Semi-supervised Stacked Label Consistent Autoencoder for Reconstruction and Analysis of Biomedical Signals

Anupriya Gogna, Angshul Majumdar*, *Senior Member, IEEE*, and Rabab Ward, *Fellow, IEEE*

***Abstract*—*Objective:* An autoencoder-based framework that simultaneously reconstruct and classify biomedical signals is proposed. Previous work has treated reconstruction and classification as separate problems. This is the first study that proposes a combined framework to address the issue in a holistic fashion. *Methods:* For telemonitoring purposes, reconstruction techniques of biomedical signals are largely based on compressed sensing (CS); these are "designed" techniques where the reconstruction formulation is based on some "assumption" regarding the signal. In this study, we propose a new paradigm for reconstruction—the reconstruction is "learned," using an autoencoder; it does not require any assumption regarding the signal as long as there is sufficiently large training data. But since the final goal is to analyze/classify the signal, the system can also learn a linear classification map that is added inside the autoencoder. The ensuing optimization problem is solved using the Split Bregman technique. *Results:* Experiments were carried out on reconstructing and classifying electrocardiogram (ECG) (arrhythmia classification) and EEG (seizure classification) signals. *Conclusion:* Our proposed tool is capable of operating in a semi-supervised fashion. We show that our proposed method is better in reconstruction and more than an order magnitude faster than CS based methods; it is capable of real-time operation. Our method also yields better results than recently proposed classification methods. *Significance:* This is the first study offering an alternative to CS-based reconstruction. It also shows that the representation learning approach can yield better results than traditional methods that use hand-crafted features for signal analysis.**

***Index Terms*—Body area network (BAN), classification, deep learning, reconstruction.**

I. Introduction

TELEMONITORING via body area networks (BANs) has been receiving considerable interest recently. The goal is to acquire biomedical signals like electrocardiogram (ECG), EEG, Pulsepleithismograph (PPG), etc., and transmit it wirelessly to a base station for manual or automated analysis. Such a system is of benefit to developed and developing countries. It helps in health monitoring of the elderly and the differently abled. It can also be employed for mass scale data monitoring of subjects in developing countries when large portions of the country do not have access to some medical facilities.

A major challenge in telemonitoring is to develop an energy efficient BAN. At a sensor node there are three energy sinks—sensing, processing, and transmission of the data. When compressed sensing (CS) is used to reduce the amount of data to be transmitted, the energy required for processing becomes relatively small compared to the other two. In this case, transmission would require the largest amount of energy followed by sensing. The standard procedure is to compress the signal prior to transmission. Transform coding based techniques are computationally expensive for a sensor node; therefore most studies use CS instead [1]–[6].

CS based solutions project a portion of the acquired signal (say 1 s) onto a random matrix (Gaussian, sparse binary, binomial) such that the size of the projected data is smaller than the length of the 1-s samples. Since CS only requires a matrix-vector product, it is computationally cheap. There are several studies that propose energy and computationally efficient hardware for the same [6]–[10]. The compressed data are wirelessly transmitted to a base station where they are reconstructed using CS techniques for further analysis and monitoring. There can be several variants of the basic CS technique for reconstruction; in [1], [3], and [6] variants of sparse Bayesian learning are used [11]. The more standard CS approach for recovery is used in [12]–[16]; in [17]–[21], the signals were reconstructed using interchannel and intrachannel correlations.

All these studies have concentrated on the reconstruction of the signal with the assumption that the signals will be analyzed retrospectively. Such off-line analysis can only be done for noncritical applications like emotion assessment [22], [23]. However in most applications of biomedical signal analysis this is not the case; for example in seizure detection [24], [25], the analysis/monitoring should be carried in real time.

For systems that are time-critical, CS will fail. CS requires solving an optimization problem iteratively. The time required to solve the reconstruction problem on a standard PC is much larger than the time duration of the signal. In this study, we reconstruct the signal based on a deep learning approach. This

Manuscript received May 23, 2016; revised August 13, 2016 and September 1, 2016; accepted September 10, 2016. Date of publication November 29, 2016; date of current version August 18, 2017. This work was supported by the Qatar National Research Fund (a member of Qatar Foundation) under NPRP Grant 7-684-1-127. *Asterisk indicates corresponding author.*

A. Gogna is with the Indraprasatha Institute of Information Technology.

*A. Majumdar is with the Indraprasatha Institute of Information Technology, Delhi110020, India (e-mail: angshul@iiitd.ac.in).

R. Ward is with the University of British Columbia.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2016.2631620

is based on the extensions of the seminal paper by Hinton and Salakhutdinov [26], which showed that autoencoders can "learn" to compress the information content of signals. Later studies [27], [28] showed that autoencoders can be used for denoising. In this study, we go a step further and show that autoencoders can be used for solving inverse problems such as in signal reconstruction. Unlike CS-based techniques, our proposed method requires only a few matrix vector products and hence can operate in real time.

In biomedical signal processing applications, the end goal is not signal reconstruction but signal classification. Problems such as estimation of emotional state and detecting seizures can all be translated to classification problems. Usually handcrafted features [25], [29]–[31] or statistical features [24], [32], [33] are extracted from the signals and standard classification tools like neural networks and support vector machines are employed. In recent times, systematic studies have shown that image analysis (e.g., biometrics [34] and speech processing, e.g., speech recognition via deep learning [35]) automatically generate deeply learned features that yield far superior results compared to hand-crafted or statistical features. The main challenge of deep learning is the requirement for large amount of training data. It is difficult to find such large data in biomedical signal analysis problems; probably that is the reason behind the relative rarity of papers in this area.

In the telemonitoring scenario there are two tasks—signal reconstruction and signal analysis. In this study, we propose a combined solution to the two. The standard autoencoder is self-supervised, i.e., the input and the output are the same; they are unsupervised in the sense that they do not require any training data. In this study, we propose a semi-supervised stacked autoencoder. Along with reconstruction, it will also learn a linear mapping to the labels of the different classes in the data (when class information is available); for signals having no class information, it will just learn to reconstruct. Such an autoencoder serves the dual purpose. It can be used for automated signal analysis and also for manual monitoring of the reconstructed signals—both in real time.

The rest of the paper is organized into several sections. The following literature review section discusses prior CS-based reconstruction techniques and the basics of autoencoders. The proposed architecture and its implementation is described in detail in Section III. Thorough experimental evaluation is carried out in Section IV. The conclusions of this study are discussed in Section V.

II. Literature Review

A. CS-Based Reconstruction

In this study, we will talk mostly about EEG reconstruction since majority of the work has been on this area. But the techniques discussed are generic enough to be applied to any other biomedical signal.

One of the earliest works that applied CS for EEG signal compression and transmission is [12]. It projected the EEG signal onto an independent identically distributed (i.i.d.) Gaussian basis for compression and used CS to recover the EEG signal by exploiting the signal's sparsity in the Gabor domain. The compression can be expressed as

$$b = \Phi z \tag{1}$$

where z is the EEG signal, Φ is the projection/compression matrix, and b is the compressed data. It was assumed that the data are sparse in some domain Ψ so that the sparse coefficients (α) could be recovered by l_1-norm minimization

$$\min_{\alpha} \|\alpha\|_1 \text{subject to } b = \Phi\Psi\alpha. \tag{2}$$

Mohsina and Majumdar [15] showed that a better way to recover the signals is to use an analysis prior formulation instead of (2)

$$\min_{z} \|\Psi z\|_1 \text{subject to } b = \Phi z \tag{3}$$

EEG is always acquired by multiple channels; ECG too is acquired from several channels. The aforementioned techniques reconstruct one channel at a time. The possibility of exploiting interchannel correlation in order to improve EEG signal reconstruction was mentioned in [12], but there was no concrete formulation. This problem is partially addressed in [16]. They do not explicitly model the interchannel correlation but frame a joint reconstruction problem where the signals from all the channels are reconstructed simultaneously.

Let "i" denote the channel number, then the compression for this channel can be represented as

$$b_i = \Phi z_i. \tag{4}$$

This can be organized as follows:

$$\begin{bmatrix} b_1 \\ \dots \\ b_C \end{bmatrix} = \begin{bmatrix} \Phi & \dots & 0 \\ \dots & \dots & .. \\ 0 & \dots & \Phi \end{bmatrix} \begin{bmatrix} z_1 \\ \dots \\ z_C \end{bmatrix}. \tag{5}$$

The concatenated solution will be sparse in wavelet domain; this sparsity of the signals from all channels is exploited in [16]. At a first glance, this looks like a joint reconstruction problem, but a closer look reveals that this is actually as good as channel by channel reconstruction; this formulation (5) does not exploit any structure across the channels. Other studies assume a block structure of the EEG signals [1], [3], [6], [11] in a transform domain (DCT or wavelet).

A recent work proposed CS techniques for EEG signal compression and transmission, but instead of sending the raw EEG signals, it subtracted the mean from all the signals, thereby reducing the number of bits to be transmitted [13]. However, such a scheme will not be energy efficient, since in order to compute the mean EEG signal, the nodes need to communicate with each other—and such communication consumes considerable energy.

In another work [14], it was shown that for certain specific tasks like seizure detection, instead of sending the full signal, it is possible to send some distinct features, which can be further analyzed at the base station for possible risks. Such a technique requires more computation than standard CS techniques, but the number of features to be transmitted are very few. Unfortunately, such a technique cannot be generalized for other applications.

Biomedical signals are inherently correlated. Prior studies hinted at using the interchannel correlation but did not propose

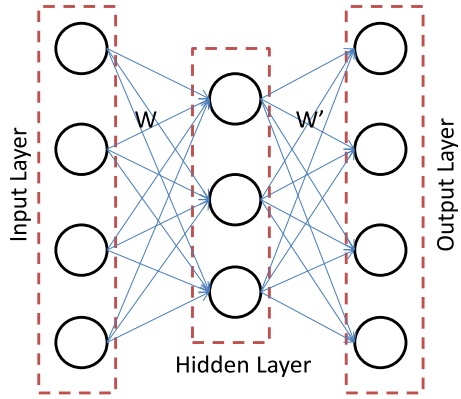

Fig. 1. Single-layer autoencoder.

any formulation to exploit this information. In [18], the common interchannel sparsity pattern was exploited to capture the correlation. Instead of (5), the organization is

$$\left[b_1 \mid \ldots \mid b_C \right] = \Phi \left[z_1 \mid \ldots \mid z_C \right]. \tag{6}$$

The signals from different channels being correlated will share a common sparsity pattern in the transform domain. Thus, the matrix $\Psi[\, z_1 \mid \ldots \mid z_C \,]$ will be row-sparse and hence can be recovered by $l_{2,1}$-minimization

$$\min_{Z} \|Z\|_{2,1} \text{subject to } B = \Phi Z \tag{7}$$

where $B = [\, b_1 \mid \ldots \mid b_C \,]$ and $Z = [\, z_1 \mid \ldots \mid z_C \,]$.

Here, the $l_{2,1}$-norm is defined as the sum of the l_2-norm of the rows. The outer l_1-norm (summation) promotes sparsity in the selection of rows. The inner l_2-norm promotes a dense solution in the selected row.

In [17], a separate take on correlation is proposed. The authors argued that if the signals are correlated, they will be linearly dependent; therefore, when stacked as columns, they will form a low-rank matrix, i.e., Z (7) will be of low rank. This property was exploited in [17]; a matrix-completion-based formulation was utilized to recover the signal ensemble

$$\min_{Z} \|Z\|_{\text{NN}} \text{subject to } B = \Phi Z. \tag{8}$$

The nuclear norm (defined as the sum of singular values) is the closest convex surrogate of rank.

Some recent studies [19], [20] exploited the Blind Compressive Sensing (BCS) formulation; here, instead of using a fixed sparsifying basis like wavelet/Gabor, it is learned from the data.

B. Autoencoder

An autoencoder (as seen in Fig. 1) consists of two parts—the encoder maps the input to a latent space, and the decoder maps the latent representation to the data. For a given input vector (including the bias term) x, the latent space is expressed as

$$h = Wx. \tag{9}$$

Here, the rows of W are the link weights from all the input nodes to the corresponding latent node. The mapping can be linear (9), but in most cases, it is nonlinear (sigmoid, tanh etc.)

$$h = \phi\left(Wx\right). \tag{10}$$

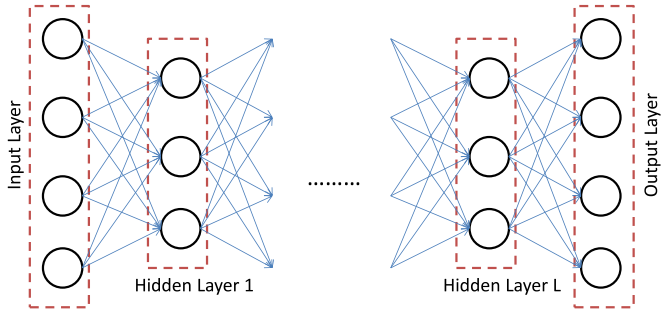

Fig. 2. Stacked autoencoder.

The decoder portion reverse maps the latent features to the data space

$$x = W'\phi\left(Wx\right). \tag{11}$$

Since the data space is assumed to be the space of real numbers, there is no activation function here.

During training, the problem is to learn the encoding and decoding weights—W and W'. This is achieved by minimizing the Euclidean cost

$$\arg\min_{W,W'} \|X - W'\phi\left(WX\right)\|_F^2. \tag{12}$$

Here, $X = [\, x_1 \mid \ldots \mid x_N \,]$ consists of all the training samples stacked as columns. The problem (12) is clearly nonconvex. However, it is solved easily by gradient descent techniques since the sigmoid function is smooth and continuously differentiable.

There are several extensions to the basic autoencoder architecture. Stacked/deep autoencoders [26], [27] have multiple hidden layers (see Fig. 2). The corresponding cost function is expressed as follows:

$$\arg\min_{W_1 \ldots W_{L-1}, W'_1 \ldots W'_L} \|X - g \circ f\left(X\right)\|_F^2 \tag{13}$$

where $g = W'_1\phi(W'_2 \ldots W'_L(f(X)))$ and $f = \phi(W_L\phi(W_{L-1} \ldots \phi(W_1 X)))$.

Solving the complete problem (13) is computationally challenging. The weights are usually learned in a greedy fashion—one layer at a time [36], [37].

Stacked denoising autoencoders (SDAE) [27] are a variant of the basic autoencoder where the input consists of noisy samples and the output consists of clean samples. Here the encoder and decoder are learned to denoise noisy input samples. The learned features appear to be more robust when learned by SDAE.

In a recent work, a marginalized denoising autoencoder was proposed [38], which does not have any intermediate nodes but learns the mapping from the input to the output. This formulation is convex (unlike regular autoencoders); the trick here is to marginalize over all possible noisy samples so that the dataset need not be augmented like SDAE. Such an autoencoder was used for domain adaptation.

Another variation for the basic autoencoder is to regularize it, i.e.

$$\arg\min_{(W)s} \|X - g \circ f\left(X\right)\|_F^2 + R\left(W, X\right). \tag{14}$$

The regularization can be a simple Tikhonov regularization—however, that is not used in practice. It can be a sparsity promoting term [28], [39] or a weight decay term (Frobenius norm of the Jacobian) as used in the contractive autoencoder [40]. The regularization term is usually chosen so that (14) is differentiable and hence minimized using gradient descent techniques.

In a recent work [41] a group-sparse autoencoder is proposed. Here the regularization term is an $l_{2,1}$-norm on each class. The idea is that by enforcing a similar sparsity signature in each class, one can enforce some level of supervision in autoencoding. This study is different from ours.

III. Proposed Approach

In this section, we discuss our proposed label consistent autoencoder. This will be divided into two sections. In Section III-A, we discuss why it is possible to recover (reconstruct) the signals using autoencoders. In Section III-B, we address how to classify and reconstruct the signal simultaneously; we add a class-label consistency penalty and complete the formulation.

A. Reconstruction

We are interested in solving a linear inverse problem of the form: $y = Ax$. For a determined or an over-determined system the solution is linear. Even for an underdetermined system the minimum energy solution is linear. However, a CS solution is nonlinear. First we discuss the reason behind this nonlinearity.

Ideally to obtain a sparse solution one would like to solve the l_0-minimization problem, but as is well known, this is NP hard

$$\min_x \|x\|_0 \text{subject to } y = Ax \tag{15}$$

One way to solve (15) is to employ greedy techniques such as orthogonal matching pursuit (OMP) [42]. This is a greedy approach which detects one support (non-zero position in x) at a time and estimates its value. The full algorithm is given by,

Here Ω is the support set of x; A_Ω consists of only those columns in A that are in the support set. After the iterations are over, we get a solution with the values at some nonzero positions. To get the full x, one needs to fill in the other positions with 0 values.

OMP is a nonlinear operation. In every iteration, one needs to compute the "max" during the support detection stage—this is a highly nonlinear operation. Extensions of OMP, such as StOMP [43], are also nonlinear.

So far, we have talked about greedy algorithms for sparse recovery. The more popular technique is to relax the NP-hard l_0-norm to its closest convex surrogate the l_1-norm. This enjoys stronger theoretical guarantees. In practice, the solution is obtained via

$$\min_x \|y - Ax\|_2^2 + \lambda\|x\|_1. \tag{16}$$

Consider the simplest technique to solve (16)—iterative soft thresholding algorithm (ISTA) [44]. Every iteration (say k) consists of two steps. The first step is the Landweber iteration (17)

Algorithm OMP.

Input: y, A, k (support)
Initialize: $r = y, \Omega = \emptyset$ **(support set)**
Repeat for k iterations
Compute Correlation: $c = abs(A^T r)$
Detect Support: $l = \arg\max_i c_i$
Update Support: $\Omega = \Omega \cup l$
Estimate values at support Ω: $x_\Omega = \min_x \|y - A_\Omega x_\Omega\|_2^2$
Compute residual: $r = y - A_\Omega x_\Omega$
End

and the second step is the soft thresholding (18)

$$b = x_{k-1} + \sigma A^T (y - Ax_{k-1}) \tag{17}$$

$$x_k = \text{sign}(b) \max\left(0, |b| - \frac{\lambda\sigma}{2}\right) \tag{18}$$

where the step size σ is inverse of the maximum Eigenvalue of A^TA.

The first step (17) is a simple gradient descent step—it is a linear operation. But the second step involves thresholding and is, hence, a nonlinear operation.

To summarize, all CS recovery techniques are nonlinear inversion operations.

CS has been used extensively in the past decade in a variety of signal processing applications, ranging from biomedical signal reconstruction to medical imaging, seismic imaging, and astronomy to name a few. However, CS cannot be used for real-time reconstruction, since the solution is iterative and hence time consuming. In our problem, the reconstruction should be in real time; this precludes use of such sophisticated inversion techniques. Instead of designing the inversion techniques (discussed in Section II), we will "learn" to reconstruct.

It is well known that neural networks act as universal function approximators. Given enough training data, the nonlinear activation functions learn to represent arbitrary functions; this was proven by Cybenko [47] and Hornik [48], respectively. Please check."?>. A more fundamental work on this topic dates back to Kolmogorov [49]. Please check."?>where he showed that a continuous function of many variables can be approximated by a superposition of continuous functions of one variable. We make use of this universal functional approximation property of neural networks to "learn" a CS-like inversion operation with autoencoders.

The basic idea is simple, an approximate inverse of a linear system is obtained by

$$x' = A^T y = A^T Ax. \tag{19}$$

The inversion is approximate since x' is a noisy version of the actual solution x. In CS, the noise is progressively removed by soft-thresholding [49]. As mentioned earlier, it has been shown that autoencoders can also be used for denoising [28]. Basically, one prepares a large number of noisy versions of the signal and input them to the autoencoder; at the output are the corresponding clean versions. The autoencoder "learns" a mapping from

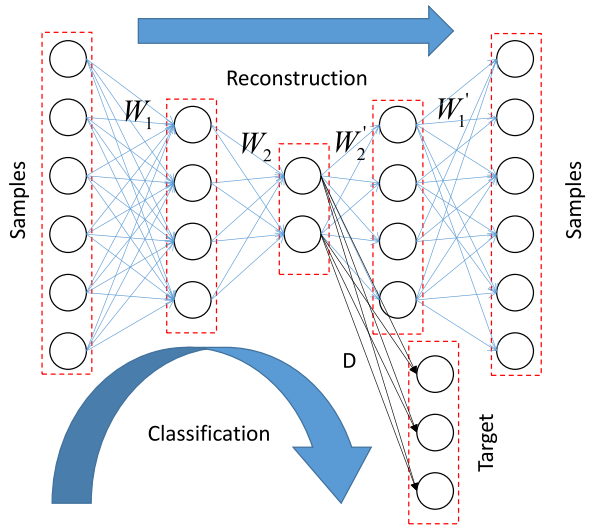

Fig. 3. Proposed label consistent autoencoder.

the noisy input to the clean output. When a new noisy signal is later applied to the input, a clean version of it is obtained at the output. In [28] it was claimed that the simple denoising autoencoder can yield decent denoising results—sometimes, even at par with dictionary-learning-based techniques.

In this study, the "noisy" signals are obtained from the approximate (x'); these are fed as input to the autoencoder for training. The corresponding clean signals are the output. During training, the autoencoder learns to produce "clean" signals. The training time can be large, but during the actual operation of the system, one only needs two (for a single-layer autoencoder) matrix vector products; therefore, it is superfast.

B. Combined Classification and Reconstruction

Whatever we have discussed so far can be done using a classical autoencoder. But such an autoencoder cannot classify. One can follow the usual deep learning approach where after learning the autoencoder, the decoder is removed and a soft-max layer is attached and the full architecture is fine-tuned for classification. Such a deep neural network would classify but could not reconstruct. As mentioned before, our problem demands both—reconstruction and classification. Existing architectures cannot achieve both simultaneously. Some of the tasks may be automated (and would not require reconstruction) but many others would require manual monitoring. This would require a novel autoencoder that can simultaneously reconstruct and classify.

Our proposed architecture is shown in Fig. 3. It is a two-layer stacked autoencoder. For all deep learning tasks, the features from the deepest layer are used; therefore, we propose to learn a linear map from the innermost layer to the class labels; this constitutes the class-label consistency penalty. This idea has been used in the past for discriminative restricted Boltzmann machine [50] and label-consistent dictionary learning [51]. The mathematical expression is given by

$$\min_{W_1', W_2', W_1, W_2, D} \left\| X - W_1' \phi \left(W_2' \phi \left(W_2 \phi \left(W_1 X \right) \right) \right) \right\|_F^2 + \lambda \left\| T - D \phi \left(W_2 \phi \left(W_1 X \right) \right) \right\|_F^2 . \tag{20}$$

Here, X is the training samples, T the class labels, and D the linear map. It is not possible to learn this architecture using off-the-shelf backpropagation techniques. This is because there are two outputs; therefore, there is no unique unambiguous way to backpropagate the errors. We will solve it using the Split Bregman technique. The current formulation (20) requires all the training samples to have corresponding class labels. In most practical scenarios, this would not be the case. We will have semi-supervised data, i.e., a portion of the training samples will have class labels while the rest of it will not. We need to allow for semi-supervision on our formulation. This leads to

$$\min_{W_1', W_2', W_1, W_2, D} \left\| X - W_1' \phi \left(W_2' \phi \left(W_2 \phi \left(W_1 X \right) \right) \right) \right\|_F^2 + \lambda \left\| T - D \phi \left(W_2 \phi \left(W_1 X_S \right) \right) \right\|_F^2 . \tag{21}$$

We assume the training data to be $X = [X_U | X_S]$, where the subscripts denote unsupervised or supervised.

C. Derivation

In this study, we solve (21) by a Bregman-type variable splitting [52]. In the first step, we substitute

$$Z^1 = \phi \left(W_2' \phi \left(W_2 \phi \left(W_1 X \right) \right) \right) . \tag{22}$$

The proxy variable Z^l has two parts—unsupervised and supervised, i.e., $Z^1 = [{Z^1}_U | {Z^1}_S]$. This allows us to express (21) as follows:

$$\min_{W_1', W_2', W_1, W_2, D, Z^1} \left\| X - W_1' Z^1 \right\|_F^2 + \lambda \left\| T - D \phi \left(W_2 \phi \left(W_1 X_S \right) \right) \right\|_F^2 \quad \text{s.t.} Z^1 = \phi \left(W_2' \phi \ \left(W_2 \phi \left(W_1 X \right) \right) \right) . \tag{23}$$

We can form the Lagrangian from (23), but the exact Lagrangian would enforce equality between the proxy and the variables in every iteration. This is not required; for practical purposes we only need the proxy and the variables to converge at the solution. Therefore, one can relax the Lagrangian to the augmented Lagrangian instead

$$\min_{W_1', W_2', W_1, W_2, D, Z^1} \left\| X - W_1' Z^1 \right\|_F^2 + \lambda \left\| T - D \phi \left(W_2 \phi \left(W_1 X_S \right) \right) \right\|_F^2 + \mu_1 \left\| Z^1 - \phi \left(W_2' \phi \left(W_2 \phi \left(W_1 X \right) \right) \right) \right\|_F^2 . \tag{24}$$

In the augmented Lagrangian formulation, one starts with a small value of μ—this relaxes the equality constraint. For each value of μ, (24) is solved and then the value of μ is increased to enforce equality progressively. As one can see this is not an elegant approach; increasing the value of μ is at best heuristic. The most elegant solution is to incorporate a Bregman relaxation variable B_1, this automatically adjusts for the equality. One does not need to tune the values of μ. The Split Bregman

formulation is

$$\min_{W_1',W_2',W_1,W_2,D,Z^1} \left\|X - W_1' Z^1\right\|_F^2 + \lambda \left\|T - D\phi\right. \\ \left.(W_2\phi(W_1 X_S))\right\|_F^2 \\ + \mu_1 \left\|Z^1 - \phi\left(W_2'\phi\left(W_2\phi\left(W_1 X\right)\right)\right) - B_1\right\|_F^2. \quad (25)$$

We apply the Split Bregman technique on the substitution $Z^2 = \phi(W_2\phi(W_1 X))$, leading to

$$\min_{W_1',W_2',W_1,W_2,D,Z^1,Z^2} \left\|X - W_1' Z^1\right\|_F^2 + \lambda \left\|T - DZ_S^2\right\|_F^2 \\ + \mu_1 \left\|Z^1 - \phi\left(W_2' Z^2\right) - B_1\right\|_F^2 \\ + \mu_2 \left\|Z^2 - \phi\left(W_2\phi\left(W_1 X\right)\right) - B_2\right\|_F^2. \quad (26)$$

As in Z^1, Z^2 has two parts—supervised (denoted by subscript S) and unsupervised (denoted by subscript U). In the third level, we substitute $Z = \phi(W_1 X)$. This leads to the final formulation

$$\min_{W_1',W_2',W_1,W_2,D,Z^1,Z^2,Z} \left\|X - W_1' Z^1\right\|_F^2 + \lambda \left\|T - DZ_S^2\right\|_F^2 \\ + \mu_1 \left\|Z^1 - \phi\left(W_2' Z^2\right) - B_1\right\|_F^2 \\ + \mu_2 \left\|Z^2 - \phi\left(W_2 Z\right) - B_2\right\|_F^2 + \mu \left\|Z - \phi\left(W_1 X\right) - B\right\|_F^2. \quad (27)$$

Even though not exactly separable, (27) can be segregated into a number of subproblems

$$P1: \min_{W_1'} \|X - W_1' Z^1\|_F^2$$

$$P2: \min_{W_2'} \|Z^1 - \phi(W_2' Z^2) - B_1\|_F^2 \equiv \\ \|\phi^{-1}(Z^1 - B_1) - W_2' Z^2\|_F^2$$

$$P3: \min_{W_2} \|Z^2 - \phi(W_2 Z) - B_2\|_F^2 \equiv \\ \|\phi^{-1}(Z^2 - B_2) - W_2 Z\|_F^2$$

$$P4: \min_{W_1} \|Z - \phi(W_1 X) - B\|_F^2 \equiv \\ \|\phi^{-1}(Z - B) - W_1 X\|_F^2$$

$$P5: \min_{Z^1} \|X - W_1' Z^1\|_F^2 + \mu_1 \|Z^1 - \phi(W_2' Z^2) - B_1\|_F^2$$

$$P6: \min_{Z^2} \lambda \|T - DZ_S^2\|_F^2 + \mu_1 \|Z^1 - \phi(W_2' Z^2) - B_1\|_F^2 \\ + \mu_2 \|Z^2 - \phi(W_2 Z) - B_2\|_F^2$$

$$P7: \min_{Z} \mu_2 \|Z^2 - \phi(W_2 Z) - B_2\|_F^2 \\ + \mu \|Z - \phi(W_1 X) - B\|_F^2 \; P8: \min_{D} \|T - DZ_S^2\|_F^2.$$

Subproblems $P1$ and $P8$ are a simple least squares problems having a closed form solutions. Subproblems $P2$–$P4$ can be recast as linear least squares problems (shown earlier) and hence can be solved analytically as well. Subproblem $P5$ can be rearranged as follows:

$$\min_{Z^1} \left\| \begin{pmatrix} X \\ \sqrt{\mu_1}\left(\phi\left(W_2' Z^2\right) + B_1\right) \end{pmatrix} - \begin{pmatrix} W_1' \\ \sqrt{\mu_1} I \end{pmatrix} Z^1 \right\|_F^2. \quad (28)$$

This turns out to be a simple least squares problem as well. Similarly one can recast $P7$ as a least squares problem in the following manner:

$$\min_{Z} \mu_2 \left\| \begin{pmatrix} \phi^{-1}\left(Z^2 + B_2\right) \\ \sqrt{\mu}\left(\phi\left(W_1 X\right) + B\right) \end{pmatrix} - \begin{pmatrix} W_2 \\ \sqrt{\mu} I \end{pmatrix} Z \right\|_F^2. \quad (29)$$

Subproblem $P6$ can be expressed in two parts

$$P6: \arg\min_{Z_S^2, Z_U^2} \lambda \left\|T - DZ_S^2\right\|_F^2 + \mu_1 \left\|Z^1 - \phi\right. \\ \left.\left(W_2'\left[{Z^2}_U \left| {Z^2}_S\right.\right]\right) - B_1\right\|_F^2 \\ + \mu_2 \left\|\left[{Z^2}_U \left| {Z^2}_S\right.\right] - \phi\left(W_2 Z\right) - B_2\right\|_F^2.$$

The variables ${Z^2}_U, {Z^2}_S$ are separable. Hence, $P6$ can be segregated as follows:

$$\min_{Z_U^2} \mu_1 \left\|Z^1 - \phi\left(W_2' {Z^2}_U\right) - B_1\right\|_F^2 + \mu_2 \left\|{Z^2}_U - \phi\right. \\ \left.(W_2 Z) - B_2\right\|_F^2 \quad (30)$$

$$\min_{Z_S^2} \lambda \left\|T - DZ_S^2\right\|_F^2 + \mu_1 \left\|Z^1 - \phi\left(W_2' {Z^2}_S\right) - B_1\right\|_F^2 \\ + \mu_2 \left\|{Z^2}_S - \phi\left(W_2 Z\right) - B_2\right\|_F^2. \quad (31)$$

As we have been doing so far, we can recast (30) and (31) as least squares problems (32) and (33), respectively

$$\min_{Z_U^2} \left\| \begin{pmatrix} \sqrt{\mu_1}\phi^{-1}\left(Z^1 + B_1\right) \\ \sqrt{\mu_2}\left(\phi\left(W_2 Z\right) + B_2\right) \end{pmatrix} - \begin{pmatrix} \sqrt{\mu_1} W_2' \\ \sqrt{\mu_2} I \end{pmatrix} Z_U^2 \right\|_F^2 \quad (32)$$

$$\min_{Z_S^2} \left\| \begin{pmatrix} \sqrt{\mu_1}\phi^{-1}\left(Z^1 + B_1\right) \\ \sqrt{\mu_2}\left(\phi\left(W_2 Z\right) + B_2\right) \\ \sqrt{\lambda} T \end{pmatrix} - \begin{pmatrix} \sqrt{\mu_1} W_2' \\ \sqrt{\mu_2} I \\ \sqrt{\lambda} D \end{pmatrix} Z_S^2 \right\|_F^2. \quad (33)$$

The last part is to update the Bregman relaxation variables. This accounts for the automatic adjustments between the variables and their proxies at convergence. The relaxation variables are updated using simple gradient descent

$$B_1 \leftarrow Z^1 - \phi\left(W_2' Z^2\right) - B_1$$
$$B_2 \leftarrow Z^2 - \phi\left(W_2 Z\right) - B_2$$
$$B \leftarrow Z - \phi\left(W_1 X\right) - B.$$

We impose two stopping criteria. Iterations continue till the objective function reaches some local minima, i.e., there is no significant change in successive iterations. Or, they continue for a fixed number of iterations. Our algorithm requires specifying several hyper-parameters. We found that they are very robust to a wide range of values; in this study, we put $\mu_1 = \mu_2 = \mu = 0.01$

IV. Experimental Evaluation

A. ECG Arrhythmia Classification and Reconstruction

In this study, five types of beat classes of arrhythmia as recommended by Association for Advancement of Medical Instrumentation (AAMI) were analyzed from ECG signals, namely,

nonectopic beats, supraventricular ectopic beats, ventricular ectopic beats, fusion betas, and unclassifiable and paced beats. The classification experiments are carried out on the MIT-BIH Arrhythmia dataset from www.physionet.org. First, we carry out reconstruction and classification using the aforesaid database. This is a fully supervised problem, i.e., all the samples have class labels.

It is well known in deep learning that "more the merrier." However, in real life, supervised samples are few; but it is easy to have a large number of unsupervised samples. Therefore, in the second set of experiments, we augment the aforesaid (supervised) dataset MIT-BIH with the European ST-T dataset (cardiac ischemia) dataset from www.physionet.org. No class information from the second dataset is used; it is only used for semisupervised learning.

The MIT-BIH Arrhythmia database contains 48 half-hour recordings of two channel ambulatory ECG, obtained from 47 subjects in the years 1975 and 1979 by the Beth-Israel Hospital Arrhythmia Laboratory at Boston. Twenty-four-hour ambulatory ECG recordings were collected from a mixed population of size 4000 having inpatients (around 60%) and outpatients (around 40%). The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a 10 mV range. Two or more cardiologists independently annotated each record; consensus was made to obtain the computer-readable reference annotations for each beat.

The European society of cardiology has provided a standard ST-T database consisting of 90 annotated samples of ambulatory ECG recordings from 79 subjects having myocardial ischemia disease. The subjects were 70 men aged from 30 to 84 years and some women aged from 55 to 71 years. Additional selection criteria were established in order to obtain a representative selection of ECG abnormalities in the database, including baseline ST segment displacement resulting from conditions such as hypertension, ventricular dyskinesia, and effects of medication. Each record is of 2 h duration and contains two signals. Each is sampled at 250 samples per second with 12-bit resolution over a nominal 20 mV input range.

As a preprocessing step the MIT-BIH dataset is downsampled to 250 Hz from its native 360 Hz; this is to ensure parity between the two datasets. Both datasets are normalized. The quantization level remains as it is. The MIT-BIH protocol is converted to the AAMI/ANSI standard. This leads to five classes—nonectopic beat (*N*), supraventricular ectopic beats (*S*), ventricular ectopic beats (*V*), fusion beat (*F*), and unknown beat (*Q*). Owing to the relative sparsity of samples in the *F* and *Q* classs they are merged with *V*; this is following the AAMI2 protocol proposed in [53].

We train our proposed label-consistent autoencoder with one second (250 points) length samples. The outer hidden layer has 125 nodes and the inner hidden layer has 63 nodes. The class of the entire duration is assigned to the sample during training. During testing, the test ECG sequence is broken down into 1-s samples (X_{test}) and passed through the trained model. The target for this is obtained by $T_{\text{test}} = D\phi(W_2\phi(W_1 X_S))$. Practically, T_{test} will not have ones and zeroes, they would be real numbers in between. We take the row averages of *T* and assign the class of X_{test} to the class having the maximum value in the corresponding row.

TABLE I
TRAIN AND TEST SET DETAILS

Dataset	*N*	*S*	*V*	*F*	*Q*	Total	# Rec
Train	45 844	943	3788	415	8	50 998	22
Test	44 238	1836	3221	388	7	49 690	22
Total	90 082	2779	7009	803	15	100 688	44

TABLE II
ECG RECONSTRUCTION RESULTS

Technique	50% Compression	25% Compression
BSBL	0.121 ± 0.056	0.262 ± 0.114
Prop.	0.140 ± 0.014	0.190 ± 0.026
Prop. Aug.	$\mathbf{0.089 \pm 0.006}$	$\mathbf{0.122 \pm 0.018}$

For the experimental protocol, we follow [54]; this is repeatable protocol. The division into test set and training set is shown in Table I. The record number (#) of the patient used for training are 101, 114, 112, 207, 223, 106, 115, 124, 208, 230, 108, 116, 201, 209, 109, 118, 203, 215, 112, 119, 205, and 220 and for testing are—100, 117, 210, 221, 233, 103, 121, 212, 222, 234, 105, 123, 213, 228, 111, 200, 214, 231, 113, 202, 219, and 232.

In this study, we emulate a health-monitoring scenario. We assume that the ECG signals are acquired and are compressed by projecting them onto a lower dimension by a sparse binary matrix [1]. The compressed data is sent to a base station for reconstruction and analysis. As a benchmark for reconstruction, we use the Block Sparse Bayesian Learning (BSBL) algorithm [11] with wavelet as the sparsifying transform; this has been used in extensively in the past for reconstructing compressively sampled biological signals [1], [3], [6]. The reconstructed signal is then used for classification. Here, we compare with two recent techniques optimum-path forest (OPF), support vector machine (SVM) [54], probabilistic neural network (PNN) [55], and extreme learning machine (ELM) [56]; all of them use hand-crafted features. The best results are obtained from the feature extraction technique proposed in [57]; hence, we use the same for our comparison. In the aforesaid references, detailed comparison have been done with other techniques and these were shown to yield the best results; hence, we compare with these studies.

As mentioned before, two sets of experiments have been carried out. In the first set, only the MIT-BIH database was used. In the second set, the database was augmented with unsupervised samples from the European ST-T database. The results for reconstruction are shown in Table II and those from classification are shown in Table III. For classification, reconstructed signals from 50% compression have been used.

For reconstruction, Normalized Mean Squared Error is the error metric.

$$\text{NMSE} = \frac{\|\text{groundtruth} - \text{reconstructed}\|_2}{\|\text{groundtruth}\|_2}.$$

TABLE III
ECG CLASSIFICATION RESULTS (AAMI2 PROTOCOL)

Classifier	Acc.	*F*		*S*		*V*	
		Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
OPF	86.5	91.2	**56.8**	11.0	97.4	62.4	90.8
SVM	90.1	**98.8**	31.9	0	97.6	41.7	95.4
PNN	93.8	94.6	55.3	15.9	97.0	48.9	89.6
ELM	89.2	95.6	39.8	0	97.0	50.2	95.2
Prop.	92.0	96.8	54.5	13.6	**100**	48.6	93.6
Prop. Aug.	**96.9**	**98.8**	56.6	**19.0**	**100**	**65.2**	**96.2**

Prop. = MIT-BIH; Prop. Aug = MIT-BIH + European ST-T

TABLE IV
EEG RECONSTRUCTION RESULTS

Technique	50% Compression	25% Compression
BSBL	0.112 ± 0.062	0.240 ± 0.084
Prop.	0.192 ± 0.024	0.292 ± 0.096
Prop. Aug.	$\mathbf{0.060 \pm 0.006}$	$\mathbf{0.096 \pm 0.012}$

TABLE V
ECG CLASSIFICATION RESULTS

	Details	EMD	DTSTF	LPE	Prop.	Prop. Aug.
A	Eyes open	**88**	**88**	86	86	**88**
B	Eyes closed	98	98	98	96	**100**
C	Interictal (epileptic focus)	94	**96**	92	92	**96**
D	Interictal (Hipocam. region)	95	95	92	93	**96**
E	Ictal state	92	**94**	90	92	**94**

We report the mean reconstruction error and the deviations. Classification accuracy (Acc.) is the most important measure for performance; but it is a standard practice to report sensitivity (Sens.) and specificity (Spec.); the standard definitions apply for all the metrics.

In Table II, the results are shown for the MIT-BIH database only. Even though we augment the dataset for our technique, we do not report the results for European ST-T; this is to keep all the results in sync. We can observe that the proposed technique improves with additional data and can yield results even better than sophisticated CS techniques. The reconstruction time required by the BSBL algorithm (takes $\sim$ 12 s) is about 40 times more than our proposed autoencoder (takes $\sim$ 0.3 s to reconstruct signals of 1 s durations) based approach. Therefore, not only do we recover the signal more accurately, we are faster than required for real-time operation.

In classification we see that our proposed technique (even without augmentation) yields competitive results. It is among the top two results. But with augmentation, the results improve even more. We always perform the best in terms of accuracy. For a few isolated cases, our specificity and sensitivity are marginally low. One should note that, the results in Table III cannot be directly compared with [54]; this is because in the prior work the ground-truth signal is used whereas in the current work the reconstructed signal is used. Therefore, there is bound to be some fall in accuracy.

B. EEG Classification and Reconstruction

A publicly available EEG dataset, made available by the University of Bonn [58] is used in this study. The EEG database consists of five sets (A–E). Each set contains 100 single-channel EEG segments, each with a duration of 23.6 s. Sets A and B have been recorded using the standard International 10–20 system for surface EEG recording. Five healthy volunteers were participated in these tests with eyes open (A) and eyes closed (B). For sets C, D, and E, five epileptic patients were selected for presurgical evaluation of epilepsy by using intracranial electrodes. Depth electrodes were implanted symmetrically to record EEG from the epileptogenic zone (D) and from hippocampal formation of the opposite hemisphere of the brain (C). Segments of set E were taken from contacts of all electrodes. In sets C and D, segments contain interictal intervals while seizure activities occur in set E. Each epoch was sampled at 173.61 Hz, resulting in a total of 4096 samples.

Most prior studies like [24], [59]–[61] convert it to a binary classification problem—seizure versus nonseizure. In this study, we classify all the five classes A to E, as defined in [62]. We compare our proposed technique with empirical mode decomposition (EMD) [24]—using SVM, rational discrete short-time Fourier (DTSTF) transform [59]—using neural network and linear prediction error (LPE) [62]—using simple thresholding. For our proposed method, the number of nodes in the outer layer is 1024 and in the inner layer is 256.

As before, we test our proposed technique in two modes. In the first mode, we only use the given dataset. In the second one, we augment this dataset with unsupervised data. The unsupervised data is obtained from the BCI competitions II, III and IV [63]. These datasets have different sampling rates, so all of them have be subsampled to 128 Hz. The same was done for the actual dataset [58] used in the experiments. Also, the signals are normalized.

The results are generated as before. The data are compressed to 25% and 50% of their original length and reconstructed using BSBL. The reconstructed signal is processed and classified using the techniques mentioned before. For our proposed techniques, the reconstruction and classification proceeds simultaneously. The reconstruction results are shown in Table IV and the classification results in Table V. The classification results are shown for the case of 50% compression.

From Table IV, we find that the reconstruction accuracy from our proposed technique is poor when we only use the test dataset [58]. This is because there is not enough data to learn the mapping; when we augment the dataset with unsupervised data, the improvement is dramatic. It yields significantly better results than sparsity-based methods. We see a similar speed improvement. BSBL takes about a minute to reconstruct 23.6 s of data, whereas our proposed method takes only 1.8 s.

Table V shows the per-class classification accuracy. Our proposed method (without augmentation) with only the supervised dataset does not yield very good results. This is likely to be

an effect of overfitting of the autoencoder. With augmentation, the overfitting issue is resolved and we get the best overall results. It must be remembered that one cannot expect these results to match those in the published works; this is because the published papers use the groundtruth samples. Here the reconstructed samples are used. Owing to the reconstruction artifacts, the classification accuracy suffers.

V. Conclusion

This study proposes a comprehensive solution for the telehealth monitoring scenario. Prior studies addressed the problem in situ. Some studies concentrated on the acquisition and reconstruction of the signals, whereas others focused on the analysis of these signals. During analysis, it was assumed that the reconstruction at a receiver side would be "perfect." This is not true. In prior studies [17]–[21], it has been shown that reconstruction artifacts can indeed reduce the performance of automated analysis. This is mainly because prior techniques were based on the extraction of hand-crafted features, e.g., on the detection of peaks, troughs and other such features. Reconstruction artifacts corrupt these structures in the signal and hence the accuracy suffers.

There are several major contributions of this study. First, we propose a new approach for reconstruction. Prior CS-based techniques are "designed" so that they satisfy certain assumptions about the structure of the signal. In this paper, the system "learns" to reconstruct the signal, and does not require any assumption regarding the structure of the signal. As long as we have enough number of samples to train, our "learned" approach excels over prior "designed" techniques.

We employ an autoencoder for reconstruction. However, as mentioned before, reconstruction is not the final goal—signal analysis (mainly classification) is. Here, we introduce a linear map in the classifier that learns the labels of the different classes from the training samples. Thus, our proposed label consistent autoencoder simultaneously learns to reconstruct and classify. We understand that learning such a structure requires a sizable portion of the data; labeled data may not be always available. Therefore, our proposed label consistent autoencoder can work with both labeled and unlabeled data. If the data are labeled, they learn to reconstruct and map (to class labels); for samples without class lables, it only learns to reconstruct.

Usually, neural networks are trained using some backpropagation (BP) algorithm. However, our said architecture is nonlinear and hence cannot be solved with BP. We solve it using a recently proposed class of optimization technique called Split Bregman. This is the first study that provides an optimal solution to the stacked autoencoder learning problem. Prior techniques use heuristics; they solve each layer greedily and then fine-tune it. We solve the complete stacked autoencoder problem.

We have thoroughly tested our algorithm on benchmark EEG and ECG datasets. We show that our approach yields better reconstruction than state-of-the-art CS-based reconstruction technique—BSBL. Moreover, our method is over an order of magnitude faster than BSBL and is capable of real-time reconstruction (BSBL and other CS-based techniques are not). We also show that our method performs at par or better than recent biomedical signal classification techniques. The methods we compared against use hand-crafted features generated by experts having deep understanding of biomedical signals; the fact that our technique beats such methods indicate that the proposed representation learning approach can indeed be used for biomedical signal analysis in the future.

The proposed semi-supervised stacked autoencoder is suitable for simultaneously addressing the reconstruction and classification problem. However, it can also be used when there is no necessity to reconstruct. One can use the same samples at the input and the output and the corresponding class labels (if available); this would result in an autoencoder-based classifier that learns and that can be applicable to any problem. In the future, we would test how the proposed method excels on benchmark deep learning datasets.

Acknowledgment

The statements made herein are solely the responsibility of the authors.

References

[1] Z. Zhang *et al.*, "Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse bayesian learning," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 2, pp. 300–309, Feb. 2013.

[2] J. Zhang *et al.*, "Energy-efficient ECG compression on wireless biosensors via minimal coherence sensing and weighted \ell _1 minimization reconstruction," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 520–528, Feb. 2015.

[3] Z. Zhang *et al.*, "Spatiotemporal sparse bayesian learning with applications to compressed sensing of multichannel physiological signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1186–1197, Jun. 2014.

[4] J. K. Pant and S. Krishnan, "Compressive Sensing of electrocardiogram signals by promoting sparsity on the second-order difference and by using dictionary learning," *IEEE Trans. Biomed. Circuits Syst.*, vol. 8, no. 2, pp. 293–302, Aug. 2014.

[5] L. F. Polanía *et al.*, "Exploiting prior knowledge in compressed sensing wireless ECG systems," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 508–519, Feb. 2015.

[6] Z. Zhang *et al.*, "Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware," *IEEE J. Biomed. Health Inform.*, vol. 60, no. 1, pp. 221–224, Jan. 2013.

[7] F. Pareschi *et al.*, "Hardware-algorithms co-design and implementation of an analog-to-information converter for biosignals based on compressed sensing," *IEEE Trans. Biomed. Circuits Syst.*, vol. 10, no. 1, pp. 149–162, Jan. 2016.

[8] F. Chen *et al.*, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE J. Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, Mar. 2012.

[9] N. Verma and A. Chandrakasan, "An ultra low energy 12-bit rate-resolution scalable SAR ADC for wireless sensor node," *IEEE J. Solid State Circuits*, vol. 42, no. 6, pp. 1196–1205, Jun. 2007.

[10] R. Harrison and C. Charles, "A low-power low-noise CMOS amplifier for neural recording applications," *IEEE J. Solid State Circuits*, vol. 38, no. 6, pp. 958–965, Jun. 2003.

[11] Z. Zhang and B. D. Rao, "Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2009–2015, Aug. 2013.

[12] S. Aviyente, "Compressed sensing framework for EEG compression," in *Proc. IEEE Workshop Statist. Signal Process.*, 2007, pp. 181–184.

[13] S. Fauvel and R. K. Ward, "An Energy efficient compressed sensing framework for the compression of electroencephalogram signals," *Sensors*, vol. 14, no. 1, pp. 1474–1496, Jan. 2014.

[14] J. Chiang and R. K. Ward, "Energy-efficient data reduction techniques for wireless seizure detection systems," *Sensors*, vol. 14, no. 2, pp. 2036–2051, Feb. 2013.
[15] M. Mohsina and A. Majumdar, "Gabor based analysis prior formulation for EEG signal reconstruction," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 951–955, Aug. 2013.
[16] M. Hosseini Kamal *et al.*, "Compressive multichannel cortical signal recording," in *Proc. IEEE ICASSP*, 2013, pp. 4305–4309.
[17] A. Majumdar *et al.*, "Low-rank matrix recovery approach for energy efficient EEG acquisition for wireless body area network," *Sensors, Special Issue State-of-the-Art Sensor Technol. Canada*, vol. 14, no. 9, pp. 15729–15748, Sep. 2014.
[18] A. Majumdar and R. K. Ward, "Non-convex row-sparse MMV analysis prior formulation for EEG signal reconstruction," *Biomed. Signal Process. Control*, vol. 13, pp. 142–147, Aug. 2014.
[19] A. Shukla and A. Majumdar, "Row-sparse blind compressed sensing for reconstructing multi-channel EEG signals," *Biomed. Signal Process. Control*, vol. 18, no. 4, pp. 174–178, Apr. 2015.
[20] A. Majumdar and R. K. Ward, "Energy efficient EEG sensing and transmission for wireless body area networks: A blind compressed sensing approach," *Biomed. Signal Process. Control*, vol. 20, pp. 1–9, 2015.
[21] A. Shukla and A. Majumdar, "Exploiting inter-channel correlation in EEG signal reconstruction," *Biomed. Signal Process. Control*, vol. 18, no. 4, pp. 49–55, Apr. 2015.
[22] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Feb. 2010.
[23] C. A. Frantzidis *et al.*, "On the classification of emotional biosignals evoked while viewing affective pictures: An integrated data-mining-based approach for healthcare applications," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 309–318, Feb. 2010.
[24] V. Bajaj and R. B. Pachori, "Classification of seizure and nonseizure EEG signals using empirical mode decomposition," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1135–1142, Jun. 2012.
[25] A. Temko *et al.*, "EEG signal description with spectral-envelope-based speech recognition features for detection of neonatal seizures," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 6, pp. 839–847, 2011.
[26] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 13, no. 5786, pp. 504–507, 2006.
[27] P. Vincent *et al.*, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.
[28] K. H. Cho, "Simple sparsification improves sparse denoising autoencoders in denoising highly noisy images," in *Proc. ICML, 2013*, pp. 432–440.
[29] N. Wang and M. R. Lyu, "Extracting and selecting distinctive EEG features for efficient epileptic seizure prediction," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1648–1659, May 2015.
[30] M. Niknazar *et al.*, "A new framework based on recurrence quantification analysis for epileptic seizure detection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 17, no. 3, pp. 572–578, Mar. 2013.
[31] Y. P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, Jun. 2010.
[32] F. Riaz *et al.*, "EMD-based temporal and spectral features for the classification of EEG signals using supervised learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 1, pp. 28–35, Jan. 2016.
[33] M. Spüler *et al.*, "Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 6, pp. 1097–1103, Jun. 2014.
[34] D. Menotti *et al.*, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inform. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.
[35] L. Deng *et al.*, "Recent advances in deep learning for speech research at Microsoft," in *Proc. IEEE ICASSP*, 2013, pp. 8604–8608.
[36] Y. Bengio *et al.*, "Greedy layer-wise training of deep networks," in *Proc. NIPS*, 2007, p. 153.
[37] Y. Bengio, "Learning deep architectures for AI," *Foundations Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.
[38] M. Chen *et al.*, "Marginalized denoising autoencoders for nonlinear representation," in *Proc. ICML, 2014*, pp. 1476–1484.
[39] A. Makhani and B. Frey, "K-sparse autoencoder," in *Proc. ICLR, 2014*, arXiv preprint arXiv:1312.5663 (2013).
[40] S Rifai *et al.*, "Contractive auto-encoders: Explicit invariance during feature extraction," in *Proc. ICML, 2011*, pp. 833–840.
[41] A. Majumdar *et al.*, "Face recognition via class sparsity based supervised encoding," IEEE Trans. Pattern Anal. Mach. Intell., vol. PP. no. 99, [Online]. Available: http://ieeexplore.ieee.org/abstract/document/7470420/
[42] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[43] D. Donoho *et al.*, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.
[44] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
[45] I. Daubechies *et al.*, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Mathematics*, vol. 4, no. 57, pp. 1413–1457, 2004.
[46] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, 1995.
[47] G. Cybenko, "Approximations by superpositions of sigmoidal functions," *Mathematics Control, Signals, Syst.*, vol. 2, no. 4, pp. 303–314, 1989.
[48] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.
[49] A. N. Kolmogorov, "On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition, "*Doklady Akad. Nauk SSSR*, vol. 144, pp. 679–681, 1957 (*Amer. Math. Soc. Transl.*, vol. 28, pp. 55–59, 1963).
[50] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, Mar. 1995.
[51] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machine," in *Proc.* ICML, 2008, pp. 536–543.
[52] Z. Jiang *et al.*, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
[53] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, Jun. 2012.
[54] M. Llamedo and J. P. Martínez, "Heartbeat classification using feature selection driven by database generalization criteria," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 616–625, 2011.
[55] E. J. da S. Luz *et al.*, "ECG arrhythmia classification based on optimum-path forest," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3561–3573, Jul. 2013.
[56] R. J. Martis *et al.*, "ECG beat classification using PCA, LDA, ICA and discrete wavelet transform," *Biomed. Signal Process. Control*, vol. 8, no. 5, pp. 437–448, 2013.
[57] J. Kim *et al.*, "Robust algorithm for arrhythmia classification in ECG using extreme learning machine," *Biomed. Eng. Online*, vol. 8, p. 31, 2009.
[58] C. Ye *et al.*, "Arrhythmia detection and classification using morphological and dynamic features of ECG signals," in *Proc.* IEEE EMBS, 2010, pp. 1918–1921.
[59] Time Series EEG Data. [Online]. Available at: http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html. Accessed on: May 9, 2016.
[60] K. Samiee *et al.*, "Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 541–552, Feb. 2015.
[61] R. B. Pachori and S. Patidar, "Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions," *Comput. Methods Programs Biomed.*, vol. 113, no. 2, pp. 494–502, 2014.
[62] V. Joshi *et al.*, "Classification of ictal and seizure-free EEG signals using fractional linear prediction," *Biomed. Signal Process. Control*, vol. 9, pp. 1–5, 2014.
[63] S. Altunay *et al.*, "Epileptic EEG detection using the linear prediction error energy," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5661–5665, 2010.
[64] BCI Competition. [Online]. Available at: http://www.bbci.de/competition/. Accessed on: May 9, 2016.

Authors' photographs and biographies not available at the time of publication.