924	1.8666	-0.0835	1.8133	-0.0807	1.8156
	cA4	-0.0598	1.5020	-0.0744	1.7525	-0.0562	1.8313	-0.0898	1.7445	-0.0948	1.7443
	cA5	**-0.0327**	**0.8261**	**-0.0332**	**0.6397**	**-0.0049**	**0.6679**	**-0.0645**	**0.6575**	**-0.0694**	**0.6518**
8	cA1	0.0447	0.5110	0.0445	0.5111	0.0447	0.5121	0.0445	0.5111	0.0445	0.5111
	cA2	0.0440	0.5060	0.0447	0.5113	0.0450	0.5160	0.0447	0.5112	0.0446	0.5112
	cA3	0.0429	0.4890	0.0439	0.5097	0.0420	0.5230	0.0433	0.5093	0.0443	0.5107
	cA4	0.0382	0.4150	0.0409	0.4824	0.0421	0.5022	0.0377	0.4809	0.0381	0.4862
	cA5	**0.0454**	**0.2375**	**0.0460**	**0.1878**	**0.0471**	**0.2198**	**0.0330**	**0.1975**	**0.0300**	**0.1908**
9	cA1	-0.0595	1.0398	-0.0591	1.0390	-0.0594	1.0412	-0.0591	1.0390	-0.0591	1.0390
	cA2	-0.0583	1.0301	-0.0589	1.0391	-0.0599	1.0488	-0.0590	1.0390	-0.0590	1.0390
	cA3	-0.0573	0.9975	-0.0597	1.0379	-0.0657	1.0680	-0.0605	1.0378	-0.0594	1.0385
	cA4	-0.0440	0.8593	-0.0544	1.0037	-0.0433	1.0481	-0.0625	0.9993	-0.0656	0.9988
	cA5	**-0.0295**	**0.4703**	**-0.0299**	**0.3661**	**-0.0130**	**0.3743**	**-0.0454**	**0.3889**	**-0.0475**	**0.3876**

Table 6 Compare accuracy results for each coefficient and wavelet type.

Type.	haar1		db4		bior1.3		coif2		sym4	
Coff.	Acc	Loss	Acc	Loss	Acc	Loss	Acc	Loss	Acc	Loss
cD1	0.9058	0.0942	0.9148	0.0852	0.8931	0.1069	0.8933	0.1067	0.9039	0.0961
cD2	0.9271	0.0729	0.9202	0.0798	0.9161	0.0839	0.8873	0.1127	0.9098	0.0902
cD3	0.9349	0.0651	0.9230	0.0770	0.8981	0.1019	0.9069	0.0931	0.9259	0.0941
cD4	0.9239	0.0761	0.9311	0.0689	0.9041	0.0959	0.9118	0.0882	0.9278	0.0922
cD5	**0.9367**	**0.0633**	**0.9386**	**0.0614**	**0.9371**	**0.0629**	**0.9363**	**0.0637**	**0.9384**	**0.0616**
cA1	0.8630	0.1370	0.8702	0.1298	0.8798	0.1202	0.8774	0.1226	0.8558	0.1442
cA2	0.8678	0.1322	0.8438	0.1563	0.8678	0.1322	0.8678	0.1322	0.8534	0.1466
cA3	0.8582	0.1418	0.8702	0.1298	0.8654	0.1346	0.8726	0.1274	0.8630	0.1370
cA4	0.8654	0.1346	0.8678	0.1322	0.8606	0.1394	0.8630	0.1370	0.8678	0.1322
cA5	**0.8702**	**0.1298**	**0.8558**	**0.1442**	**0.8678**	**0.1322**	**0.8486**	**0.1514**	**0.8510**	**0.1490**

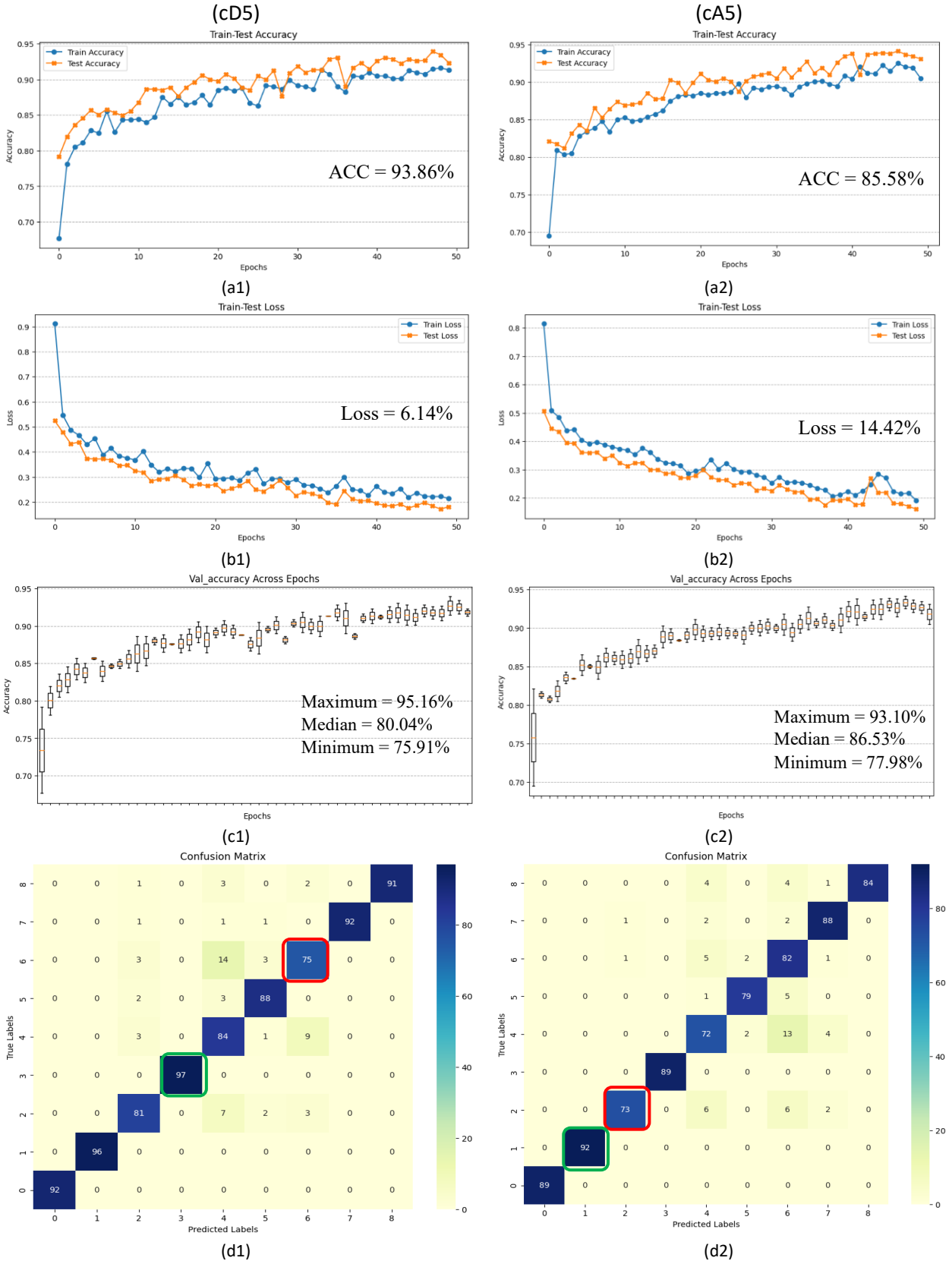


Fig. 6 The db4 wavelet features with cD5 and cA5 via LSTM.

Table 7 Compare results for precision, recall, and F1-score.

Type.	haar1			db4			bior1.3			coif2			sym4		
Coff.	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.	Pre.	Rec.	F1.
cD1	0.8021	0.8564	0.8431	0.8713	0.8510	0.8663	0.8597	0.9181	0.8532	1.0618	0.8373	0.7229	0.8528	0.8458	0.8602
cD2	0.8928	0.8799	0.8441	0.8324	0.8647	0.8401	0.6088	0.8919	0.8440	0.8371	0.8892	0.8798	0.8505	0.8319	0.8880
cD3	0.8647	0.8551	0.8441	0.8978	0.8637	0.8618	0.6343	0.8481	0.9042	0.7757	0.8245	0.8152	0.8458	0.8620	0.8926
cD4	0.8531	0.8505	0.8445	0.8515	0.8706	0.8988	0.6435	0.8528	0.9157	0.7360	0.8451	0.8086	0.8273	0.8759	0.8935
cD5	**0.8722**	**0.9496**	**0.8422**	**0.9209**	**0.9096**	**0.9382**	**0.9412**	**0.8690**	**0.9134**	**0.8352**	**0.8343**	**0.8076**	**0.9204**	**0.8458**	**0.9181**
cA1	0.8726	0.8207	0.7500	0.8549	0.8863	0.8568	0.6481	0.8343	0.8986	0.8960	0.8709	0.7639	0.8481	0.8690	0.8389
cA2	0.8447	0.7247	0.7578	0.8875	0.8892	0.8246	0.8204	0.8273	0.8296	0.8533	0.7000	0.7858	0.7736	0.8273	0.7389
cA3	0.8455	0.9065	0.8588	0.8333	0.7765	0.8614	0.8435	0.8181	0.8366	0.8254	0.8980	0.7710	0.8343	0.8296	0.8597
cA4	0.7574	0.7049	0.7500	0.8573	0.7873	0.8729	0.8458	0.8597	0.8903	0.8025	0.7294	0.8669	0.8389	0.8713	0.8343
cA5	**0.8448**	**0.8866**	**0.8725**	**0.8540**	**0.8922**	**0.8971**	**0.8250**	**0.8343**	**0.8273**	**0.8856**	**0.7304**	**0.8524**	**0.7505**	**0.7551**	**0.7528**

Note: Precision (pre.), Recall (rec), F1-score (F1)

It is noteworthy that the precise interpretation of mean or standard deviation values derived from representative data within each class remains ambiguous. Notably, cD1 and cD2 exhibit markedly low values as they pass through a high pass filter, allowing only high frequencies to traverse. However, with an increase in the level to cD3, cD4, and cD5, both average and standard deviation values witnessed an escalation. Upon comparing wavelet types for each class, discernible disparities in average and standard deviation values were observed. Nonetheless, upon closer inspection, certain similarities persisted across classes, such as the resemblance between cD5 values for classes C6 and C7, as well as C8 and C9, particularly in terms of average values for db4. Regarding cAs, characterized by the low pass filter coefficient that permits only low frequencies to pass through, the resultant values are generally higher than those of cD. Notably, cA1, cA2, and cA3 exhibit similar characteristics or comparable average and standard deviation values, while similarities are also evident between cA4 and cA5 across classes and various wavelet types. For instance, similarities were identified in cD5 values for classes C6 and C8 using the coif2 wavelet type. Therefore, while the analysis of data using statistical calculations aids in feature extraction, it is essential to acknowledge the inherent limitations in clarity and specificity.

Thirdly, the data pertaining to each component, ranging from cD1 to cD5 and cA1 to cA5, of every wavelet type were inputted into the LSTM model for classification purposes. The outcomes of this classification process are presented in Table 6. Notably, variations in accuracy and loss were observed across different levels of cDs and cAs, with cD4 and cD5 exhibiting the highest accuracy, particularly when employing the db4 wavelet type. However, upon closer examination of each wavelet type, it was noted that db4 and sym4 demonstrated similar physical characteristics, resulting in the separated components at level 5 being comparable. During the experiment, it was concluded that both small waves are conducive for utilization in the decomposition analysis of PQDs when juxtaposed with other wavelets. Furthermore, graphical representations comparing the training and testing of each cDs and cAs were plotted, as depicted in Fig. 6 (a1) and (a2). An exemplary db4 wavelet of cD5 and cA5 is showcased, wherein both feature sets exhibited different average accuracy. However, a discernible discrepancy was observed in the data set's distribution, with cD5 displaying a relatively larger shift, suggesting the presence of similarities in form among certain elements within each class, indicative of disturbance moments or events. The accuracy and loss graphs, illustrated in Fig. 6 (b1) and (b2), respectively, revealed contrasting trends for both cD5 and cA5 feature sets, with results converging close to zero. Furthermore, Box Plots were employed to analyze maximum median and minimum values, aiding in decision-making regarding the suitability of cD5 and cA5 for application as shown in Fig. 5 (c1) and (c2). Moreover, the results for each class were plotted to compare the true labels and the predicted labels, as depicted in Fig. 5 (d1) and (d2), utilizing a confusion matrix with gradients based on accuracy. It was observed that for cD5, the model most accurately predicted C-4 (Interruption with Harmonics) at 97%, while the least accurate prediction was for C-7 (Flicker with Swell) at 75%**.** Conversely, for cA5, the model most accurately predicted C-2 (Sag with Harmonics) at 92%, while the greatest discrepancy was observed for C-3 (Swell with Harmonics), predicted correctly only 73% of the time. Notably, the model exhibited numerous prediction errors in classes C-3 and C-7, both of which encompass swell, a feature exhibiting similarity with other classes such as flicker and oscillatory, and crucially, exhibiting fluctuations corresponding to load usage behavior.

In addition to the accuracy metric, we also employed the primary metrics used in the performance matrix to evaluate the classification performance for PQDs scenarios. The detailed results are presented in Table 7. Data from all nine classes were averaged to calculate the performance metrics for each wavelet type. Precision was used to measure the correctly classified PQDs instances, Recall measured the proportion of true positive instances, and F1-score measured the harmonic mean between precision and recall. While the overall results exhibited some dispersion due to variations in the occurrence of TP events, a crucial variable, we consistently observed that cD5 (cD5: precision=92.09%, Recall=90.96%, F1-score 93.82%) and cA5 (cA5: precision=85.40%, Recall=89.22%, F1-score=89.71%) for db4 maintained superior performance.

5. Conclusion

In this study, we synthesized multiple PQDs through the implementation of 900 sets encompassing 9 distinct classes of mathematical equations. These PQDs were subsequently subjected to decomposition using DWT configured to level 5. Subsequently, the characteristics of each wavelet type: haar1, db4, bior1.3, coif2, and sym4, were compared and analyzed utilizing statistical features before being fed into LSTM model to classify the characteristics of each class. The experimental results demonstrate our ability to synthesize and adjust parameters to generate multiple PQDs exhibiting diverse values in accordance with the IEEE standard 1159-2019. Notably, the DWT, set to level 5, facilitated decomposition into detailed coefficients (cD1 to cD5) and approximation coefficients (cA1 to cA5), with cD5 and cA5 derived from the db4 wavelet proving to be suitable components for input into the LSTM model. Throughout the process of characterizing PQDs, we employed k-fold cross-validation (with k=9) to partition and shuffle each dataset within the training model for learning and result evaluation. The utilization of db4 wavelets for cD5 and cA5 yielded a different accuracy rate of 93.86% and 85.58%, respectively. However, it is pertinent to acknowledge certain limitations encountered during the experiment, including the simplicity of the techniques employed and the utilization of specific synthesized data, resulting in a less-than-optimal accuracy rate. Addressing these limitations will be imperative for future research endeavors in this domain.

Acknowledgements

This work was supported by the Faculty of Engineering, Academic Year 2023, Prince of Songkla University, Thailand, and in part by the Department of Electrical and Biomedical Engineering. The authors would like to thank the Meter and Transformer Section, Provincial Electricity Authority Ranot Branch, for providing equipment and materials used in the experiment.

References

[1] Chung, I. Y., Won, D. J., Kim J. M., Ahn S. J. and Moon S. I., Development of a network-based power quality diagnosis system. *Electric Power Systems Research.* 77(8) (2007) 1086–1094, doi: https://doi.org/10.1016/j.epsr.2006.09.011.

[2] Mohanty, S.R., Ray P. K., Kishor N. and Panigrahi B. K., Classification of disturbances in hybrid DG system using modular PNN and SVM. *International Journal of Electrical Power & Energy Systems.* 44(1) (2013) 764–777, doi: https://doi.org/10.1016/j.ijepes.2012.08.020.

[3] Khetarpal, P. and Tripathi, M. M., A critical and comprehensive review on power quality disturbance detection and classification. *Sustainable Computing Informatics and Systems.* 28 (2020) 400-417, doi: https://doi.org/10.1016/j.suscom.2020.100417.

[4] IEEE Recommended Practice for Monitoring Electric Power Quality, *IEEE Std 1159-2019 (Revision of IEEE Std 1159-2009)*, 2 (2019) 1-98, doi: https://doi.org/10.1109/IEEESTD.2019.8796486.

[5] Provincial Electricity Authority, Regulations on electrical networksSystem connection requirements, <https://www.pea.co.th/Portals/0/Document/vspp/PEA%20Interconnection%20Code%202016.pdf> (2016).

[6] Provincial Electricity Authority, Power quality research department, research division, Electrical System Research and Development Department, <https://www.pea.co.th.> (2023).

[7] Saxena, D., Verma, K. S. and Singh, S., Power quality event classification: an overview and key issues. *International Journal of Engineering, Science and Technology*. 2(3) (2010) 186-199, doi: https://doi.org/10.4314/ijest.v2i3.59190.

[8] S. Upadhyaya and S. Mohanty, Power quality disturbance detection using wavelet-based signal processing. in *Annual IEEE India Conference (INDICON. (2013)* 1-6, doi: https://doi.org/10.1109/UPEC.2016.8114077.

[9] Dekhandji, F. Z., Detection of power quality disturbances using discrete wavelet transform. in *5th International Conference on Electrical Engineering - Boumerdes (ICEE-B).* (2017) 1-5, doi: https://doi.org/10.1109/ICEE-B.2017.8192080.

[10] Diego, R. I. and Barros, J., Subharmonic measurement using DFT and wavelet-packet transform in an IEC extended framework, *Measurement.* 43 (2010) 1603–1608, doi: https://doi.org/10.1016/j.measurement.2010.09.012.

[11] Pan, M. Xu, Chen, T., J. and Li, Z., A harmonic detection method for distributed generation connected grid system using DWT and HHT. *Journal of Electrical Engineering & Technology.* 14(4) (2019) 1495–1503, doi: https://doi.org/10.1007/s42835-019-00190-1.

[12] Parvez, M., Aghili, A. I., Sarwat, S., Rahman, R. and Alam, F., Online power quality disturbance detection by support vector machine in smart meter. *Journal of Modern Power Systems and Clean Energy.* 7 (2018) 1328–1339, doi: https://doi.org/10.1007/s40565-018-0488-z.

[13] Mahela, O. P., Khan, B., Alhelou, H. H., and Tanwar, S., Assessment of power quality in the utility grid integrated with wind energy generation. *IET Power Electronics*. 13(13) (2020) 2917–2925, doi: https://doi.org/10.1049/iet-pel.2019.1351.

[14] Mahela, O. P., Khan, B., Alhelou, H. H and Siano, P., Power quality assessment and event detection in distribution network with wind energy penetration using stockwell transform and fuzzy clustering. *IEEE Transactions on Industrial Informatics.* 16(11) (2020) 6922–6932, doi: https://doi.org/10.1109/tii.2020.2971709.

[15] Exploring Technologies. Discrete wavelet transform (DT) of 1-D Signals, <https://www.exptech.co.in/.> (2022).

[16] Rico-Medina, V., Reyes-Archundia, E., Gnecchi, J. A. G., Olivares-Rojas, J. C. and Del-Carmen, M., Garcia-Ramirez, C., Analysis of the appropriate decomposition level based on discrete wavelet transform for detection of power quality disturbances. in *IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. (2022) 1-6, doi: https://doi.org/10.1109/ropec55836.2022.10018687

[17] Chawda, G. S., Shaik, A. G., Shaik, M., Padmanaban, S., Nielsen, J. B. H., Mahela, O. P. and Kaliannan, P., Comprehensive review on detection and classification of power quality disturbances in utility grid with renewable energy penetration. *IEEE Access*. 8 (2020) 146807–146830, doi: https://doi.org/10.1109/access.2020.3014732.

[18] Addison, P. *The illustrated wavelet transform handbook*, <https://doi.org/10.1887/0750306920> (2002).

[19] Siddiqi, H. *Applied functional analysis: numerical methods, Wavelet methods, and image processing*. 1st edn, CRC Press, 2003.

[20] Semmlow, J. L., *Biosignal and biomedical image processing: MATLAB-based* applications, <http://ci.nii.ac.jp/ncid/BA66341356> (2022).

[21] Cui, J., Gao, Q. and Li, D., Improved long short-term memory network based short term load forecasting. in *2019 Chinese Automation Congress (CAC)*. (2019), 4428-4433, doi: https://doi.org/10.1109/CAC48633.2019.8996379.
[22] Karasu, S. and Sarac, Z., Classification of power quality disturbances by 2D-Riesz transform, multi-objective grey wolf optimizer and machine learning methods. *Digital Signal Processing*. 101 (2020) 102711, doi: https://doi.org/10.1016/j.dsp.2020.102711.
[23] MathWorks. *MATLAB version: 9.13.0 (R2022b)*, <https://www.mathworks.com> (2023).
[24] Janthong, S. and Phukpattaranont, P., Classification of double power quality disturbances using discrete wavelet transform and bi-directional recurrent neural networks. in *PEACON & Innovation 2023: Moving towards to digital grid and green energy, Bangkok Thailand*. (2023), 50-59.
[25] Zhang, B., Chen, H., Feng, G. and Zhang, H., Application of wavelet multi-resolution analysis to harmonics detection based on MATLAB in power system. in *IEEE International Conference on Automation and Logistics.* (2007), 137-143, doi: https://doi.org/10.1109/ICAL.2007.4338545.

Appendix

Table 8 Differentiate the decomposition at level 5, each wavelet and class.

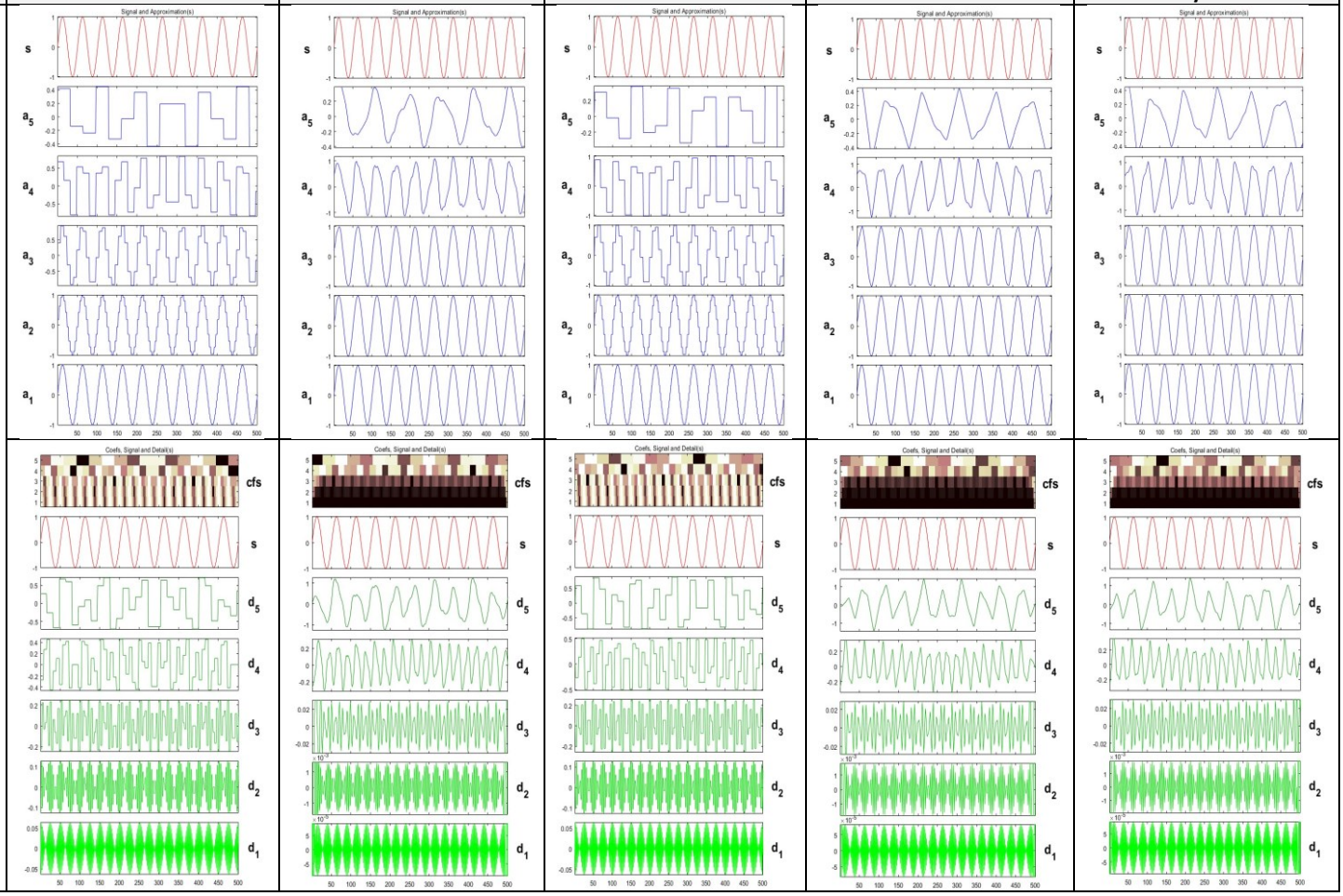
C.	haar1	db4	bior1.3	coif2	sym4
1					

Table 8 Differentiate the decomposition at level 5, each wavelet and class (Cont.).

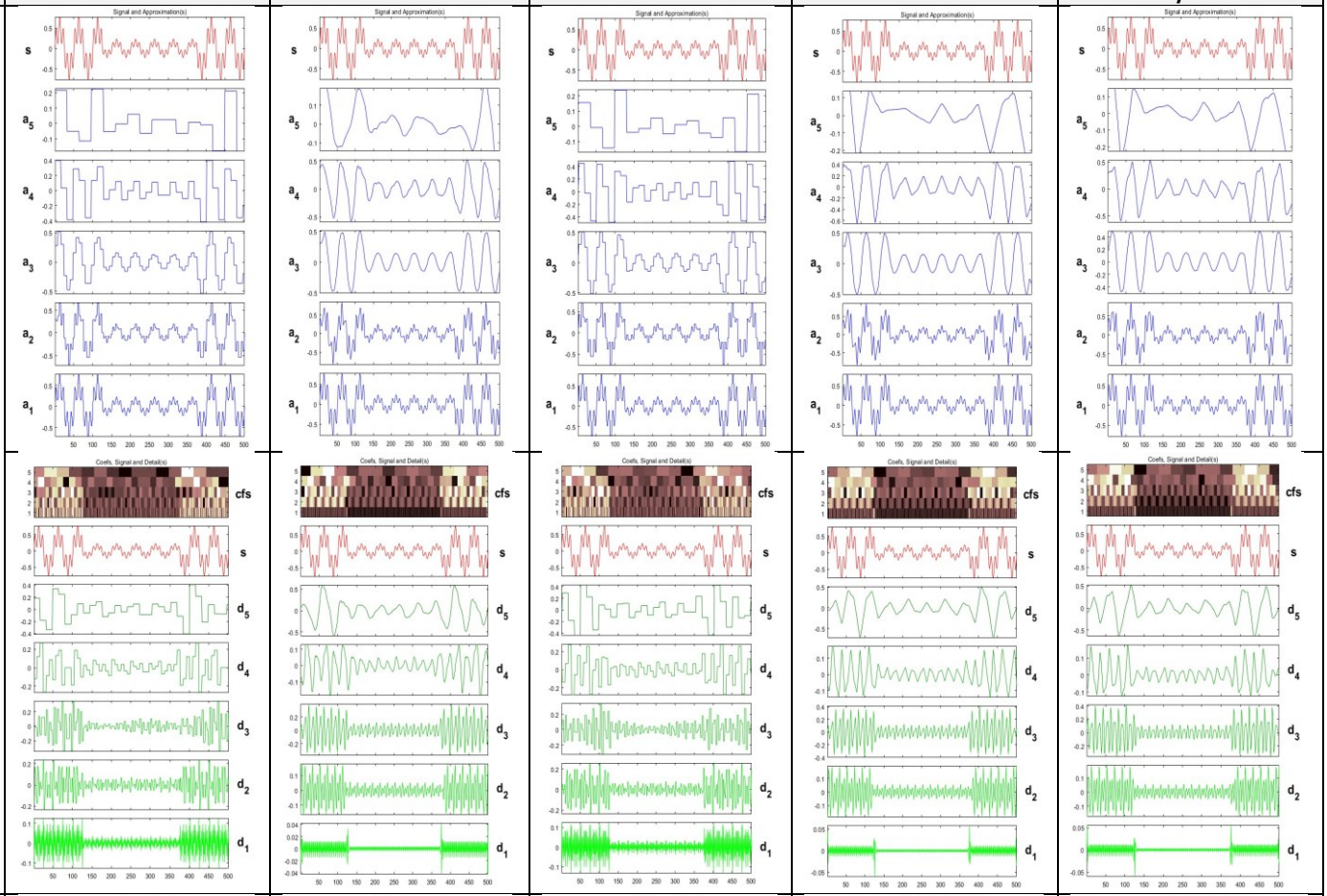
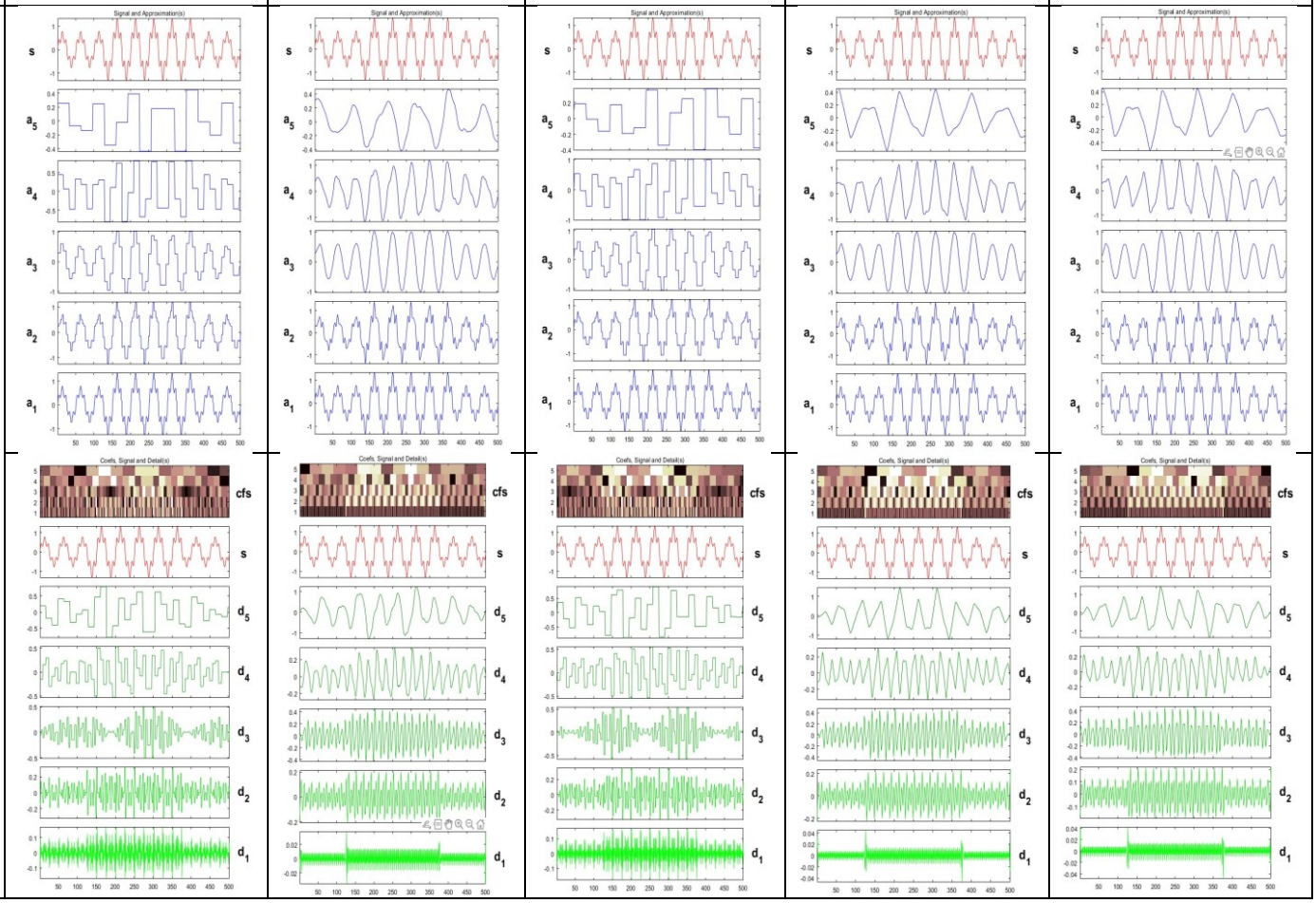
C.	haar1	db4	bior1.3	coif2	sym4
2					
3					

Table 8 Differentiate the decomposition at level 5, each wavelet and class (Cont.).

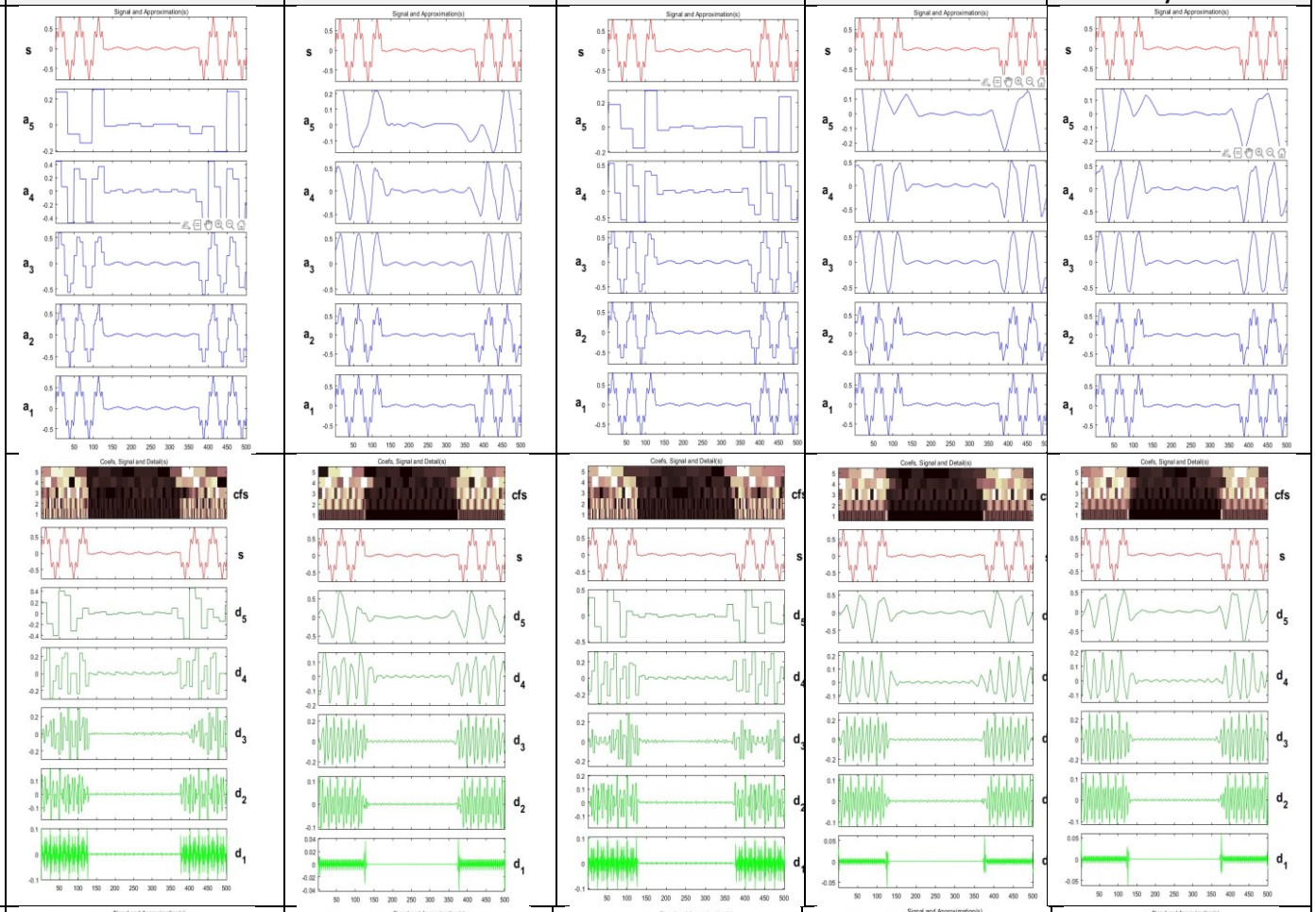
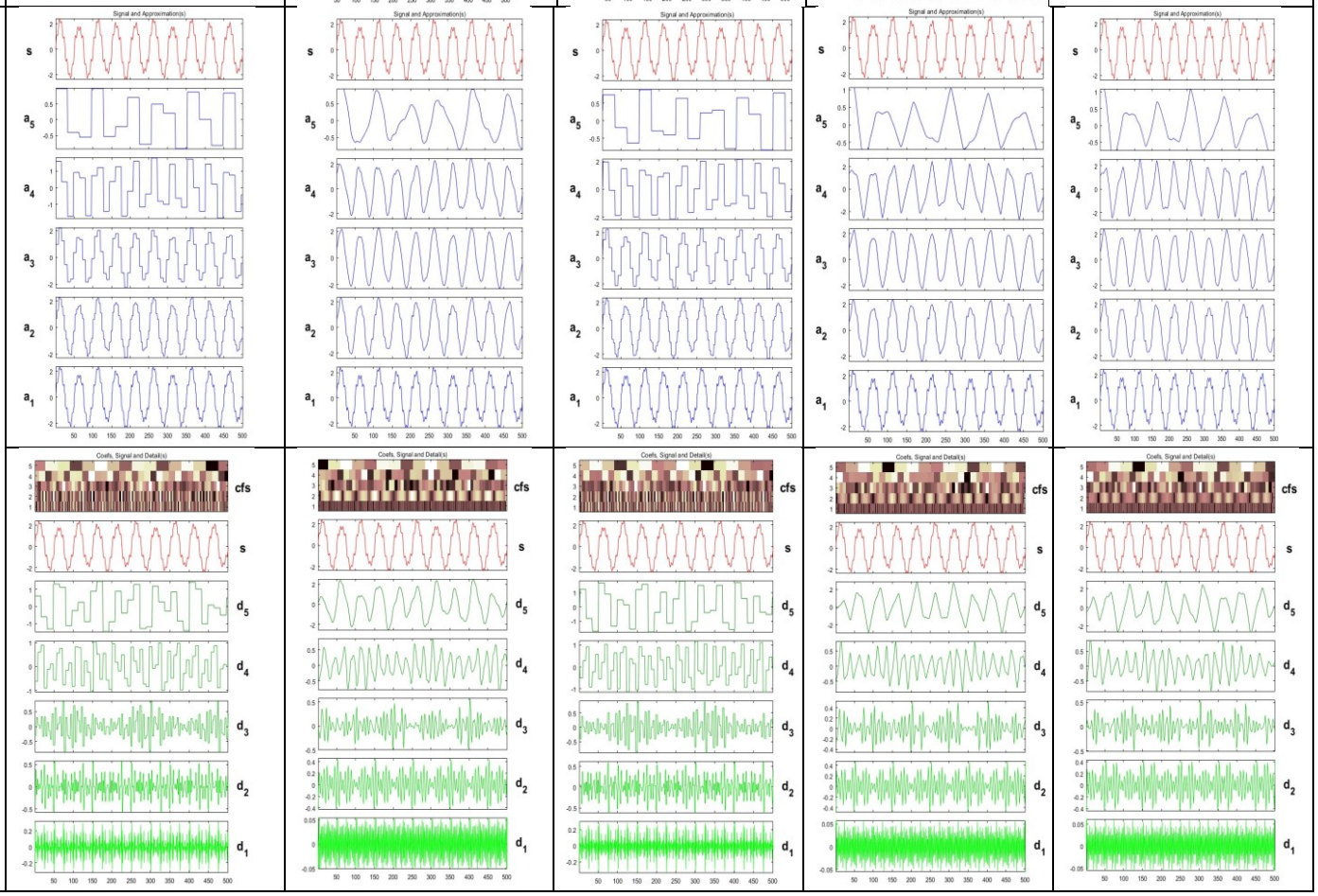
C.	haar1	db4	bior1.3	coif2	sym4
4					
5					

Table 8 Differentiate the decomposition at level 5, each wavelet and class (Cont.).

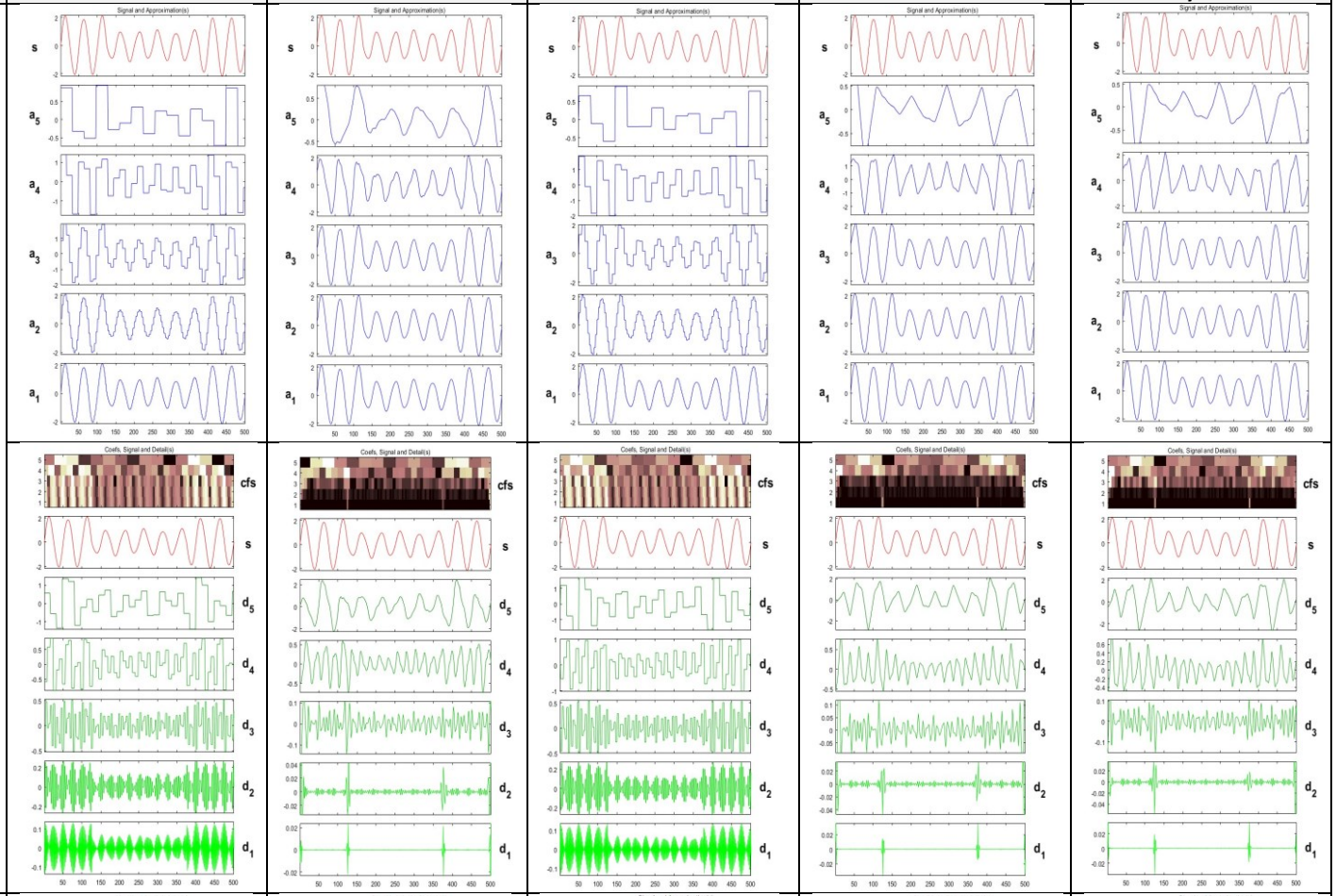
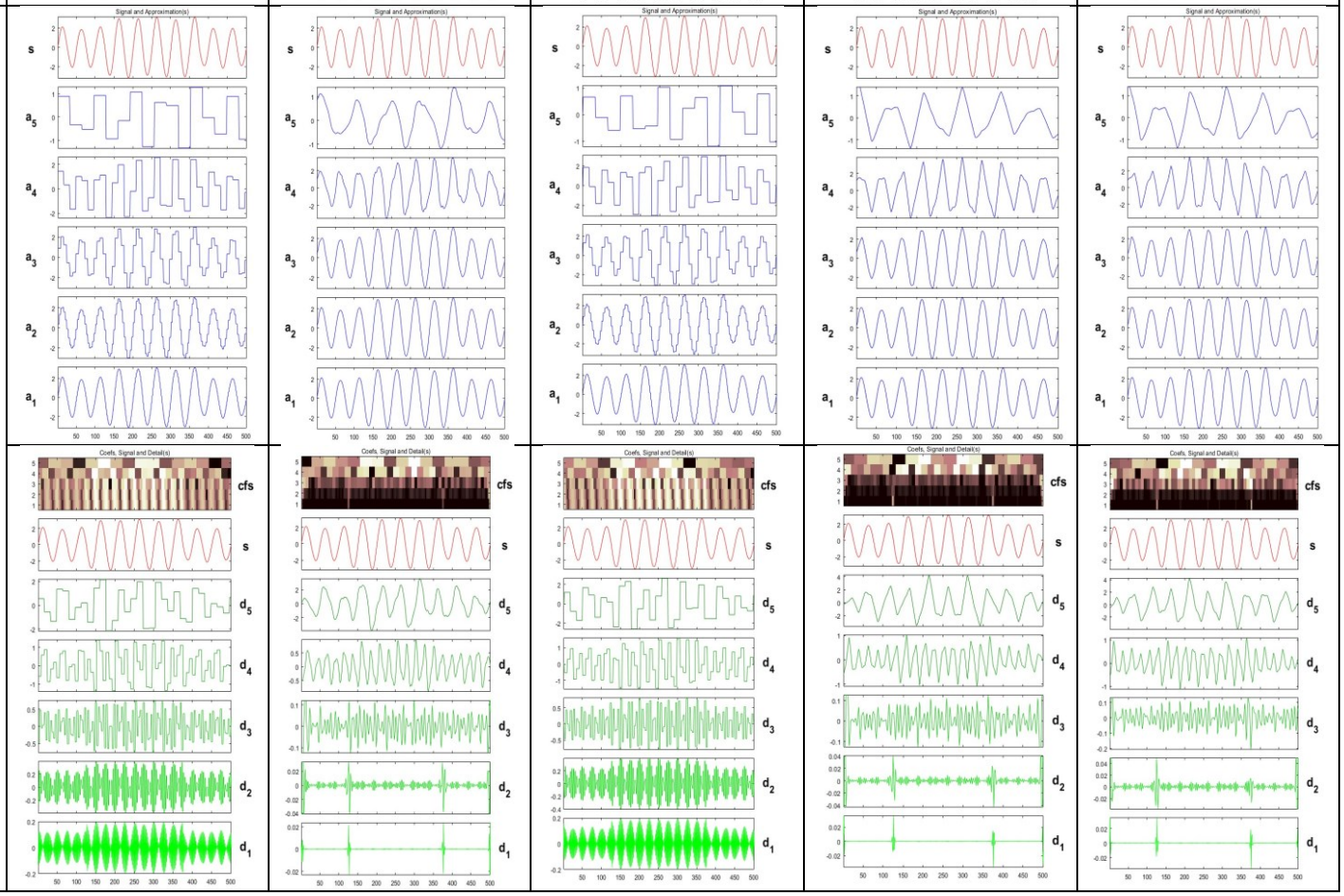
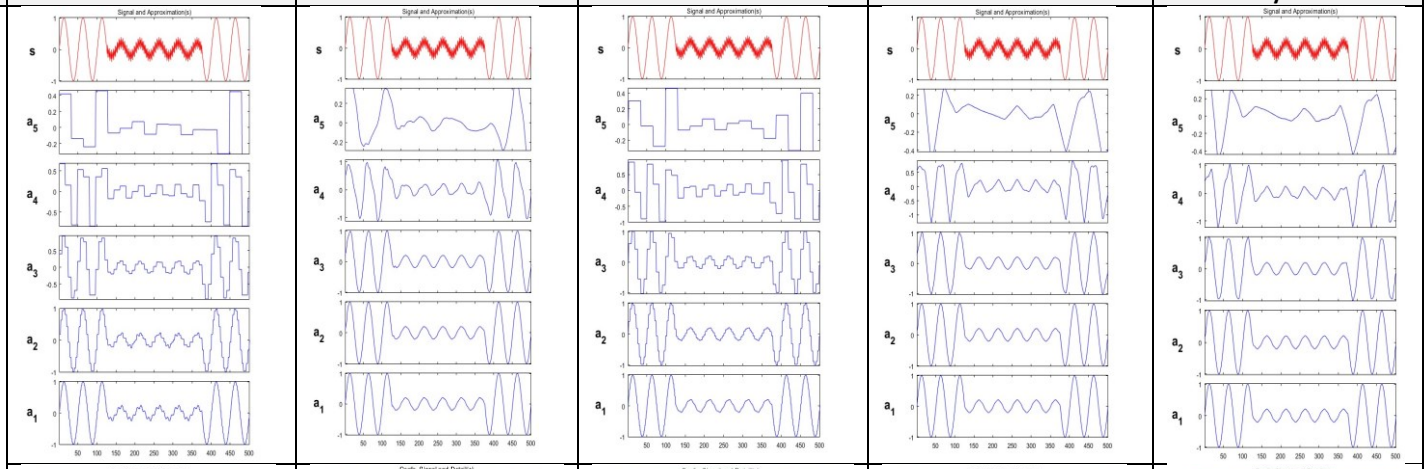
C.	haar1	db4	bior1.3	coif2	sym4
6					
7					

Table 8 Differentiate the decomposition at level 5, each wavelet and class (Cont.).

C.	haar1	db4	bior1.3	coif2	sym4
8					
9