Sakarya University Journal of Science, 22(4): 1142-1151, 2018.

SAKARYA UNIVERSITY
JOURNAL OF SCIENCE

e-ISSN: 2147-835X
http://www.saujs.sakarya.edu.tr

Received	Revised	Accepted	DOI
19.03.2018	02.07.2018	18.04.2018	10.16984/saufenbilder.407686

Robust ECG data compression method based on ε-insensitive Huber loss function

Ömer Karal*[1] İlyas Çankaya[1]

Abstract

Electrocardiogram (ECG) signals are continuously monitored for early diagnosis of heart diseases. However, a long-term monitoring generates large amounts of data at a level that makes storage and transmission difficult. Moreover, these records may be subject to different types of noise distributions resulting from operating conditions. Therefore, an effective and reliable data compression technique is needed for ECG data transmission, storage and analysis without losing the clinical information content. This study proposes the ε-insensitive Huber loss based support vector regression for the compressing of ECG signals. Since the Huber loss function is a mixture of quadratic and linear loss functions, it can properly take into account the different noise types in the data set. Compression performance of the proposed method has been assessed using ECG records from the MIT-BIH arrhythmia database. Experimental results demonstrate that the proposed loss function is an attractive candidate for compressing ECG data.

Keywords: Data Compression, Electrocardiogram, Huber loss function, Support Vector Regression

1. INTRODUCTION

ElectroCardioGram (ECG) plays a very important role in the diagnosis and analysis of heart diseases in patient. In order to detect any heart diseases in advance, the ECG signals are continuously recorded, stored and transmitted over digital communication networks. However, these types of records produce large amounts of data that will make storage and transmission difficult. Moreover, such records may be subject to unknown complex noise due to environment. The above-mentioned problems can be overcome by effectively compressing ECG signals while preserving the clinical information content in the reconstructed signal.

Until now, many algorithms have been proposed for compressing ECG signals. Existing data compression algorithms can be roughly classified into three classes: (i) direct, (ii) parameter extraction, and (iii) transform-based methods [1]. Direct data compression methods, also known as time domain techniques, attempt to remove redundancies in the actual samples of ECG signal. However, it fails to achieve a high data rate in terms of preserving clinically important contents. Some of the techniques for this category are amplitude zone time coding (AZTEC) algorithm [2], turning point (TP) algorithm [3], coordinate reduction time system (CORTES) algorithm [4]. Parameter extraction methods are based on extracting some dominant parameters (features) from the raw signal for use in the reconstruction process. Typical examples of this category can be shown as Linear Prediction (LP) based algorithm [5], vector quantization (VQ) based algorithm [6], and template matching (TM) based algorithm [7].

[1] Ankara Yıldırım Beyazıt Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Elektrik-Elektronik Mühendisliği, Ankara- karal@ybu.edu.tr, icankaya@ybu.edu.tr

However, since parameter extraction methods are irreversible they are therefore do not have widespread use in practice [8]. In transform-based compression methods, the raw signal is expressed as the weighted sum of basis functions. The weights (coefficients) of these basis functions are properly coded and transmitted instead of the original signal [9]. Transform-based techniques are usually preferred for ECG data compression since they achieve higher compression gain and are more insensitive to noise in the original ECG signals. The most important examples in this category can be listed as Discrete Cosine Transform (DCT) [10], Fourier Transform (FT) [11], Wavelet Transform (WT) [12]. For a detailed review of these methods, see [13-15].

Wavelet transform-based techniques have recently attracted considerable interest owing to some important features such as time frequency localization, and energy compression [13]. However, in wavelet transform-based techniques, there is a proportional relationship between block size and compression ratio. As the block size increases, the compression ratio increases for a specific error criterion; but the computation time and storage requirements of adaptive wavelet coding schemes will also increase in the same way. For these reasons, how to determine the block size in wavelet transform is still a fundamental problem [13].

More recently, unlike the existing transform-based compression methods, Karal [16] has shown that ECG data can be compressed optimally according to the given error tolerance using Support Vector Regression (SVR) method. In addition, it is noted that the proposed SVR technique performs better than the well-known FT, DCT, and WT techniques.

In SVR theory, loss functions play an important role because they represent the properties of the error distribution in the data set [17]. According to the Bayesian approach, there is a powerful relationship between loss functions and error distributions. If the error distribution in the data set is known, the corresponding optimal loss function can be derived using the Bayesian approach. For example, square loss function is optimal for Gaussian error distribution. Therefore, different loss functions lead to the creation of different optimization costs [17]. Classical SVR method [16] uses ε-insensitive Laplacian (called as Vapnik) loss function which is optimal for ε-insensitive Laplacian noise distribution. However, ECG recordings may be subject to different noise distributions resulting from operating conditions. In this case, classical SVR method will not be optimal. To address the aforementioned problems, this study proposes the ε-insensitive Huber loss based SVR for the compressing of ECG signals. This loss function is known as a robust loss function in the literature [18] since it is a mixture of square and linear loss functions.

The presented ε-insensitive Huber loss function has a significant advantage because it can properly take into account different noise distributions in terms of robustness. It also provides sparsity (compression) in the solution presentation by ignoring small noisy training samples falling into the ε-insensitive region. The compressed signal is expressed as the weighted sum of basis functions. Unlike other transformation-based compression methods, the number, position, and shape of these functions are automatically determined by the SVR algorithm, which is based on the solution of the quadratic optimization problem.

Rest of this paper is organized as follows. In section 2, the ε-insensitive Huber loss based SVR formulation is explicitly given. In section 3, the experimental results get from the compression of

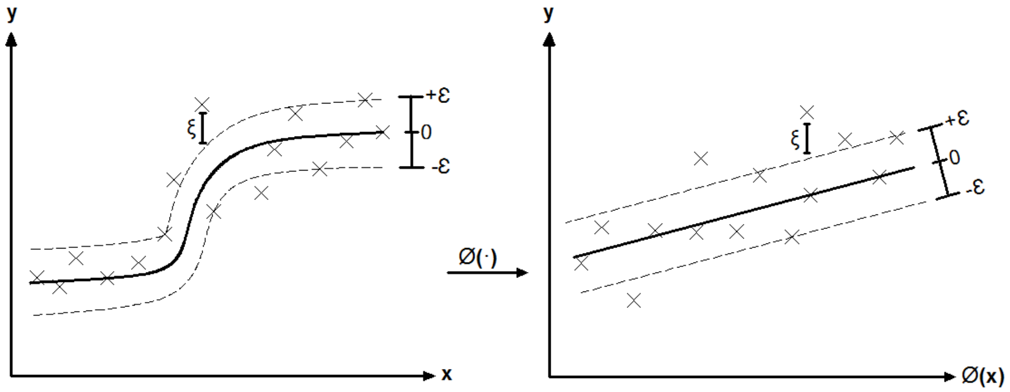

Figure 1. In the SVR method, the training samples represented by nonlinear regression in the input space are expressed by linear regression in the high dimensional feature space using nonlinear basis functions.

ECG data with Huber-SVR are shown. In the last section, evaluations are made on the results obtained from the experiments.

2. SUPPORT VECTOR REGRESSION

Support vector regression (SVR) provides a good generalization ability since it tries to minimize both the empirical risk minimization and the structural risk minimization principle. Therefore, SVR has been used in many fields such as biomedical [19-21], time series forecasting [22-24], and renewable energy [25-27].

Given N pairs of training samples $D=\{(\mathbf{x}_s, y_s)|s=1,\cdots N\}$, in which $\mathbf{x}_s \in R^n$ is the sth input vector, and $y_s \in R$ is the actual output for the input $\mathbf{x}_s$. In the classical SVR formulation, training samples are moved to a high dimensional space by means of nonlinear function $\varphi(\cdot): R^n \rightarrow R^m$ and then a linear model is applied (see Figure 1).

$$f(\mathbf{x}, \boldsymbol{w}) = \boldsymbol{w}^T \varphi(\mathbf{x}) + b \qquad (1)$$

where $f(\cdot)$ is the estimation of y_s, $\boldsymbol{w} \in R^m$ is the model parameter (weight) vector, and b is a threshold to be determined in the function. The SVR optimization problem consists of two parts: minimizing errors (represented as a loss function $L_\varepsilon(y_s, f(\mathbf{x}_s))$) and minimizing the model parameters (weights, $\boldsymbol{w}$) representing the model.

$$\min_{\boldsymbol{w}\in R^m, b\in R} J(\boldsymbol{w}, b) = \left\{ \frac{1}{2} \| \boldsymbol{w} \|_2^2 + C \sum_{s=1}^{N} L_\varepsilon(e_s) \right\} \qquad (2)$$

where, $C \in R^+$ is a positive number user defined parameter, and $L(e_s)_\varepsilon = L(y_s - f(\mathbf{x}_s))$ is the ε insensitive loss function which is symmetric convex. It also has two discontinuities at $\pm\varepsilon \geq 0$ in the first derivative and is zero in the predetermined ε value. Small noisy training samples falling into the ε-insensitive zone of the loss function are not included in the solution presentation. Therefore, SVR yields sparse (compressed) model in the solution representation.

There are many choices for the loss function. The classical SVR method uses the ε-insensitive Laplace (Vapnik) loss function, which is called the ε-insensitive absolute loss function (3). Vapnik's loss function ignores the errors lower than the user specified ε value (see Figure 2). The mathematical definition of it is given as follows.

$$|y_s - f(\mathbf{x}_s)|_\varepsilon = \begin{cases} 0 & for\, |y_s - f(\mathbf{x}_s)| < \varepsilon \\ |y_s - f(\mathbf{x}_s)| - \varepsilon & otherwise \end{cases} \qquad (3)$$

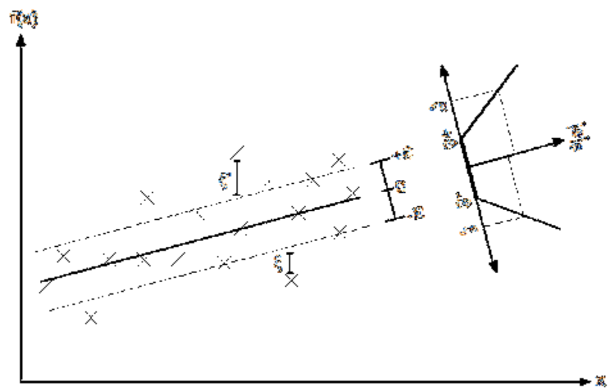

Figure 2 ε insensitive Laplace (Vapnik) loss function.

From Bayesian perspective, Vapnik's loss function is optimal for the ε-insensitive Laplace noise distribution. However, in some practical applications, the observed data may be subject to different noise distributions depending on the operating conditions. In order to deal with different error distributions, this paper introduces the ε-insensitive Huber loss function for the compressing of ECG signals. It is defined as follows [18].

$$L_{\varepsilon}(e_s)=\begin{cases} e_s-\dfrac{\mu}{2} & if\ \ e_s \ge \mu \\ \dfrac{1}{2\mu}(e_s-\varepsilon)^2 & if\ \ \varepsilon \le e_s < \mu \\ 0 & if\ \ -\varepsilon < e_s < \varepsilon \\ \dfrac{1}{2\mu}(e_s+\varepsilon)^2 & if\ -\mu < e_s < -\varepsilon \\ e_s+\dfrac{\mu}{2} & if\ -\mu \le e_s \end{cases} \tag{4}$$

By substituting (4) into (2),

$$\min_{w\in R^m, b\in R}\left\{\frac{1}{2}\|w\|^2 + C\sum_{s=1}^{N}\left[L(\xi_s)+L(\xi_s^*)\right]\right\} \tag{5}$$

$$\text{subject to}\begin{cases} y_s - w^T\varphi(\mathbf{x}_s) - b \le \varepsilon + \xi_s \\ -y_s + w^T\varphi(\mathbf{x}_s) + b \le \varepsilon + \xi_s^* \\ \xi_s, \xi_s^* \ge 0,\ \varepsilon \ge 0\ ,\ s\in\{1,\ldots,N\}, \end{cases} \tag{6}$$

where, ξ_s and $\xi_s^{'}$ are called slack variables to deal with positive and negative errors outside the epsilon-insensitive zone, respectively. The value of the slack variables is considered to be quadratic if it is between ε and μ, while it is considered linear if it is greater than μ (see Figure 3).

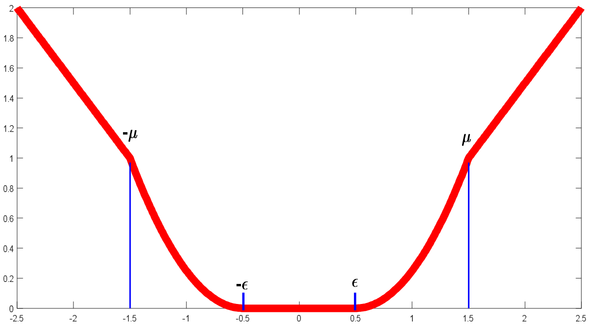

Figure 3. ε-insensitive Huber loss function

To solve primal optimization problem under the constraints in (5), the Lagrange function is constructed as in the following.

$$\begin{aligned}\min_{\substack{w\in R^m, b\in R \\ \alpha_s,\alpha_s^{'},\gamma_s,\gamma_s^{'},\xi_s,\xi_s^{'}\ge 0 \\ s\in\{1,\ldots,N\}}} & J(w,b,\alpha_s,\alpha_s^{'},\gamma_s,\gamma_s^{'},\xi_s,\xi_s^{'}) = \\ & \frac{1}{2}w^T w + C\sum_{s=1}^{N}\left[L(\xi_s)+L(\xi_s^*)\right] \\ & -\sum_{s=1}^{N}\alpha_s(\varepsilon+\xi_s-y_s+w^T\varphi(\mathbf{x}_s)+b) \\ & -\sum_{s=1}^{N}\alpha_s^{'}(\varepsilon+\xi_s^{'}+y_s-w^T\varphi(\mathbf{x}_s)-b) \\ & -\sum_{s=1}^{N}(\gamma_s\xi_s+\gamma_s^{'}\xi_s^{'})\end{aligned} \tag{7}$$

where, the positive variables α_s, α_s^* and γ_s, γ_s^* are Lagrange multipliers (dual variables) associated to each training sample. To find optimal solution of the unconstraint optimization problem in (7), Karush-Kuhn-Tucker conditions are applied and then saddle points of Lagrangian function are determined as shown in the following.

$$\frac{\partial J}{\partial b}=\sum_{s=1}^{N}(\alpha_s^*-\alpha_s)=0 \tag{8}$$

$$\nabla_w J = w - \sum_{s=1}^{N}(\alpha_s-\alpha_s^*)\varphi(\mathbf{x}_s)=0 \tag{9}$$

$$\frac{\partial J}{\partial \xi_s}=C\frac{\partial L(\xi_s)}{\partial \xi_s}-\alpha_s-\gamma_s=0 \tag{10}$$

$$\frac{\partial J}{\partial \xi_s^*}=C\frac{\partial L(\xi_s^*)}{\partial \xi_s^*}-\alpha_s^*-\gamma_s^*=0 \tag{11}$$

If the equations (8) - (11) are substituted in (7), the primal variables $(w,b,\xi_s,\xi_s^{'})$ are removed and the dual optimization problem is obtained in terms of Lagrange multipliers $(\alpha_s,\ \alpha_s^{'})$.

$$\begin{aligned}\max_{\alpha\in R^N} J(\alpha_s,\alpha_s^*) = & -\frac{1}{2}\sum_{s=1}^{N}\sum_{r=1}^{N}(\alpha_s-\alpha_s^*)\mathrm{K}(\alpha_r-\alpha_r^*) \\ & -\varepsilon\sum_{s=1}^{N}(\alpha_s+\alpha_s^*)+\sum_{s=1}^{N}y_s(\alpha_s-\alpha_s^*) \\ & -\frac{1}{2C}\sum_{s=1}^{N}\mu((\alpha_s^2)+(\alpha_s^*)^2)\end{aligned} \tag{12}$$

$$\text{Subject to } \sum_{s=0}^{N}(\alpha_s^*-\alpha_s)=0,\ \text{and } 0\le\alpha_s^*,\alpha_s\le C \tag{13}$$

where, K denotes the kernel matrix whose entries are the kernel functions $\mathrm{K}(\mathbf{x}_s, \mathbf{x}_r)$ as described as the inner product of two samples $\varphi(\mathbf{x}_s)$ and $\varphi(\mathbf{x}_r)$ in the kernel space.

$$\mathrm{K} = [\mathrm{K}(\mathbf{x}_s, \mathbf{x}_r)]_{s,r} = [\varphi^{\mathrm{T}}(\mathbf{x}_s)\cdot\varphi(\mathbf{x}_r)]_{\mathrm{s,r}} = \begin{bmatrix} \mathrm{K}(\mathbf{x}_1,\mathbf{x}_1) & \mathrm{K}(\mathbf{x}_1,\mathbf{x}_2) & \cdots & \mathrm{K}(\mathbf{x}_1,\mathbf{x}_N) \\ \mathrm{K}(\mathbf{x}_2,\mathbf{x}_1) & \mathrm{K}(\mathbf{x}_2,\mathbf{x}_2) & \cdots & \mathrm{K}(\mathbf{x}_2,\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{K}(\mathbf{x}_N,\mathbf{x}_1) & \mathrm{K}(\mathbf{x}_N,\mathbf{x}_2) & \cdots & \mathrm{K}(\mathbf{x}_N,\mathbf{x}_N) \end{bmatrix} \quad (14)$$

The optimization problem in (12) is a quadratic programming problem, so its solution gives a global (unique) minimum. After solving of (12), the optimal Lagrange multipliers (support vectors, α_s and α_s^*) are obtained, and then the optimal model parameter $\boldsymbol{w}$ (9) can be written as follow.

$$\boldsymbol{w} = \sum_{\mathbf{x}_s \in SV} (\alpha_s - \alpha_s^*)\varphi(\mathbf{x}_s) \quad (15)$$

The linear model in (1) for the test sample x can be expressed as

$$f(\mathbf{x}) = \sum_{\mathbf{x}_s \in SV} (\alpha_s - \alpha_s^*)\mathrm{K}(\mathbf{x}_s, \mathbf{x}) + b \quad (16)$$

where, SV is the set of training samples corresponding to $\alpha_s - \alpha_s^* \neq 0$ (called a support vector). $\boldsymbol{w}$ does not need to be explicitly calculated when evaluating $f(\mathbf{x})$. As seen from (16), the operations needed for SVR model can be carried out directly in the primal space with the kernel matrix without mapping the training samples from the primal space to the high dimensional space with the help of nonlinear functions. This process is known as the “kernel trick” in the literature [28]. It considerably reduces calculation time needed for solving optimization problem.

3. EXPERIMENTAL RESULTS

In this section, various experimental results are presented for the ε-insensitive Huber loss function in the SVR framework. Furthermore, the performance of the proposed ε-insensitive Huber loss is compared with the Vapnik loss under the SVR framework. These experiments are performed in Matlab 2016a environment installed on a personal computer with Intel Core I5 processor 3.0 GHz, 10 GB RAM and 64 bit Windows 10 operating system. As a kernel function, Radial Basis Function (RBF) $\mathrm{K}(\mathbf{x}_s, \mathbf{x}) = \exp(-\|\mathbf{x} - \mathbf{x}_s\|_2^2 / 2\sigma^2)$ is chosen. The user defined optimal parameters (ε, μ, σ, and C) are determined from the sets {0.01, 0.012, 0.014, 0.015, 0.016, 0.018, 0.02, 0.025, 0.03}, {15, 10, 5, 2, 1, 0.5, 0.25, 0.125, 0.1, 0.05, 0.025}, {0.01, 0.011, 0.012, 0.013, 0.014, 0.015}, {1, 2, 4, 8, 16, 32, 64, 128, 256} and respectively, using 5-fold cross validation technique. The performance of the learned SVR networks is evaluated by the three important metrics i.e., compression ratio (CR), reconstruction quality (Root Mean Square Error (RMSE) and Percent Root Mean Square Difference (PRMSD)). Their definitions are given as

$$\mathrm{CR} = \frac{\#\mathrm{TSROS}}{\#\mathrm{SVNCS}} \quad (17)$$

where, #TSROS denotes the number of training samples required for the original signal and #SVNCS specifies the number of support vectors needed for the compressed signal.

$$\mathrm{PRD} = \sqrt{\frac{\sum_{s=1}^{N}[(y_s - f(\mathbf{x}_s)]^2}{\sum_{s=1}^{N}[(y_s)]^2}} \times 100 \quad (18)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{s=1}^{N}(y_s - f(\mathbf{x}_s))^2} \quad (19$$

In the experiments, ECG signals selected from 48 half-hour excerpts of two-channel ambulatory ECG recordings in the MIT-BIH database were used [29]. These recordings were digitized at 360 samples per second per channel. In order to visualize compressing ECG signal, a period ECG signal (normal sinus rhythm, see Figure 4) is extracted from the ECG recordings.

To indicate that the proposed ε-insensitive Huber loss function is suitable for different noise models, the ECG data set is contaminating with Gaussian $\left(1/\sqrt{2\pi}\right)exp(-(x-\mu)^2/2\tau^2)$ and Cauchy noise $1/\pi\tau[1+(x-\mu/\tau)^2]$ distributions with the mean μ=0 and the variance τ=0,01.

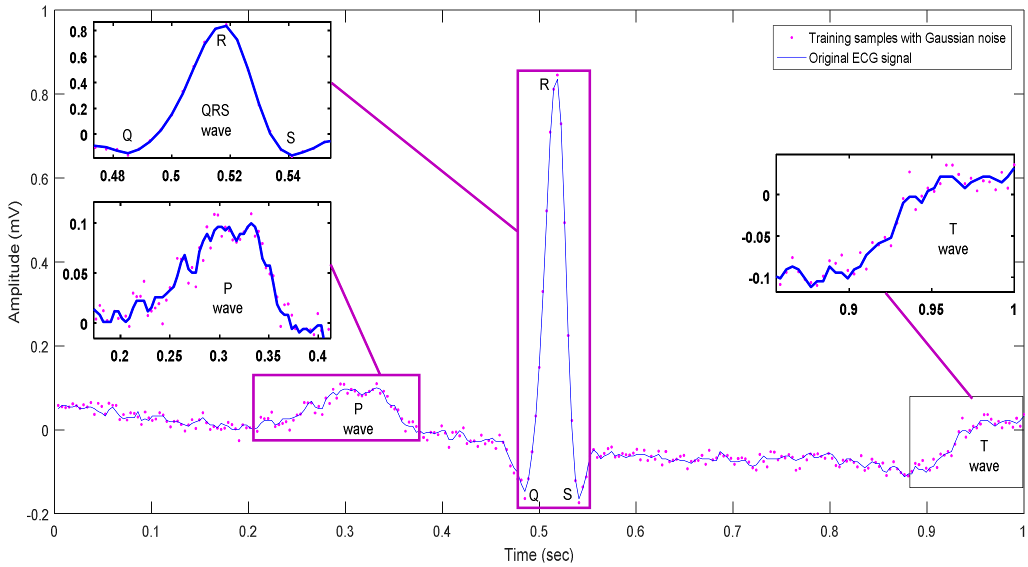

Figure 4. ε-insensitive Huber loss function Original ECG signal (dark blue), the Gaussian noise added ECG signal (magenta dots) and the P, QRS and T wave forms (violet rectangles) that make up the original ECG signal

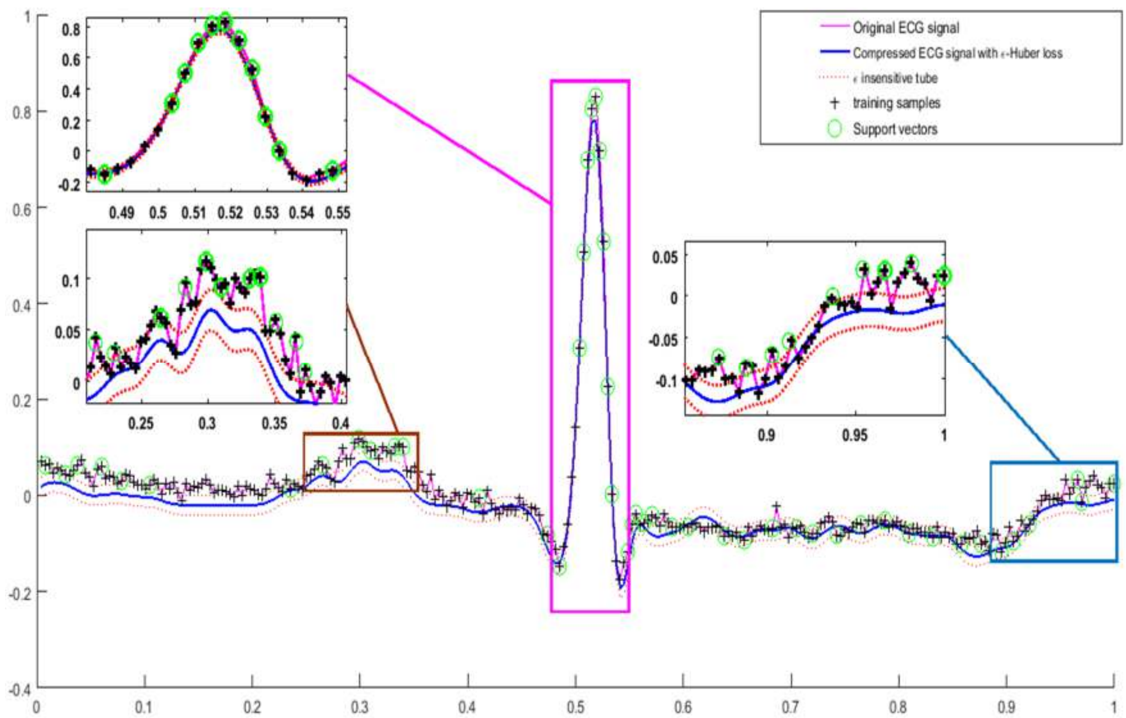

Figure 5. ECG (dark blue) signal compressed by ε-insensitive Huber loss based SVR. The areas separated by the red dashed line below and above the compressed ECG signal show the ε-insensitive regions. The '+' signs in the black color represent the training samples, and the '+' signs in the green circles represent the support vectors (the training samples that yield the compressed ECG signal).

Figures 5 shows the compressed version of the ECG (dark blue) signal contaminated with Gauss noise, using the ε-insensitive Huber loss function with parameters ($\varepsilon = 0.012$, $C = 16$ and $\mu = 0.25$).

As shown in Figure 5, the examples (200 black '+' signs) in the ε-insensitive regions of the Huber loss function are not considered and are therefore not included in the solution. In other words, only samples outside the ε-insensitive regions of the Huber loss function (68 support vectors, black '+' signs in green circles) appear in the solution presentation. Thus, the reconstructed ECG signal contains fewer samples than the original signal, which results in a sparse (compressed) solution.

Table 1. Experimental results of ε-insensitive Huber loss-based SVR for compressing ECG signals polluted with Gaussian noise distribution

Model Parameters		#TS	#SV	w	RMSE	PRD	CR
ε=0.010	C=16 σ=0.012 μ=0.25	268	82	0,81	0,017	13,83	3,26
ε=0.012			68	0,77	0,018	14,81	3,94
ε=0.015			55	0,79	0,021	17,33	4,87
ε=0.020			43	0,76	0,026	21,47	6,23
ε=0.025			37	0,75	0,032	26,12	7,24
ε=0.030			32	0,73	0,038	30,70	8,37

Table 2. Experimental results of ε-insensitive Huber loss-based SVR for compressing ECG signals polluted with Cauchy noise distribution

Model Parameters		#TS	#SV	w	RMSE	PRD	CR
ε=0.010	C=256 σ=0.012 μ=0.025	268	73	0,95	0,028	22,11	3,67
ε=0.012			60	0,97	0,028	22,72	4,46
ε=0.014			48	0,96	0,029	23,64	5,58
ε=0.016			40	0,91	0,031	24,72	6,70
ε=0.018			37	0,99	0,033	25,40	7,24
ε=0.020			27	0,99	0,034	27,26	9,92

The user-defined ε parameter provides us to control the selection of samples (support vectors) that are directly related to sparseness. In terms of the number of training samples (#TS), the number of support vectors (#SV), the smoothness (w), root mean square error (RMSE), percent root mean square difference (PRD), and compression ratio (CR), detailed analysis results of the compressed versions of ECG signals contaminated with Gauss and Cauchy noises for various ε values are listed in Table 1 and Table 2, respectively.

As can be seen from Table 1 and Table 2, as the value of the compression parameter (ε) increases, the compression ratio also increases, but also increases the PRD and RMSE values, leading to distortions in the compressed signal.

In the next experiments, the proposed SVR model (ε-insensitive Huber loss function) is compared with classical SVR model (Vapnik loss function) on ECG data sets contaminated by Gaussian and Cauchy noise distributions. As shown in Table 3,

Table 3. Comparison of classical SVR (Vapnik loss) and proposed SVR (ε-insensitive Huber loss) performance in compressing ECG data polluted with Gaussian noise distribution

Model parameters		#TS	Loss function	#SV	w	RMSE	PRD	CR
C=1, μ=15	ε=0.015 σ=0.012	268	ε-Huber	91	0,90	0,019	15,49	2,94
			ε-Vapnik	93	0,99	0,019	15,07	2,88
C=2, μ=10			ε-Huber	86	0,93	0,019	15,35	3,11
			ε-Vapnik	90	1,11	0,019	15,07	2,97
C=4, μ=5			ε-Huber	82	0,95	0,019	15,22	3,26
			ε-Vapnik	88	1,22	0,019	15,19	3,04
C=8, μ=2			ε-Huber	72	0,94	0,019	15,34	3,72
			ε-Vapnik	87	1,15	0,019	15,07	3,08
C=16, μ=1			ε-Huber	63	0,95	0,019	15,25	4,25
			ε-Vapnik	77	1,16	0,019	15,14	3,48
C=32, μ=0.5			ε-Huber	45	0,96	0,019	15,54	5,95
			ε-Vapnik	69	1,21	0,019	15,19	3,88
C=64, μ=0.25			ε-Huber	26	0,92	0,019	15,40	10,30
			ε-Vapnik	61	1,27	0,019	15,36	4,39
C=128, μ=0.125			ε-Huber	13	0,95	0,019	15,35	20,61
			ε-Vapnik	48	1,51	0,019	15,14	5,83

the proposed SVR and classical SVR models exhibit different behaviors at different C values for the same ε=0.015 and σ=0.012 values in the ECG dataset contaminated with Gaussian noise. In particular, as the C value increases, the proposed SVR model performs better than the classical SVR model for different μ values. That is to say, in the same RMSE and PRD values, depending on the μ value, the proposed SVR model gives a higher compression ratio while at the same time producing a smoother model. For example, for the same RMSE=0.019, PRD=15.40, C=32, ε=0.015 and σ=0.012 values, the proposed SVR model (with μ=0.5) yields SV=45, ***w***=0.96, and CR=5.95 while the classical SVR model yields SV=69, ***w***=1.21, and CR=3.88. Moreover, if the C and μ are set to 64 and 0.25 respectively, the proposed SVR model produces SV=26, ***w***=0.92, and CR=10.30 while the classical SVR model produces SV=61, ***w***=1.27, and CR=4.39.

In the case of using Cauchy noise distribution (Table 4), the proposed SVR (with ε-insensitive Huber loss function) provides a better compression ratio (CR) and smoothness (***w***) than the classical SVR in the same RMSE and PRD values, depending on the μ value, thus confirming the superiority of the proposed SVR model.

These experimental results show that the proposed SVR model is more robust than the classical SVR model even if the ECG recordings are subjected to different noise distributions due to ambient conditions.

4. CONCLUSION AND FUTURE WORK

Long-term monitored ECG recordings lead to a large volume of data that makes storage and transmission difficult. A number of algorithms have been proposed to effectively compress these records. Recently, support-vector based algorithms have attracted considerable attention in regression (compression) problems because they attempt to reduce not only the experimental measurement error but also the upper limit of the generalization error. However, their performance depends on the loss function used, i.e., error distribution, as shown in Table 3 and 4. The classic SVR method assumes that ECG data has a Laplacian error distribution. However, long-term ECG recordings may be subject to different error distributions, so classical SVR may not be the appropriate choice. In order to cope with different error distributions, this study proposes the ε-insensitive Huber loss based SVR for the compressing of ECG signals.

Table 4. Comparison of classical SVR (Vapnik loss) and proposed SVR (ε-insensitive Huber loss) performance in compressing ECG data polluted with Cauchy noise distribution

Model parameters		#TS	Loss function	#SV	***w***	RMSE	PRD	CR
C=1, μ=15	ε=0.01 σ=0.012	268	ε-Huber	212	1,19	0,028	22,47	1,26
			ε-Vapnik	216	1,33	0,027	21,40	1,24
C=2, μ=10			ε-Huber	207	1,28	0,027	21,94	1,29
			ε-Vapnik	212	1,73	0,028	22,29	1,26
C=4, μ=1			ε-Huber	203	0,94	0,028	22,37	1,32
			ε-Vapnik	219	1,97	0,029	23,38	1,22
C=8, μ=0.25			ε-Huber	190	0,79	0,029	23,41	1,41
			ε-Vapnik	217	2,55	0,029	23,20	1,23
C=16, μ=0.125			ε-Huber	188	0,96	0,028	21,65	1,42
			ε-Vapnik	215	4,63	0,029	22,77	1,24
C=32, μ=0.1			ε-Huber	172	0,88	0,028	22,55	1,55
			ε-Vapnik	218	6,49	0,028	22,12	1,22
C=64, μ=0.05			ε-Huber	132	0,90	0,029	22,33	2,03
			ε-Vapnik	220	21,31	0,028	22,09	1,21
C=128, μ=0.025			ε-Huber	69	0,81	0,028	22,63	3,88
			ε-Vapnik	221	27,50	0,028	21,79	1,23

Computer simulations for various values of the compression parameter (ε) show that the ε-insensitive Huber loss based SVR provides an optimal compression ratio in the ECG data set contaminated by Gaussian and Cauchy error distributions. Furthermore, in compressing ECG data, computer simulations have shown that the proposed SVR (with Huber loss) performs better than classical SVR against Gauss and Cauchy error distributions. For example, for the same RMSE=0.019, PRD=15.35, C=128, ε=0.015 and σ=0.012 values the proposed SVR model (with μ=0.125) yields SV=13, ***w***=0.95, and CR=20.61 while the classical SVR model yields SV=48, ***w***=1.51, and CR=5.83 for the Gaussian error distribution.

As a future study, the proposed SVR model can be applied to other biomedical signals such as Electromyography (EMG), Electroencephalograpy (EEG) since it does not require any preprocessing algorithm.

References

[1] S. M. Jalaleddine, C. G. Hutchens, R. D. Strattan, and W. A. Coberly, "ECG data compression techniques-a unified approach," *IEEE Transsctions on Biomedical Engineering*, vol. 37, no. 4, pp. 329-343, 1990.

[2] J. R. Cox, F. M. Nolle, H. A. Fozzard, and G. C. Oliver, "AZTEC, a preprocessing program for real-time ECG rhythm analysis," *IEEE Transactions on Biomedical Engineering*, vol. 2, pp. 128-129, 1968.

[3] W. C. Mueller, "Arrhythmia detection program for an ambulatory ECG monitor. Biomedical sciences instrumentation," vol. 14, pp. 81-85, 1978.

[4] J. P. Abenstein, and W. J. Tompkins, "A new data-reduction algorithm for real-time ECG analysis," *IEEE Transactions on Biomedical Engineering*, vol. 1, pp. 43-48, 1982.

[5] G. Nave and A. Cohen, "ECG compression using long-term prediction," *IEEE transactions on Biomedical Engineering*, vol. 40, no. 9, pp. 877-885, 1993.

[6] C. P. Mammen and B. Ramamurthi, "Vector quantization for compression of multichannel ECG," *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 9, pp. 821-825, 1990.

[7] C. Paggetti, M. Lusini, M. Varanini, A. Taddei, and C. Marchesi, "A multichannel template based data compression algorithm," *1994 IEEE Conference on Computers in Cardiology*, pp. 629-632, 1994.

[8] J. Ma, T. Zhang, and M. Dong, "A novel ECG data compression method using adaptive fourier decomposition with security guarantee in e-health applications," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 986-994, 2015.

[9] M. Karczewicz and M. Gabbouj, "ECG data compression by spline approximation," *Signal Processing*, vol. 59, no. 1, pp. 43-59, 1997.

[10] R. Benzid, A. Messaoudi, and A. Boussaad, "Constrained ECG compression algorithm using the block-based discrete cosine transform," *Digital Signal Processing*, vol. 18, pp. 56-64, 2008.

[11] B. S. Reddy and I. S. N. Murthy, "ECG data compression using Fourier descriptors," *IEEE Transactions on Biomedical Engineering*, vol. 4, pp. 428-434, 1986.

[12] P. S. Addison, "Wavelet transforms and the ECG: a review," *Physiological measurement*, vol. 26, no. 5, pp. 155-199, 2005.

[13] M. S. Manikandan and S. Dandapat, "Wavelet-based electrocardiogram signal compression methods and their performances: a prospective review," *Biomedical Signal Processing and Control*, vol. 14, pp. 73-107, 2014.

[14] B. Singh, A. Kaur, and J. Singh, "A review of ecg data compression techniques," *International journal of computer applications*, vol. 116, no. 11, pp. 39-44, 2015.

[15] A. A. Shinde and P. Kanjalkar, "The comparison of different transform based methods for ECG data compression," *2011 IEEE International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN),* pp, 332-335, 2011.

[16] O. Karal, "Destek Vektör Regresyon ile EKG Verilerinin Sıkıştırılması," *Journal of the Faculty of Engineering and*

Architecture of Gazi University (Accepted), 2018.

[17] O. Karal, “Maximum likelihood optimal and robust Support Vector Regression with lncosh loss function,” *Neural networks*, vol. 94, pp. 1-12, 2017.

[18] Z. Shi, and M. Han, “Support vector echo-state machine for chaotic time-series prediction,” *IEEE Transactions on Neural Networks*, vol. 18, no. 2, pp. 359-372, 2007.

[19] M. B. Huber, S. L. Lancianese, M. B. Nagarajan, I. Z. Ikpot, A. L. Lerner, and A Wismuller, “Prediction of biomechanical properties of trabecular bone in MR images with geometric features and support vector regression,” *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 1820-1826, 2011.

[20] H. Mahmoodian, L. Ebrahimian, “Using support vector regression in gene selection and fuzzy rule generation for relapse time prediction of breast cancer,” *Biocybernetics and Biomedical Engineering*, vol. 36, pp. 466-472, 2016.

[21] M. Valizadeh and M. R. Sohrabi, “The application of artificial neural networks and support vector regression for simultaneous spectrophotometric determination of commercial eye drop contents,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 193, pp. 297-304, 2018.

[22] Y. Yaslan and B. Bican, “Empirical mode decomposition based denoising method with support vector regression for time series prediction: a case study for electricity load forecasting,” *Measurement*, vol. 103, pp. 52-61, 2017.

[23] N. Nava, T. D. Matteo, and T. Aste, “Financial Time Series Forecasting Using Empirical Mode Decomposition and Support Vector Regression,” *Risks*, vol. 6, no. 7, pp. 1-22, 2018.

[24] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, “Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes,” *Energy and Buildings*, vol. 130, pp. 519-531, 2016.

[25] A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, “Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system,” *Sustainable Energy Technologies and Assessments*, vol.25, pp. 146-160, 2018.

[26] M. Guermoui, A. Rabehi, K. Gairaa, and S. Benkaciali, “Support vector regression methodology for estimating global solar radiation in Algeria,” *The European Physical Journal Plus*, vol. 133, no. 22, pp. 1-9, 2018.

[27] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, and A. Stojcevski, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912-928, 2018.

[28] B. Schölkopf and A. J. Smola, “Learning with kernels: support vector machines, regularization, optimization, and beyond,” *MIT press*, 2002.

[29] G. B. Moody and R. G. Mark, “The impact of the MIT-BIH arrhythmia database,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, pp. 45-50, 2001.