

CLASSIFICATION ACCURACY ENHANCEMENT BASED MACHINE LEARNING MODELS AND TRANSFORM ANALYSIS

Hanan A. R. Akkar¹, Wael A. H. Hadi², Ibraheem H. Al-Dosari^{3,*}, Saadi M. Saadi⁴, Aseel Ismael Ali⁴

¹Department of Electrical Engineering, University of Technology, Baghdad, Iraq

²Department of Communications Engineering, University of Technology, Baghdad, Iraq

³Department of Computer Communications Engineering, Al-Rafidain University College, Baghdad, Iraq

⁴Ministry of Education, Minister's Office, Institute of Gifted Students, Baghdad, Iraq

*E-mail of corresponding author: ibraheemdoser77@gmail.com

Resume

The problem of leak detection in water pipeline network can be solved by utilizing a wireless sensor network based an intelligent algorithm. A new novel denoising process is proposed in this work. A comparison study is established to evaluate the novel denoising method using many performance indices. Hardy-rectified thresholding with universal threshold selection rule shows the best obtained results among the utilized thresholding methods in the work with Enhanced signal to noise ratio (SNR) = 10.38 and normalized mean squared error (NMSE) = 0.1344.

Machine learning methods are used to create models that simulate a pipeline leak detection system. A combined feature vector is utilized using wavelet and statistical factors to improve the proposed system performance.

Article info

Received 29 February 2020

Accepted 26 August 2020

Online 8 February 2021

Keywords:

zigbee,
wavelet transform,
statistical features,
wireless sensor network (wsn),
classification

Available online: <https://doi.org/10.26552/com.C.2021.2.C44-C53>

ISSN 1335-4205 (print version)

ISSN 2585-7878 (online version)

1 Introduction

Many signals in the real-time processing need filtering prior to feature extraction or further processing. The wavelet analysis tools play an important role in signal processing theory, also wavelet transformed coefficients are considered as features of signals, these features can be used for classification and recognition for some classes or objects respectively [1]. In this work a dataset for water pipelines network are used to train some machine learning based models. A comparison study for different classifier is established. Some performance measures are used to validate the comparison test. Although there exist some popular wavelet thresholding methods, such as soft and hard, however researcher still seek for the new thresholding method to improve the wavelet based denoising or compression algorithms. As a contribution to the knowledge of art, a novel thresholding method is proposed which try to improve the denosied signal in terms of increasing improved signal to noise ratio (ISNR) factor or minimize the mean squared error (MSE). The principal core for the proposed wavelet rectified thresholding method is demonstrated in the principle of cutting the details wavelet coefficients in the ripple manner rather than popular flat method. Basically, based on the frequency domain analysis this new thresholding technique will reserve some of the energy in the details wavelet coefficient and hence

improve the denoised signal after re-transforming back the threshold wavelet coefficients into time domain [2]. In the last decade many proposals are suggested by researchers for monitoring the water pipelines network, some of these methods are based on distributed sensor motes and Wireless sensor network to read the status for the fluid inside the pipeline and monitor the network accordingly. Other methods used the statistical features for the gathered data and use them for diagnosing the system status at different scenarios [3]. In this work, different features are used to construct the feature vector, some of them based on statistical values and others used wavelet decomposition coefficients, the work also utilized some classifiers, which used the correlation between two features to enhance the classification accuracy. The real time signals are gathered from wireless pressure sensors, which are installed along a water pipeline at different locations. Artificial leak valves are inserted within the pipeline to make manual leak with different size ranging from small leak (less than 12.7mm) and through medium leak (between 0.5" and 1") continuing up to the large leak (more than 25.4mm) [4]. These measured pressure signals are processed and their extracted features are fed to different classifier models to diagnose the existence of a system leak. Three main classes are considered for this problem (large, medium, and small); classification confusion matrix is used for classification comparison test.



2 Theoretical background

2.1 Support vector machine

The support vector machine (SVM) is a machine learning tool that can be used for solving problems of seeking for the minimizing certain parameters such as MSE. The principle idea in the SVM theory is to transform the given data to other space using some transformation kernel. New representation for the data has been gotten, such that a hyper plane was used to classify the data into classes with different proposed intervals. Classification algorithms represent one of supervised learning which has ability to create a new model based on a given dataset. This predictive model can be used for further response estimation when the new test data is applied to the predictive model.

2.2 Nearest neighbor

This technique uses the principle of distance calculation between certain data point from the dataset and the closet predication. There are many distances' measure suggestion using different aspects such as Euclidian, Spearman, Jaccard, correlation, cosine, and Chebyshev distance. The KNN can be used for solving different kind of problems in machine learning such as classification, clustering, and regression. It has wide application in new area such as signal processing, data mining and database analysis.

2.3 Zigbee protocol

The IEEE STD 802.15.4 standard describes the physical layer (PHY) and medium access control (MAC) sub layer particulars for low-information rate remote availability with fixed, compact and moving gadgets with no battery or constrained battery utilization prerequisites, normally working in the individual working space of 10m. It is predicted that, contingent upon the application, a more drawn-out range at a lower information rate might be an adequate tradeoff. The IEEE 802.15.4 standard (2003) characterizes the gadget types that can be utilized in a low rate wireless personal area network (LR-WPAN), which are Full Functional Device (FFD) and Reduced Functional Device (RFD). The RFD can be utilized in basic applications in which they do not have to transmit a lot of information and they need to discuss just with a particular FFD. The FFD can fill in as a personal area network (PAN) organizer, as a facilitator, or as a straightforward gadget. It can speak with either another FFD or a RFD [5].

3 Proposed system

In this work, a new technique for the leak detection and classification is suggested, this technique combines

two methods from machine learning science. Dataset for the problem to be solved is provided from the real-time measurements for a water pipeline network. Artificial leaks are created with different sizes at various locations along the pipeline. After data gathering process from wireless pressure sensor nodes, the data is conditioning using denoising process based wavelet transform method. A new proposed Hardy-Rectified thresholding is suggested by authors of this work, a universal threshold and five decomposing levels are used to improve the SNR at the input to the proposed classifier. Both domains (statistical and wavelet) features are extracted and correlated, and then the new correlated features are used to construct a feature matrix that is used for training different classifier models. Comparison test is held between different classifiers types, in order to summarize the obtained results for about 23 classifiers with different criteria and kernels [6]. The Zigbee protocol and its transceiver modules are adopted for data transmission between wireless sensor nodes in the network. A comparison study between different modulation types is established, the results show a sample for the signal between transmitter and receiver of the two nodes with a proper choice of modulation scheme and signal to noise ratio.

4 Proposed wavelet thresholding mathematical models

Denoising for signals and images can be achieved using different methods, wavelet based method can be considered as a one of the well-known methods for denoising, which can be summarized by three main steps: firstly the wavelet decomposition for the signal or image into two coefficients known as approximations and details, then thresholding the details coefficients by proper threshold and finally reconstruct the approximations and threshold details coefficients to reconstruct the de-noised signal or image. There are many thresholding functions used in the literature, such as soft and hard thresholding, however in this work a new proposed thresholding function, known as rectified sine thresholding function, is proposed with two models (Softy and Hardy) in order to enhance the denoising process. The models of the thresholding techniques are given below:

Softy-rectified thresholding

$$Q_j = \begin{cases} [sign(W_j)(|W_j - \lambda|)] \\ + a \text{ rectfied } \sin(b\pi W_j) & |W_j| \geq \lambda \\ 0 & |W_j| < \lambda \end{cases}, \quad (1)$$

Hardy- rectified thresholding

$$Q_j = \begin{cases} [W_j] + a \text{ rectfied } \sin(b\pi W_j) & |W_j| \geq \lambda \\ 0 & |W_j| < \lambda \end{cases}, \quad (2)$$

Hard thresholding

$$Q_j \begin{cases} W_j & |W_j| \geq \lambda \\ 0 & |W_j| < \lambda \end{cases}, \quad (3)$$

Soft thresholding

$$Q_j = \begin{cases} [\text{sign}(W_j)(|W_j - \lambda|)] & |W_j| \geq \lambda \\ 0 & |W_j| < \lambda \end{cases}, \quad (4)$$

Softy-sawtooth thresholding

$$Q_j = \begin{cases} [\text{sign}(W_j)(|W_j - \lambda|) \\ + a \text{ sawtooth}(b\pi W_j)] & |W_j| \geq \lambda \\ 0 & |W_j| < \lambda \end{cases}, \quad (5)$$

Hardy-sawtooth thresholding

$$Q_j = \begin{cases} [W_j] + a \text{ sawtooth}(b\pi W_j) & |W_j| \geq \lambda \\ 0 & |W_j| < \lambda \end{cases}, \quad (6)$$

where:

W_j - the input signal to wavelet thresholding at level j ,

λ - threshold,

Q_j - the output signal from wavelet thresholding at level j .

5 Results and discussion

Referring to the obtained results, Tables 1 through 3 explain the procedure followed by the wavelet denoising process. In Table 1 a comparison has been done for different wavelet mother functions from the coiflet family, which yields the success for coif2 wavelet mother function with hardy-rectified thresholding, while other parameters are set to: the thresholding rule as universal and 5 decomposing levels are used with two thresholding methods (hard and hardy-rectified). After success for coif 2 in the first stage, it is then used (as shown in Table 2) with different decomposing levels to search for the optimal level. However, level 5 is the best level, due to its highest ISNR and lowest MSE. In Table 3 different kinds for thresholding methods (soft, hard, softy-rectified, and hardy-rectified) have been used with various threshold selection rules (huresure, sqtwolog, rigrsure and minimaxi). The hardy-rectified thresholding with sqtwolog as threshold (universal) selection rule is the best obtained results from Table 3. Different classifiers are used with the same dataset,

which is divided to three parts - 60% for training the model, 30% for validating it and 10% for testing. The confusion matrix is used for comparison between different classifiers [7]. Referring to Table 4, the best classification accuracy is obtained with linear discriminant model, however, fine tree classifier model appeared to be a model with the highest prediction speed and minimum training time. Figure 1 shows the novel wavelet thresholding method (hardy-rectified and softy-rectified) as compared to the conventional methods (hard and soft); 22 machine learning models are trained with the same dataset (see Table 4), Figures 2 through 6 explain the performance for the best classifier model (linear discriminant). Figure 2 explains the confusion matrix between the true and predicted classes for 100 samples of different leak sizes. Figure 3 shows the percentage confusion matrix with positive predictive value and false discovery rate for each class. Figure 4 show the scatter plot for the prediction model with three leak size (large, medium, and small); where the cross mark indicates the misclassification point. From Receiver operating characteristic (ROC) curve in Figure 5 the classifier shows 98% true positive rate with 2% false positive rate [8]. Table 5 explains a comparison between the proposed method and different wavelet thresholding methods in recent researches, like the mixed method [9], improved method [10] hierarchical method [11] and adaptive method [12]. Finally, Figure 6 shows the classifier model with parallel coordinates for the four predictors and different standard deviation ranges.

6 Conclusions

In this work different classifiers from the machine learning toolbox in MATLAB program are used to solve the problem of the leak detection and size prediction in the water pipeline network. Pressure sensors are utilized and installed properly along the water pipelines for data gathering prior to feature extraction. Different features had been used for classification problem, statistical features and the wavelet-based features are the most popular

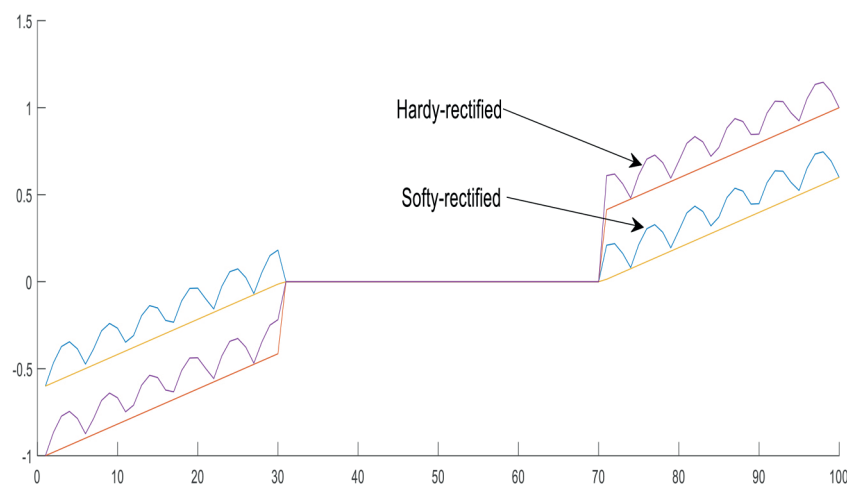


Figure 1 Proposed hardy and softy rectified thresholding methods

used features, but in this work the correlated features, which represent the combination between both features are utilized, as well. Simulation results for the novel wavelet-based rectified thresholding method open the way for further studies and research in deployment of this new technique in wavelet denoising or compression in signals and images. Although the classifiers have different classification accuracies; most of them show

excellent accuracy with acceptable predication speed and satisfactory training time. The Zigbee protocol has good bit error rate (BER) results for different values of the SNR as explained by Figure 7. So, it can be used for wireless sensor network (WSN) deployment and especially with OQPSK 2450MHz, which shows excellent transceiver communications and minimum absolute error at the receiver side, see Figure 8.

True Class	Large	49	1	
	Medium	2	48	
	Small			50
		Large	Medium	Small
		Predicted Class		

Figure 2 Confusion matrix for the best model

True Class	Large	96%	2%	
	Medium	4%	98%	
	Small			100%
Positive Predictive Value		96%	98%	100%
False Discovery Rate		4%	2%	
		Large	Medium	Small
		Predicted Class		

Figure 3 PPV and FDR for the best model

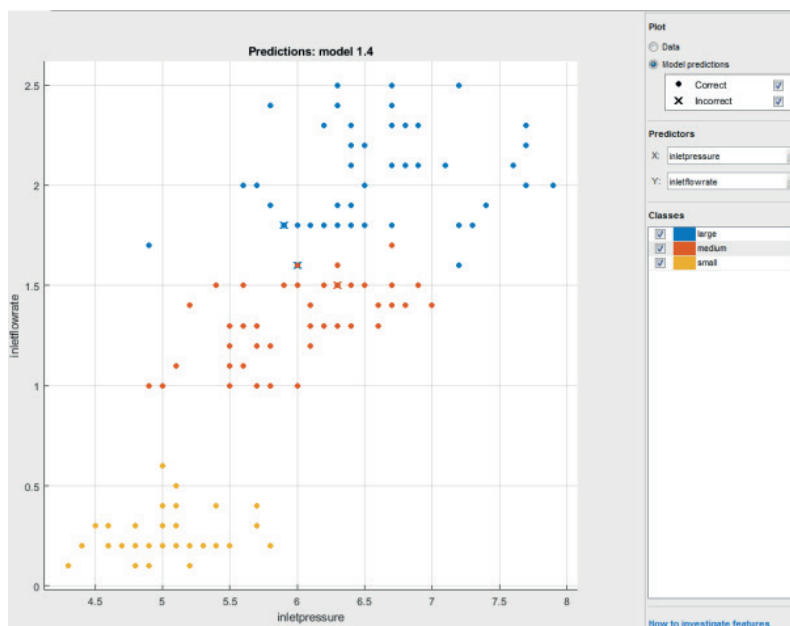


Figure 4 Scatter plot for the best model

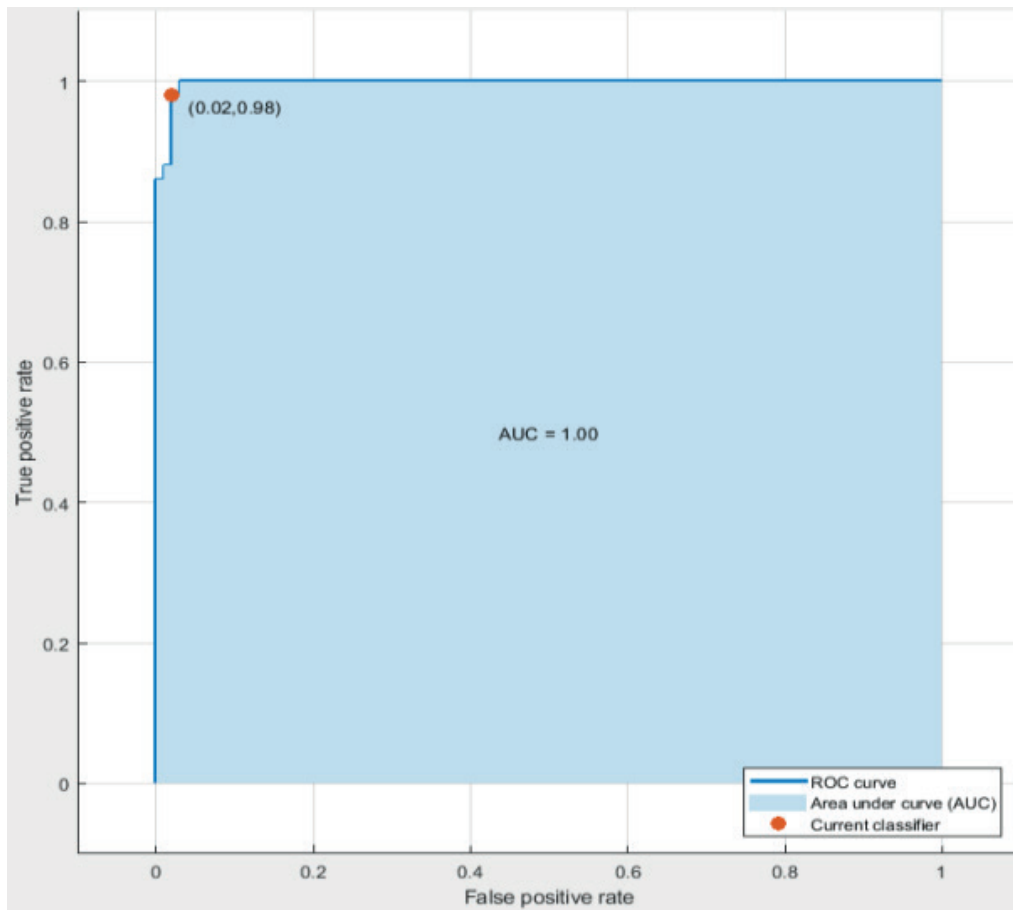


Figure 5 ROC for the best model

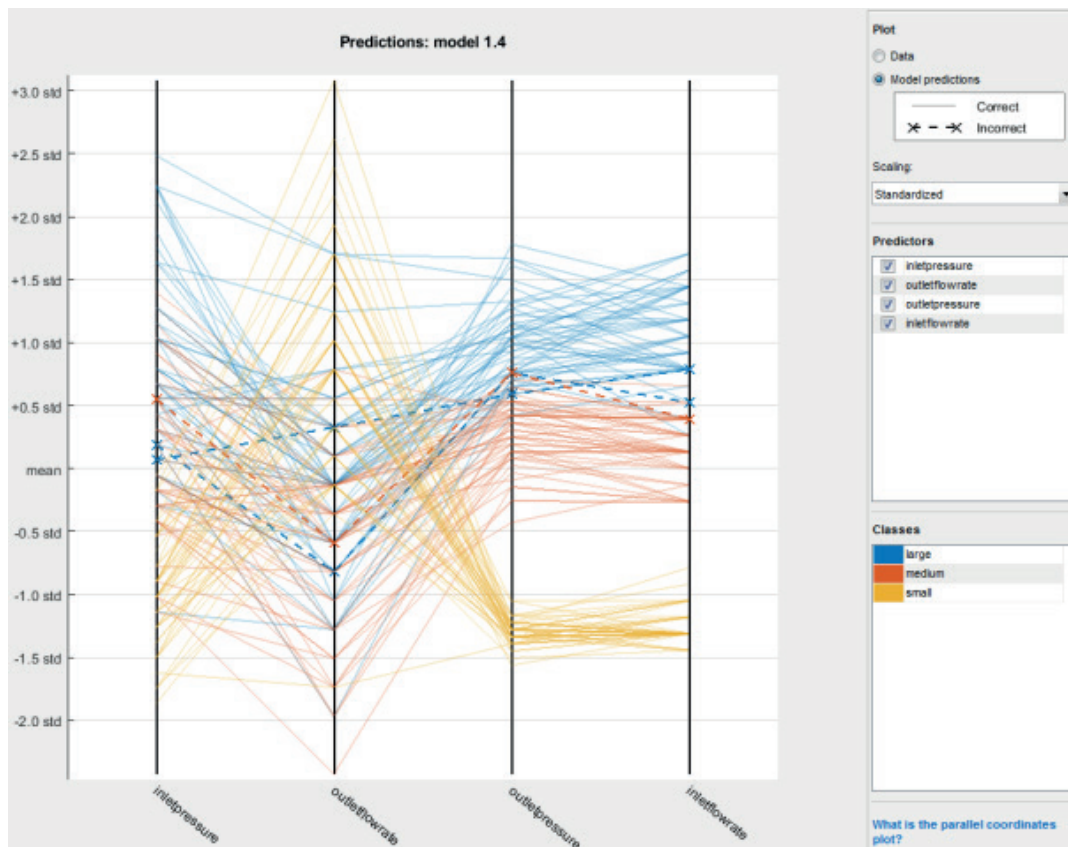


Figure 6 Parallel coordinates plot for the best model

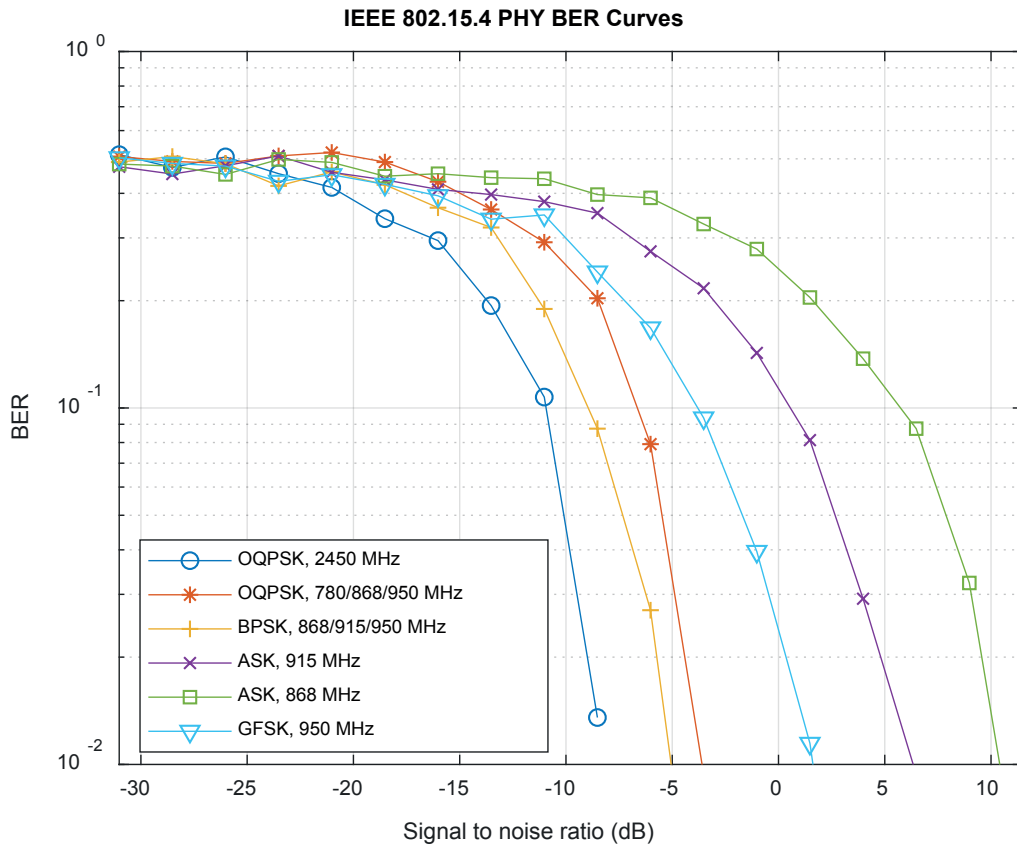


Figure 7 BER versus SNR for different modulation types in the Zigbee transceiver

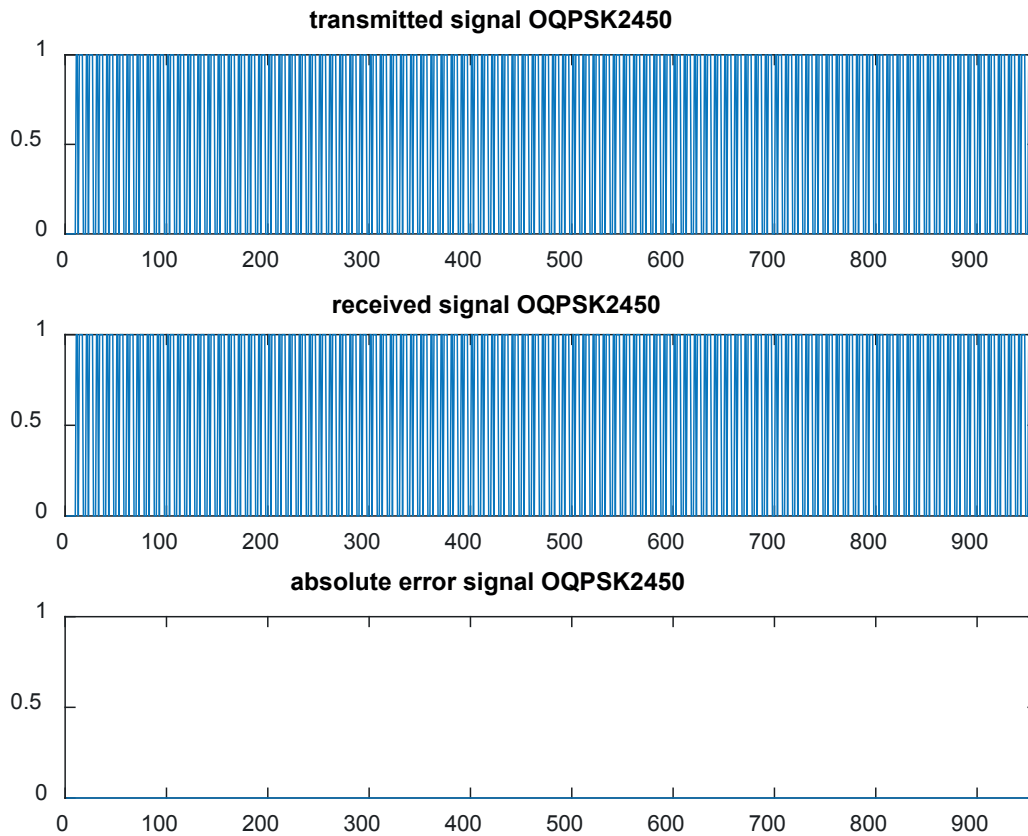


Figure 8 Zigbee transceiver with OQPSK 2450

Table 1 Signal denoising results with various Coiflet wavelets

type of signal		lev	rule of threshold			
noisy signal		5	square-root-log / hardy-rectified			
family of wavelet		coif 5	coif 4	coif 3	coif 2	coif 1
measure index	Norm MSE	0.1408	0.1473	0.1364	0.1345	0.15
	Root MSE	0.2329	0.2456	0.2236	0.2193	0.2506
	Enhanced SNR	9.3589	8.4989	10.0509	10.3875	8.1815
	PDR	0.1669	0.1727	0.1627	0.1607	0.175

type of signal		lev	rule of threshold			
noisy signal		5	square-root-log / hard			
family of wavelet		coif 5	coif 4	coif 3	coif 2	coif 1
measure index	Norm MSE	0.141	0.1473	0.1364	0.1348	0.1503
	Root MSE	0.2333	0.2456	0.2235	0.2201	0.2511
	Enhanced SNR	9.3274	8.4972	10.0549	10.3265	8.1474
	PDR	0.1671	0.1728	0.1626	0.161	0.1753

Table 2 Denoising results using the Coiflet 2 wavelet with various levels of decomposition

type of signal		method of thresholding			type of wavelet	rule of threshold	rescaling for threshold	
noisy signal		hard			coif 2	square root log	estimated noise for each level	
level of decomposition	1	2	3	4	5	6	7	
Measured index	NMSE	0.2452	0.183	0.1497	0.1375	0.1348	0.1349	0.3284
	RMSE	0.37	0.3007	0.2501	0.226	0.2201	0.2202	0.4407
	ESNR	2.8075	5.5157	8.2126	9.8666	10.3265	10.3173	0.7111
	PDR	0.2298	0.198	0.1748	0.1638	0.161	0.1611	0.2622

type of signal		method of thresholding			type of wavelet	rule of threshold	rescaling for threshold	
noisy signal		hardy-rectified			coif 2	square root log	estimated noise for each level	
decomposition level	1	2	3	4	5	6	7	
Measured index	NMSE	0.2452	0.183	0.1497	0.1375	0.1345	0.1348	0.328
	RMSE	0.37	0.3007	0.2501	0.226	0.2193	0.22	0.4404
	ESNR	2.8075	5.5157	8.2126	9.8666	10.3875	10.3306	0.7184
	PDR	0.2298	0.198	0.1748	0.1638	0.1607	0.161	0.2621

Table 3 Denoising results using the Coiflet 2 with various rules of threshold

rule of threshold	square-root-log		regular-sure		mini-maxi		
thresholding method	hard	soft	hard	soft	hard	soft	
performance	norm MSE	0.1347	0.1441	0.2109	0.1432	0.1834	0.1385
	root MSE	0.22	0.2396	0.3345	0.2379	0.3013	0.2282
	enhanced SNR	10.3264	8.887	4.0853	9.0005	5.4826	9.6955
	PDR	0.1609	0.1699	0.2135	0.1692	0.1982	0.1647

rule of threshold	square-root-log		regular-sure		mini-maxi		
thresholding method	hardy-rectified	softy-rectified	hardy-rectified	softy-rectified	hardy-rectified	softy-rectified	
performance	norm MSE	0.1344	0.1422	0.2022	0.1406	0.177	0.1393
	root MSE	0.2192	0.2359	0.3246	0.2327	0.2926	0.23
	enhanced SNR	10.3874	9.1432	4.4796	9.3634	5.8877	9.5587
	PDR	0.1606	0.1682	0.2089	0.1668	0.1943	0.1656

Table 4 Classification learning models comparison

model type	accuracy (%)	prediction speed (obs/ms)	training time (s)
fine tree	96	6.4	0.99
medium tree	96	5	1.25
coarse tree	96	4.7	1.17
linear discriminant	98	2.2	3.21
quadratic discriminant	97.3	1.5	2.85
linear SVM	97.3	1.4	6.11
quadratic SVM	97.3	1.5	5.85
cubic SVM	95.3	3	6.55
fine Gaussian SVM	92.7	2	6.43
medium Gaussian SVM	95.3	2.7	6.21
coarse Gaussian SVM	96	2.6	5.24
fine KNN	94.7	1.8	6.04
medium KNN	95.3	3.6	5.9
coarse KNN	66	3.2	5.74
cosine KNN	86	2.7	6.48
cubic KNN	95.3	3.4	6.33
weighted KNN	96.7	3.4	6.23
boosted trees	33.3	2.9	7.67
bagged trees	95.3	0.48	10.01
subspace discriminant	96	0.35	12.44
subspace KNN	96.7	0.24	14.33
RUSboosted trees	33.3	6.1	12.78

Table 5 Comparison between different thresholding methods in recent researches

reference	variance of the simulated signal	signal type	thresholding method	enhancement SNR
[9]	30	ECG	hard	7.2
[9]	30	ECG	soft	11.2
[9]	30	ECG	garrote	10.5
[9]	30	ECG	scad	9.5
[9]	30	ECG	mixed	8.9
[10]	15	battery	improved	14.99
[10]	15	battery	donoho	14.17
[10]	15	battery	standard	15.77
[10]	15	blocks	improved	15.18
[10]	15	blocks	donoho	14.53
[10]	15	blocks	standard	15.38
[10]	15	bumps	improved	14.5
[10]	15	bumps	donoho	15.67
[10]	15	bumps	standard	15.67
[11]	15	Lena	global	10.2
[11]	15	Lena	hierarchical	10.7
[11]	20	Lena	global	7.2
[11]	20	Lena	hierarchical	7.8
[12]	10	echo	soft	19.6
[12]	10	echo	hard	12.6
[12]	10	echo	proposed	22.6
[12]	10	echo	adaptive	21.8
this work	30	pressure	hardy-rectified	10.38
this work	30	pressure	softy- rectified	9.55
this work	30	pressure	hard	10.32
this work	30	pressure	Soft	9.69

References

- [1] RASHID, S., QAISAR S., SAEED H., FELEMBAN, E. Performance analysis of leak detection algorithm in long range pipeline networks using transform analysis. In: IEEE Conference on Systems, Process and Control ICSPC '13: proceedings. IEEE. 2013. p. 10-15.
- [2] SHAMSI, M. RAZZAGHI, M., NAZARZADEH, J., SHAFIEE, M. Haar wavelets method for solving pocklington's integral equation. *Kybernetika* [online]. 2004, 4, p. 491-500. ISSN 0023-5954, eISSN 1805-949X. Available from: <http://www.kybernetika.cz/content/2004/4/491/paper.pdf>
- [3] SAEED, H., AIL, S., RASHID, S., QAISAR, S., FELEMBAN, E. Reliable monitoring of oil and gas pipelines using wireless sensor network (WSN). In: IEEE 9th International System of Systems Engineering Conference SOSE: proceedings. IEEE. 2014. eISBN 978-1-4799-5227-4. Available from: <https://doi.org/10.1109/SYSESE.2014.6892493>
- [4] HAIDINE, A., EL HASSANI, S., AQQAL, A. Evaluation of deployment and technology scenarios for metropolitan wireless networks in smart distribution grid. *Journal of Engineering Science and Technology*. 2019, 14(2), p. 998-1018. ISSN 1823-4690.
- [5] AKKAR, H., AL-SAMARRIE, A., SAEED, A. Simulation design of back propagation neural system of sensor network trained by particle swarm optimization. *International Journal of Scientific and Engineering Research*. 2016, 7(4), p. 576-582. ISSN 2229-5518.
- [6] WIBOWO, A., SETIANTO D., WISNU, W. P., JATMIKO, W. Optimization of backpropagation using Nguyen-Widrow and stimulus-sampling for breast cancer classification and feature selection. *Journal of Engineering Science and Technology*. 2019, 14(6), p. 3437-3456. ISSN 1823-4690.

- [7] AKKAR, H., JASIM, F. Intelligent Training Algorithm for Artificial Neural Network EEG Classifications. *International Journal of Intelligent Systems and Applications* [online]. 2018, **10**(5), p. 33-40. ISSN 2074-904X, eISSN 2074-9058. Available from: <https://doi.org/10.5815/ijisa.2018.05.04>
- [8] HUSSAIN, R. H., JABR, Z. F., SALEH, S. R. Improved localisation algorithm for wireless sensor networks by using XBEE. *Journal of Engineering Science and Technology*. 2019, **14**(6), p. 3470-3480. ISSN 1823-4690.
- [9] KOTESWARARAO, M. *Analysis of de-noising on different signals using new thresholding function' spaces-2018*. Department of ECE, 2018.
- [10] HUI, L., WEIDA, W., CHANGLE, X., LLJIN, H., HAIZHAO, N. A de-noising method using the improved wavelet threshold function based on noise variance estimation. *Mechanical Systems and Signal Processing* [online]. 2018, 99, p. 30-46. ISSN 0888-3270. Available from: <https://doi.org/10.1016/j.ymssp.2017.05.034>
- [11] ZHANG, J., ZHU, Q., ZHANG, J., SONG, L., WANG, J. A novel algorithm for threshold image de-noising based on wavelet construction. *Cluster Computing* [online]. 2019, **22**, p. 12443-12450. ISSN 1386-7857, eISSN 1573-7543. Available from: <https://doi.org/10.1007/s10586-017-1655-0>
- [12] XIAOBIN, X. Single pulse threshold detection method with lifting wavelet denoising based on modified particle swarm optimization. *Infrared Physics and Technology*. 2018, **88**, p. 174-83. ISSN 1350-4495. Available from: <https://doi.org/10.1016/j.infrared.2017.11.023>