Wavelet Set Theory Enhanced Wavelet Entropy Clustering for Pattern Recognition

Shivani Saxena, Prasadini Mahapatra, and Ahsan Z. Rizvi

***Abstract*— The Wavelet Entropy Clustering (WEC) technique enhances the pattern recognition (PR) methods in science and engineering domains. WEC helps identify hidden patterns in time series (TS) data. State-of-the-art WEC methods rely on advanced algorithms and wavelet function selection. Selecting pre-defined wavelets for PR might not capture all the patterns in the TS data. This limitation affects the accuracy of the TS data analysis. Existing data-driven-based PR methods handle complex TS data. They construct wavelets tailored to the data. This approach improves the accuracy of PR in complex TS datasets. Balancing all the desired wavelet properties like orthogonality and compact support becomes challenging. This inherent trade-off in achieving all properties limits the effectiveness of data-driven methods. Wavelet set theory (WST) offers a framework for addressing these challenges. WST tailors wavelets from a TS signal. Tailor wavelets perform better signal decomposition, capture complex features, and conduct multi-scale analysis. In this work, we propose *WaveClust*, a new approach which combines WST into the WEC framework. *WaveClust* extracts hidden features from TS data using multi-level wavelet packet entropy. We compare the performance of *WaveClust* with existing wavelet-based methods. *WaveClust* outperforms conventional approaches in PR tasks. It improves the ability of PR tasks to identify complex patterns in TS data. *WaveClust* aids researchers in analyzing complex TS data for scientific and engineering applications.**

***Index Terms*— Pattern recognition, Time-series data analysis, Wavelet entropy clustering, Wavelet set theory.**

I. Introduction

Pattern recognition (PR) [1-6] drives advanced technologies in various scientific and engineering fields. PR aims to identify and classify hidden patterns within these applications [1-6]. Most of the applications generate time series (TS) data [7]. Recognizing the inherent patterns in these TS data is crucial. A fundamental challenge lies in dealing with the complex TS data. The complex TS data include non-stationarity, noise, and high dimensionality [8]. Analyzing TS data with these complexities raises the need for a PR algorithm.

PR methods [1-6] rely on extracting features from the TS data. These algorithms use the extracted features to identify and classify hidden patterns from the TS data. Researchers call this strategy "extract and classify". Wavelet entropy clustering (WEC) [9-15] is a well-established technique for this approach. It is a versatile technique used in various fields. These fields include biomedical research [9, 10], monitoring strategies [11, 12], imaging [13], and fault detection [14, 15]. It utilizes wavelets and entropy mathematical functions to analyze TS data. Wavelets decompose the TS signal into different scales. It allows for a detailed analysis of the signal at various levels of granularity. Entropy measures the complexity within these decomposed scales to identify the patterns [15-17].

WEC technique helps in classifying and studying complex phenomena of the TS data. It proves its effectiveness in several key areas of TS data analysis. These areas include signal denoising [18, 19, 20], feature extraction [21, 22, 23], and PR [24, 25]. This versatility makes WEC a valuable tool across diverse disciplines. Researchers applied WEC for tasks like classifying fish feeding sounds to identify species [26], detecting pavement damage through smartphone sensors [27], and even analyzing human emotions using neural networks [28]. WEC is also applicable to study community-acquired pneumonia [29, 30, 31] and identify pavement quality [32].

WEC tackles two crucial challenges in TS data analysis. These are noise reduction and multi-scale pattern identification. It provides cleaner data and the ability to identify patterns at different scales. WEC leads to a more accurate analysis of TS data in detail. TS signals can be tricky to analyze. Sometimes, they hold meaningful patterns at both local and global levels. They appear across different time scales. Capturing the full range of these complex patterns using WEC remains a challenge [33-40]. This challenge arises from the crucial role of selecting a pre-defined wavelet function. The pre-defined wavelets might not always align well with the TS data [41, 42, 43]. This misalignment with the characteristics of the signal affects the effectiveness of the traditional analysis.

Data-driven approaches (DDA) [45-49] address the issue in traditional WEC methods. It builds wavelets based on the specific characteristics of the TS data. This strategy offers a degree of customization in the wavelet design. Various methods exist for wavelet design. These methods include multi-resolution analysis (MRA) [33, 34], lifting schemes [35, 36], and filter banks [37, 38, 39, 40, 44]. In [45], researchers use a template-based method for the specific wavelet design. Machine learning (ML) techniques are employed to design wavelets [46, 47, 48]. In these techniques, they learn and identify the best wavelet for the TS signal analysis. Likewise, evolutionary methods [49] are also used to design the custom wavelet. These approaches enhance the PR task

Shivani Saxena and Prasadini Mahapatra are associated with the Department of Computer Sciences and Engineering, Institute of Advanced Research, Gandhinagar, India.

Ahsan Z. Rizvi is associated with the Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, Umeå, Sweden.

in TS signal analysis. However, DDA faces its limitations. The limitation lies in achieving all wavelet properties [41] during wavelet construction. These properties are orthogonality, compact support, perfect reconstruction, and scaling [41]. These properties are important for efficient signal decomposition. They also aid in capturing hidden patterns across different time scales in the TS data. Balancing all these properties in wavelet design is a challenging task [50, 24, 51, 52]. Designing wavelets for complex, non-stationary TS signals is a tricky process. It's hard to achieve all ideal properties at once. They often lead to suboptimal performance of wavelet analysis for complex signals. This limitation in DDA methods motivates the exploration of Wavelet Set Theory (WST) [50, 24, 25, 51, 52] as a solution.

WST offers a mathematical framework for designing and analyzing wavelets. It provides a systematic and efficient way to develop wavelets for specific applications. These applications include signal denoising or compression [51, 52]. WST ensure wavelet properties during wavelet design. It supports orthogonality and compact support [50, 24, 51, 52]. These make the resulting wavelets well-suited for TS data analysis [53, 54, 55]. WST offers faster processing time [51]. It is also efficient in the calculation of wavelet coefficients [52, 56]. Overall, this tailored approach enhances PR capabilities in TS data. In this study, we use the advantages of WST and propose *WaveClust*, a new approach for PR in TS data.

WaveClust algorithm classifies TS data by capturing patterns at multi-scale. It combines the Wavelet Packet Entropy (WPE) [57] with classical clustering algorithms. The *WaveClust* algorithm calculates the features from the TS data. These features are energy entropy (e_E) [57], power entropy (p_E) [57], and slope entropy (s_E) [58]. The classical clustering algorithms classify these features. These algorithms include k-means [59, 60], density-based spatial clustering of applications with noise (DBSCAN) [61, 62], and expectation maximization (EM) [62]. Figure 1 illustrates the experimental process with a flow chart.

The *WaveClust* method tackles TS data analysis by taking a multi-step approach. First, it divides the TS signal into smaller chunks for more manageable analysis. Then, for each chunk, it creates a custom-made wavelet function using WST. The least-square optimization method then optimizes the custom wavelet. This strategy ensures that the wavelet has all the properties. The LSO process then filters out wavelets that don't meet the wavelet properties. This careful selection process helps guarantee the best possible feature extraction. Finally, the chosen wavelet decomposes each chunk of data using WPT. This decomposition reveals hidden features within the data at different scales. These extracted features help to identify patterns and structures within the TS data. In the final step, clustering algorithms classify the extracted features.

The resulting results are used for statistical comparison with existing methods. To assess *WaveClust*'s effectiveness, the researcher compares its results with

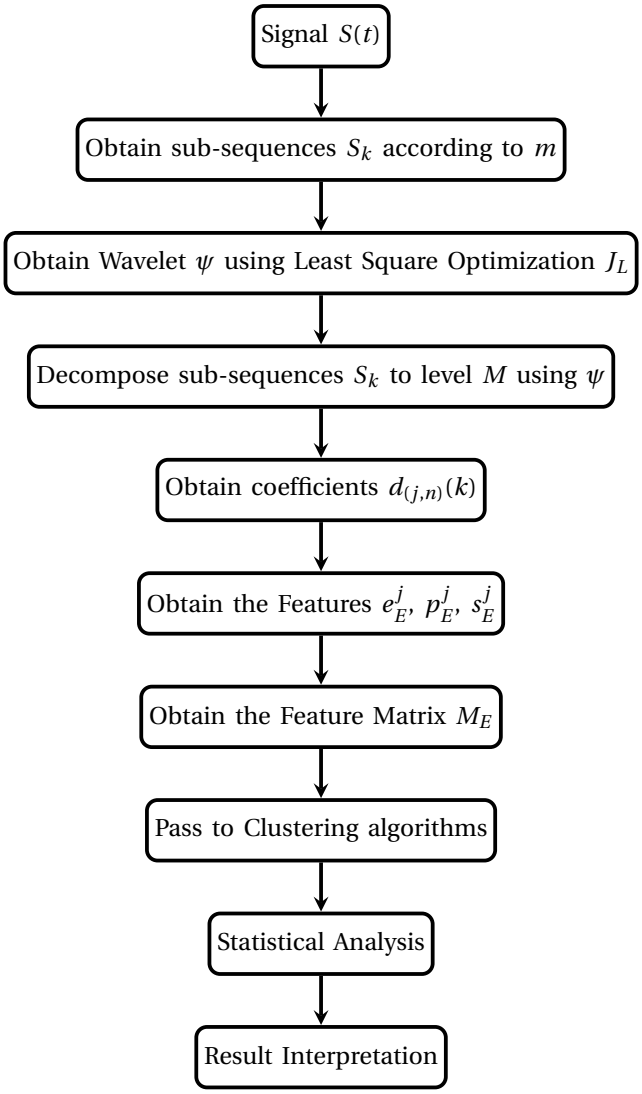

Fig. 1: Flowchart of the proposed methodology for TS Classification.

pre-defined wavelets. These wavelets are haar, db2, db4, db8, coif4, sym4, and bior2.8. The researcher also compares the results with established methods. The results show the fitness of *WaveClust* algorithm. It captures patterns and structures in TS data. The success of *WaveClust* lies in its ability to tailor the analysis of TS data. This result highlights *WaveClust*'s potential as a valuable tool for various fields. These include image and signal processing, bioinformatics, and finance.

II. Methodology

A wavelet set is a valuable tool for determining a wavelet [50]. A function ψ defines wavelets, where $\psi \in L^2(\mathbb{R})$ [24]. The dyadic dilates and translates $\psi \in L^2(\mathbb{R})$ form an orthonormal basis. In any subspace of $L^2(\mathbb{R})$, an orthonormal basis is considered, if Ψ is a measurable subset of $(\mathbb{R})$. The wavelet set Ψ is obtained from the signal $s(t)$, where $s(t)$ is in the rectangular window function $w(t)$. The wavelet set Ψ is defined as a wavelet set if the characteristic function $\chi_\Psi \in L^2(\mathbb{R})$. This condition employs in equation 1, the wavelet generated by $s(t)$ is defined.

$$w(t) = \begin{cases} 1; & -\frac{N-1}{2} \leq t \leq \frac{N-1}{2} \\ 0; & otherwise \end{cases} \quad (1)$$

$$||\chi_\Psi|| = \sqrt{m\left(\int_{-\frac{N-1}{2}}^{\frac{N-1}{2}} s(t)dt\right)} \quad (2)$$

Wavelet sets are a powerful tool for analyzing signals. They are defined by a function ψ in $L^2(\mathbb{R})$, which generates an orthonormal basis through dyadic dilations and translations. To ensure that ψ is a wavelet set, the characteristic function χ_Ψ must belong to $L^2(\mathbb{R})$. This function is obtained by using a rectangular window function $s(t)$ and a wavelet set Ψ, which satisfies certain conditions, such as being measurable and forming a disjoint partition for $\mathbb{R}$.

The Fourier transform of an orthonormal dyadic wavelet is given by $\frac{\chi_\Psi}{\sqrt{m(\Psi)}}$, where $m(\Psi) < \infty$. Inverse Fourier Transform can be used to reconstruct the wavelet. Various wavelet families exist, such as $Haar$, $Daubechies$, $Coiflets$, and $Symlets$. This article utilizes several wavelets, including $Haar$, $db2$, $db4$, $db8$, $sym4$, $coif4$, and $bior2.8$.

The wavelet function is constructed using the least square optimization technique, which minimizes the sum of squared errors (SSE) for a given pattern, as described in [39] and [40]. This technique optimizes the construction of the wavelet by finding the best set of parameters that fits the input signal. Specifically, the least square method considers a set of input-output pairs (s_i, y_i) and some parameters β, and defines a function $f(s)$ as $f(s) = S\beta$, where S is a matrix that depends on the input signal. The least square method then minimizes the function $f(s_i)$ by minimizing the $SSE(s_i) = y_i - f(s_i)$, which measures the difference between the predicted output and the actual output. The objective function J_L for the least square method is defined as follows:

$$J_L = \frac{1}{2}\sum_i SSE(s_i)^2 \quad (3)$$

Here, y is a vector of the actual outputs, S is a matrix that depends on the input signal, β is a vector of the parameters to be optimized, and $||\cdot||_2$ denotes the Euclidean norm. The least square method solves for the optimal parameters β that minimize the objective function J_L, which in turn leads to the construction of the wavelet function. Using equation 3, the function $f(t)$ is optimized, and the wavelet function $\psi(t)$ is obtained. The window function $w(t)$ defines a continuous and linearly independent function $f = s_1, s_2, ..., s_N$. Through a finite linear combination, the function $f(t)$ is optimized, and $\psi(t)$ is defined as

$$\psi(t) = \sum_i \beta_i s_i \quad (4)$$

where β_i are the coefficients, τ_i is the translation factor, and N is the total number of basis functions in the linear combination. This linear combination is optimized to minimize the sum of squared errors. The resulting $\psi(t)$ is a wavelet function that can be used for signal analysis and processing.

To summarize, in order for a function $\psi(t)$ to be considered a wavelet, it must meet the following conditions: $\psi(t) \in L^1 \cap L^2$ and $t\psi(t) \in L^1$. The admissibility condition is then used to ensure that the function is in L^2, which is defined as

$$\int_0^\infty \frac{|\hat{\psi}(\lambda)|^2}{\lambda} d\lambda = \int_{-\infty}^0 \frac{|\hat{\psi}(\lambda)|^2}{\lambda} d\lambda < +\infty \quad (5)$$

where $\hat{\psi}$ is the Fourier Transform of ψ. By plugging $\lambda = e^{-\nu}$ in Equation 5, we obtain

$$\int_{-\infty}^{+\infty} |\psi(-e^{-\nu})|^2 d\nu \quad (6)$$

The discretization of Equation 6 is done logarithmically, where the number of discretization points between $a = 1$ and $a = 10$ is the same as the number of discretization points between $a = 10$ and $a = 100$. Substituting $\omega = a_0^m \delta$ and $\frac{d\omega}{\omega} = \frac{d\delta}{\delta}$, we obtain the following equation:

$$\int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega = \int_1^{a_0} \left\{ \sum_{m=-\infty}^{\infty} |\hat{\psi}(a_0^m \delta)|^2 \right\} \frac{d\delta}{\delta}; \quad (7)$$

where $0 < A \leq \sum^\infty m = -\infty |\hat{\psi}(a_0^m \delta)|^2 \leq B < \infty$. The sum of dilated spectra (S_D) is the name given to the bounded quantity $S_D(\psi, a_0)(\lambda) = \sum m = -\infty^{+\infty} |\hat{\psi}(a_0^m \delta)|^2$, and the Frame Property is the boundary condition. This property is also employed in inversion. The construction of a new wavelet can be summarized by the following steps:

1) Optimize the function $f(t)$ using equation 3 and obtain the function $\psi(t)$.
2) Ensure that $\psi(t)$ meets the admissibility condition defined in equation 5.
3) Discretize the admissibility condition logarithmically using equation 7.
4) Ensure that the sum of dilated spectra S_D is bounded between A and B.
5) Apply the frame property to obtain the boundary condition.
6) Invert the wavelet to obtain the scaling function.

Figure 2 provides a graphical illustration of the steps involved in constructing a new wavelet ψ using the admissibility condition and compact support properties. The construction of such a wavelet is crucial in computing the Multilevel Wavelet Packet Entropy (M_E) of a signal, which extracts features from the signal $s(t)$ (defined in para2) using the wavelet ψ and the Wavelet Packet Decomposition (WPD). The WPD of the signal $s(t)$ at different scales $(j, j = 1, 2, ..., M)$ using wavelet ψ (eq. 4) with the band (n) and surge (k) parameters is defined as:

$$d_{(j,n)}(k) = \sqrt{2} \int_{-\infty}^{\infty} s(t) \psi_n(2^{-j}t - k) dt \quad (8)$$

Multiple entropies are calculated for the signal $s(t)$, including energy entropy (e_E), power entropy (p_E), and slope entropy (s_E). The signal $s(t)$ is divided into $k =$

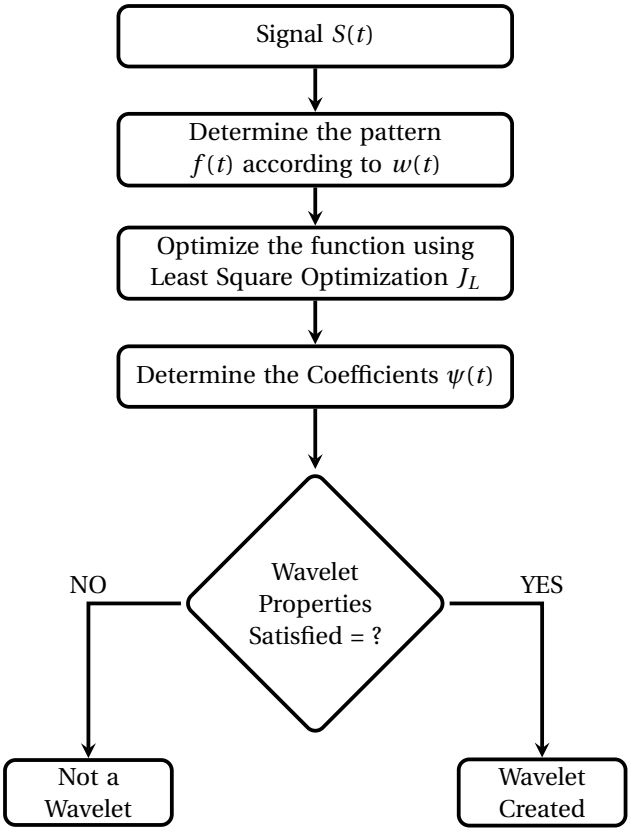

Fig. 2: Construction of Wavelet using Least Square Optimization.

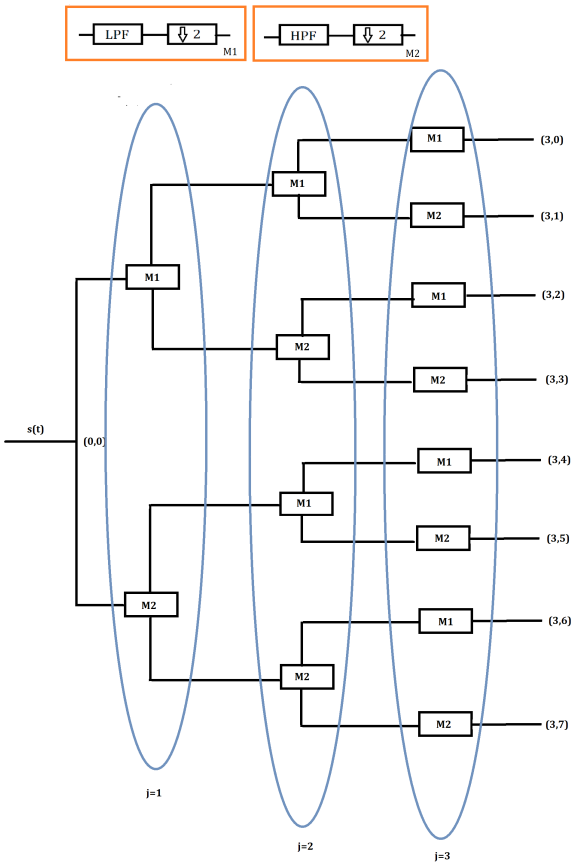

Fig. 3: Wavelet Packet Decomposition of signal $s(t)$ at level 3.

$N-m+1$ sub-sequences using a window $w(t)$, where m is the size of the window $(2<m<N)$. The resulting signal sub-sequences are defined as:

$$S_k = \{s_k, s_{k+1}, \ldots, s_N\} \tag{9}$$

All sub-sequences $S_1, S_2, \ldots, S_k$ contain m elements. The Wavelet Packet Decomposition (WPT) is applied to each sub-sequence S_k using equation 8, which results in 2^m coefficient nodes at each decomposition level. At each sub-band, the energy entropy (e_E), power entropy (p_E), and slope entropy (s_E) are calculated. The WPT coefficient obtained after decomposition is defined as the set of all calculated entropy values at all sub-bands.

$$d_{(j,n)}(k) = \sqrt{2}\int_{-\infty}^{\infty} S_k \psi_n(2^{-j}t-k)dt \tag{10}$$

The energy $E_{(j,n)}$ of the decomposed signal $d_{(j,n)}(k)$ at each sub-band is calculated, and the total energy E_T at each level of decomposition is computed using $E_T = \sum_n E_{(j,n)}$. For each sub-band, the relative energy $R_{e(j,n)}$ is defined as $R_{e(j,n)} = \frac{E_{(j,n)}}{E_T}$. Finally, the energy entropy (e_E) is defined as:

$$e_E^j = -\sum_n R_{e(j,n)}\ ln\ (R_{e(j,n)}) \tag{11}$$

Similarly, the power entropy (p_E) is calculated as

$$p_E^j = -\sum_n R_{p(j,n)}\ ln\ (R_{p(j,n)}) \tag{12}$$

The slope $S_{(j,n)}$ of decomposed signal $d_{(j,n)}(k)$ is calculated for each sub-band. Let S_1, S_2, ... , S_k be obtained at scale (j,n) and contain m elements, such as $S_1 = \{s_1, s_2, \ldots, s_m\}$. Subtract the two adjacent sub-sequences to create new k sub-sequences H_k and is defined as

$$H_k = \{h_k, h_{k+1}, \ldots, h_{N-1}\} \tag{13}$$

H_k contains $m-1$ elements with $h_k = s_{k+1} - s_k$. Two thresholds T_{h1} and T_{h2} are determined and compared to all the sequences obtained from eq. 13 with different combinations of T_{h1} and T_{h2} $(-T_{h2}, -T_{h1}, T_{h1}$, and $T_{h2})$ and form new sequence L_k.

$$L_k = \begin{cases} -2; & h_k < -T_{h2} \\ -1; & -T_{h2} < h_k < -T_{h1} \\ 0; & -T_{h1} < h_k < T_{h1} \\ 1; & T_{h1} < h_k < T_{h2} \\ 2; & h_k > T_{h2} \end{cases} \tag{14}$$

Following the comparison, the new sequence is defined as

$$L_k = \{l_k, l_{k+1}, \ldots, l_{N-1}\} \tag{15}$$

The L_k sequence contains five distinct values in it (-2, -1, 0, 1, and 2). Different combinations are formed

$g = 5^{m-1}$. And each type of combination is referred to as $r_1, r_2, \ldots, r_g$, with frequency R_s. It calculates each type of combination at each level of decomposition (j,n). For example, when $m = 4$, there will be 125 L_k combinations, such as $\{-2,-2,-2\}$, $\{-2,-2,-1\}$, …, $\{0,0,0\}$, $\{0,0,1\}$, …, $\{2,2,1\}$, $\{2,2,2\}$ and is defined as

$$R^g_{s(j,n)} = \frac{r_g}{k} \tag{16}$$

The Slope Entropy s_E is defined as

$$s^j_E = -\sum_g R^g_{s(j,n)} \; ln \; (R^g_{s(j,n)}) \tag{17}$$

The features of the signal $s(t)$ are calculated for each sub-band, and M_E is obtained using equation 18. The resulting matrix is then used as input for the clustering algorithms that classify the signal. M_E is defined as follows:

$$M_E = \begin{bmatrix} e^{1,j}_E & p^{1,j}_E & s^{1,j}_E \\ e^{2,j}_E & p^{2,j}_E & s^{2,j}_E \\ \vdots & \vdots & \vdots \\ e^{N,j}_E & p^{N,j}_E & s^{N,j}_E \end{bmatrix} \tag{18}$$

The feature matrix M_E is then passed to various clustering algorithms for signal classification. The clustering algorithms used in this study include DBSCAN, Agglomerative, Fuzzy-C Means, Mean-Shift, BIRCH, and Gaussian Mixture Model. These algorithms are defined in equations 19, 20, and 21.

$$J_F = \sum_{i=1}^{K} \sum_{j=1}^{n} \mu^m_{i,j} ||M^j_E - A_i||^2 \tag{19}$$

$$J_M = \frac{1}{nh^d} \sum_{n=1}^{N} k\left(\frac{M_E - M_{E(i)}}{h}\right) \tag{20}$$

$$Q(\theta^*, \theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk})[ln\pi_k + lnN(x_n|\mu_k, \Sigma_k)] \tag{21}$$

Statistical analysis is performed to compare the results of different wavelet-based clustering techniques. The performance of the techniques is evaluated using various metrics, including the silhouette measure (S_m), the Jaccard similarity index (J), the Rand Index (R), and the Area under the ROC Curve (AUC).

The S_m is calculated using Cohesion and Separation. It is used to ensure that data in a cluster is compact and tight. The clustering algorithm with the highest S_m and the lowest SSE is chosen as the best method. The sum of Cohesion and Separation is SSE. If a is the cohesion value and b is the separation value, then S_m is calculated as follows:

$$S_m = \frac{b-a}{max(b,a)} \tag{22}$$

The S_m result is determined by: -1 (very poor clustering), 0 (good clustering with room for improvement), and 1. (clustering is very good). The labels are assigned to different data points. They calculate the Jaccard similarity index (J). To compare the results, various parameters are used. TN: data points from the same class but from different clusters. TP: Data points belonging to the same class and cluster. FN: Data points in the same class but in different clusters, FP: Data points in the same class but in the same cluster. Table I displays the confusion matrix for the given data. The F-score (F), Precision (P), Accuracy (A), and J are then calculated.

TABLE I: Confusion matrix

	Class 1	**Class 2**
Positive Prediction	True positive (TP)	False positive (FP)
Negative Prediction	False negative (FN)	True negative (TN)

$$J = \frac{TP}{TP+TN+FP+FN} \quad P = \frac{TP}{TP+FP}$$
$$A = \frac{TP+TN}{TP+TN+FN+FP} \quad F = 2 \times \frac{P \times S}{P+S} \tag{23}$$

The Rand Index (R) is used to compare the similarities of clustering algorithms.

$$R = \frac{TP+TN}{TP+TN+FP+FN} \tag{24}$$

The Area under the ROC Curve (AUC) measures the binary classifier's performance. It displays the classifier model's performance. The curve represents the relationship between the True Positive Rate and the False Positive Rate. Sensitivity (S) is the ratio of true positives out of positive conditions, while Specificity (SP) is the ratio of false positives out of negative conditions. The AUC is calculated as follows.

$$S = \frac{TP}{TP+FN} \qquad SP = \frac{TN}{FP+TN} \tag{25}$$

III. Result and Discussion

WaveClust's performance is evaluated and compared with existing methods using a dataset of 48 ECG signals [63, 64], consisting of 25 normal and 23 abnormal ECG signals. The pre-processing of the signal is done to remove the DC components from the signal [65 - 66], and WPT is applied to the newly obtained signal, decomposed to level 5.

Three types of entropies e^j_E, p^j_E, and s^j_E are used as the features of the ECG signals. The combination type m is set to 4, and the threshold parameters T_{h1} and T_{h2} for the s^j_E entropy are set to 0.40 and 0.80, respectively. The number of classes c is set to 2 for all clustering algorithms, and a window size of 8000 is found to be best suited for ECG signal clustering.

Different wavelets such as $haar$, $db2$, $db4$, $db8$, $coif4$, $sym4$, $bior2.8$, and *WaveClust* wavelet are used with different window sizes 100, 500, 1000, 5000, 8000, 10000, 20000, and 50000, and their sensitivity is calculated and compared. Figure 4 shows the sensitivity

with different window sizes. The average sensitivity for different clustering algorithms for the window size 8000 with the mentioned wavelets is as follows: 0.9264 for $haar$, 0.9668 for $db2$, 0.9687 for $db4$, 0.8958 for $db8$, 0.9411 for $coif4$, 0.9471 for $sym4$, 0.9543 for $bior2.8$, and 0.9845 for *WaveClust* wavelet.

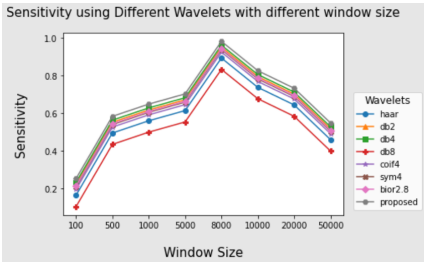

Fig. 4: Sensitivity value using different Wavelets with different window sizes to cluster ECG Signals using *WaveClust*.

Different parameters are employed to compare the performance of clustering algorithms. One such parameter is the S_m score, which is used to determine the optimal number of clusters for data clustering. In this study, the S_m score is calculated for 2, 3, 4, 5, and 6 clusters, resulting in scores of 0.8994, 0.5675, 0.5580, 0.3827, and 0.3681, respectively. Among these scores, the S_m value for 2 clusters is found to be the highest, indicating that the clustering algorithm performs best when the data is divided into two clusters.

Fig. 5: S_m score with different number of clusters.

Figure 6 displays the overall accuracy of *WaveClust* with different wavelets. Among the clustering algorithms used, FCM performs the best. The accuracy of FCM with various wavelets including $Haar+FCM$, $db2+FCM$, $db4+FCM$, $db8+FCM$, $coif4+FCM$, $sym4+FCM$, $bior2.8+FCM$, and $WaveClust+FCM$ are 90.32 %, 96.70 %, 96.80 %, 84.26 %, 93.47 %, 94.68 %, 94.68 %, and 98.94 % respectively.

The graph in Figure 7 displays the AUC scores for various wavelets used in ECG signal clustering. The *WaveClust* wavelet yields the highest AUC of 0.9756, while the $db8$ wavelet yields the lowest AUC of 0.7333. The values of J, R, and $F-1$ for different combinations of wavelets and clustering algorithms are provided in the appendix section (see Table II). According to the table, the FCM clustering algorithm outperforms the others in terms of classifying ECG signals across all wavelets. The average J value for clustering algorithms

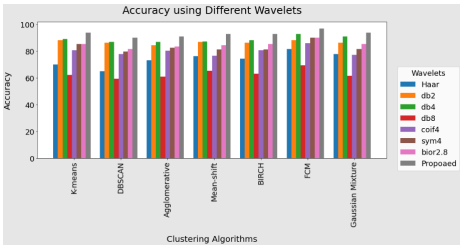

Fig. 6: *WaveClust* Accuracy with different existing Wavelets and Clustering Algorithms to classify the ECG signals.

with $haar$, $db2$, $db4$, $db8$, $coif4$, $sym4$, $bior2.8$, and *WaveClust* wavelets is 0.8110, 0.9077, 0.9226, 0.7261, 0.8571, 0.8794, 0.8973, and 0.9494, respectively. The highest J value is achieved with the *WaveClust* wavelet, followed by $db4$ and $db2$. Similarly, the average R value for clustering algorithms with $haar$, $db2$, $db4$, $db8$, $coif4$, $sym4$, $bior2.8$, and *WaveClust* wavelets is 0.8735, 0.9404, 0.9464, 0.8095, 0.8988, 0.9107, 0.9241, and 0.9687, respectively.

In addition, the sensitivity and accuracy of *WaveClust* are compared with existing methods, such as SVM [67], ANN+PSO+FFNN [48], U-Net [68], KNN [69], CNN [47], and Shape+Feature [45]. The sensitivity of Shape and Feature, CNN, ANN+PSO+FFNN, U-Net, KNN, and *WaveClust* is 98.35, 95.47, 92, 92.81, and 98.94, respectively. The accuracy of Shape and Feature, CNN, ANN+PSO+FFNN, U-Net, KNN, SVM, and *WaveClust* is 97.94, 98.25, 93.60, 98.49, 96.26, 96.00, and 98.94, respectively.

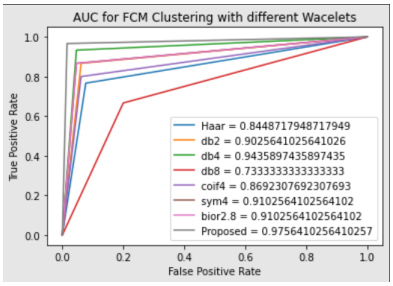

Fig. 7: FCM Clustering AUC with different Wavelets.

IV. Conclusion

In this paper, we propose *WaveClust*, a new approach to identify patterns in TS data. The method uses multilevel WPE to extract features and identify patterns in TS data. *WaveClust* clusters TS data using WEC. It solves the limitations of the state-of-art WEC technique. *WaveClust* uses WST with entropy. It provides a systematic and rigorous mathematical framework. It creates wavelets tailored to the type of signal under consideration. It also ensures all wavelet properties. These are orthogonality, scaling properties, compact support, and perfect reconstruction. LSO optimizes the constructed wavelet. Further, the constructed wavelet decomposes the TS data using WPT. It extracts entropy features at each decomposition level of the TS data. It

then obtains a wavelet coefficient matrix, representing the data on a new basis. Clustering algorithms classify this coefficient matrix. The resulting clusters identify data points with similar characteristics or patterns.

Researchers tested the efficiency of *WaveClust* with classical clustering algorithms. They also compare the results with existing wavelets and established methods. The results demonstrated that *WaveClust* outperformed established methods. It achieves better sensitivity and accuracy. The findings suggest that *WaveClust* can identify patterns in TS data. It can be a valuable tool in various applications where PR task is essential. These are image and signal processing, bioinformatics, and finance. Future work on *WaveClust* can involve testing it on real-world datasets. Further, it can aid its potential applications in diverse fields.

Author Contribution

SS, PM, and AR have designed the study, developed the methodology, performed the analysis, and written the manuscript. They have contributed equally to this work.

Funding

This research is the outcome of the SERB Project: **MTR/2019/000432**. We are thankful to SERB for providing financial support.

REFERENCES

1) Datao Xu. et al., "A new method proposed for realizing human gait pattern recognition: Inspirations for the application of sports and clinical gait analysis," Gait Posture (Elsevier), vol. 107, no. 1, pp. 293-305, 2024.
2) Ngou, B.P.M. et al., "Evolutionary trajectory of pattern recognition receptors in plants," Nat. Commun. vol. 15, no. 1, pp. 308, 2024.
3) M. Flasiński et al., "Multi-Derivational Parsing of Vague Languages - the New Paradigm of Syntactic Pattern Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 1, no. 1, pp.1-7, 2024.
4) Xinrong Zhi et al., "Hybrid tactile sensor array for pressure sensing and tactile pattern recognition," Nano Energy (E;sevier), Vol. 125, no. 1, pp. 109532, 2024.
5) Evans, C.G. et al., "Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly," Nature, vol. 625, no. 1, pp. 500–507, 2024.
6) Gupta, R. et al., "Novel similarity measure between hesitant fuzzy set and their applications in pattern recognition and clustering analysis," J. Eng. Appl. Sci., vol. 71, no. 1, pp. 1-5, 2024.
7) Ahmed, S. et al., "Transformers in Time-Series Analysis: A Tutorial," Circuits Syst Signal Process., vol. 42, no. 1, pp. 7433–7466, 2023.
8) Runge, J. et al., "Causal inference for time series," Nat Rev Earth Environ., vol. 4, no. 1, pp. 487–505, 2023.
9) Haq QMu et al., "A Hybrid Hand-Crafted and Deep Neural Spatio-Temporal EEG Features Clustering Framework for Precise Emotional Status Recognition," Sensors, vol. 22, no. 14, pp. 5158, 2022.
10) Y. -T. Chen et al., "An Effective Entropy-Assisted Mind-Wandering Detection System Using EEG Signals of MM-SART Database," IEEE J. of Biomedical and Health Informatics, vol. 26, no. 8, pp. 3649-3660, 2022.
11) Elvira-Ortiz DA et al., "An Entropy-Based Condition Monitoring Strategy for the Detection and Classification of Wear Levels in Gearboxes," Entropy, vol. 25, no. 3, pp. 424, 2023.
12) R. Lu et al., "HTDet: A clustering method using information entropy for hardware Trojan detection," Tsinghua Science and Technology, vol. 26, no. 1, pp. 48-61, 2021.
13) X. Yang et al., "Entropy-based Thunderstorm Imaging System with Real-time Prediction and Early Warning," IEEE Trans. on Instrumentation and Measurement, vol. 1, no. 1, pp. 1-6, 2022.
14) Z. Wang et al., "Identification of Series Arc Fault Occurred in the Three-Phase Motor With Frequency Converter Load Circuit via VMD and Entropy-Based Features," IEEE Sensors J., vol. 22, no. 24, pp. 24320-24332, 2022.
15) J. Hu et al., "TW-Co-MFC: Two-level weighted collaborative fuzzy clustering based on maximum entropy for multi-view data," Tsinghua Science and Technology, vol. 26, no. 2, pp. 185-198, 2021.
16) H. Wei et al., "Comparative research on noise reduction of transient electromagnetic signals based on empirical mode decomposition and variational mode decomposition," Radio Science, vol. 56, no. 10, pp. 1-19, 2021.
17) G. Yao et al., "Clustering of Typical Wind Power Scenarios Based on K-Means Clustering Algorithm and Improved Artificial Bee Colony Algorithm," IEEE Access, vol. 10, no. 1, pp. 98752-98760, 2022.
18) Y. Dong et al., "Bidirectional Denoising Autoencoders-Based Robust Representation Learning for Underwater Acoustic Target Signal Denoising," IEEE Trans. on Instrumentation and Measurement, vol. 71, no. 1, pp. 1-8, 2022.
19) S. Tan et al., "High-Quality Signal Denoising and Deep Feature Representation and Its Application to the ENPEMF," IEEE Geoscience and Remote Sensing Letters, vol. 19, no. 1, pp. 1-5, 2022.
20) R. Ranjan et al., "Motion Artifacts Suppression From EEG Signals Using an Adaptive Signal Denoising Method," IEEE Trans. on Instrumentation and Measurement, vol. 71, no 1, pp. 1-10, 2022.
21) Z. Song et al., "MSFYOLO: Feature fusion-based detection for small objects," IEEE Latin America

Trans., vol. 20, no. 5, pp. 823-830, 2022.
22) Z. Yang et al., "Road Extraction From Satellite Imagery by Road Context and Full-Stage Feature," IEEE Geoscience and Remote Sensing Letters, vol. 20, no. 1, pp. 1-5, 2023.
23) Junhao Zhu et al., "ECG heartbeat classification based on combined features extracted by PCA, KPCA, AKPCA and DWT," IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), vol. 1, no. 1, pp. 1-7, 2022.
24) Dai, X. et al., "Wavelet sets in Rn," I. J. Fourier Anal. Appl., vol. 3, no. 4, pp. 451–456, 1997.
25) Prasadini Mahapatra et al., "Construction of MRA and non-MRA wavelet sets on Cantor dyadic group," Elsevier Masson SAS, vol. 1, no. 1, pp. 1-8, 2021.
26) Qi R et al., "Effects of Different Culture Densities on the Acoustic Characteristics of Micropterus salmoide Feeding," Fishes. vol. 8, no. 3, pp. 126, 2023.
27) Parastoo Kamranfar et al., "Pavement Distress Recognition via Wavelet-Based Clustering of Smartphone Accelerometer Data," J. of Computing in Civil Engineering Vol. 36, no. 4, pp. 1-8, 2022.
28) Haq QMu et al., "A Hybrid Hand-Crafted and Deep Neural Spatio-Temporal EEG Features Clustering Framework for Precise Emotional Status Recognition," Sensors, vol. 22, no. 14, pp. 5158, 2022.
29) S. N. V. Bramareswara Rao, et al., "Fault Detection in Cluster Microgrids of Urban Community using Multi-Resolution Technique Based Wavelet Transforms," I. J. of Renewable Energy Research, Vol. 12, no.3, pp. 1-7, 2022.
30) Rahul Dubey et al., "Ball-Bearing Fault Classification Using Comparative Analysis of Wavelet Coefficient based on Entropy Measurement", IETE J. of Research, vol. 1, no. 1, pp. 1-8, 2022.
31) Wang, SH. et al., "Community-Acquired Pneumonia Recognition by Wavelet Entropy and Cat Swarm Optimization," Mobile Netw Appl, vol. 1, no. 1, pp. 1-8, 2022.
32) Mahlberg JA et al., "Pavement Quality Evaluation Using Connected Vehicle Data," Sensors, vol. 22, no. 23, pp. 9109, 2022.
33) J. Morlet, "Sampling theory and wave propagation," Proc. 51st Annu. Proc., Meeting Soc. Explor. Geophys., vol. 1, no. 1, pp. 233–261, 1981.
34) M. M. Jam et al., "Design and evaluation of optimal orthogonal wavelet with the least length of wavelet filters using spectral matching," IEEE Access, vol. 6, no. 1, pp. 57414–57424, 2018.
35) M. Xu et al., "Graph Neural Networks With Lifting-Based Adaptive Graph Wavelets," IEEE Trans. on Signal and Information Processing over Networks, vol. 8, no. 1, pp. 63-77, 2022.
36) Daniela Lanz et al., "Graph-Based Compensated Wavelet Lifting for Scalable Lossless Coding of Dynamic Medical Data", IEEE Tran. on Image Processing, vol. 29, no. 1, pp. 1-7, 2019.
37) M. B. Nagare et al., "On the Design of Biorthogonal Halfband Filterbanks With Almost Tight Rational Coefficients", IEEE Trans. on Circuits and Systems II: Express Briefs, vol. 67, no. 4, pp. 1-7, 2020..
38) David B. H. et al., "Almost Tight Rational Coefficients Biorthogonal Wavelet Filters," IEEE Signal Processing Letters, vol. 25, no. 6, pp. 1-12, 2018.
39) Yiran Oiao et al., "Research on Hybrid Electric Vehicle Torque Observer Based on Discrete Wavelet Transform and Least Squares Support Vector Machine Joint Optimization Theory," Chinese Automation Congress (CAC), vol. 1, no. 1, pp. 1-8, 2019.
40) Yajun Tian et al., "Super-Resolution Optimal Basic Wavelet Transform and Its Application in Thin-Bed Thickness Characterization," IEEE Trans. on Geoscience and Remote Sensing, vol. 60, no. 1, pp. 1-8, 2022.
41) M. Wasimuddin et al., "Stages-Based ECG Signal Analysis From Traditional Signal Processing to Machine Learning Approaches: A Survey," IEEE Access, vol. 8, no. 1, pp. 177782-177803, 2020.
42) J. Wang et al., "Towards Interpretable Arrhythmia Classification With Human-Machine Collaborative Knowledge Representation," IEEE Trans. on Biomedical Engineering, vol. 68, no. 7, pp. 2098-2109, 2021.
43) L. Zhang et al., "Two-Phase Multivariate Time Series Clustering to Classify Urban Rail Transit Stations," IEEE Access, vol. 8, no. 1, pp. 167998-168007, 2020.
44) Digambar V. Puri et al., "Automatic detection of Alzheimer's disease from EEG signals using low-complexity orthogonal wavelet filter banks," Biomedical Signal Processing and Control, Vol. 81, no. 1, pp. 1-8, 2023.
45) S. Lee et al., "Efficient Template Cluster Generation for Real-Time Abnormal Beat Detection in Lightweight Embedded ECG Acquisition Devices," IEEE Access, vol. 9, no. 1, pp. 70596-70605, 2022.
46) Huan Wang et al., "Interpretable convolutional neural network with multilayer wavelet for Noise-Robust Machinery fault diagnosis," Mechanical Systems and Signal Processing, Vol. 195, no. 1, pp. 1-9, 2023,
47) L. H. Wang et al., "Automated classification model with OTSU and CNN method for premature ventricular contraction detection," IEEE Access, vol. 9, no. 1, pp. 156581-156591, 2021.
48) S. Sahay et al., "Detection and Classification of ECG Signal through Machine learning," I. J. of Innovative Technology and Exploring, vol. 8, no. 10, pp. 3221–3227, 2019.
49) Singhai, P. et al., "ECG Signal Compression Based on Optimization of Wavelet Parameters and Threshold Levels Using Evolutionary Techniques," Circuits Syst Signal Process., vol. 42, no. 1, pp. 3509–3537, 2023.

50) Swiftset, "Wavelet Set", msc 65T60, 2013.
51) S. Mallat, "Wavelet set theory: advantages for wavelet construction," IEEE Signal Processing Magazine, vol. 38, no. 5, pp. 51-59, 2021.
52) H. Wang et al., "Wavelet set analysis of nonstationary signals," IEEE Trans. on Instrumentation and Measurement, vol. 67, no. 10, pp. 2321-2329, 2018.
53) A. Cohen et al., "Biorthogonal Bases of Compactly Supported Wavelets," Comm. Pure Appl. Math., vol. 45, no. 5, pp. 485-560, 1992.
54) Y. Meyer, "Orthonormal Wavelet Sets That Generate Continuous Wavelets," Applied and Computational Harmonic Analysis, vol. 6, no. 1, pp. 19-42, 1999.
55) D. Donoho et al., "Adapting to Unknown Smoothness via Wavelet Shrinkage," J. of the American Statistical Association, vol. 90, no. 432, pp. 1200-1224, 1995.
56) J. Bertrand et al., "Wavelets on the Interval and Fast Wavelet Transforms," IEEE Trans. on Signal Processing, vol. 41, no. 12, pp. 3394-3412, 1993.
57) Achmad Rizal et al., "Phonocardiogram Classification using Multilevel Wavelet Packet Entropy and Random Forest," I. Conf. on Science and Technology (ICST), vol. 6, no. 1, pp. 1-5, 2022.
58) Cuesta-Frau, D., "Slope Entropy: A New Time Series Complexity Estimator Based on Both Symbolic Patterns and Amplitude Information," Entropy, vol. 21, no. 1, oo. 1167, 2021.
59) P. Bradley et al., "Refining Initial Points for K-means clustering," 15th Int. Conf. on Machine Learning, vol. 15, no. 1, pp. 91-99, 1998.
60) T. Kanungo, "An efficient k-means clustering algorithm: An analysis and implementation," IEEE Trans. on Pattern Analysis and Machine Learning, vol. 24, no. 7, pp. 881-892, 2000.
61) Han, J. et al., "Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2001.
62) Mitchell T., "Machine Learning?, McGraw Hill Publishers, 1997.
63) G. B. Moody et al., "PhysioNet: a web-based resource for the study of physiologic signals," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 70-75, 2001.
64) G. B. Moody et al., "The impact of the MIT-BIH arrhythmia database," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45-50, 2001.
65) Z. Guo et al., "Multiscale Wavelet Transfer Entropy With Application to Corticomuscular Coupling Analysis," IEEE Trans. on Biomedical Eng., vol. 69, no. 2, pp. 771-782, 2022.
66) R. Dekimpe et al., "ECG Arrhythmia Classification on an Ultra-Low-Power Microcontroller," IEEE Trans. on Biomedical Circuits and Systems, vol. 16, no. 3, pp. 456-466, 2022.
67) C. Venkatesan et al., "ECG signal preprocessing and SVM classifier-based abnormality detection in remote healthcare applications," IEEE Access, vol. 6, no. 1, pp. 9767–9773, 2018.
68) S. L. Oh et al., "Automated beat-wise arrhythmia diagnosis using modified U-Net on extended electrocardiographic recordings with heterogeneous arrhythmia types," Computers in Biology and Medicine, vol. 105, no. 1, pp. 92-101, 2019.
69) J. Chen et al., "Smart heart monitoring: early prediction of heart problems through predictive analysis of ECG signals," IEEE Access, vol. 7, no. 1, pp. 120831-120839, 2019.

APPENDIX

TABLE II: ECG Clustering performance comparision using *WaveClust* with different existing wavelets.

Wavelets	Statistics	K-means	DBSCAN	Hierarchy	Mean Shift	BIRCH	FCM	Gaussian mixture	Average
Haar	**P**	0.90	0.90	0.94	0.93	0.94	0.94	0.94	0.93
	S	0.90	0.90	0.92	0.95	0.93	0.95	0.93	0.93
	Acc	0.82	0.82	0.86	0.89	0.88	0.90	0.88	0.86
	SP	0.38	0.53	0.62	0.54	0.55	0.38	0.44	0.49
	F	0.90	0.90	0.93	0.94	0.93	0.95	0.94	0.93
	J	0.78	0.74	0.79	0.82	0.82	0.88	0.84	0.81
	R	0.83	0.83	0.88	0.90	0.89	0.91	0.89	0.87
db2	**P**	0.96	0.98	0.95	0.98	0.98	0.98	0.98	0.97
	S	0.98	0.96	0.97	0.96	0.98	0.99	0.95	0.97
	Acc	0.94	0.94	0.91	0.94	0.96	0.97	0.93	0.94
	SP	0.33	0.60	0.29	0.60	0.71	0.71	0.50	0.54
	F	0.97	0.97	0.96	0.97	0.98	0.98	0.96	0.97
	J	0.92	0.91	0.90	0.91	0.91	0.92	0.91	0.91
	R	0.94	0.94	0.92	0.94	0.96	0.97	0.93	0.94
db4	**P**	0.98	0.97	0.96	0.98	0.98	0.99	0.98	0.97
	S	0.97	0.98	0.97	0.96	0.96	0.98	0.98	0.97
	Acc	0.95	0.95	0.93	0.94	0.94	0.97	0.96	0.95
	SP	0.50	0.57	0.33	0.50	0.50	0.67	0.50	0.51
	F	0.97	0.97	0.96	0.97	0.97	0.98	0.98	0.97
	J	0.93	0.91	0.91	0.92	0.92	0.95	0.94	0.92
	R	0.95	0.95	0.93	0.94	0.94	0.97	0.96	0.95
db8	**P**	0.87	0.86	0.85	0.84	0.91	0.91	0.87	0.87
	S	0.90	0.87	0.86	0.97	0.86	0.91	0.90	0.90
	Acc	0.79	0.76	0.74	0.82	0.80	0.84	0.79	0.79
	SP	0.47	0.42	0.37	0.36	0.53	0.50	0.47	0.45
	F	0.88	0.86	0.85	0.90	0.89	0.91	0.88	0.88
	J	0.72	0.70	0.69	0.75	0.73	0.78	0.72	0.73
	R	0.81	0.78	0.76	0.83	0.81	0.85	0.81	0.81
coif4	**P**	0.95	0.94	0.93	0.94	0.94	0.98	0.94	0.95
	S	0.93	0.94	0.93	0.94	0.95	0.96	0.93	0.94
	Acc	0.89	0.89	0.87	0.89	0.90	0.93	0.88	0.89
	SP	0.43	0.55	0.25	0.50	0.44	0.67	0.44	0.47
	F	0.94	0.94	0.93	0.94	0.95	0.97	0.94	0.94
	J	0.86	0.83	0.85	0.84	0.86	0.90	0.84	0.86
	R	0.90	0.90	0.88	0.90	0.91	0.94	0.89	0.90
sym4	**P**	0.96	0.95	0.98	0.93	0.94	0.98	0.95	0.96
	S	0.96	0.93	0.92	0.97	0.93	0.97	0.95	0.95
	Acc	0.91	0.89	0.90	0.90	0.88	0.95	0.91	0.91
	SP	0.33	0.50	0.60	0.40	0.29	0.50	0.50	0.45
	F	0.96	0.94	0.95	0.95	0.94	0.97	0.95	0.95
	J	0.90	0.85	0.88	0.86	0.86	0.93	0.88	0.88
	R	0.92	0.90	0.91	0.91	0.89	0.95	0.92	0.91
bior2.8	**P**	0.98	0.93	0.98	0.95	0.98	0.98	0.97	0.96
	S	0.97	0.97	0.91	0.96	0.95	0.97	0.97	0.95
	Acc	0.95	0.90	0.89	0.91	0.92	0.95	0.93	0.92
	SP	0.67	0.33	0.33	0.17	0.60	0.50	0.57	0.45
	F	0.97	0.95	0.94	0.95	0.96	0.97	0.97	0.96
	J	0.91	0.88	0.89	0.90	0.90	0.93	0.90	0.90
	R	0.95	0.91	0.90	0.91	0.93	0.95	0.94	0.92
WaveClust	**P**	0.99	0.98	0.97	0.98	0.98	1.00	0.99	0.98
	S	0.99	0.97	0.98	0.99	0.99	0.99	0.99	0.98
	Acc	0.98	0.95	0.95	0.97	0.97	0.99	0.98	0.97
	SP	0.67	0.50	0.40	0.50	0.50	1.00	0.67	0.60
	F	0.99	0.97	0.97	0.98	0.98	0.99	0.99	0.98
	J	0.96	0.93	0.93	0.95	0.95	0.98	0.96	0.95
	R	0.98	0.95	0.95	0.97	0.97	0.99	0.98	0.97

***Note:** P = Precision, S = Sensitivity, Acc = Accuracy, SP = Specificity, F = F-1 Score, J = accard similarity index, R = Rand Index.