IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 12, 2017, accepted October 31, 2017, date of publication December 18, 2017,
date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2771536

Semi-Supervised Deep Blind Compressed Sensing for Analysis and Reconstruction of Biomedical Signals From Compressive Measurements

VANIKA SINGHAL[1], ANGSHUL MAJUMDAR[2], (Senior Member, IEEE), AND RABAB K. WARD[3], (Fellow, IEEE)

[1]Department of Computer Science and Engineering, Indraprastha Institute of Information Technology, New Delhi 110020, India
[2]Department of Electronics and Communication Engineering, Indraprastha Institute of Information Technology, New Delhi 110020, India
[3]Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC V6T1Z4, Canada

Corresponding author: Vanika Singhal (vanikas@iiitd.ac.in)

ABSTRACT In this paper, the objective is to classify biomedical signals from their compressive measurements. The problem arises when compressed sensing (CS) is used for energy efficient acquisition and transmission of such signals for wireless body area network. After reconstruction, the signal is analyzed via certain machine learning techniques. This paper proposes to carry out joint reconstruction and analysis in a single framework; the reconstruction ability is obtained inherently from our formulation. We put forth a new technique called semi-supervised deep blind CS that combines the analytic power of deep learning with the reconstruction ability of CS. Experimental results on EEG classification show that the proposed technique excels over the state-of-the-art paradigm of CS reconstruction followed by deep learning classification.

INDEX TERMS Classification, compressed sensing, deep learning, EEG, reconstruction.

I. INTRODUCTION

In the recent past, several studies like [1]–[4] proposed compressed sensing based techniques for energy efficient acquisition of biomedical signals for wireless body area network (WBAN). Here the goal is to sample the biomedical signal (EEG, ECG, PPG etc.) and transmit it to a remote base station. In any WBAN there are three sinks of power consumption – acquisition, processing and transmission; the last one being the most power hungry. Like all wireless sensor network, power is a premium in WBA; therefore, every effort is made to save power. The only way to achieve this is by compressing the signal.

Standard compression protocols based on transform coding are too computationally complex to be implemented at the sensor nodes. A more computationally efficient way to compress the signal in situ, is by projecting it onto a random matrix. The compressed signal is more energy efficient to transmit. The original signal can be recovered from the compressed one using compressive sampling techniques. In recent times, some deep learning based techniques for compression and quantization has been proposed [5], but compressed sensing still continues to be the primary choice for the said task.

There have been several studies in developing algorithms for signal recovery [1], [2] and there have been papers proposing circuits for such energy efficient acquisition [3], [4]. However, it must be remembered that signal reconstruction is not the end of the information processing pipeline; the final objective is to analyze the signal. The analysis can be either manual or automated. Today, there are several tasks that can has been automated. For example, seizure detection [6], [7] and brain computer interface [8], [9] based on EEG signals; arrhythmia classification [10], [11] and sleep apnea [12], [13] from ECG signals.

So far, signal acquisition-reconstruction and signal analysis were considered two domains. To the best of our knowledge there has been only a single attempt to combine these fields by two of the authors of this papers [14]; but that too has limited applicability – it can handle sub-sampled signals but not compressed ones (produced by projection onto a compressed random matrix). This work proposes an integrated approach for signal reconstruction and analysis from any type of compressed / sub-sampled measurements.

The proposed work is based on the deep learning framework. However, traditional deep learning tools like stacked autoencoders, deep belief networks and convolutional neural networks cannot optimally handle missing data. In [15], we introduced the topic of unsupervised deep blind compressed sensing. This work proposes to extend this

rudimentary concept to a deep neural network capable of handling missing inputs.

Although our main goal is signal analysis (classification) our formulation yields the reconstructed signal as an intermediate by-product. The reconstruction quality is better than compressed sensing based solutions as can be seen in the results section.

II. LITERATURE REVIEW

A. ENERGY EFFICIENT ACQUISITION AND RECONSTRUCTION

In order to compress the acquired biomedical signal, it is projected onto a random matrix. In practice it is usually a sparse binary random matrix [1]–[4]. A more efficient way to compress the signal is to partially sample it [16] – this has the added benefit of reducing signal acquisition cost. Either of these acquisition protocols can be mathematically expressed as follows–

$$y = Ax \tag{1}$$

where x is the signal, A is the random projection / compression matrix and y is the compressed representation.

The compressed representation (y) is transmitted to the base station; here one needs to recover the original signal for further analysis. The recovery is achieved by compressed sensing (CS). Since biomedical signals are sparse in transform domains like Gabor or wavelet, CS exploits this property to recover them from the compressed domain. The recovery is formulated as follows–

$$\min_{x} \ \|\Psi x\|_1 \ \text{subject to } y = Ax \tag{2}$$

Here Ψ is the sparsifying transform.

There are many variants to this basic formulation. Owing to limitations in space, we do not discuss or refer them.

B. BLIND COMPRESSED SENSING

In CS, it is assumed that the sparsifying basis is given; for biomedical signals it is usually Gabor or wavelet. However it has been known that significantly better results are obtained by learning the basis adaptively from data. This leads to blind compressed sensing (BCS) [17] formulation.

For a series of acquired signals 'X', BCS learns a dictionary / basis 'D' so that the signals can be expressed by sparse coefficients 'Z'. This is expressed as follows –

$$X = DZ \tag{3}$$

Here the signals are stacked as columns of X.

In the given scenario, the signals are compressed. Incorporating the BCS model into the compression scenario (1) leads to,

$$Y = AX = ADZ \tag{4}$$

Here A is defined the same as before; Y is the matrix formed by stacking the measured compressed data as columns.

The basis and the coefficients are learnt by solving the following optimization problem –

$$\min_{D,Z} \ \|Y - ADZ\|_F^2 + \lambda \|Z\|_1 + \mu \|D\|_F^2 \tag{5}$$

In essence BCS combines compressed sensing with dictionary learning; it reconstructs the signal by learning a basis during the process. In the past it has been used for feature extraction from compressive samples using the concept of doubly sparse dictionary learning [18].

BCS / dictionary learning is based on the concept of matrix factorization; it essentially factors the data matrix into a dictionary matrix D and a coefficient matrix Z. With the success of deep learning techniques in various fields of signal analysis, recent studies proposed 'deep matrix factorization' [19]. Here the data matrix is factored into multiple levels of dictionaries and a coefficient matrix. For three levels it is shown as,

$$X = D_1 D_2 D_3 Z \tag{6}$$

In this formulation, the non-linearity between the different layers of dictionaries is imposed by non-negativity constraints. In a related work [20] on deep dictionary learning, the non-linearity is more explicitly incorporated so as to handle different types of activations.

Our prior work [15] incorporated the deep matrix factorization into the BCS framework, leading to–

$$Y = AX = AD_1 D_2 D_3 Z \tag{7}$$

The prior aforesaid study, used the formulation for unsupervised feature extraction.

III. SUPERVISED DEEP BLIND COMPRESSED SENSING

Deep learning is successful when the full data is available. When the data is partially sampled or is compressed, deep learning is unable to handle it. The issues is not confined to deep learning; in fact, most off-the-shelf techniques in machine learning are unable to perform when the data is only partially observed [21].

Given the problem of analyzing the signals from compressed measurements, there can be two options. In the first one, we can reconstruct the biomedical signals and then apply deep learning / machine learning on the reconstructed signals. The problem with this approach is that the compressed sensing (CS) reconstruction process introduces artifacts, and these artifacts affect the learning process.

The second option is to directly analyze the signals from the compressed / partially observed domain. This is the approach followed in this paper. However, instead of a using deep BCS as an unsupervised feature extraction tool [15], we propose to incorporate a classifier into the deep BCS methodology.

A. FORMULATION

In [15] the unsupervised version of deep BCS was proposed. The formulation was given by –

$$\min_{D_1,D_2,D_3,Z} \ \|Y - AD_1 D_2 D_3 Z\|_F^2 + \lambda \|Z\|_1 \tag{8}$$

In the proposed supervised formulation proposed below (9), we learn a linear map from the feature / coefficient space Z to the target class labels. This is incorporated in the following formulation,

$$\min_{D_1,D_2,D_3,Z,M} \|Y - AD_1D_2D_3Z\|_F^2 + \lambda\|Z\|_1 + \mu\|T - MZ\|_F^2 \tag{9}$$

Here T corresponds to the binary class labels and M the linear classification map. The parameter μ controls the relative importance between the data consistency term and the classification term.

In this work, we want to go a step further and show how (9) can be modified for semi-supervised formulation, i.e. when there is a combination of labeled and unlabeled data.

$$\begin{aligned}\min_{D_1,D_2,D_3,Z,M} \quad & \|[Y_U|Y_L] - AD_1D_2D_3\,[Z_U|Z_L]\|_F^2 \\ & + \lambda\,\|[Z_U|Z_L]\|_1 + \mu\,\|T - MZ_L\|_F^2 \end{aligned} \tag{10}$$

Here the subscripts U and L stand for unlabeled and labeled respectively and $Z = [Z_U|Z_L]$. Basically, the unsupervised portion of the formulation is applicable to both labeled and unlabeled data whereas the supervised portion is applied only to the labeled data. The derivation of the algorithm (training) is given the following section.

During the testing stage, given the test measurement y_{test}, the task is to assign a class label to it. For that, the first task is to generate the coefficient z_{test}. This is achieved by using the dictionaries (learnt) during the training phase as basis.

$$\min_{z_{test}} \|y_{test} - AD_1D_2D_3z_{test}\|_F^2 + \lambda\,\|z_{test}\|_1 \tag{11}$$

Once the coefficient / feature is generated the learnt linear classifier (M) is used to assign a class label. The test feature is simply multiplied by M. The output is never a binary label, but we assign the position of the highest magnitude coefficient in Mz_{test} as the class label.

B. DERIVATION

In the unsupervised formulation [15], the multiple levels of dictionaries and the coefficients were solved in a brute force fashion using conjugate gradient. The problem with that approach is that the constraints on the dictionary and the coefficients at each level cannot be added. Therefore in this work, we employ a variable splitting technique that will lead to a more elegant solution.

First we express $X = D_1D_2D_3Z$; where X is the complete signal. With this substitution, (10) can be expressed as,

$$\begin{aligned}\min_{D_1,D_2,D_3,Z,M,X} \quad & \|[Y_U|Y_L] - A\,[X_U|X_L]\|_F^2 + \lambda\,\|[Z_U|Z_L]\|_1 \\ & + \mu\,\|T - MZ_L\|_F^2 \;\text{ such that } X = D_1D_2D_3Z \end{aligned} \tag{12}$$

Here $X = [X_U|X_L]$.

Ideally one would like to formulate the Lagrangian for the newly introduced constraint. However, the Lagrangian would enforce equality at each step. This is not required in practice, one would only need to enforce the constraint at convergence. This is the reason, we follow the augmented Lagrangian approach instead [22], [23]. This is given in (13)

$$\begin{aligned}\min_{D_1,D_2,D_3,Z,M,X,Z_1} \quad & \|[Y_U|Y_L] - A[X_U|X_L]\|_F^2 + \lambda\|[Z_U|Z_L]\|_1 \\ & + \mu\,\|T - MZ_L\|_F^2 + \eta\,\|X - D_1D_2D_3Z\|_F^2 \end{aligned} \tag{13}$$

In the next step we substitute introduce another proxy variable: $Z_1 = D_2D_3Z$. The corresponding augmented Lagrangian is expressed as,

$$\begin{aligned}\min_{D_1,D_2,D_3,Z,M,X,Z_1,Z_2} \quad & \|[Y_U|Y_L] - A\,[X_U|X_L]\|_F^2 \\ & + \lambda\,\|[Z_U|Z_L]\|_1 \\ & + \mu\|T - MZ_L\|_F^2 + \eta\|X - D_1Z_1\|_F^2 + \eta_1\|Z_1 - D_2D_3Z\|_F^2 \end{aligned} \tag{14}$$

In the final step, we substitute, $Z_2 = D_3Z$. This leads to,

$$\begin{aligned}\min_{D_1,D_2,D_3,Z,M,X,Z_1,Z_2} \quad & \|[Y_U|Y_L] - A\,[X_U|X_L]\|_F^2 \\ & + \lambda\,\|[Z_U|Z_L]\|_1 \\ & + \mu\,\|T - MZ_L\|_F^2 + \eta\,\|X - D_1Z_1\|_F^2 \\ & + \eta_1\,\|Z_1 - D_2Z_2\|_F^2 + \eta_2\,\|Z_2 - D_3Z\|_F^2 \end{aligned} \tag{15}$$

The final form (15) can be solved by using the technique of alternating direction method of multipliers (ADMM) [24], [25]. The basic idea in ADMM is to update each variable by assuming the others to be constant. This leads to the following sub-problems.

$$\begin{aligned}
&\text{P1:} \min_{D_1} \|X - D_1Z_1\|_F^2 \\
&\text{P2:} \min_{D_2} \|Z_1 - D_2Z_2\|_F^2 \\
&\text{P3:} \min_{D_3} \|Z_2 - D_3Z\|_F^2 \\
&\text{P4:} \min_{Z} \eta_2\,\|Z_2 - D_3Z\|_F^2 + \lambda\,\|Z\|_1 \\
&\text{P5:} \min_{M} \|T - MZ_L\|_F^2 \\
&\text{P6:} \min_{X} \|Y - AX\|_F^2 + \eta\,\|X - D_1Z_1\|_F^2 \\
&\text{P7:} \min_{Z_1} \eta\,\|X - D_1Z_1\|_F^2 + \eta_1\,\|Z_1 - D_2Z_2\|_F^2 \\
&\text{P8:} \min_{Z_2} \eta_1\,\|Z_1 - D_2Z_2\|_F^2 + \eta_2\,\|Z_2 - D_3Z\|_F^2
\end{aligned}$$

Apart from sub-problem P4, all others are simple least squares problems having a closed form solution in the form of pseudoinverse. P4 is an l_1-minimization problem. This can be solved using Iterative Soft Thresholding Algorithm (ISTA) [26]. The steps for ISTA are shown below.

Initialize: Z=0

Till convergence repeat

$$B = Z(k-1) + \frac{1}{\alpha}D_3^T\left(Z_2 - D_3Z(k-1)\right)$$

α = maximum Eigenvalue of $D_3^TD_3$

$$Z(k) = signum(B)\cdot\max\left(0, |B| - \frac{\lambda}{2\eta_2}\right)$$

Here 'k' represents the iteration number.

This concludes the basic training algorithm. But there needs to be some additional checks. In order to prevent degenerate solutions where the dictionaries are of high values and the coefficients of small values or vice versa, we need to normalize the columns of the dictionaries after every iteration.

Furthermore in the current version, we have not incorporated any non-linearity. The prior studies on deep matrix factorization incorporated Rectified Linear Unit (ReLU) type non-linearity [18], [19]. The non-linearity helps in two ways. First, it prevents collapsing of the multiple layers of dictionaries into one layer. Second, the non-linearity improve function approximation capability. We too introduce ReLU type non-linearity in this work. The coefficients in each stage X, Z_1, Z_2 an Z_3 are constrained to be positive; this is achieved by simply putting the negative values in them to be zeroes in every iteration.

C. RECONSTRUCTION ABILITY

The basic formulation for deep blind compressed sensing is given in (7); we repeat it for the sake of convenience.

$$Y = AX = AD_1D_2D_3Z$$

Here $X = D_1D_2D_3Z$, i.e. from the learnt dictionaries D_1, D_2, D_3 and the learnt coefficients Z we can always reconstruct the signal X.

The reconstruction can be done both for training and testing. For reconstruction of training samples, the methodology is exactly the same as [27]. The learning is expressed as follows,

$$\min_{D_1,D_2,D_3,Z,M} \|[Y_U|Y_L] - AD_1D_2D_3\,[Z_U|Z_L]\|_F^2 + \lambda\,\|[Z_U|Z_L]\|_1 + \mu\,\|T - MZ_L\|_F^2$$

Once the learning is over, the original data can be recovered simply by using the formula: $X = D_1D_2D_3Z$.

For testing, the formulation for generating the features is given by (11), repeated here for the reader's convenience.

$$\min_{z_{test}} \|y_{test} - AD_1D_2D_3z_{test}\|_F^2 + \lambda\,\|z_{test}\|_1$$

Once the feature is obtained, the full signal can be recovered via

$$x_{test} = D_1D_2D_3z_{test} \tag{16}$$

Thus we see that by using our formulation we can recover signals in the physical domain for both training and testing. This is a by-product of our formulation.

D. CONTRIBUTIONS

There are two fundamental contributions of this work.

- This is the only deep learning tool that can learn from compressive samples. No other deep framework (stacked autoencoder, deep belief network or convolutional neural network) can infer from compressed or missing data directly (without reconstruction).
- The second contribution is a by-product of our formulation. Ours is the first work that proposes a technique for matrix completion (from under-sampled projections) using a deep learning technique.

IV. EXPERIMENTAL RESULTS

A. SEIZURE DETECTION FROM EEG

A publicly available EEG dataset, made available by the University of Bonn [28] is used in this work. 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.

There are not many studies in deep learning based seizure classification to compare our method with. In [29] and [30], it was found that DBN with logistic regression yields good detection results; it was better than traditional classifiers like nearest neighbor and support vector machine. In [31] a convolutional neural net (CNN) with long short term memory (LSTM) units yielded good detection rate. The paper [32] used stacked autoencoder (SAE) with logistic regression for classification.

In this work, our goal is to classify the signals from compressive measurements; sparse binary projection matrices are used for the purpose. Traditional deep learning is not able to learn (optimally) from compressed measurements. Therefore while comparing with other deep learning techniques, we first recover the original signal from the compressive measurements via a CS; in this work we use [1].

With the advent of compressed sensing in the past decade, researchers in neural networks explored their efficacy in compressed measurement domain [33]–[35]. However these studies are not directly related to our work. In [33], a neural network is used for regression from compressed domain. In [34] and [35], they propose a neural network solution for reconstructing the signal from compressed measurements. A very recent work [36] addresses the problem of classification from compressed domain; however it solves the problem piecemeal. First a standard technique is used for recovering the data in the physical domain and then a standard deep learning tool is applied on the recovered data. This is the same procedure we use in our study for comparison.

It is well known that deep learning (DBN [30], CNN-LSTM [32], SAE [33]) benefits from pre-training with

TABLE 1. Effect of depth at 4:1 compression on class-wise accuracy.

	Details	Proposed – 1 layer	Proposed – 2 layer	Proposed – 3 layer	Proposed – 4 layer
A	Eyes open	89	91	**92**	90
B	Eyes closed	92	95	**97**	95
C	Inter-ictal (epileptic focus)	91	95	**96**	94
D	Inter-ictal (Hipocam. region)	92	95	**96**	94
E	Ictal state	94	96	**98**	96

TABLE 2. Effect of depth at 2:1 compression on class-wise accuracy.

	Details	Proposed – 1 layer	Proposed – 2 layer	Proposed – 3 layer	Proposed – 4 layer
A	Eyes open	91	93	**94**	93
B	Eyes closed	98	100	**100**	100
C	Inter-ictal (epileptic focus)	95	97	**98**	96
D	Inter-ictal (Hipocam. region)	96	97	**99**	97
E	Ictal state	97	100	**100**	100

TABLE 3. EEG classification results (%age) at 4:1 compression.

	Details	DBN	CNN–LSTM	SAE	Proposed
A	Eyes open	85	85	84	**92**
B	Eyes closed	94	94	94	**97**
C	Inter-ictal (epileptic focus)	90	92	89	**96**
D	Inter-ictal (Hipocam. region)	91	91	90	**96**
E	Ictal state	88	90	89	**98**

unlabeled data; these pre-trained models are fine-tune with the actual training data from [28]. Therefore we use datasets from the BCI competitions II and III; these datasets are resampled to fit the University of Bonn dataset. Our proposed method does not require pre-training; it uses both labeled and unlabeled data in one single formulation.

We have experimented with four architectures of depths 1, 2, 3, and 4. The configurations are 200, 200-100, 200-100-50 and 200-100-50-25 respectively. The practice of reducing the number of nodes by half in subsequent layers is common in deep learning. The value of $\mu = 1$ has been used to give equal importance to the reconstruction and regression costs. The hyper-parameters η, η_1 and η_2 have been fixed at unity. This is because, they correspond to different layers of the deep learning architecture. Since there is no reason to favor one layer over the other, we keep them to be unity.

In the first set of experiments we show the influence of depth on results; these are in Tables 1 (4:1 compression) and II (2:1 compression). The results can be explained. It is known in deep learning that as we go deeper, more abstract and robust representation is learnt this helps in inference and analysis. But there is a trade-off between depth and over-fitting; as one goes deeper more and more parameters need to be learnt. With limited training data, the deeper models tend to overfit. This is reason we see that when we go from levels 1 to 3, there is a gradual increase in accuracy but in the 4^{th} layer there is a dip.

The configurations of the other techniques have been taken from the corresponding papers; we skip repeating them for the sake of brevity. These methods (DBN, CNN-LSTM and

TABLE 4. EEG classification results (%age) at 2:1 compression.

	Details	DBN	CNN–LSTM	SAE	Proposed
A	Eyes open	89	89	89	**94**
B	Eyes closed	97	98	96	**100**
C	Inter-ictal (epileptic focus)	92	95	92	**98**
D	Inter-ictal (Hipocam. region)	94	95	93	**99**
E	Ictal state	92	94	92	**100**

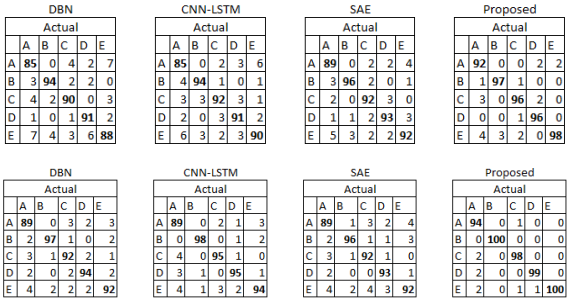

FIGURE 1. Confusion Matrices for EEG. Top – For 4:1 Compression. Bottom – For 2:1 Compression. For every matrix, the columns are the actual classes and the rows are the predicted classes.

SAE) uses 9 hand-crafted features – area under wave, normalized decay, line length, mean energy, average peak amplitude, average valley amplitude, peak variation, and root mean square. These features are extracted from all the channels and used as input to the aforesaid deep learning classifiers.

The comparative experimental results with for 4:1 and 2:1 compression is shown in Tables 3 and 4 respectively. The corresponding confusion matrices are also shown in Fig. 1. Our method, always yields the best results. The results from CNN-LSTM combination are a distant second. DBN and SAE almost yield the same performance. The CNN-LSTM performs slightly better than DBN and SAE because of its ability to capture dynamical attributes. As expected, the results with higher compression ratio is lower than lower compression ratio.

It must be noted that these results cannot be compared with the prior published work because, the prior works used the complete signal and where as this one uses reconstructed values. The reconstruction introduces error which reduces the performance. It is interesting to note other methods uses hand-crafted features and we use raw compressive measurements; yet we perform significantly better than the state-of-the-art.

As discussed before, our method can inherently reconstruct the signals from the compressive measurements. Therefor we compare the reconstruction performance of our proposed method with CS [1]. For visual evaluation a sample is shown in Fig. 1. One can see in Fig. 2. that our proposed method does a much better reconstruction than standard CS. For numerical corroboration we show the mean and the standard deviation of the normalized mean squared error in the following table.

The reconstruction results (Table 5) indicate that for both the training and the testing stage, we perform significantly better than CS. The results for both training and testing are

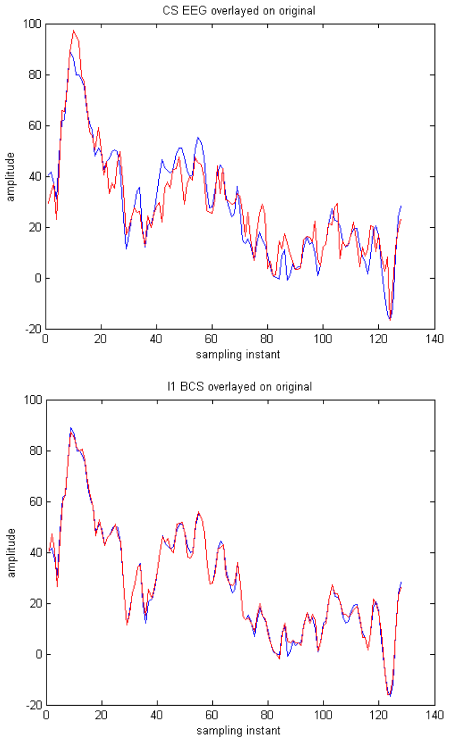

FIGURE 2. EEG reconstruction. Top – CS [1] Reconstruction. Bottom – Proposed.

TABLE 5. EEG reconstruction results.

Technique	2:1 Compression	4:1 Compression
Compressed Sensing	0.112±0.062	0.240±0.084
Proposed – training	**0.056±0.008**	**0.102±0.014**
Proposed – testing	**0.056±0.006**	**0.102±0.011**

equally accurate, but the testing stage has lesser variability. This is likely because the dictionaries are pre-learnt.

B. ARRHYTHMIA CLASSIFICATION FROM ECG

In this study, five types of beat classes of arrhythmia as recommended by Association for Advancement of Medical Instrumentation (AAMI) were analyzed from electrocardiogram (ECG) signals namely: non-ectopic beats, supra-ventricular 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.

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. It has been shown in [14] that augmenting the supervised learning problem (MIT-BIH) with unsupervised data from the ST-T dataset (cardiac ischemia www.physionet.org) improves the reconstruction and classification results. No class information from the second dataset is used; it is only used for semi-supervised learning. We follow the same protocol in this paper.

TABLE 6. Train and test set details.

Dataset	N	S	V	F	Q	Total	# Rec
Train	45844	943	3788	415	8	50998	22
Test	44238	1836	3221	388	7	49690	22
Total	90082	2779	7009	803	15	100688	44

The MIT-BIH Arrhythmia database contains 48 half hour recordings of two channel ambulatory ECG, obtained from 47 subjects in the year 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 hours 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 pre-processing step the MIT-BIH dataset is down-sampled to 250 Hz from its native 360 Hz; this is to ensure parity between the two datasets.

For classification experiments, the MIT-BIH protocol is converted to the AAMI / ANSI standard. This leads to 5 classes - Non ectopic beat (N), Supra-ventricular 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 class they are merged with V; this is following the AAMI2 protocol proposed in [37].

For the experimental protocol we follow [37]; this is repeatable protocol (also followed in [14]). The division into test set and training set is shown in Table 6. The record number (#) of the patient used for training are 101,114,112, 207,223,106,115,124,208,230,108, 116,201,209,109,118,203,215,112,119,205,220; for testing are – 100,117,210,221,233,103,121,212,222,234,105,123, 213,228, 111,200,214,231,113,202,219,232.

The prior study [14] showed that stacked autoencoders beat other shallow classifiers like optimum path forest, support vector machine, probabilistic neural network and extreme learning machine for the said task. Therefore we do not compare with the such shallow methods any more.

A recent study [38], proposed deep neural networks learned by stacked autoencoder (SAE) and deep belief

TABLE 7. ECG classification results (%age) on 4:1 compression.

Classifier	Acc.	F		S		V	
		Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
1 layer	89.9	80.6	55.9	11.8	81.5	75.6	88.1
2 layer	93.6	87.8	60.2	18.9	89.8	84.1	92.4
3 layer	**98.0**	**100**	**67.2**	**23.0**	**100**	**90.1**	**100**
4 layer	98.0	100	67.1	23.0	100	89.7	98.4

TABLE 8. ECG classification results (%age) on 2:1 compression.

Classifier	Acc.	F		S		V	
		Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
1 layer	94.9	95.0	66.4	22.9	88.4	85.9	93.8
2 layer	96.8	96.2	68.0	26.2	92.1	88.5	97.2
3 layer	**100**	**100**	**70.5**	**28.0**	**100**	**94.1**	**100**
4 layer	100	100	70.2	28.0	100	93.7	100

TABLE 9. ECG classification results (%age) on 4:1 compression.

Classifier	Acc.	F		S		V	
		Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
DBN	59.7	60.2	25.9	0	51.3	16.8	28.1
SAE	85.6	87.3	33.9	0	82.6	39.8	81.6
Proposed	**98.0**	**100**	**67.2**	**23.0**	**100**	**90.1**	**100**

network (DBN). We compare against [38]. For this task, the SAE and the DBN are pre-trained with unlabeled data (ST-T) and fine-tuned with the labeled samples (MIT-BIH).

Note that neither [38] nor [14] can act directly on compressed measurements. The method proposed in [14] can only operate when the samples are missing, but not when they are compressed. In this work we assume that the data has been compressed in situ by a sparse binary matrix. Therefore before applying [38] we recover the samples from the compressive measurement via compressed sensing.

As done in the previous set of experiments, we study the variation in accuracy with the number of levels for the proposed method. We use four depths – 1 layer (250 atoms), 2 layer (250-125 atoms), 3 layer (250-125-125 atoms) and 4 layers (250-125-125-63 atoms). These configurations gave the best results for each depth. The results are shown in Tables 7 and 8.

The results show the common trend envisaged before. From one layer to three layers the results improve. But above 3 layers, the results do not; it either remains the same or dips slightly.

Next we carry out comparative experiments with other techniques. The configuration for the SAE and the DBNs are obtained from [38], since they are supposed to yield the best results. For our proposed method, we use a three layer architecture. The number of basis are 250-125-125. The value of $\lambda = 0.01$ and $\mu = 1$ is used as before.

Experimental results are reported on 4:1 (Table 9) and 2:1 (Table 10) compression. 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.

TABLE 10. ECG classification results (%age) on 2:1 compression.

Classifier	Acc.	F		S		V	
		Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
DBN	94.8	95.7	66.5	**23.9**	**100**	86.9	**100**
SAE	93.8	96.9	62.8	20.8	**100**	83.2	**100**
Proposed	**100**	**100**	**70.5**	**28.0**	**100**	**94.1**	**100**

TABLE 11. ECG reconstruction results.

Technique	2:1 compression	4:1 compression
CS	0.082±0.036	0.118±0.098
Proposed – training	**0.037±0.027**	**0.085±0.061**
Proposed – testing	**0.037±0.026**	**0.085±0.060**

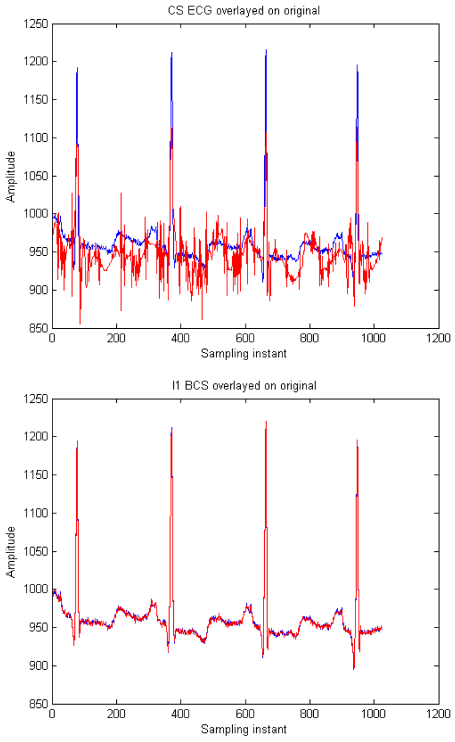

FIGURE 3. ECG reconstruction. Top – CS [1] Reconstruction. Bottom – Proposed.

The results from Tables 9 and 10 show that our proposed technique yields better results than CS reconstruction followed by deep neural network classification [29]. This points to the fact that, there is indeed benefit in learning the reconstruction and analysis operations in a joint fashion.

Finally we compare the reconstruction ability of our proposed technique (Table 11). For comparison we employ the compressed sensing technique proposed in [1]. The results are shown in the following table. We can see that our proposed method yields significantly superior results compared to compressed sensing. Between the training and the testing stages, the testing stage yields better results. This has also been seen for EEG reconstruction.

For visual evaluation we have shown one sample reconstructed by our proposed technique and CS in Fig. 3.

One clearly sees that our proposed technique follows the original signal perfectly, but CS introduces lot of artifacts. This is the reason why the performance of deep learning methods deteriorate. The artifacts introduced by CS reconstruction hampers their performance.

V. CONCLUSION

In this work our objective is to classify biomedical signals from their compressive measurements. This requirement arises in wireless body area networks [39], where it is essential to compress the signal at the sensor nodes with minimal computation. In the past the process of analyzing the compressed signal was a two-stage process. In the first stage, the signal was reconstructed by some compressed sensing technique. In the second stage, some classifier was used to analyze the reconstructed samples. The main issue with this approach is that the compressed sensing reconstruction introduced certain artifacts, which hampered the classification performance.

This work combines the two stages (reconstruction and classification) in a single formulation. By eliminating a separate reconstruction stage, we minimize the errors and artifacts and hence improve the results. A by-product of our formulation is that it can inherently reconstruct the signals. Results show that our method excels over the state-of-the-art in both reconstruction and classification.

This work garners the reconstruction ability of compressed sensing with the power of deep learning. Although shown here for EEG and ECG signals the methodology is generic and is applicable to a wide variety of problems. For example, it can be applied for object recognition problems where parts of the object may be occluded (can be assumed as missing data). It can be used in hyperspectral image analysis where there are missing values owing to the malfunctioning of the sensors.

In fact the methodology can also work on fully sampled problems as well. In that case instead of the projection matrix, we need to use an identity operator.

ACKNOWLEDGEMENT

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

REFERENCES

[1] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "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. Chiang and R. K. Ward, "Energy-efficient data reduction techniques for wireless seizure detection systems," *Sensors*, vol. 14, no. 2, pp. 2036–2051, 2014.

[3] Z. Zhang, T.-P. Jung, S. Makeig, and B. D. Rao, "Compressed sensing of EEG for wireless telemonitoring with low energy consumption and inexpensive hardware," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 221–224, Jan. 2013.

[4] F. Pareschi, P. Albertini, G. Frattini, M. Mangia, R. Rovatti, and G. Setti, "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, Feb. 2016.

[5] B. Sun, H. Feng, K. Chen, and X. Zhu, "A deep learning framework of quantized compressed sensing for wireless neural recording," in *IEEE Access*, vol. 4, pp. 5169–5178, 2016.

[6] G. Chen, "Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features," *Expert Syst. Appl.*, vol. 41, no. 5, pp. 2391–2394, Apr. 2014.

[7] S. Altunay, Z. Telatar, and O. Erogul, "Epileptic EEG detection using the linear prediction error energy," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5661–5665, Aug. 2010.

[8] N. Robinson, A. P. Vinod, K. K. Ang, K. P. Tee, and C. T. Guan, "EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2123–2132, Aug. 2013.

[9] A. Myrden and T. Chau, "A passive EEG-BCI for single-trial detection of changes in mental state," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 4, pp. 345–356, Apr. 2017.

[10] O. Castillo, P. Melin, E. Ramírez, and J. Soria, "Hybrid intelligent system for cardiac arrhythmia classification with fuzzy K-nearest neighbors and neural networks combined with a fuzzy system," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 2947–2955, Feb. 2013.

[11] P. Cheng and X. Dong, "Life-threatening ventricular arrhythmia detection with personalized features," *IEEE Access*, vol. 5, pp. 14195–14203, 2017.

[12] M. Bsoul, H. Minn, and L. Tamil, "Apnea MedAssist: Real-time sleep apnea monitor using single-lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–427, May 2011.

[13] F. Lin *et al.*, "SleepSense: A noncontact and cost-effective sleep monitoring system," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 1, pp. 189–202, Feb. 2017.

[14] A. Majumdar, A. Gogna, and R. Ward, "Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2196–2205, Sep. 2017.

[15] S. Singh, V. Singhal, and A. Majumdar, "Deep blind compressed sensing," in *Proc. Data Compress. Conf.*, Apr. 2017, p. 459.

[16] S. Gleichman and Y. C. Eldar, "Blind compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6958–6975, Oct. 2011.

[17] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[18] S. Li and H. Qi, "Compressed dictionary learning for detecting activations in fMRI using double sparsity," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2014, pp. 434–437.

[19] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller, "A deep matrix factorization method for learning attribute representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 417–429, Mar. 2017.

[20] S. Tariyal, A. Majumdar, R. Singh, and M. Vatsa, "Deep dictionary learning," *IEEE Access*, vol. 4, pp. 10096–10109, 2016.

[21] M. Saar-Tsechansky and F. Provost, "Handling missing values when applying classification models," *J. Mach. Learn. Res.*, vol. 8, pp. 1623–1657, Jul. 2007.

[22] D. JakovetiÄĞ, J. M. F. Moura, and J. Xavier, "Linear convergence rate of a class of distributed augmented Lagrangian algorithms," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 922–936, Apr. 2015.

[23] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Philadelphia, PA, USA: SIAM, 2014.

[24] R. Nishihara, L. Lessard, B. Recht, A. Packard, and M. I. Jordan. (2015). "A general analysis of the convergence of ADMM." [Online]. Available: https://arxiv.org/abs/1502.02009

[25] Y. Wang, W. Yin, and J. Zeng. (2015). "Global convergence of ADMM in nonconvex nonsmooth optimization." [Online]. Available: https://arxiv.org/abs/1511.06324

[26] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, Nov. 2004.

[27] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1132–1145, Jun. 2013.

[28] *Time Series EEG Data*. Accessed: Aug. 2016. [Online]. Available: http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html

[29] A. Page, J. T. Turner, T. Mohsenin, and T. Oates, "Comparing raw data and feature extraction for seizure detection with deep learning methods," in *Proc. AAAI FLAIRS*, 2014, pp. 1–4.

[30] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: Fast classification and anomaly measurement," *J. Neural Eng.*, vol. 8, no. 3, p. 036015, 2015.

[31] P. Thodoroff, J. Pineau, and A. Lim. (Jul. 2016). "Learning robust features using deep learning for automatic seizure detection." [Online]. Available: https://arxiv.org/abs/1608.00220
[32] A. Supratak, L. Li, and Y. Guo, "Feature extraction with stacked autoencoders for epileptic seizure detection," in *Proc. IEEE EMBC*, Aug. 2014, pp. 4184–4187.
[33] Y. Zhang, Y. Li, J. Sun, and J. Ji, "Estimates on compressed neural networks regression," *Neural Netw.*, vol. 63, pp. 10–17, Mar. 2015.
[34] A. Fabisch, Y. Kassahun, H. Wöhrle, and F. Kirchner, "Learning in compressed space," *Neural Netw.*, vol. 42, pp. 83–93, Jun. 2013.
[35] L. Vidya, V. Vivekanand, U. Shyamkumar, and D. Mishra, "RBF-network based sparse signal recovery algorithm for compressed sensing reconstruction," *Neural Netw.*, vol. 63, pp. 66–78, Mar. 2015.
[36] J. Zhao, Y. Lv, Z. Zhou, and F. Cao, "A novel deep learning algorithm for incomplete face recognition: Low-rank-recovery network," *Neural Netw.*, vol. 94, pp. 115–124, Oct. 2017.
[37] E. J. S. da Luz, T. M. Nunes, V. H. C. de Albuquerque, J. P. Papa, and D. Menotti, "ECG arrhythmia classification based on optimum-path forest," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3561–3573, Jul. 2013.
[38] M. M. Al Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, and R. R. Yager, "Deep learning approach for active classification of electrocardiogram signals," *Inf. Sci.*, vol. 345, no. 1, pp. 340–354, Jun. 2016.
[39] H. Viittala, L. Mucchi, M. Hämäläinen, and T. Paso, "ETSI SmartBAN system performance and coexistence verification for healthcare," *IEEE Access*, vol. 5, pp. 8175–8182, 2017.

VANIKA SINGHAL received the master's degree from The LNM Institute of Information Technology, Jaipur, in 2014. She is currently pursuing the Ph.D. degree with the Indraprastha Institute of Information Technology, New Delhi. She was a Lecturer with Manipal University Jaipur, India, from 2014 to 2016. Her research interests include deep learning.

ANGSHUL MAJUMDAR received the bachelor's degree from Bengal Engineering College, Shibpur, in 2005, and the master's and Ph.D. degrees from The University of British Columbia in 2009 and 2012, respectively. Since 2012, he has been an Assistant Professor with the Indraprastha Institute of Information Technology, Delhi, India. His research interests include optimization, with applications in signal processing and machine learning. His current research interests include deep learning. He is the Founding Chair of the IEEE SPS Delhi Chapter and is serving as the Chair for the IEEE SPS Chapter's Committee from 2016 to 2018. He has authored or co-authored over 150 papers in reputable conferences and journals, and has authored one book *Compressed Sensing-Based Magnetic Resonance Imaging Reconstruction* (Cambridge University Press) and has co-edited a book *MRI: Physics, Reconstruction and Analysis* (CRC Press).

RABAB K. WARD is currently a Professor Emeritus with the Electrical and Computer Engineering Department, The University of British Columbia (UBC), Canada. Her research interests are mainly in the areas of signal, image, and video processing. She is a fellow of the Royal Society of Canada, the Institute of Electrical and Electronics Engineers, the Canadian Academy of Engineers, and the Engineering Institute of Canada. She is the President of the IEEE Signal Processing Society. She has received the UBC Applied Science Dean's Medal of Distinction in 2016, the Killam Award for Excellence in Mentoring in 2013, the Career Achievement Award of CUFA, BC, the organization representing all professors and academic staff at BC's doctoral universities in 2011, the UBC Engineering Co-op Faculty Member of the Year Award in 2010, the IEEE Signal Processing Society top award—The Society Award in 2008, the YWCA Woman of Distinction Award in 2008, the British Columbia's top engineering award—The RA McLachlan Memorial Award in 2006, and the UBC Killam Research Prize in 1998.