

Automatic EEG-based Alertness Classification using Sparse Representation and Dictionary Learning

¹Mohammad Hadra , ¹Iman Abdelrahman

¹Dep. Of Electrical and Electronic Engineering, University of Khartoum , Khartoum, Sudan;
mahmd_hadra@yahoo.com; maaly@uofk.edu

ABSTRACT

Automation of human alertness identification has been widely investigated in recent decades. Many applications can benefit from automatic alertness state identification, such as driver fatigue detection, monotonous task workers' vigilance detection and sleep studies in the medical field. Many researchers have tried to exploit different types of behavioural aspects in vigilance detection, such as eye movement, head position and facial expression. On the other hand, some biomedical signals like ECG, EEG and heart rhythm are also exploited; however, there is a consensus of the superiority of EEG signal in alertness classification due to its close relation with different human vigilant states. In this paper, we propose an automatic method for vigilance detection using a single EEG channel along with sparse representation and dictionary learning. We used Discrete Wavelet Packet Transform to extract the features related to different human vigilance states. We use well-known other classifiers to compare the performance of our proposed method. Results of classification with sparse representation and dictionary learning produced better accuracy results than the other methods.

Keywords: Sparse representation, Dictionary learning, optimization, OMP, 1-norm, reconstruction error, EEG, LDA, RF.

1 Introduction

Keeping vigilant for a long time, especially in monotonous tasks, is an arduous practice. Falling asleep while driving, monitoring heavy machinery and guarding valuable assets have become a typical trait for many people which may result in catastrophic consequences.

The National Highway Transportation and Safety Administration (NHTSA) states that drowsy driving accounts for 1550 deaths, 40000 injuries and more than 56000 crashes every year in the United States, This varies between 2.2% - 2.6% of the fatal crashes[1].

A study conducted in Sultan Qaboos University concludes that daytime sleeping while driving is a fairly common practice among young Omani drivers [2].

Emerging technologies in computer systems and sensing devices ease the quest to automate human vigilance monitoring. Real-time reporting of the lack of vigilance can prevent catastrophic consequences. Automatic human vigilance detection has been widely investigated. Neurophysiological signal,

physiological signals and behavioural-based symptoms such as eye movement and face expressions were used as indicators for vigilance state.

In [3], M.Tagldin et al. state that EEG is the golden standard to monitor and classify the human vigilance states due to its close correlation with the different vigilance states.

Z. Zhang et al. [4] developed a complete active safety model to control vehicle speed based on vigilance detection, a wireless wearable EEG used for data collection. They used Wavelet de-noising for noise and artifact removal, and FFT was used to extract features, a combination of sparse representation and k-mean singular value decomposition (KSVD) used in classification. Results of classification passed to the brake system. Their final results prove the validity of their proposal in simulated and realistic conditions. In [5], Yu H et al. used sparse representation to classify the EEG signal. The classification was based on the idea of minimum residual after signal reconstruction. Driving simulated room with EEG cap were used for data collection, Continuous Wavelet Transform (CWT) and Independent Component Analysis (ICA) utilized for data pre-processing. Their method was compared with Support Vector Machine classifier. Results prove that the proposed method outperforms Support Vector Machine (SVM). In [1], N. Gurudath et al. used the Discrete Wavelet Transform (DWT) to extract the feature of the EEG signal. They used first-order statistical moments such as means, median, variance and standard deviation as input features for a combination of unsupervised learning (K-mean Clustering) and ANN as a classifier. In [6], B. G. Lee et al. used EEG signal along with the respiration signal to develop mobile healthcare for automatic driving sleep-onset detection. They used the Wavelet Packet Transform (WPT) to extract the features from both signals, then a mutual information technique (MI) was used to select the most prominent features. SVM was utilized as a classifier.

Multi-channel was used as well for sleep stage classification. In [7], J. Shi et al. introduced a new method called Collaborative Representation (CR) and used it to represent data of multi-channels.

In [8], T.Da Silveira et al. used WPT to analyze a single-channel EEG. They extracted the known EEG rhythm patterns (Delta, Theta, Alpha, Beta); spectral power-based indices were computed for the mentioned pattern, then statistical Wilcoxon signed-rank test was used to detect the drowsy state. In [9], the authors used a streamlined set of features calculated from the analysis of the EEG signal by Discrete Wavelet Transform. They trained a Random Forest predictor to classify the data epochs into one of the Rechtschaffen and Kales' stages.

In [10], Belakhdar et al. address the problem of interpersonal variability between the different subjects in the range of Alpha sub-band rhythm. They represent the Alpha sub-band by only one frequency instead of using the whole range 8 to 12, i.e. by deriving the Individual Alpha Frequency (IAF) for each subject. This process helps in increasing the accuracy and reduce the number of features, which in turn reduces the computational burden and helps in hardware implementation.

The rest of this paper organize as follow: Section one is about data acquisition and pre-processing. In section two, we demonstrate the materials used and the method of work. Section three is the results and discussion of our work.

2 Data Acquisition and Pre-processing

We acquired the data used in this research, i.e. EEG signal, from a public database, named the Sleep-EDF database (expanded) on the Physionet. [11]. This data set has been widely used in the literature. The database is a collection of 61 Polysomnography (PSG) obtained from 1987–2002 including the older Sleep EDF database recordings before 1991. The recordings include data from 20 healthy subjects numbered 00 through 19, which had ten male and ten female participants between the ages of 25 and 34. Two PSGs with lengths of approximately 20 hours were recorded each day, including nighttime, for two days at the subjects' homes. The PSG recordings contain EEG Fpz-Cz, EEG Pz-Oz, EOG horizontal, submental chin EMG and event marker signals. Most files also contain oro-nasal respiration and rectal body temperature, the EEG signal used in this work sampled at 100 Hz. The associated hypnogram files (a graph represents the alertness stages as a function of time) contain sleep patterns recorded from each subject. Well-trained technicians manually scored all hypnograms according to the R&K manual [12]. These patterns consist of sleep stages W, S1, S2, S3, S4, REM, M (movement time) and (?) (question mark symbol) as an indication that technicians have not scored it.

In this works we used (W), (S1) and (S2) only, which are mapped to wakefulness, drowsy and sleep states, respectively [13]. The annotation of the PSG is based on a 30 second length segment, i.e. each state mentioned in the hypnogram file is referred to a 30 second length of the subject state. In the rest of this paper, we call each segment as “epoch”. We formed an overlapped epoch $Y \in \mathbb{R}^{4096}$ to cover any travelling state and to fulfil the constrained of power of 2 for WPT.

We used 70% of the data for training and 30% for testing.

In this work, we use the channel Pz-Oz due to its close relation to the alertness state and better accuracy over Fpz-Oz[14].

3 Materials and Methods:

Due to the high dimension of data in the EEG signal, we used feature extraction as a technique for dimensionality reduction. We decomposed the signal by using the Discrete Wavelet Packet Transform. We select Wavelet coefficients representing the frequency rhythm related to human vigilance as features, (we described it in detail in feature extraction and selection part) that done in light of table 1 which denotes the relation of each frequency rhythm band with vigilance state

Table 1. Frequency Rhythms Related to Human Vigilance

Freq. Range	Alertness State
Delta δ (1-4Hz)	Deep Sleep, sudden lapses of attention
Theta θ (4-8Hz)	Deep relaxation, balance and stability, deeper breathing and relaxing lungs.
Alpha α (8-13 Hz)	Wake and relaxed, calm state of mind
Beta β (13- 30 Hz)	Conscious and awake, active concentration, busy and anxious state, motor activity.

Then a set of features were selected from these extracted features, i.e., relative Power Spectrum Density (PSD) for each sub-band, relative length of sub-band, and relative max value for each sub-band.

We used feature vectors to build the dictionary for each class. Then we conducted optimization with sparsity constraints for each of these dictionaries.

Sparse Representation (SR) as a signal representation method was utilized here for classification. The technique, in brief, is to represent each feature vector of data over all dictionaries of the different classes. Then, reconstruct it. We assigned the feature vector to the class of the dictionary that yields the minimum reconstruction error. We used the same set of features again with the well-know Linear Discriminant Analysis (LDA) and Random Forest (RF) classifiers to estimate and compare the performance of classification with the proposed SR with Dictionary Learning (DL) method.

3.1 Feature Extraction and Selection:

3.1.1 Feature Extraction:

EEG signal is a Non-stationary signal with different frequency properties[8]. Conventional signal decomposition methods such as Fourier analysis may not be a suitable option for such signals. We should look for multi-resolution analysis in both time and frequency domains. Such capability is available in wavelet transforms.

Wavelet Packet Decomposition or Wavelet Packet Transform WPT, or just wavelet packet is a wavelet transform where the discrete-time signal is passed through more filters than the ordinary discrete wavelet transform. However, WPT can provide more precise frequency resolution than Wavelet Decomposition [15]. Wavelet Decomposition passes the previous approximation coefficient only through low and high pass filter, and WPT passes both previous approximate and detail coefficient through low and high pass filters yielding new coefficients in a new level.

Depending on the number of decomposition levels required, the process of passing through the filters iterated.

For an “n” number of levels of decomposition using WPT, the decomposition results produce 2^n different sets of coefficients (or nodes).

In this work, we use WPT to decompose our signal. We decomposed the signal to six levels. (Fig 1 , 2)

In terms of the suitable wavelet family or wavelet basis for this work, Haar mother wavelet reported having excellent performance in a time domain computer-generated signal in computational time context [15]. Also, Haar reported as the most straightforward wavelet[16], and it avoids any special boundary treatment[17]. In this work, we used Haar basis (Mother Wavelet Function).

Decomposing a signal using WPT allows breaking down the whole signal in a specific frequency range between 0 and half of the signal sample rate [18] which is 100 Hz in our data.

Decomposing EEG signal using WPT provides better insight into the frequency distribution over time. The wavelet packet decomposition at the j th level of a given signal gives 2^j sets of sub-band coefficients of length $\{P_{j,m}(n) | k = 1, 2, \dots, N/2^j\}$. [6].

We formulated frequency ranges yields as below:

$$\left[\frac{(m-1)Fs}{2^{j+1}}, \frac{mFs}{2^{j+1}} \right] \quad (1)$$

Where F_s is the sampling frequency, and $m = 0, 1, \dots, 2^{j-1}$. In each level of decomposition, signal transformed into two components, each component with a frequency range half of the upper-level frequency. We decomposed our signal into six levels to extract the frequency components. Frequency component holds information related to human vigilance, as stated in table 1 were extracted as features. WPT coefficient holding the above features are as illustrated in table 2.

Table 2 . Wavelet coefficients (Node Order) for each sub-band

Appropriated Wavelet Coefficient: Node order and #	Frequency Range Covered	Rhythm Name
$N_{6,1} \text{ -- } N_{6,5}$	0 Hz – 3.906 Hz	Delta
$N_{6,6} \text{ -- } N_{6,10}$	4.687Hz– 7.812 Hz	Theta
$N_{6,11} \text{ -- } N_{6,17}$	8.593Hz– 13.281 Hz	Alpha
$N_{6,18} \text{ -- } N_{6,45}$	14.062Hz– 35.156 Hz	Beta

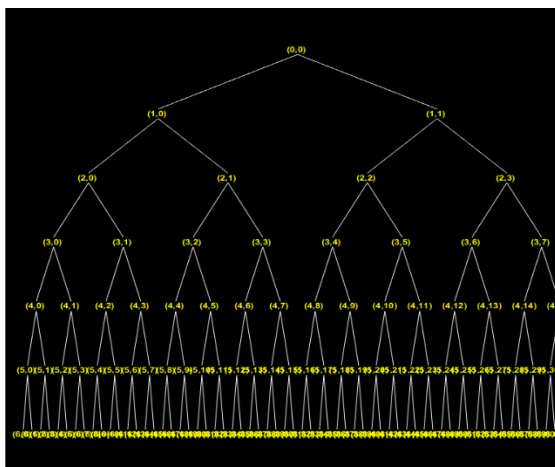


Fig. 1: Wavelet Packet tree for 6 level analysis

```
Wavelet Packet Object Structure
=====
Size of initial data :    [4096  108]
Order                  :      2
Depth                  :      6
Terminal nodes         :    [63  64  65  66  67  68  69  70  71]
-----
Wavelet Name           :haar
Low Decomposition filter :[ 0.7071  0.7071]
High Decomposition filter :[-0.7071  0.7071]
Low Reconstruction filter :[ 0.7071  0.7071]
High Reconstruction filter :[ 0.7071 -0.7071]
-----
Entropy Name           :shannon
Entropy Parameter       :0
```

Fig. 2: Wavelet Packet Object Structure

3.1.2 Feature Selection

In recent literature, most of the similar work done in this area used statistical values such as mean and max and used them as selected features. In this work, we choose a new set of features that is: the relative Power Spectrum Density (PSD), The sub-band length and the sub-band Max point.

Power Spectrum Density (PSD) is the measure of the power distributed over a signal that contains frequency component, PSD can be a discriminative measure for any frequency sub-band since the amount

of power has a close correlation with the frequency component in this rhythm. We also select the sub-band length as a feature. **The sub-band length** can represent another discriminative view of each sub-band. Calculating the relative length for each sub-band gives another look of the sub-band in the whole signal, i.e. concentration of each sub-band across the entire signal.

The max point in each sub-band also can give another discriminative view. Relative max point is calculated and selected as a feature in this work. We selected twelve features in total as input features for our next classification stage.

3.2 Sparse Representation Classification:

3.2.1 Dictionary construction and Learning:

Two options are available for constructing an over-complete dictionary: either to use a pre-specified set of functions such as wavelets, curvelets. (off-the-shelf dictionary), OR build and learn a dictionary from the available pre-processed annotated data. In this work, we adopt the second option. We used 70% of the pre-processed data to build a dictionary. We made three dictionaries: Sleep, Drowsy and Awake. The set of the selected feature was stacked together to form the initial dictionaries. The resulting dictionary is an m by n matrix, m is the number of features chosen which is 12 and n is the number of data epochs used for training which is 70 epochs.

Each column vector in this dictionary is $[d_{1,n}, d_{2,n}, \dots, d_{12,n}]^T$, $n = 1$ to 70 is known as dictionary atoms, in our dictionary we have 70 atoms. Then, each dictionary optimized using K-mean Singular Value Decomposition algorithm (K-SVD). K-SVD algorithm is a dictionary learning algorithm. It assumes that there is an initial dictionary, and it iteratively optimizes it for sparse coding by the generalization of the K-Mean clustering method via singular value decomposition, as mentioned in [19]. Learning iterations formulated as a constrained optimization problem in the following:

$$\{x_{opt}, D_{opt}\} = \underset{x}{\operatorname{argmin}} \frac{1}{2} \|Y - Dx\|_2^2 + \lambda J(x) \quad (2)$$

In this work, we set the learning numbers of iterations to 25, and the regularization factor Lamda was set to 0.99. We used Orthogonal Matching Pursuit (OMP) as a sparse coding algorithm, according to the recommendation in [19].

$$\begin{bmatrix} d_{1,1} & d_{1,2} & \cdot & \cdot & \cdot & d_{1,70} \\ d_{2,1} & d_{2,2} & \cdot & \cdot & \cdot & d_{2,70} \\ d_{3,1} & d_{3,2} & \cdot & \cdot & \cdot & d_{3,70} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ d_{11,1} & d_{11,2} & \cdot & \cdot & \cdot & d_{11,70} \\ d_{12,1} & d_{12,2} & \cdot & \cdot & \cdot & d_{12,70} \end{bmatrix}$$

Fig. 3: Dictionary structure.

3.2.2 Sparse Coding (Representation):

Sparse representation is one of the emerging methods of representation in the signal processing field. It is worth representing a feature vector with another vector, most of whose points are zeros over a dictionary. Such representations have a wide range of applications in image procession, medical imaging, classification, and compression.

The problem of sparse coding formulated as follows:

$$\hat{\alpha} = \operatorname{argmin} \Psi(x) \quad s.t \ y = Dx \quad (3)$$

Where : Ψ is the measure of sparsity, D is an over-complete dictionary, and $\hat{\alpha}$ is the required sparse vector. The sparse representation method with 0-norm minimization can obtain the fundamental sparse solution of $\hat{\alpha}$ over the matrix D . The problem is still a non-deterministic polynomial-time hard (NP-hard) problem, and the solution is difficult to approximate.[20] A different measure of sparsity was investigated in the literature, in [3], they investigate using the Gini Index as sparsity measure for sparse coding. However, recent literature demonstrates that if the constrain of sufficient sparsity is fulfilled, using 1-norm as a sparsity measure can be equivalent to use 0-norm.[21][22] Accordingly, in this work, we use 1-norm as sparsity measure.

3.2.3 Classification :

Reconstruction error is used as a classification indicator in this work, assuming that sparse coding of the data vector using the same class of dictionary in which the data belong to, should yield a minimum reconstruction error. We conduct Sparse Coding for each data segment using the pre-prepared dictionaries, reconstruction error estimated for each dictionary, data assigned to dictionary class with minimum error. We formulated this in below formula:

$$\text{class}(y) = \operatorname{argmin} (\|y - D_i x_i\|_2^2) \quad i = 1, 2, 3 \quad (4)$$

$D1$ = Drowsy Dictionary, $D2$ = Sleep Dictionary , $D3$ = Awake Dictionary , y = Tested data and x_i is the produced sparse vector in each sparse coding iteration.

3.3 Classification with other methods

To compare the accuracy of our proposed method with existing classifiers, we used two different well-known classifiers: Linear Discriminant Analysis (LDA) and Random Forest (RF). Classifiers were used with the same set of features in SR classification. Those two classifiers cover different techniques and methods used in classification. LDA is a projection method [23]; It projects the features in a new dimension to maximize between-class separability and reduce within-class variability, LDA classifier produced perfect classification results for linearly separated data.

Random forest is an ensemble classification technique, based on a combination of a large set of decision trees. The forest is built by combining the predictions of several trees, each of which is trained in isolation [24].

4 Results and Discussion:

We used six subjects in this work. We mainly selected the subjects that contain a sufficient drowsy state duration. As is commonly known, drowsy state duration is the shortest duration compared to awake and

sleep states. Considering the data balancing among subject, we used the same duration (data length) for each vigilant state. The subjects selected were:

SC4011, SC4041 , SC4052 , SC4102 , SC4122, SC4182.

Table 3 shows the results of accuracy per subject using SR with DL classification. The average accuracy is 95.17%.

Table 3: Result using SR and DL

	Subject	Drowsy	Sleep	Awake
1	SC4011	88.20%	96.56%	100%
2	SC4041	87.90%	96.80%	100%
3	SC4052	87.90%	97.60%	100%
4	SC4102	88.60%	97.83%	100%
5	SC4122	88.90%	98.60%	100%
6	SC4182	87.60%	96.65%	100%

Table 4 shows the accuracy result for each subject using LDA classifier. The average accuracy is 89.2%.

Table 4: Result using LDA classifier

	Subject	Drowsy	Sleep	Awake
1	SC4011	82.20%	90.00%	100%
2	SC4041	81.90%	91.00%	98%
3	SC4052	78.00%	87.00%	96%
4	SC4102	80.60%	90.56%	100%
5	SC4122	77.00%	80.60%	95%
6	SC4182	84.60%	93.65%	100%

Table 5 shows the accuracy result for each subject using RF classifier. The average accuracy is 90.49%.

Table 5: Result using RF classifier

	Subject	Drowsy	Sleep	Awake
1	SC4011	82.20%	92.56%	98%
2	SC4041	81.90%	92.20%	99%
3	SC4052	83.00%	90.00%	100%
4	SC4102	80.60%	90.56%	100%
5	SC4122	82.90%	80.60%	97%
6	SC4182	84.60%	93.65%	100%

From the above tables, the average accuracy is 95.17%, 90.49% and 89.23% for SR, RF and LDA classifiers respectively. SR with DL classifier outperforms the other two classifiers in term of accuracy. SR as an emerging feature representation method is better exploring the underlying structure of the signal since it represents the signal as a linear combination of the dictionary atoms. This property makes classification with reconstruction error more accurate than conventional LDA model and ensemble RF classification methods. Although the accuracy of SR with DL is better than LDA and RF, it is not enough to evaluate the overall performance. Further investigation and performance measure need to be investigated to proper

rank our proposed method relative to other methods. Time and resource required for each technique is an essential measure to specify the best and suitable classifier.

Results also show that drowsy state gains the lowest accuracy relative to wake state and sleep state for all the three classifiers. This is because drowsiness is a transient state in nature, and it is more challenging to identify.

5 Conclusion and Future Work:

In this work, we proposed a method for EEG based alertness classification using sparse representation and dictionary learning. For comparison purpose, the same set of features used as input features in two other well-known classifiers. Classification using SR with DL outperforms the other two classifiers in term of accuracy. Further investigation should be addressed to specify if SR with DL may be the better alternative for the conventional classifiers. Using multi-EEG channel is an issue that should be investigated in future work as well. A framework for utilizing all available EEG channel is required. A new multi-view dictionary with multi-signal may be an exciting improvement for classification of alertness using EEG signal.

REFERENCES

- [1] N. Gurudath and H. Bryan Riley, "Drowsy driving detection by EEG analysis using Wavelet Transform and K-means clustering," *Procedia Comput. Sci.*, vol. 34, pp. 400–409, 2014, doi: 10.1016/j.procs.2014.07.045.
- [2] S. Q. University, "No Title." [Online]. Available: <https://anwaar.squ.edu.om/en/2018/10/28/study-says-daytime-sleepiness-common-among-young-omani-drivers/>.
- [3] M. Tageldin, T. Al-Mashaikki, H. Bali, and M. Mesbah, "EEG Sparse Representation Based Alertness States Identification Using Gini Index," in *Neural Information Processing*, 2018, pp. 478–488.
- [4] Z. Zhang *et al.*, "A vehicle active safety model: Vehicle speed control based on driver vigilance detection using wearable EEG and sparse representation," *Sensors (Switzerland)*, vol. 16, no. 2, pp. 1–26, 2016, doi: 10.3390/s16020242.
- [5] H. Y. H. Yu, H. L. H. Lu, T. O. T. Ouyang, H. L. H. Liu, and B.-L. L. B.-L. Lu, "Vigilance detection based on sparse representation of EEG," *Eng. Med. Biol. Soc. (EMBC), 2010 Annu. Int. Conf. IEEE*, vol. 2010, pp. 2439–2442, 2010, doi: 10.1109/IEMBS.2010.5626084.
- [6] B. G. Lee, B. L. Lee, and W. Y. Chung, "Mobile healthcare for automatic driving sleep-onset detection using wavelet-based EEG and respiration signals," *Sensors (Switzerland)*, vol. 14, no. 10, pp. 17915–17936, 2014, doi: 10.3390/s141017915.
- [7] J. Shi, X. Liu, Y. Li, Q. Zhang, Y. Li, and S. Ying, "Multi-channel EEG-based sleep stage classification with joint collaborative representation and multiple kernel learning," *J. Neurosci. Methods*, vol. 254, pp. 94–101, 2015, doi: 10.1016/j.jneumeth.2015.07.006.
- [8] T. L. T. Da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Automated drowsiness detection through wavelet packet analysis of a single EEG channel," *Expert Syst. Appl.*, vol. 55, pp. 559–565, 2016, doi: 10.1016/j.eswa.2016.02.041.
- [9] T. L. T. da Silveira, A. J. Kozakevicius, and C. R. Rodrigues, "Single-channel EEG sleep stage classification based on a streamlined set of statistical features in wavelet domain," *Med. Biol. Eng. Comput.*, vol. 55, no. 2, pp.

- 343–352, 2017, doi: 10.1007/s11517-016-1519-4.
- [10] I. Belakhdar, W. Kaaniche, R. Djemal, and B. Ouni, "Single-channel-based automatic drowsiness detection architecture with a reduced number of EEG features," *Microprocess. Microsyst.*, vol. 58, no. December 2017, pp. 13–23, 2018, doi: 10.1016/j.micpro.2018.02.004.
 - [11] "Sleep-EDF Database Expanded v1.0.0." [Online]. Available: <https://physionet.org/content/sleep-edfx/1.0.0/>. [Accessed: 15-Feb-2020].
 - [12] "Rechtschaffen, A. and Kales, A. (1968) A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. Public Health Service, US Government Printing Office, Washington DC. - References - Scientific Research Publishi." .
 - [13] A. Kales, *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Washington DC: United States Government Printing Office, 1968.
 - [14] C. T. Lin *et al.*, "EEG-based subject- and session-independent drowsiness detection: An unsupervised approach," *EURASIP J. Adv. Signal Process.*, vol. 2008, 2008, doi: 10.1155/2008/519480.
 - [15] M. Y. Gokhale and D. K. Khanduja, "Time Domain Signal Analysis Using Wavelet Packet Decomposition Approach," *Int. J. Commun. Netw. Syst. Sci.*, 2010, doi: 10.4236/ijcns.2010.33041.
 - [16] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin, "Wavelets for computer graphics: a primer, part 2," *IEEE Comput. Graph. Appl.*, no. September, 1995, doi: 10.1109/38.391497.
 - [17] A. Kozakevicius and A. A. Schmidt, "Wavelet transform with special boundary treatment for 1D data," *Comput. Appl. Math.*, vol. 32, no. 3, pp. 447–457, Oct. 2013, doi: 10.1007/s40314-013-0050-6.
 - [18] I. Daubechius, "Ten Lectures of Wavelets," *Philadelphia, PASociety Ind. Appl. Math. Anal.*, pp. 1544–1576, 1992, doi: 10.1137/1.9781611970104.
 - [19] M. Aharon, M. Elad, and A. M. Bruckstein, "The {K-SVD}: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representations," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
 - [20] Z. Zhang, S. Member, Y. Xu, and S. Member, *07102696*, vol. 3. 2015.
 - [21] T. R. Gillett and A. L. Holven, "Grading of Soft Sugars Photoelectric Reflectance Measurements," *Ind. Eng. Chem.*, vol. 35, no. 2, pp. 210–214, 1943, doi: 10.1021/ie50398a018.
 - [22] D. L. Donoho, "For most large underdetermined systems of equations, the minimal ℓ_1 -norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, 2006, doi: 10.1002/cpa.20131.
 - [23] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing - Proceedings of the IEEE Workshop*, 1999, pp. 41–48, doi: 10.1109/nnsp.1999.788121.
 - [24] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.