Multimodal deep learning approach for joint EEG-EMG data compression and classification

Ahmed Ben Said, Amr Mohamed, Tarek Elfouly
Computer Science
and Engineering Department
Qatar Univrsity
2713, Doha, Qatar
Email: {abensaid, amrm, tarekfouly}@qu.edu.qa

Khaled Harras
Computer Science Department
Carnegie Melon university-Qatar
24866, Doha, Qatar
Email: kharras@qatar.cmu.edu

Z. Jane Wang
Electrical and Computer
Engineering Department
University of British Columbia
Vancouver, BC, Canada
Email: zjanew@ece.ubc.ca

Abstract—In this paper, we present a joint compression and classification approach of EEG and EMG signals using a deep learning approach. Specifically, we build our system based on the deep autoencoder architecture which is designed not only to extract discriminant features in the multimodal data representation but also to reconstruct the data from the latent representation using encoder-decoder layers. Since autoencoder can be seen as a compression approach, we extend it to handle multimodal data at the encoder layer, reconstructed and retrieved at the decoder layer. We show through experimental results, that exploiting both multimodal data intercorrelation and intracorelation 1) Significantly reduces signal distortion particularly for high compression levels 2) Achieves better accuracy in classifying EEG and EMG signals recorded and labeled according to the sentiments of the volunteer.

Index Terms—mHealth, deep learning, compression, classification

I. INTRODUCTION

Healthcare has always been considered as a strategic priority worldwide. The increasing number of elderly and chronic disease patients has made the physical contact between the caregiver and patients more and more difficult. Following the fast development of wireless technologies, the interoperability between healthcare entities and mobile has grown. The development of complex devices has stimulated the creation of many mobile health or 'mHealth' applications and wearable devices for fitness tracker, sleep monitoring [1]. . . The mHealth industry is predicted to grow $12 billions by 2018 [2].

Motivated by the myriad of biomedical sensors, mobile phones and applications, the scientific communities have standardized the system that focuses on the acquisition of vital signs such as electroencephalogram (EEG) and Electromyogram (EMG) by body area sensor networks (BASN) under IEEE 802.15.6 [3]. Thus, a typical mHealth BASN system consists of sensors collecting the data, a Personal Data Aggregator (PDA) and remote server. However, due to network limitation, data delivery through the network can be hindered. Consequently, we need to optimize every bit of data being sent. One of the possible pre-processing stages is to encode the data in the PDA i.e. mapping x_i to compressed data z_i. At the server level, the received data z_i is decoded i.e. mapped to $\hat{x_i}$ which approximates the original data x_i. Many successful algorithms have been proposed for time series compression. Srinvasan et al. [4] designed a 2-D lossless EEG compression where the signal is arranged in 2-D matrix as a preprocessing. Compression is achieved through a two-stage coder composed of Set Partitioning In Hierarchical Trees (SPIHT) [5] layer and Arithmetic Coding (RC) layer. Hussein et al. [6] proposed a scalable and energy efficient EEG compression scheme based on Discret Wavelet Transform (DWT) and Compressive Sensing (CS) in wireless sensors. Several parameters have been considered to control the total energy consumption of the encoder and transmitter. The optimal configuration of these parameters is chosen based on optimization scheme where the total power consumption should not exceed a certain threshold. In [7], authors applied CS technique for EEG signal compression. Since the multichannel EEG signals have common sparse support in the transform domain, they stack the sparse transform coefficients as columns of matrix. Thus, the recovery problem becomes row-sparse and solved through Bregman algorithm [8]. Majumdar et al. [9] argued that CS is not efficient for EEG compression because there is no sparsifying basis that fulfills the requirements of incoherence and sparsity. Instead, authors formulated the problem as a rank deficiency problem solved by a Bregman-derived algorithm.

Following the development of wireless BASN, vital signs data have become abundant. mHealth systems are now capable of collecting data from different modalities (EEG, EMG, etc). Although, they may seem totally different, these data can describe the same phenomena. For example, in case of schizophrenic person, when a stimulus is presented, a peak in the EEG registration is witnessed while the functional Magnetic Resonance Imaging (fMRI) data shows activations in the temporal lobe and the middle anterior cingulate region [10]. Thus, both modalities are very likely to be correlated. Each modality has its advantages and limitations but analyzing multiple modalities offers better understanding of the investigated phenomena. The aforementioned methods, although exhibit good performance, do not exploit the correlation among multiple modalities.

Deep learning approach has emerged as one of the possible techniques to exploit the correlation of the data from multiple modalities. Ngiam et al. [11] proposed a multimodal deep

learning approach for cross modality feature learning from video and speech data. Srivastava et al. [12] built a multimodal deep belief network [13] to learn multimodal representation from image and text data for image annotation and retrieval tasks. In [14], authors designed a deep Boltzmann machine [15] based architecture to extract a meaningful representation from multimodal data for classification and information retrieval task. Liu et al. [16] proposed a multimodal autoencoder [17] approach for video classification based on audio, image and text data where the intra-modality semantic for each data is separately learning by a stacked autoencoder. Next, the learned features are concatenated and fed to another deep autoencoder with a softmax layer for classification.

Few research attempts have addressed the possible application of autoencoder for biomedical and mHealth applications. In [18], Yann Ollivier proved that there is a strong relationship between minimizing the code length of the data and minimizing reconstruction error that an autoencoder seeks. Tan et al. [19] used a stacked autoencoder for mammogram image compression. Training is conducted on image patches instead of the whole images. In [20], authors applied the autoencoder for Electrocardiogram (ECG) compression. Comparison results with various classic compression methods showed that this special type network is reliable for signal compression. However, the problem of multimodal data compression in context of mHealth is still not well-investigated.

We propose in this paper a multimodal approach for data compression and feature learning. The encoding-decoding scheme can be achieved through a stack of autoencoders. Our approach exploits the intracorrelation as well as the intercorrelation among multiple modalities to achieve efficient compression and classification.

The rest of the paper is organized as follows: in section II, we present the autoencoder architecture. Section III is dedicated for presenting our multimodal approach for joint EEG and EMG data compression and classification. Experimental results are illustrated and discussed in section IV and we conclude in the last section.

II. BACKGROUND

A. Autoencoder

An autoencoder, illustrated in Fig. 1, is a special type of neural network consisting of three layers. The data are first fed into the input layer, propagated to a second layer called the hidden or bottleneck layer and then reconstructed at a third layer called the reconstruction layer.

The encoder transforms the set of data vectors $x \in R^X$ into hidden representation $h \in R^H$ via an activation function f:

$$h = f(Wx + b) \qquad (1)$$

The decoder transforms back the hidden representation h to reconstruction data $r \in R^X$ via an activation function g:

$$r = g(W'h + b') \qquad (2)$$

The Parameters $W \in R^{X \times H}$ and $W' \in R^{H \times X}$ are called weight matrices. $b \in R^H$ and $b' \in R^X$ are called

the bias vectors. f and g are typically hyperbolic function $tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ or sigmoid function $sigmoid(x) = 1/(1 + e^{-x})$. In practice, we use tight weight configuration i.e. $W' = W^T$. Autoencoder seeks the optimal set of parameters $\Theta = \{W, b, b'\}$ that minimizes the reconstruction error $J_\Theta(x, r)$. This error is generally the Squared Euclidean distance $L(x, r) = ||x - r||^2$ or cross-entropy loss $L(x, r) = -\sum_{i=1}^{X} x_i log(r_i) + (1 - x_i) log(1 - r_i)$. When using affine activation function and squared error loss, autoencoder essentially performs Principal Component Analysis (PCA) [1]. Minimization is generally carried out via gradient descent algorithm. In other words, the purpose of this minimization is to obtain $r \approx x$ i.e. an approximation to the identity function. But, by constraining the system by limiting the number of hidden units at the hidden layer, we are forcing the system to learn a compressed version of the data. Furthermore, to prevent overfitting, i.e. just learning the identity function, another constraint is often added: a weight term that regularizes $J_\Theta(x, r)$. Then, we have:

$$J_\Theta(x, r) = L(x, r) + \lambda ||W||_2^2 \qquad (3)$$

Where λ is the parameter that controls the amount of regularization.

B. Stacked autoencoder

Stacked autoencoder (SAE), illustrated in Fig 2, is a neural network which consists of multiple layers of autoencoders. The output of each layer is fed to the next layer. SAE is trained via a greedy layer-wise training [21]. Specifically, it is done one layer at a time. At each layer, we consider the autoencoder composed of the current layer and its previous one which is the output of the previous layer. Once $N-1$ layers are trained, we can compute the output of th N^{th} layer wired to it. This unsupervised stage is followed by a supervised fine-tuning of the parameters where a softmax layer is added on top of the SAE.
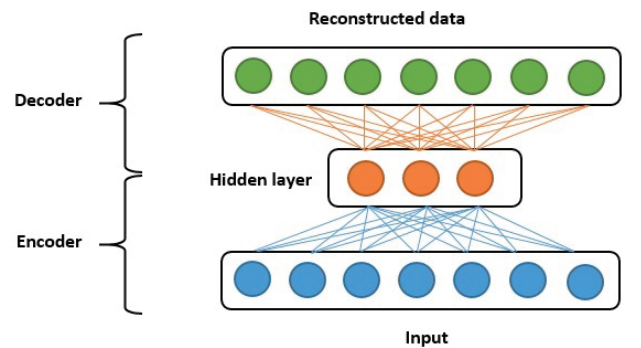

Fig. 1. Autoencoder: the encoder maps the data to hidden representation. The decoder maps the encoded features to reconstruct the data

---

[1]It will find the same subspace as PCA but the projection direction does not essentially correspond to the principal components directions
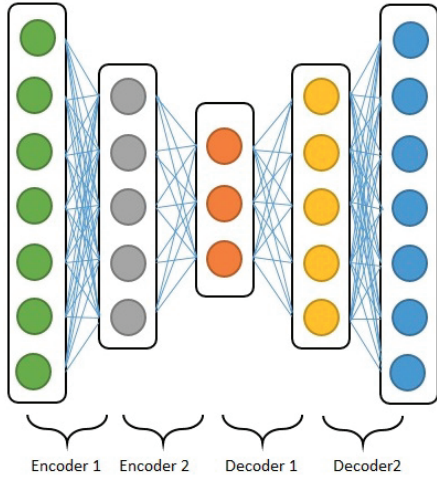
Fig. 2. Stacked autoencoder: the output of each layer is the input of the next layer

III. MULTIMODAL AUTOENCODER FOR EEG-EMG COMPRESSION AND CLASSIFICATION

Fig. 3 exhibits the multimodal autoencoder architecture. It consists of two pathways for EEG and EMG. Each pathway represents a unimodal stacked autoencoder dedicated to learn the intra-modality correlation of the data while the joint layer merge the higher level features.

A. Unimodal data pre-training

SAE is applied separately for each modality, we use the sigmoid activation function and the Squared Euclidean distance as loss function regularized by a weight term. We apply also tied weight configuration. The output of the ith layer is obtained as follows:

$$
\begin{aligned}
z_1 &= sigmoid(W_1 x_1 + b_1) \qquad i = 1 \\
z_i &= sigmoid(W_i x_i + b_i) \qquad i = 2..N
\end{aligned}
\tag{4}
$$

The SAE is trained using the greedy-layer wise training approach where we feed the latent representation of the autoencoder found below to the current layer. This deep architecture makes the system more scalable and efficient while progressively extracting higher level features from the high dimensional data.

B. Deep multimodal learning

The single modal pre-training does not involve inter-modality correlation which can contribute in better representation of the higher level features. It especially allows encoding the multiple modalities in a single shared representation obtained by the joint layer. The output of this layer encompasses the contribution of each modality in the code which represents the compressed data. The joint representation is obtained as follows:

$$
z = \sum_{i \in \{e,m\}} sigmoid\big(W_{N+1}^i z_{N+1}^i + b_{N+1}^i\big)
\tag{5}
$$

Where e and m refer to EEG and EMG respectively. Furthermore, we train the multimodal autoencoder with an augmented noisy data where additional examples are added leading to samples with only one single modality. In practice, we add zeros values examples for one modality while keeping the original values for the other modality and vice-versa. Thus, one third of the training data is EEG only, another one third is EMG only and the rest has both EEG and EMG data. This strategy, inspired from Nigiam et al [11], follows the denoising autoencoder paradigm [22] and is justified by twofold:

- Correlation among multiple modalities is very likely to be non-linear.
- This non-linearity often leads to hidden units being activated by one single modality.

Therefore, the original and corrupted inputs are propagated independently to the higher layers which are then trained progressively to reconstruct the clean presentation from both inputs.

C. Fine-tuning

The compressed data can be used for classification task, that is, to fine-tune the layers with respect to a supervised criterion by plugging the bottleneck layer to a softmax classifier [23]:

$$
\hat{p} = \frac{exp(Wy + b)}{\sum_{l=1}^{L} exp(W^l y + b^l)}
\tag{6}
$$

Where $\hat{p}$ is the predicted object label, y represents the compressed data and L is the number of classification labels. Therefore, the overall objective function to minimize is:

$$
\Im(x, r, p, \hat{p}) = J_\Theta(x, r) + L(p, \hat{p})
\tag{7}
$$

Where p is the true label and $L(p, \hat{p})$ can be an entropic loss function.

IV. EXPERIMENTAL RESULTS

mHealth systems acquire, process, store, secure and transport the medical data. Data delivery should be as efficient and optimized as possible in terms of energy consumption and bandwidth usage. A typical system consists of mHealth wearable device that senses vital signs. These data are collected by a PDA and should be transmitted to a remote server handled by a medical entity [24]. At the server level, a multimodal autoencoder is already trained and the optimal configuration is already found. This configuration is also known by the PDA which should apply it on the collected data for compression. We present in this section several experimental results where we compare our compression scheme with some state of the art compression methods. Furthermore, we compare our multimodal strategy with the unimodal one to highlight the importance of exploiting the intermodality correlation.

A. Dataset

We conduct our experiments on the DEAP dataset [25]. It consists of EEG, EMG and multiple physiological signals recorded from 32 participants during 63 seconds at 128 Hz. During experiments, volunteers watched 40 music videos and
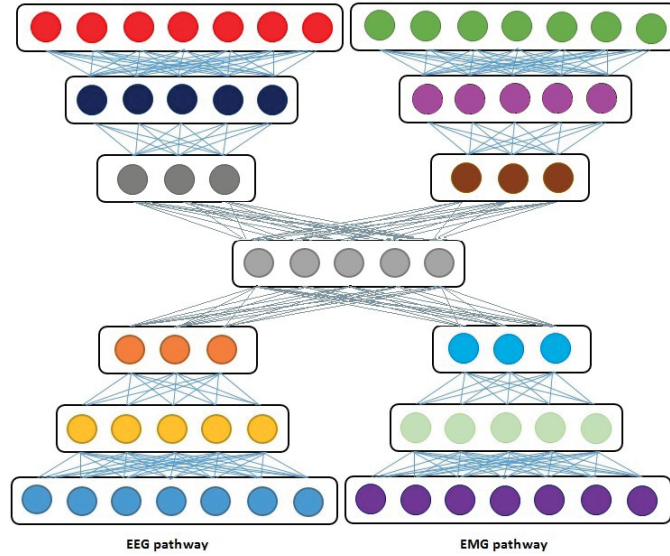
Fig. 3. Multimodal stacked autoencoder with EEG and EMG pathways. The input of each pathway is the set of vectors of each data with the number of units corresponding to the data dimension

rate them on a scale from 1 to 9 with respect to four criteria: likeness (dislike, like), valence (ranges from unpleasant to pleasant), arousal (ranges from uninterested or bored to excited) and dominance (ranges from helpless and weak feelings to empowered feeling).

Signals are segmented into 6 seconds segments, whitened and normalized between 0 and 1. For both EEG and EMG data, we have 23040 samples of 896 dimensionality. These data should then be divided into training and testing sets.

B. Compression tasks

We compare our compression method with the Discrete Wavelet Transform (DWT) [26], Compressed Sensing (CS) [27] and the 2D compression approach which is based on SPIHT and FastICA [28]. For the latter algorithm, we use two configurations of 3 and 6 independent components denoted 2D-SPIHT-3-ICs and 2D-SPIHT-6-ICs. We evaluate performance using compression ratio (CR) and residual distortion (D).

$$CR = \left(1 - \frac{m}{n}\right) * 100 \qquad (8)$$

$$D = \frac{||r - x||}{||x||} * 100 \qquad (9)$$

Where m and n are the length of the compressed and original signals (number of samples). D is the percentage root-mean-square difference between the compressed and original signals. For each data pathway, we use a two-layer SAE. Table I presents the numbers of hidden units for each layer of the multimodal autoencoder as well as the DWT thresholds and their corresponding CRs. We divide data to 50% training and testing.

Fig. 4 and 5 exhibit distortion variation with respect to different CR values for EEG and EMG. The findings show that for higher compression ratios, the multimodal approach

TABLE I
MULTIMODAL AUTOENCODER AND DWT CONFIGURATION AND THE RESULTING COMPRESSION RATIO

Multimodal autoencoder	DWT threshold	CR (%)
896-806	EEG: 0.025 ; EMG: 0.019	10
896-716	EEG: 0.05 ; EMG: 0.04	20
896-627	EEG: 0.085 ; EMG: 0.06	30
896-537	EEG: 0.13 ; EMG: 0.10	40
896-448	EEG: 0.29 ; EMG: 0.51	50
440-358	EEG: 0.66 ; EMG: 0.64	60
440-268	EEG: 0.75 ; EMG: 0.69	70
440-179	EEG: 0.83 ; EMG: 0.74	80
380-89	EEG: 0.92 ; EMG: 0.78	90

performs better than DWT and CS. For example, for CR=80%, our approach is able to reconstruct EEG and EMG with distortions of 12% and 13.85% respectively while CS distorts EEG by 22% and EMG by 17.21%. With 2D-SPIHT-3-ICs, EEG and EMG distortions are 33.7% and 35.7% respectively while with 2D-SPIHT-3-ICs, EEG and EMG are distorted by 33% and 33.5%. DWT exhibits low performance with 68% and 73.12% for EEG and EMG respectively. Although DWT, CS and the 2D approach perform better for low compression levels, the proposed method presents stable performance for different compression levels. This can be explained by the capacity of the underlying architecture to exploit the statistics of the data to achieve better compression.

We further examine the effect of training/testing data partition on the compression results. Fig. 6 and Fig. 7 illustrate the whisker diagrams for EEG and EMG signals respectively. We can clearly deduce that more training data result in less distortion. This confirms a known deep learning rule of thumb stating the more training data we have, the better the results are.
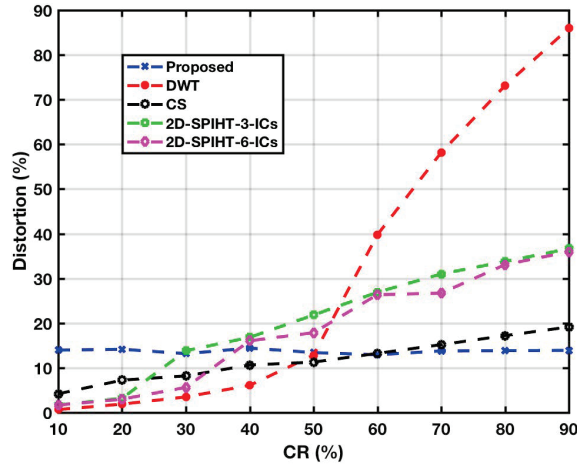
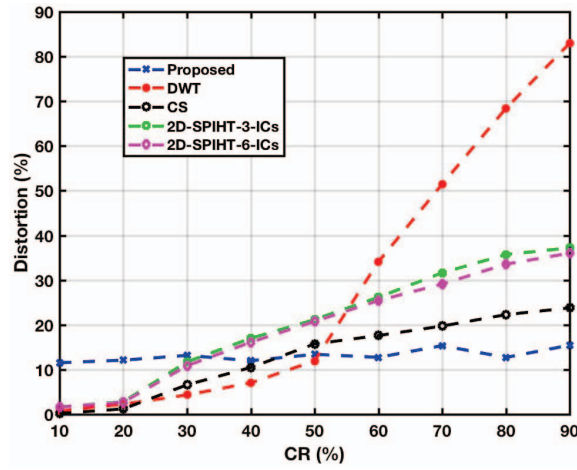Fig. 4. EEG Distortion (%) with respect to the Compression Ratio (%).

Fig. 5. EMG Distortion (%) with respect to the Compression Ratio (%).

C. Classification task

The objective of this experiment is to demonstrate the importance of the multimodal approach. We conduct binary classification of the EEG and EMG with respect to two of the four labeling possibilities: dominance and arousal. We follow the same approach as in [25]: video ratings are thresholded into two classes. On the scale of 1 to 9, we simply place the threshold in the middle. We compare our approach with two-layer SAE and Deep Boltzmann Machine (DBM) architectures [15] with the softmax classifier on top of them. For SAE, we use the sigmoid activation function. We choose 75% training-testing partition. Figures 8 and 9 illustrate the classification results with respect to the dominance and arousal respectively. By exploiting the inter-modality correlation, the proposed approach achieves the best results with 78.1%. The single-modality approaches are less accurate.

These findings confirm that, when available, multiple modalities can offer better understanding of the underlying phenomena even if the data exhibit different characteristics.

D. Discussion

In a typical mHealth system, a client-server architecture is the common choice where the system relies on the available networks to deliver the data. In general, the healthcare giver generally relies on multiple vital signs for an accurate diagnosis. The proposed approach is flexible in the sense that, if an additional modality is collected by the PDA via a wearable device, can be easily incorporated in the architecture presented in Fig. 3, compressed and classified. The deep neural network can be trained offline. Once it achieves good performance, the optimal configuration (weights and biases) is applied at the client side for efficient data delivery. However, it is worth-noting that our approach it less efficient for low compression ratio.

V. CONCLUSION

We have presented a deep learning approach for multimodal data compression and classification. Our strategy focuses on exploiting the inter and intra correlation among multiple modalities to enhance the compression and classification of data in context of mHealth application.

The core of the proposed method is based on the classic autoencoder which has been originally designed for encoding-decoding data. For each modality presented, we dedicate a stacked autoencoder to extract high level abstraction of the data by modeling the intra-correlation. A joint layer is added on top of each encoding part of the stacked autoencoders to model data intercorrelation.

We have conducted compression and classification experiments. Comparison with DWT and CS have shown that our approach performs better with high compression ratio. We have also demonstrated the effectiveness of the multimodal approach for classification of EEG and EMG. Comparison with some unimodal algorithms e.g. Deep Botzmann Machines and stacked autoencoders shows that the multimodal autoencoder leads to better classification accuracy.

In future work, we will investigate the possible application of Convolutional Neural Network. Furthermore, we intend to
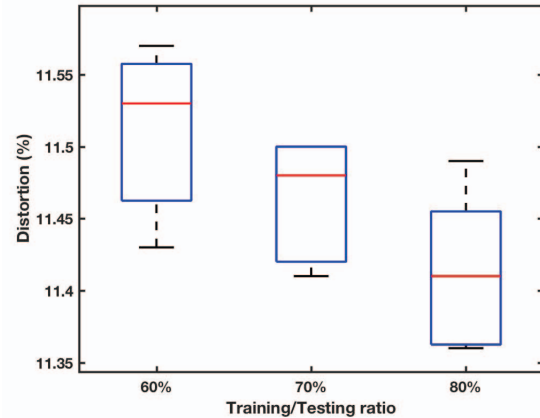

Fig. 6. Whisker diagram of EEG data with various training/testing partitions
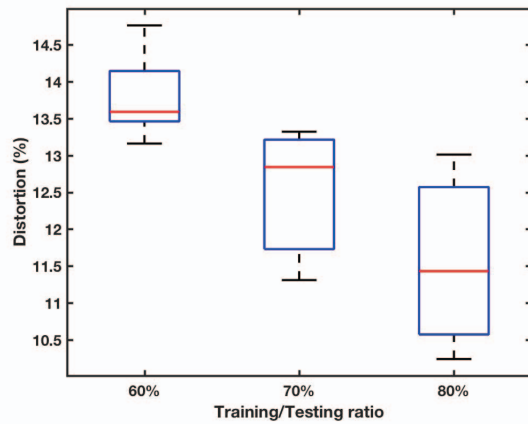
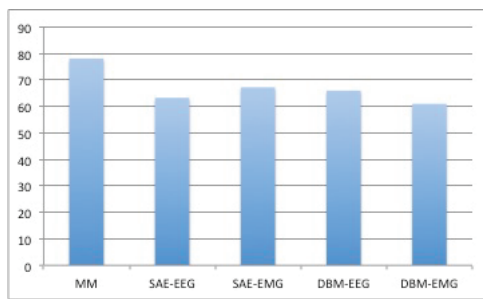Fig. 7. Whisker diagram of EMG data with various training/testing partitions

Fig. 8. Classification accuracy with respect to the dominance label: the multimodal approach (MM) achieves the best performance with 78.1%
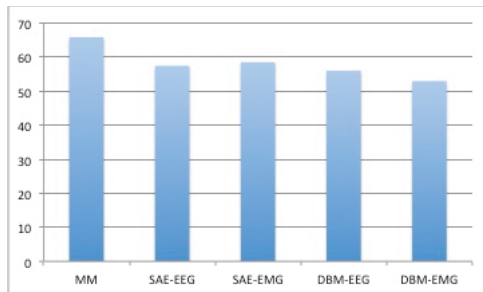

Fig. 9. Classification accuracy with respect to the arousal label: the multimodal approach (MM) achieves the best performance with 65.9%

make the autoencoder-based compression scheme adaptive by including the network resource in the choice of the neural network architecture.

ACKNOWLEDGMENT

REFERENCES

[1] "mhealth app market sizing 2015 - 2020 data report to size opportunities in the mhealth app market."

[2] [Online]. Available: http://mhealthintelligence.com/news/the-history-of-mobile-health-from-cell-phones-to-wearables

[3] Ieee 802.15 wpan task group 6 body area networks, 2012. [Online]. Available: http://standards.ieee.org/findstds/standard/802.15.6-2012.html

[4] K. Srinivasan, J. Dauwels, and M. R. Reddy, "A two-dimensional approach for lossless eeg compression," *Biomedical Signal Processing and Control*, 2011.

[5] A. Said and W. A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, 1996.

[6] R. Hussein, A. Mohamed, and M. Alghoniemy, "Scalable real-time energy-efficient eeg compression scheme for wireless body area sensor network," *Biomedical Signal Processing and Control*, 2015.

[7] A. Shukla and A. Majumdar, "Row-sparse blind compressed sensing for reconstructing multi-channel eeg signals," *Biomedical Signal Processing and Control*, 2015.

[8] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for L1-minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences*, 2008.

[9] A. Majumdar, A. Gogna, and R. Ward, "A low-rank matrix recovery approach for energy efficient eeg acquisition for a wireless body area network," *Sensors*, 2014.

[10] N. Correa, Y.-O. Li, T. Adali, and V. D. Calhoun, "Examining associations between fmri and eeg data using correlation analysis," in *In proceeding of 5th IEEE international symposium on biomedical imaging: From nano to macro,*, 2008.

[11] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning*, 2011.

[12] N. Srivastava and R. Salakhutdinov, "Learning representations for multimodal data with deep belief nets," in *Proceedings of the 29th International Conference on Machine Learning*, 2012.

[13] G. E. Hinton. (2009) Deep belief networks. [Online]. Available: www.scholarpedia.org/article/Deep_belief_networks

[14] N. Srivastava and R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," in *Advances in neural information processing systems*, 2012.

[15] R. Salakhutdinov and G. E. Hinton, "Deep boltzman machines," in *In Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2009.

[16] Y. Liu, X. Feng, and Z. Zhou, "Multimodal video classification with stacked contractive autoencoders," *Signal Processing*, 2016.

[17] G. E. Hinton and R. Salakhutdinov, "Reducing the dimensionality ofdata with neural networks," *Science*, 2006.

[18] Y. Ollivier, "Auto-encoders: reconstruction versus compression," 2014, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01104268

[19] C. C. Tan and C. Eswaran, "Using autoencoders for mammogram compression," *Journal of Medical Systems*, 2011.

[20] D. D. Testa and M. Rossi, "Lightweight lossy compression of biometric patterns via denoising autoencoders," *IEEE Signal Processing Letters*, vol. 22, pp. 2304–2308, 2015.

[21] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Inproceedings of Advances in Neural Information Processing Systems*, 2006.

[22] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *International Conference on Machine Learning*, 2008.

[23] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, pp. 1–127, 2009.

[24] A. Awad, A. Mohamed, and C.-F. Chiasserini, "User-centric network selection in multi-rat systems," in *2016 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, April 2016, pp. 97–102.

[25] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, , T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis ;using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, pp. 18–31, 2012.

[26] S. Mallat, *A Wavelet Tour of Signal Processing, Third Edition*, A. Press, Ed., 2008.

[27] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[28] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 45, pp. 411–430, 2000.