Swarm Intelligence Algorithm Based on Plant Root System in 1D Biomedical Signal Feature Engineering to Improve Classification Accuracy

Rui Gong[1,2,*], Kazunori Hase[2]

[1]Organization of Liberal Arts Education, Mejiro University, Tokyo, Japan

[2]Faculty of Systems Design, Tokyo Metropolitan University, Tokyo, Japan

Received 13 November 2022; received in revised form 02 February 2023; accepted 16 February 2023

DOI: https://doi.org/10.46604/aiti.2023.11169

Abstract

The classification accuracy of one-dimensional (1D) biomedical signals is limited due to the lack of independence of the extracted features. To address this shortcoming, the study applies a swarm intelligence algorithm based on plant root systems (PRSs) to feature engineering. Some basic features of 1D biomedical signals are integrated into a digitized soil, and a root matrix is generated from this digitized soil and the PRS algorithm. The PRS features are extracted from the root matrix and used to classify the basic features. Following classification with the same biomedical signals and classifier, the accuracy of the added PRS set is generally higher than that of the base set. The result shows that the proposed algorithm can expand the application of 1D biomedical signals to include more biomedical signals in classification tasks for clinical diagnosis.

Keywords: one-dimensional biomedical signal, feature engineering, plant root system algorithm, swarm intelligence

1. Introduction

One-dimensional (1D) biomedical signals are widely used in the medical field and are crucial for clinical diagnosis and treatment. Electrocardiograms (ECGs) are commonly used to diagnose heart-related diseases, and electroencephalograms (EEGs) are used to determine sleep quality or mental state [1]. Electrical-based biomedical signals perform well, and magnetic, vibration, and acoustic-based biomedical signals are popularly used in clinical applications [2-3]. However, with increasingly shorter sampling intervals and longer signal lengths, analyzing 1D biomedical signals without the assistance of a computer has become a research limitation.

Machine learning classification is a cutting-edge computer-assisted 1D technique for analyzing biomedical signals [4]. Supervised classification is a machine-learning technique that helps doctors improve diagnostic accuracy by training classifiers with features of biomedical signals [5]. This technique is effective for complex and large datasets, and a well-designed classifier training model combined with high-quality dataset features frequently generates unexpectedly good outcomes. However, feature extraction methods are not as versatile as model building; thus, related research has received less attention.

Traditional and classical methods are often used for the extraction of 1D biomedical signals. The curve length of the signal in the time domain and the total spectral power in the frequency domain are common classification features of ECG signals [6]. Meanwhile, Shannon entropy is utilized as a classification feature of electromyography (EMG) signals in clinical settings [7].

* Corresponding author. E-mail address: r.gong@mejiro.ac.jp

Although these features may seem adequate, the naive Bayes classifier, which is the basis of machine learning theory, requires each condition to be independent during the classification process. In other words, features of biomedical signals extracted from classical feature engineering techniques should not be correlated [8]. This independence between features is undoubtedly challenging to achieve because both time- and frequency-domain features can always be connected by certain operations and their inverse. Therefore, new features with strong independence are needed to characterize 1D biomedical signals.

After investigation, it was found that swarm intelligence computing might solve the above challenges. When a feature is extracted using swarm intelligence, the new feature has a distinguishing independence because each member of the swarm participates in making the choice. Moreover, the operations and transformations of the choice process are irreversible. The ant colony algorithm is a popular swarm intelligence algorithm, while the bee colony algorithm has been developed based on the swarming and social nature of bees [9]. Currently, dozens of swarm intelligence algorithms have been applied to feature selection [10]. However, they are all based on social animals, and the independence of extracted features is still lacking. Therefore, this study focuses on "social plants." The root systems of plants are complex enough to form group intelligence as well.

Holker et al. [11] hypothesized that the root system of plants is a type of swarm intelligence system. The root apices communicate with each other through electrical signals or chemical pheromones and build a massive "brain-like" system underground. Within such a complex root system, independent roots acquire water and nutrients through cooperation. To support proper growth and development, the root system competes with other root systems, beginning a "war" when it has an advantage and retreating when it is at a disadvantage [11].

A swarm intelligence algorithm based on a plant root system (PRS) is proposed and applied to feature engineering in this study. Specifically, the proposed PRS approach imitates the growth of plant roots, in which the root apices cooperate to absorb as much of the sustaining nutrients as possible.

In the proposed approach, the base features used by the selected machine learning model to classify biomedical signals are equivalent to the nutrients in digitized soil. These base features are obtained using common feature extraction methods. The sum of the nutrients absorbed during growth also serves as the first auxiliary feature, and the growth area of the root system is the second auxiliary feature. These auxiliary features are referred to as auxiliary PRS features. During model training, the most significant advantage of the PRS features based on swarm intelligence is that they are ideally independent (with a low correlation with the other features) and highly correlated with the dependent variable. The independence of the PRS features is derived from the brain-like structure of the root system, where the co-determination of the root tips reduces the correlation with the base features.

As an advantage, PRS auxiliary features of biomedical signals are obtained from the proposed algorithm. After appending these features to the training set, the correlation between the features decreases and leads to reduced bias and variance and improved accuracy in machine learning classification. Most importantly, the PRS algorithm is universally applicable for the feature engineering of most 1D biomedical signals. This technology also enables 1D biomedical signals to become more competitive in the clinical field by improving the accuracy of machine learning classification.

2. Method

2.1. Extraction of base features

The growth of natural plants requires suitable soil; similarly, the extraction of PRS features requires digitized soil (a feature matrix) composed of base features. When extracting base features, time-domain features are generally preferred over frequency-domain and nonlinear features. Time-domain features are more reliable than nonlinear features for involving a lower computational burden for nonstationary 1D biomedical signals. [12]. Moreover, time-domain features are more dependable than frequency-domain features because they avoid the risk of spectral leakage caused by signal decomposition [13].

Time-domain feature selection is essential for base feature extraction. The number of features needs to be carefully considered: a small number of features limits the growth of the root system, while a large number of features may adversely affect classifier performance [14]. For the base features, this study selected 12 features that have relatively low computational complexity and have been successfully used for classification [15-16]. The definitions of the base features are listed in Table 1, which x_n represents the 1D biomedical signal in the segment $\bar{x}$, N is the length of the signal, and $\bar{x}$ is the average value of x_n.

Table 1 Time-domain features

Base feature	Definition
Standard deviation (STD)	$\mathrm{STD}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|x_i-\frac{1}{N}\sum_{i=1}^{N}x_i\right\|^2}$
Variance of signal (VAR)	$\mathrm{VAR}=\frac{1}{N-1}\sum_{n=1}^{N}x_n^2$
Root mean square (RMS)	$\mathrm{RMS}=\sqrt{\frac{1}{N}\sum_{n=1}^{N}x_n^2}$
Skewness (SKW)	$\mathrm{SKW}=\frac{1}{N}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)^3\Big/\left(\frac{1}{N}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)^2\right)^{3/2}$
Kurtosis (KURT)	$\mathrm{KURT}=\frac{1}{N}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)^4\times\left(\frac{1}{N}\sum_{n=1}^{N}\left(x_n-\bar{x}\right)^2\right)^2$
Mean absolute value (MAV)	$\mathrm{MAV}=\frac{1}{N}\sum_{n=1}^{N}\left\|x_n\right\|$
Zero crossing (ZC)	$\mathrm{ZC}=\sum_{n=1}^{N-1}\left[\mathrm{sgn}\left(x_n\times x_{n+1}\right)\cap\left\|x_n-x_{n+1}\right\|\geq\mathrm{threshold}\right]$
Slope sign change (SSC)	$\mathrm{SSC}=\sum_{n=2}^{N-1}\left\{f\left[\left(x_n-x_{n-1}\right)\times\left(x_n-x_{n+1}\right)\right]\right\},\ f\left(x\right)=\begin{cases}1, & \text{if } x\geq\mathrm{threshold}\\ 0, & \text{otherwise}\end{cases}$
Willison amplitude (WAMP)	$\mathrm{WAMP}=\sum_{n=1}^{N-1}f\left(\left\|x_n-x_{n+1}\right\|\right),\ f\left(x\right)=\begin{cases}1, & \text{if } x\geq\mathrm{threshold}\\ 0, & \text{otherwise}\end{cases}$
Simple sign integral (SSI)	$\mathrm{SSI}=\sum_{n=1}^{N}\left\|x_n\right\|^2$
Nonlinear energy (NLE)	$\mathrm{NLE}\left(x_n\right)=\frac{1}{N-2}\sum_{i=2}^{N-1}x_n\left(i\right)^2-x_n\left(i-1\right)x_n\left(i+1\right)$
Waveform length (WL)	$\mathrm{WL}=\sum_{n=1}^{N-1}\left\|x_{n+1}-x_n\right\|$

Among the 12 features comprising the base features set, the statistical features are the most basic, including VAR, STD, RMS, SKW, KURT, and the MAV, which is similar to the average rectified value. Three features related to frequency are also included: ZC, i.e., the amplitude value of a signal that crosses the zero-amplitude axis; SSC, i.e., the frequency of a signal within the time domain; and the WAMP, i.e., the sum difference between the signal amplitudes for two adjacent samples. The last three features indicate energy and complexity: SSI represents signal segment energy, NLE approximates signal amplitude energy, and WL measures signal complexity. These twelve base features collectively represent a 1D biomedical signal series. The new features derived from these representative values closely approximate the actual measurement values.

2.2. Feature sorting

For the development of a root system, the site where the seed is planted requires nutritious soil. In this study, the base features were modeled based on the distribution of nutrients. Each element $\beta_{m,n}$ of the base feature set $B_{m\times n}\langle\ \rangle$ must be sorted. First, the dataset must be normalized using a scaling technique during the preprocessing stage to eliminate scale

differences among various elements while sorting. Min-max normalization was used to scale the data within the range [0,1] and maintain the relationships that existed among the original data [17]. The min-max normalization based on the feature set $\dot{B}_{m\times n}\langle\ \ \rangle$ can be expressed as:

$$\dot{B}_{m\times n}\langle\rangle = \frac{\beta_{m,n} - B\langle\rangle_{\min}}{B\langle\rangle_{\max} - B\langle\rangle_{\min}} \tag{1}$$

where n = 1, 2, …, 12 and m = 1, 2, …, the length of β_n.

Feature importance measurements can reduce the initial number of features by eliminating those features with a low importance score and improving classifier performance [18]. In this study, the features were sorted based on their worth. To score a feature's worth, the information associated with the feature was used. Features with higher scores were those closest to the center location. Information gain was first input into a decision tree and used to rank the priority of feature nodes; this was then expanded and applied to importance measurement.

Assuming the class label of a feature has two distinct values that define binary classes $\{C_1,C_2\}$. $\mathrm{Info}(\dot{B})$, also known as the entropy of $\dot{B}_{m\times n}\langle\ \ \rangle$, can be defined as

$$\mathrm{Info}(\dot{B}) = -p_1 \log_2(p_1) - p_2 \log_2(p_2) \tag{2}$$

where p_1, p_2 are the nonzero probabilities that an arbitrary tuple $\dot{B}_{m\times n}\langle\ \ \rangle$ belongs to classes C_1,C_2 and are estimated using $|C_{1\text{ or }2,\dot{B}}| / |\dot{B}|$.

Because each feature column $\dot{\beta}_n$ is a set of discontinuous values, these discrete values of $\dot{\beta}_n$ need to be split into two sets using an unsupervised algorithm. The K-means clustering method can partition a given dataset into k groups prespecified by the analyst [19]. In this study, this process was repeated 25 times to produce a lower within-cluster variation and a more stable result. Each $\dot{\beta}_{m\times n}$ now belongs to the split set $\{s_n^1, s_n^2\}$ and has a class label. The expected information required to classify the tuple from $\dot{B}$ based on feature $\dot{\beta}_n$ is

$$\mathrm{Info}_{\dot{\beta}_n}(\dot{B}) = \sum_{i=1}^{2} \frac{|\dot{B}_{s_n^i}|}{|\dot{B}|} \times \mathrm{Info}(\dot{B}_{s_n^i}) \tag{3}$$

$$\mathrm{Gain}(\dot{\beta}_n) = \mathrm{Info}(\dot{B}) - \mathrm{Info}_{\dot{\beta}_n}(\dot{B}) \tag{4}$$

The $\dot{\beta}_n$ with the highest information gain was chosen to replace the center of $\{\dot{\beta}_1, \dot{\beta}_2, \ldots, \dot{\beta}_{12}\}$ and the other features were sorted individually from the center to the two sides with decreasing information gain. Thus, a sorted base feature set was obtained.

2.3. Generating nutritious soil based on features

Digitized soil consisting of the sorted base feature set cannot only be wide; it must also have depth. Therefore, continuous feature discretization (from scalar to vector) is necessary. In this process, equal-width interval binning is the most common method used for discretizing data to produce nominal values from continuous features [20]. With this binning method, if feature a has values bounded by a_{max} and a_{min}, the method computes k equally sized bin widths as follows:

$$\delta = \frac{a_{\max} - a_{\min}}{k} \tag{5}$$

And it constructs bin boundaries at $a_{min} + j\delta$, where $j = 1, 2, \ldots, k-1$. To imitate the natural environment of the root system, this study set $k = 15$.

In addition, the vertical distribution of soil nutrients should follow the laws of nature. Most nutrients are concentrated in the shallowest layer and decrease with depth because nutrients return to the soil through biocycles. Thus, the discrete feature set $d_{15\times12}$ was arranged in decreasing order from top to bottom. By contrast, the horizontal distribution of nutrients is associated with crustal movement. The kernel convolution values τ_A and τ_B were used to calculate the reconstituted nutrient distributions.

Nutritional reconstructions are divided into two types: τ_A, which affects the nutrient distribution in shallower layers; and τ_B, which affects the nutrient distribution in deeper layers. The convolution kernel is then defined as:

$$\tau_A = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0.5 & 0.5 & 0.5 \end{bmatrix} \tag{6}$$

$$\tau_B = \frac{1}{2}\begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{7}$$

The nutritious soil used to generate the root system is written as $D_{15\times12}$, which is a result of the $d_{15\times12}$ convolution with kernels τ_A and τ_B. The calculated target matrix requires zero padding with a size of 1 before each convolution. The completed soil with the mineral nutrition generation process is shown in Fig. 1. The following matrix for the distribution of nutrients in the soil is referred to as the nutrient matrix. The elements of the base feature set are arranged in order of importance, from the center to either side. Each element in the set is then discretized. Finally, the soil matrix is obtained using two convolution calculations.

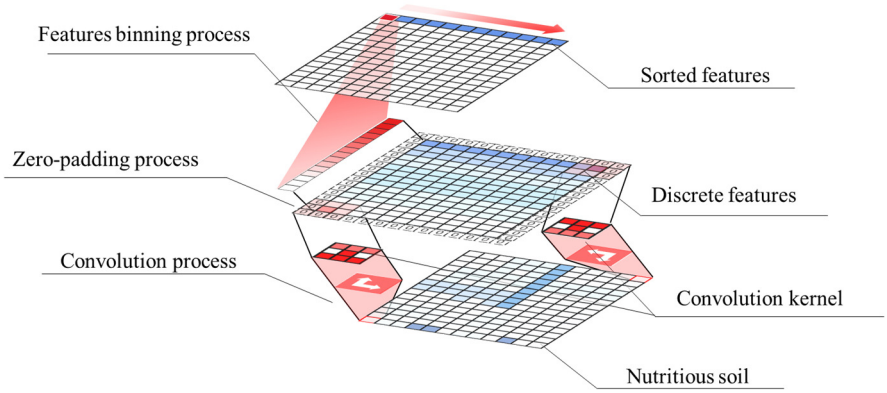


Fig. 1 Top-to-bottom order of soil matrix construction

2.4. Plant root system algorithm

In plants, either the radicle or primary root is the first organ to emerge from the seed coat. Before the first leaf grows, the energy required by the cotyledon for root development comes from the seed itself. Additional inorganic nutrients (including water) are essential. Only a small fraction of these nutrients come from the seed itself, and the rests are from the surrounding soil which contains mineral nutrients [21]. The rule that is inherent in the genes of higher plants is to develop a sufficiently large first leaf and long roots before the energy stored in the seed is exhausted. The proposed algorithm follows this rule. As organic matter cannot be synthesized by photosynthesis before the first leaf has grown, the task is to absorb more mineral nutrients by growing a sufficient large root system with limited energy (root division) in a limited time. The matrix distribution for the roots in the soil is hereafter referred to as the root matrix.

In the proposed approach, the aforementioned natural processes must be transformed into a data-processing program. First, the root and nutrient matrices form a four-dimensional tensor, as shown in Fig. 2. The initial root matrix is a zero matrix, where the upper-center location assigning values can create a radicle matrix. To absorb more nutrients, the rule of root division

is to preferentially proliferate at the global maximum in the mapping nutrient matrix. Neighboring root tips are also necessary. The radicle matrix, photosynthetic day, and daily root division are used as modification parameters. The developmental process can be characterized by the root distribution area and total nutrient absorption. Fig. 3 shows the natural and digital processes used to build a brain-like root system. The figure also depicts that the growth of digital roots is dependent on the base features in this study. By extracting the PRS features from the brain-like root system, the correlation with the base features is reduced.

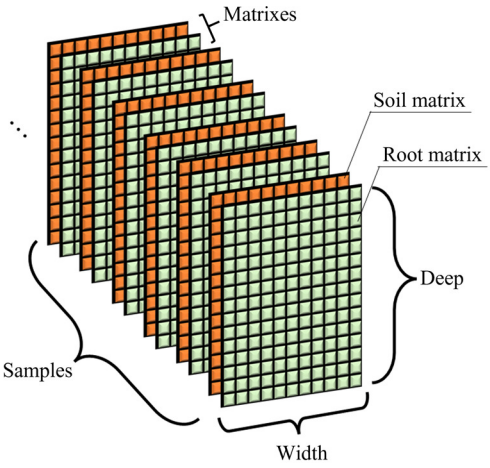


Fig. 2 Four-dimensional data tensor of biomedical signals

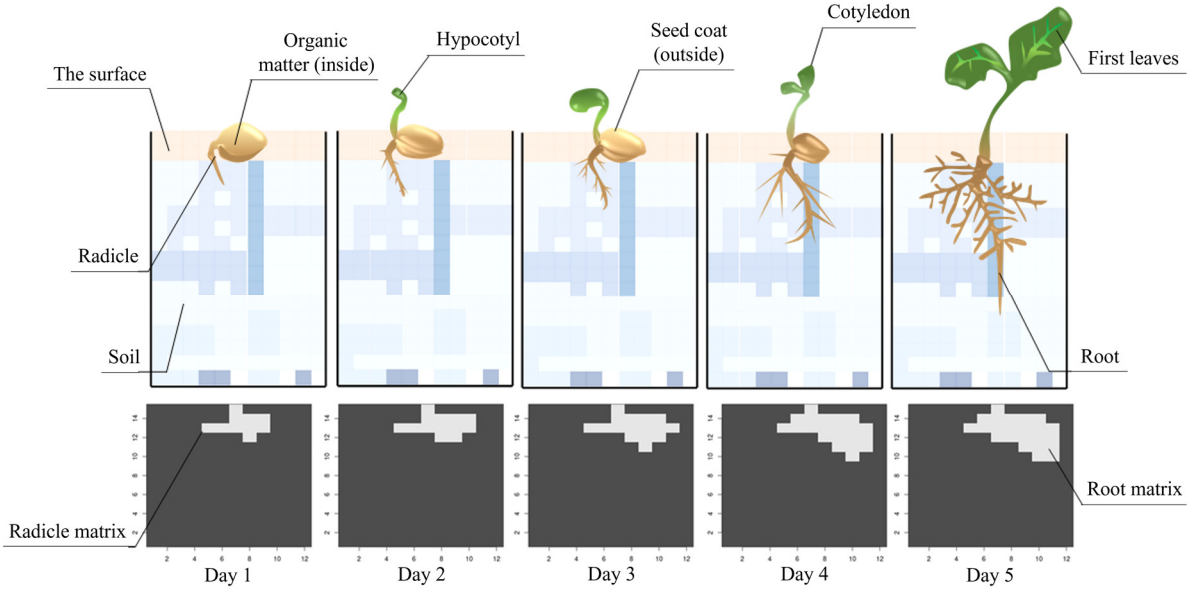


Fig. 3 Natural and digital processes for building brain-like root systems

In Table 2, the root growth process can be written as a pseudocode. In the proposed algorithm, the core steps of root development involve seeking the global maximum location of the nutrient matrix map relative to the root matrix location using a tensor of a biomedical signal base feature set. The entire root system makes decisions in a brain-like fashion to improve total nutrient absorption. The root distribution can be drawn in the coordinate system as a polygon. The area of this polygon serves as a feature. This study uses a well-known tool to calculate the polygon area proposed by Paul Bourke in 1988. The procedure provided two results, which are both referred to as PRS features: nutrient feature (NF), representing the total number of nutrients absorbed during root system development in the digitized soil; and root feature (RF), representing the area of root distribution. The calculations of NF and RF are shown in Table 2 as the nutrient amount (na) and area (A). A is calculated based on Eq. (8) and Fig. 4, wherein the polygon is closed and composed of line segments between N vertices (p_i, q_i), $i = 0$ to $N-1$.

$$A = \frac{1}{2}\sum_{i=0}^{N-1}\left(p_i q_{i+1} - p_{i+1} q_i\right) \tag{8}$$

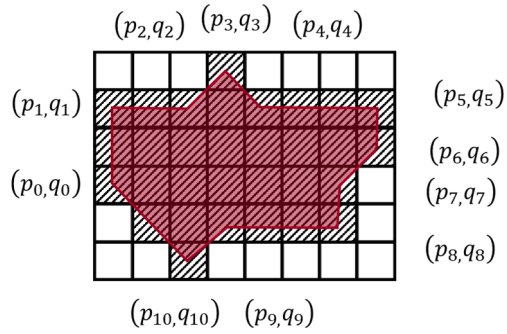


Fig. 4 Example of 10-vertex closed polygon constructed from a root matrix

Table 2 Plant root system algorithm

Algorithm 1 Plant root system		
	Pseudocode	Comments
1	Input:	
2	$\mathbb{DR}_{x\times y}$: distribution of the radicle in soil (initial root matrix);	x, y are the number of rows and columns of the distribution matrix
3	$\mathbb{DN}_{x\times y}$: distribution of nutrients in the soil (nutrient matrix);	
4	Output:	
5	na: nutrients absorbed by the roots;	
6	A: root distribution area;	
7	$na \leftarrow 0$;	
8	$\mathbb{R}_{x\times y} \leftarrow \mathbb{DR}_{x\times y}$;	$\mathbb{R}_{x\times y}$: distribution of the growing root in the soil (root matrix)
9	for $i = 1 \rightarrow day$ do	*day* is the number of days for the first leaf to appear
10	for each $\mathbb{R}_{[x^{th},y^{th}]} = 1$ do	$x^{th} \in (1,2,\dots,x)$, $y^{th} \in (1,2,\dots,y)$
11	$m \leftarrow \left\{\mathbb{DN}_{[x+1^{th},y^{th}]}, \mathbb{DN}_{[x-1^{th},y^{th}]}, \mathbb{DN}_{[x^{th},y+1^{th}]}, \mathbb{DN}_{[x^{th},y^{th}-1]}\right\}$;	$\mathbb{R}_{[x^{th},y^{th}]} \in \mathbb{R}_{x\times y}\mathbb{N}_{[x^{th},y^{th}]} \in \mathbb{DN}_{x\times y}$
12	return *m* to $\mathbb{M}$;	$\mathbb{M}$ is a matrix
13	end for each	
14	decreasing sort $\mathbb{M}$ to *M*;	*M* is a vector
15	for $j = 1 \rightarrow l$ do; $(l \leq length\ of\ M)$	l is a limit of root division in one day
16	$M[j]$ mapping back to $[x^{th}, y^{th}]$;	$[x^{th}, y^{th}]$is a location of the matrix
17	if $\mathbb{R}_{[x^{th},y^{th}]} \neq 1$ do	
18	$\mathbb{R}_{[x^{th},y^{th}]} \leftarrow 1$;	
19		
20	$na \leftarrow \begin{cases} 0, M[j] = 0 \\ \frac{M[j]}{1+\lvert M[i]\rvert} + 0.49, M[j] \neq 0 \end{cases}$	absorption rate
21		
22	end if	
23	do $na++$;	
24	return na and $\mathbb{R}_{x\times y}$	na is the first output
25	end for	
26	do $\dot{\mathbb{R}} \leftarrow \mathbb{R}$	root extraction as polygon
27	do $A \leftarrow area(\dot{\mathbb{R}})$;	$area$ of the polygon A is the second output
28	return *A;*	
29	end procedure	

Incorporating the extracted PRS features into the unprocessed base feature set produced a set for classification, which was composed of 14 features. Finally, the final feature set can be obtained by performing min-max normalization. The flowchart in Fig. 5 explains the PRS feature extraction process. The entire process is divided into two parts. The first involves creating a nutrient matrix with common features. The core idea is to develop digitized soil with an appropriate size and nutrient distribution through a mathematical calculation between features. The second part involves developing the root matrix using the nutrient matrix and seeds. The core idea is to limit the root division rule for radicles using parameters.

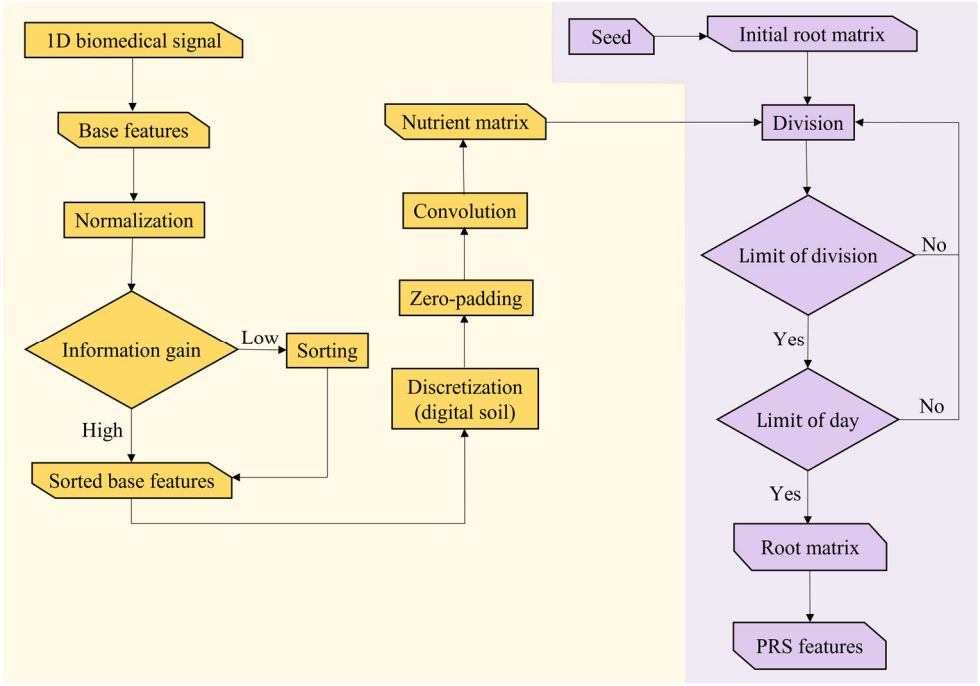


Fig. 5 The flow chart of the calculation of FRS features

3. Datasets and Evaluation Methods

The accuracies of various classifiers were considered for evaluation. The key measure for each classifier is the change in accuracy before and after adding root-based features to the original feature set. Accuracy is often used in epidemiology [22]. For a sample, a correct classification is true, and an incorrect classification is false. The accuracy is usually expressed as a percentage, calculated by dividing the total number of correct samples by the sum of all samples. A result closer to 100% will be more accurate.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{9}$$

where true positive (TP) and false positive (FP) are the number of true and false positive samples. True negative (TN) and false negative (FN), which are the number of true and false negative samples.

3.1. DatasetsU3

When selecting a dataset of biomedical signals, a nonstationary signal is preferred. Almost all biomedical signals are nonstationary because the human system is always in a state of dynamic equilibrium under the control of the brain [23]. Accordingly, three biomedical signals were selected: a vibroarthrographic (VAG) signal, the vibration signal generated by the friction between the knee cartilage during flexion-extension to detect knee joint disorders, captured by acceleration sensors near the kneecap [24]; forearm EMG, the electrical signal generated by the extensor carpi radialis longus during fist relaxation, captured by electrodes near the muscles of the small arm [25]; and the audio signal recorded by a smartphone web application to detect cough emanating from the lungs and airways of a patient who may be infected with COVID-19, captured by a microphone placed near the head [26]. All signals are binary, classified as young or old for the VAG signals, rest or fist for the EMG signals, and positive or negative for the cough sound signals. Table 3 presents the detailed characteristics of these signals.

Table 3 Properties of biomedical signals

Dataset	Signal type	Class	Sample size	Sampling rate	Length	Noise reduction
VAG of the knee joint	Vibration	2	144	2000 Hz	3-5 k	Yes
EMG of hand	Electrical	2	72	1000 Hz	3-5 k	No
Coughing sound	Sound	2	121	22.5 kHz	60-80 k	No

3.2. Evaluating PRS features

Classifiers for feature testing should be selected following the principle of interpretability, avoiding black boxes as much as possible [27]. The logistic regression (LR) model, which is widely used in classification problems and has the advantage of outputting probabilities, was chosen as the main linear classifier in this study. However, the LR classifier is also limited by these probabilities. In some cases, a result of 51% or even 99% pointing to a given class can produce mismatching, which means that the predicted class may not be accurate even though the probability of the prediction is high. [27].

In contrast to LR, a support vector machine (SVM), which focuses on maximizing the probability of group membership, attempts to find the separating hyperplane that maximizes the distance of the closest points to the border. As the feature sets have more than 12 dimensions, a polynomial kernel-based nonlinear SVM was used as the main nonlinear classifier. A linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) were used as complementary classifiers to the LR and SVM classifiers. These machine learning-based classifiers were used to test the feature set [28].

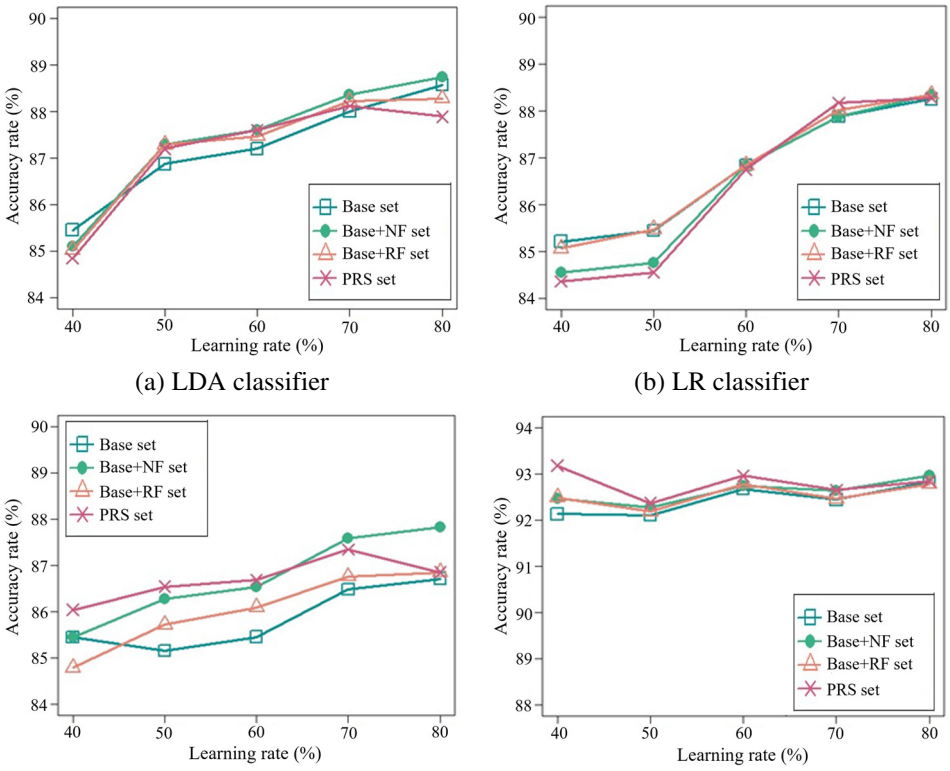
Feature extraction methods for biomedical signals typically use both time- and frequency-domain features. Because all base features in this study were time-domain features, two frequency-domain features were used to complete the comparison feature set [6]. Specifically, this study uses the maximum power spectrum density (MaxPSD) and median power spectrum density (MedPSD) as auxiliary features to the base features in the comparison set. The two frequency-feature extraction techniques were based on Fourier transforms [6]. The MaxPSD and MedPSD measures were added to the unprocessed base feature set to create a comparison set that included the same 14 features as the experimental set. The comparison set was then normalized using min-max normalization to ensure that the features were on the same scale. Finally, the accuracy of the PRS algorithm was compared to the accuracy of the comparison set. Implementations of signal processing, feature engineering, the PRS algorithm, and classifiers were supported by RStudio and related packages, and the result has been provided in subsequent sections.

4. Results

The study used four datasets to evaluate the performance of each classification model. The first dataset, referred to as the base set, contained only 12 time-domain features as listed in Table 1. The second and third datasets, named base + NF and base + RF respectively, contained one PRS feature and 12 time-domain features each, in addition to those in the base set. Finally, a fourth dataset was used that included all 14 features (the 12 time-domain features and the 2 PRS features). This dataset, referred to as PRS with all features, was used to evaluate the performance of each classification model when all features were included. All datasets including the VAG, EMG, and cough signals were input individually into the LR, SVM, LDA, and QDA classifiers. Each classifier was trained with learning rates ranging from 40% to 80% of the total samples (in increments of 10%). The remaining samples were used as test data. The classification accuracy was then calculated using Eq. (9). The results for the VAG, EMG, and cough signals are summarized in Figs. 6-8, where the vertical coordinate represents the average accuracy of 100 classifications using various learning rates with a one-way ANOVA test after the Shapiro test.

As shown in the figures, all input sets for the three types of biomedical signals exhibit an increasing trend in accuracy with a raised learning rate. However, the most notable results are that the sets including the NF or RF features show greater accuracy than merely the base set. This result is consistent with the hypothesis of this study that the new feature introduced by the algorithm as a control variable that is less correlated with other existing variables reduces the overall correlation between variables reduces bias in training, and improves classification accuracy.

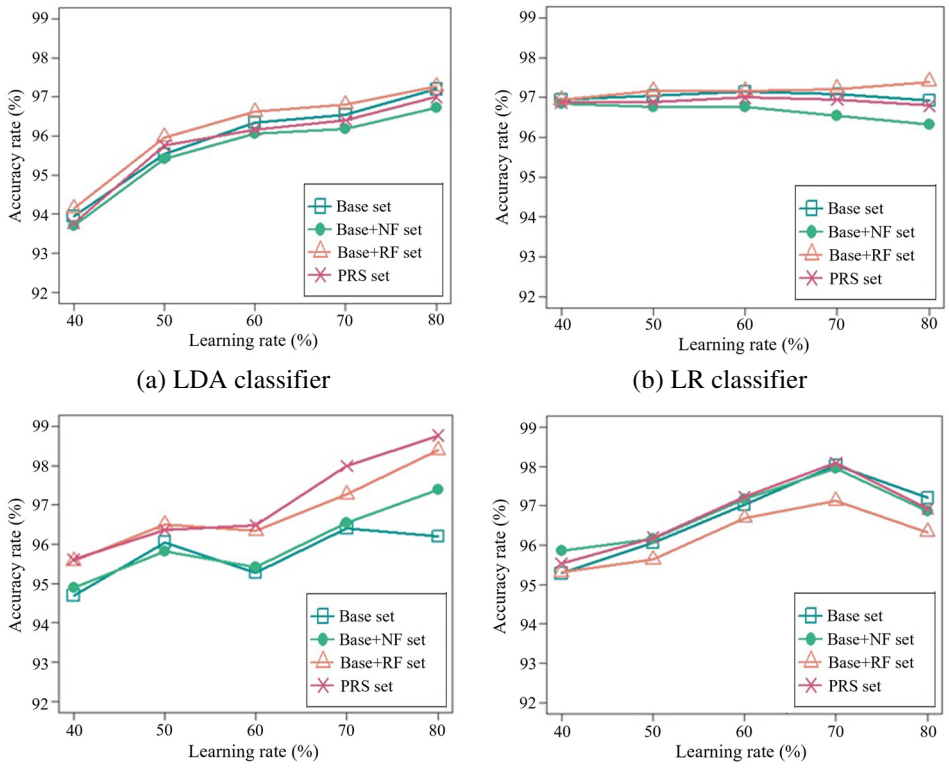
Focusing on the VAG signal in Fig. 6. First, the input sets containing the NF tend to perform better. The QDA classifier with the set containing the NF and RF has the highest classification accuracy. Further analysis indicates that the classification accuracy improves by 3.12% using the base set and 3.65% for the base+NF set when the training rate is increased to 80% for the LR classifier.



(a) LDA classifier (b) LR classifier

(c) SVM classifier (d) QDA classifier

Fig. 6 Accuracy of VAG signal for four classifiers



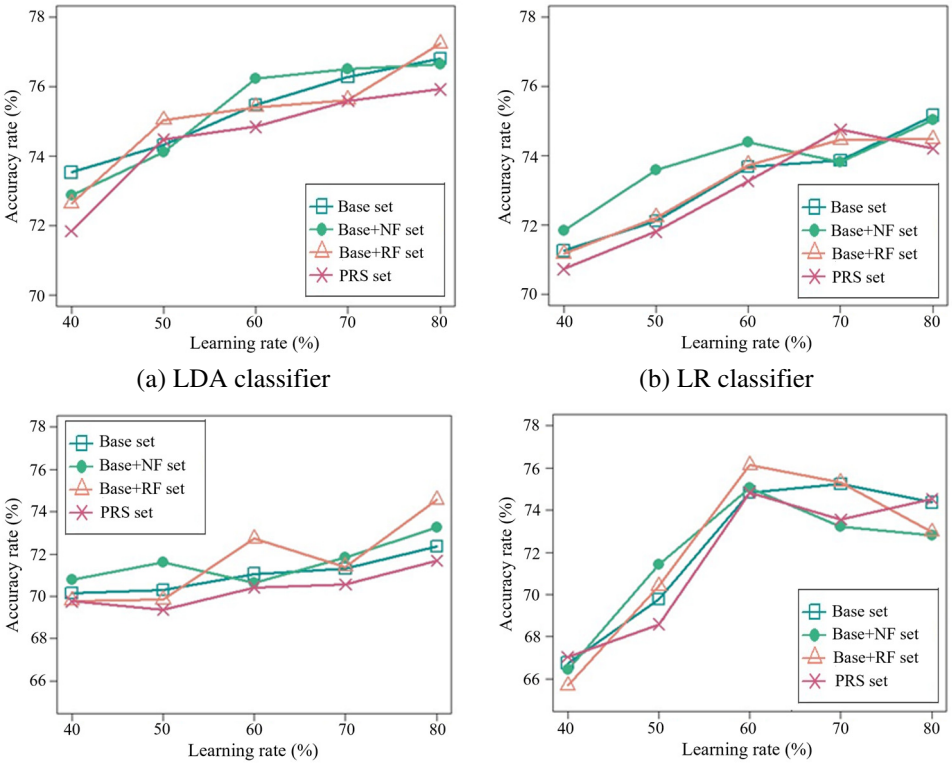
(a) LDA classifier (b) LR classifier

(c) SVM classifier (d) QDA classifier

Fig. 7 Accuracy of EMG signal for four classifiers

Fig. 7 indicates that the classification accuracy for the EMG signal is as high as 98% for all four classifiers, which is considerably higher than the maximum values for the other biomedical signals. The set containing the NF or RF shows a higher accuracy under most conditions except for the QDA classifier applied at an 80% training rate. Fig. 8 shows that the cough signal is less accurate and regular than the other signals. In this study, the largest improvement in accuracy is observed at a

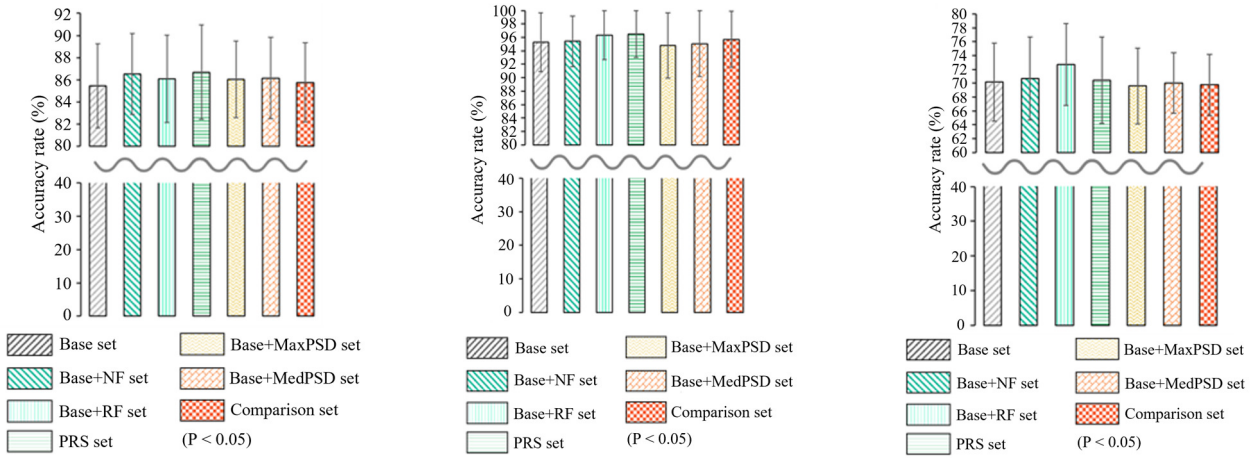
60% training rate for the QDA classifier, which is an improvement of approximately 10% compared to the initial stage. In addition, although the accuracy monotonically increased as the training rate raised, there were several monotonic changes in all biomedical signals. This result is most likely due to errors during random sampling; therefore, a repetitive sampling comparison was implemented as follows:



(a) LDA classifier (b) LR classifier

(c) SVM classifier (d) QDA classifier

Fig. 8 Accuracy of cough signal for four classifiers



(a) VAG signal (b) EMG signal (c) Cough signal

Fig. 9 Accuracy of the PRS and comparison sets with a 60% learning rate for the SVM classifier

Fig. 9 compares the accuracy of the base set, PRS-features sets, and frequency-domain feature sets using a 60% learning rate for the SVM classifier. Tukey's multiple comparison tests of the means using a 95% family-wise confidence level were conducted on the raw data shown. For the VAG signal, the average differences between the PRS-features set and the base set were greater than 3.71%, while the PRS-features sets and frequency-domain features sets were greater than 1.63%. All p-values were less than 0.05. The EMG and cough signals followed the same pattern. In summary, the results indicate that the sets containing only the base features had a consistently lower accuracy rate; however, the PRS-features sets had a consistently higher accuracy rate.

5. Discussion

For the VAG signals, the PRS features-sets showed a maximum accuracy of 93%, which is higher than that reported in previous studies [3, 24]. These results validate the findings of Rauber et al. [14] in 2015 that only appropriate features could improve the performance of the classifiers. Although the PRS-features sets did not achieve the 80% accuracy reported by Chaudhari et al. [26] in 2020, a 76% accuracy could be achieved with only one-tenth of the sample size. This study was unable to find published classification results for the EMG signals; however, the authors believe that a classification accuracy of up to 98% is quite satisfactory.

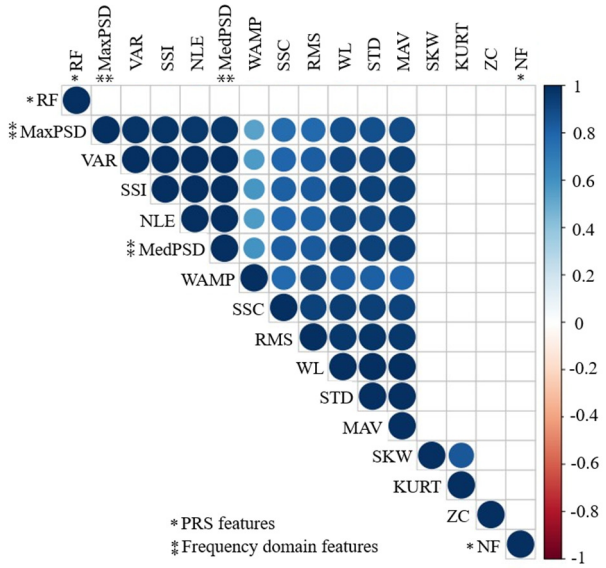


Fig. 10 Correlation between individual features of the EMG signal

The swarm intelligence approach's most significant characteristic is to extract a feature set with exceptionally low correlations by simulating the development of natural PRS. The feature correlations for the EMG signals are shown in Fig. 10, where the circle at the intersection of a feature row and the column of another feature represents the correlation between the two features. A cool-toned circle indicates a positive correlation, and a warm-toned circle indicates a negative correlation, and the smaller the circle size, the weaker the correlation. In addition, there was no negative correlation in this result, and all results were positive.

Surprisingly, the two features based on the PRS algorithm had nearly zero correlation with the other features. By contrast, the time and frequency-domain features were more or less correlated. The low correlation between the PRS features and other features was because of the different feature extraction methods. The development of a root system is a product of group decision-making, which reduces the correlation between the root matrix and the nutrient matrix in a tensor. Thus, the performance of the classifier improved after the addition of PRS features.

Another significant advantage of the PRS algorithm is the simplicity of interpretation. Although the full procedure of the proposed method might appear lengthy, the algorithm is rather lightweight. The nutrient matrix construction comprises two-thirds of the work, whereas the construction of the root matrix and the extraction of the PRS features constitute the remaining one-third. If the length of the biomedical signal does not exceed 100,000 and the sample size is less than 200, the training process can be completed within 5 min.

The proposed PRS algorithm performed better than expected. Nevertheless, this study had certain limitations. First, the improvement of the accuracy rate was modest for each classification model. As shown in Fig. 7, a mere 1% improvement in accuracy is insufficient to offset the error bar. One possible explanation for this might be the small sample size, which resulted in inadequate model training [29]. Second, although the overall accuracy using the PRS features tended to be higher than that of the base and classical feature sets for each classifier, some exceptions still existed.

These phenomena are likely related to the possibility that extracting features based on the PRS algorithm does not work well in the presence of outliers. Root growth is based on values in the soil matrix, which are obtained by performing statistical calculations on biomedical signals. Errors are passed down, causing a reduction in accuracy. This is the reason for an obtained lower accuracy at high training rates. Although the PRS was initially successfully used for feature engineering in the field of biomedical signals, it has some limitations, and further research is required. One limitation is that this study only considered a simple classification problem and did not consider multivalued classification. Moreover, the effect of imbalanced datasets on learning was not studied. More experiments are needed to verify the diversity of the signal types.

In the past, swarm intelligence and swarm intelligence algorithms, such as ant colony algorithms and bee colony algorithms, were mostly developed based on the behavior of social animals. Currently, intelligent algorithms have also been modeled on certain static systems, including rivers and the PRS used in this study. This study contributes toward renewed interest in the development of such algorithms.

6. Conclusions

In conclusion, this study has made significant strides in the field of biomedical signal processing by introducing a novel feature extraction algorithm, the PRS algorithm. Inspired by swarm intelligence and the natural growth of roots, the PRS algorithm effectively explores the feature space and identifies the most discriminative features for classification. Notably, the PRS algorithm exhibits good interpretability and low correlation with traditional feature extraction methods, setting it apart from existing techniques.

Through a series of experiments, our findings demonstrated a substantial improvement in classification accuracy when utilizing the PRS algorithm, as compared to conventional methods. These results underscore the potential of the PRS algorithm in revolutionizing the way biomedical signals are processed and analyzed, ultimately paving the way for more effective clinical applications.

This study serves as a strong foundation for future research in extending the use of biomedical signals for clinical applications. As part of our ongoing investigation, we plan to validate the PRS algorithm with a broader range of biomedical signals, assess its performance in diverse scenarios, and explore its potential applicability in 2D signal applications. By building upon the successes of the PRS algorithm, we aim to advance the field of biomedical signal processing and contribute to more accurate and efficient diagnostic tools in clinical settings.

Conflicts of Interest

The authors have no conflict of interest to declare.

References

[1] M. Gertsch, The ECG: A Two-Step Approach to Diagnosis, Berlin: Springer, 2003.

[2] C. J. Stam, “Use of Magnetoencephalography (MEG) to Study Functional Brain Networks in Neurodegenerative Disorders,” Journal of the Neurological Sciences, vol. 289, no. 1-2, pp. 128-134, February 2010.

[3] R. Gong, H. Ohtsu, K. Hase, and S. Ota, “Vibroarthrographic Signals for the Low-Cost and Computationally Efficient Classification of Aging and Healthy Knees,” Biomedical Signal Processing and Control, vol. 70, article no. 103003, September 2021.

[4] A. Forestiero, “Bio-Inspired Algorithm for Outliers Detection,” Multimedia Tools and Applications, vol. 76, no. 24, pp. 25659-25677, December 2017.

[5] L. Abualigah, M. A. Elaziz, N. Khodadadi, A. Forestiero, H. Jia, and A. H. Gandomi, “Aquila Optimizer Based PSO Swarm Intelligence for IoT Task Scheduling Application in Cloud Computing,” Integrating Meta-Heuristics and Machine Learning for Real-World Optimization Problems, Cham: Springer International Publishing, 2022.

[6] C. Liu and J. Li, Feature Engineering and Computational Intelligence in ECG Monitoring, Singapore: Springer, 2020.

[7] C. J. Gallego Duque, L. D. Muñoz, J. G. Mejía, and E. Delgado Trejos, "Discrete Wavelet Transform and K-NN Classification in EMG Signals for Diagnosis of Neuromuscular Disorders," XIX Symposium on Image, Signal Processing and Artificial Vision, pp. 1-5, September 2014.
[8] E. Keogh, "Naive Bayes Classifier," http://www.cs.ucr.edu/~eamonn/CE/Bayesian%20Classification%20withInsect_examples.pdf, November 05, 2006.
[9] A. Forestiero, "Heuristic Recommendation Technique in Internet of Things Featuring Swarm Intelligence Approach," Expert Systems with Applications, vol. 187, article no. 115904, January 2022.
[10] X. F. Song, Y. Zhang, Y. N. Guo, X. Y. Sun, and Y. L. Wang, "Variable-Size Cooperative Coevolutionary Particle Swarm Optimization for Feature Selection on High-Dimensional Data," IEEE Transactions on Evolutionary Computation, vol. 24, no. 5, pp. 882-895, October 2020.
[11] F. Hölker, C. Wolter, E. K. Perkin, and K. Tockner, "Light Pollution as a Biodiversity Threat," Trends in Ecology & Evolution, vol. 25, no. 12, pp. 681-682, December 2010.
[12] Q. Jiang, Y. Shen, H. Li, and F. Xu, "New Fault Recognition Method for Rotary Machinery Based on Information Entropy and a Probabilistic Neural Network," Sensors, vol. 18, no. 2, article no. 337, February 2018.
[13] D. A. Lyon, "The Discrete Fourier Transform, Part 4: Spectral Leakage," Journal of Object Technology, vol. 8, no. 7, pp. 23-34, November 2009.
[14] T. W. Rauber, F. de Assis Boldt, and F. M. Varejão, "Heterogeneous Feature Models and Feature Selection Applied to Bearing Fault Diagnosis," IEEE Transactions on Industrial Electronics, vol. 62, no. 1, pp. 637-646, January 2015.
[15] B. R. Nayana and P. Geethanjali, "Analysis of Statistical Time-Domain Features Effectiveness in Identification of Bearing Faults from Vibration Signal," IEEE Sensors Journal, vol. 17, no. 17, pp. 5618-5625, September 2017.
[16] A. Phinyomark, P. Phukpattaranont, and C. Limsakul, "Investigating Long-Term Effects of Feature Extraction Methods for Continuous EMG Pattern Classification," Fluctuation and Noise Letters, vol. 11, no. 04, article no. 1250028, December 2012.
[17] S. G. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," https://doi.org/10.48550/arXiv.1503.06462, March 19, 2015.
[18] G. Chandrashekar and F. Sahin, "A Survey on Feature Selection Methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16-28, January 2014.
[19] K. Jeyalakshmi, "Convergence of Optimization Problems," Bonfring International Journal of Data Mining, vol. 2, no. 1, pp. 13-16, March 2012.
[20] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, et al., "Feature selection: A Data Perspective," ACM Computing Surveys, vol. 50, no.6, article no. 94, November 2018.
[21] R. Crang, S. Lyons-Sobaski, and R. Wise, Plant Anatomy: A Concept-Based Approach to the Structure of Seed Plants, Cham: Springer, 2018.
[22] C. E. Metz, "Basic Principles of ROC Analysis," Seminars in Nuclear Medicine, vol. 8, no. 4, pp. 283-298, October 1978.
[23] K. C. Chua, V. Chandran, U. R. Acharya, and C. M. Lim, "Application of Higher Order Statistics/Spectra in Biomedical Signals—A Review," Medical Engineering & Physics, vol. 32, no. 7, pp. 679-689, September 2010.
[24] R. Gong, K. Hase, H. Goto, K. Yoshioka, and S. Ota, "Knee Osteoarthritis Detection Based on the Combination of Empirical Mode Decomposition and Wavelet Analysis," Journal of Biomechanical Science and Engineering, vol. 15, no. 3, p. 20-00017, 2020.
[25] S. Lobov, N. Krilova, I. Kastalskiy, V. Kazantsev, and V. A. Makarov, "Latent Factors Limiting the Performance of sEMG-Interfaces," Sensors, vol. 18, no. 4, article no. 1122, April 2018.
[26] G. Chaudhari, X. Jiang, A. Fakhry, A. Han, J. Xiao, S. Shen, et al., "Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough," https://doi.org/10.48550/arXiv.2011.13320, November 26, 2020.
[27] C. Molnar, Interpretable Machine Learning, Morrisville: Lulu, 2020.
[28] E. Ofori-Ntow Jnr, Y. Y. Ziggah, M. J. Rodrigues, and S. Relvas, "A New Long-Term Photovoltaic Power Forecasting Model Based on Stacking Generalization Methodology," Natural Resources Research, vol. 31, no. 3, pp. 1265-1287, June 2022.
[29] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, "Machine Learning: A Review of Classification and Combining Techniques," Artificial Intelligence Review, vol. 26, no. 3, pp. 159-190, November 2006.