

Wearable EEG-based Activity Recognition in PHM-related Service Environment via Deep Learning

Soumalya Sarkar¹, Kishore Reddy², Alex Dorgan³, Cali Fidopiastis³ and Michael Giering⁵ *

^{1,2,3,4,5} *United Technologies Research Center, East Hartford, CT 06118, USA*

sarkars@utrc.utc.com, reddykk@utrc.utc.com, dorgana@utrc.utc.com, fidopicm@utrc.utc.com, GierinMJ@utrc.utc.com

ABSTRACT

It is of paramount importance to track the cognitive activity or cognitive attention of the service personnel in a Prognostics and Health Monitoring (PHM) service related training or operation environment. The electroencephalography (EEG) data is one of the good candidates for cognitive activity recognition of the user. Analyzing electroencephalography (EEG) data in an unconstrained (natural) environment for understanding cognitive state and classifying human activity is a challenging task due to multiple reasons such as low signal-to-noise ratio, transient nature, lack of baseline availability and uncontrolled mixing of various tasks. This paper proposes a framework based on an emerging tool named deep learning that monitors human activity by fusing multiple EEG sensors and also selects a smaller sensor suite for a lean data collection system. Real-time classification of human activity from spatially non-located multi-probe EEG is executed by applying deep learning techniques without performing any significant amount of data preprocessing and manual feature engineering. Two types of deep neural networks, deep belief network (DBN) and deep convolutional neural network (DCNN) are used at the core of the proposed framework, which automatically learns necessary features from EEG for a given classification task. Validation on extensive amount of data, which was collected from several subjects while they were performing multiple tasks (listening and watching) in PHM service training session, is presented and significant parallels are drawn from existing domain knowledge on EEG data understanding. Comparison with machine learning benchmark techniques shows that deep learning based tools are better at understanding EEG data for task classification. It is observed via sensor selection that a significantly smaller EEG sensor suite can perform at a comparable accuracy as the original sensor suite.

1. INTRODUCTION

It is becoming an ubiquitous practice in industry for the field technicians to use wearables (with multi-modal sensor nodes) while performing PHM related service. The wearables keep track of vital statistics of the technicians to make sure that safety-critical jobs are not compromised due to some physical issues. Also Real-time tracking of the the service personnel's cognitive activity in a Prognostics and Health Monitoring (PHM) related environment is significant both for designing an effective multi-media training module and evaluating the quality of service at the PHM-critical industries. EEG data is a preferable non-invasive candidate for tracking human activity. Electroencephalography (EEG) is the process of measuring the brain's neural activity as electrical voltage fluctuations along the scalp that results from the current flows in brain's neurons (Niedermeyer & da Silva (2005)). Brain-computer interface (BCI) or Brain-Machine interface (BMI), depending on understanding brain waves, has become one of the main tools to help estimating the cognitive state of the subject in real-time. For industries, real-time understanding of individual workload, fatigue and alertness of field maintenance personnel facilitates the process of creating an efficient and safe work environment. But historically, the EEG data collection system has been bulky and EEG data is highly transient and noisy. Also the quality of data analysis is significantly dependent on proper calibration of the baseline data for a specific human, which needs to be updated periodically. Most existing state-of-the-art techniques analyze EEG data collected in a constrained environment along with a known baseline activity response. For a real work environment, an activity monitoring process along with a lean data collection system have to be developed which can process EEG data and recognize human activity in real-time without baseline knowledge.

Any proposed representation of EEG waves needs to address low signal-to-noise ratio and other interferences like muscle movement, cardiac cycles, and ocular movements. Many approaches have been proposed to represent EEG signals and classify these representations into mental tasks that the sub-

*corresponding author

Soumalya Sarkar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ject is performing. Iscan et al. (Iscan et al. (2011)) applied multiple signal processing tools based on time domain (power spectral density), frequency domain (power spectral density) and wavelet transform based features for understanding EEG data. Initial baseline values and changes in baseline have always been an issue in estimating cognitive state (Vartak et al. (2008)). There have been multiple endeavors in applying Artificial Neural Networks (ANNs) (Tsoi et al. (1993)) as the classifier for EEG data analysis while using various types of features such as relative wavelet energy (RWE) (Peters et al. (1998)), lifting-based discrete wavelet transform (LBDWT) (Subasi et al. (2005)), statistical parameters from the decomposed wavelet co-efficients (Patnaik & Manyam (2008)), kurtosis and amplitudes (Chambayil et al. (2010)), principle component analysis (PCA) (Kottaimalai et al. (2013)) based features, and multi-resolution analysis (MRA) based wavelet features (Omerhodzic et al. (2013)). Recurrent neural network (RNN) has also been used for temporal EEG classification (Güler et al. (2005); Naderi & Mahdavi-Nasab (2010)). An autoencoder neural network was implemented in (Morimoto & Sketch (n.d.)) to automatically learn features from unlabeled EEG data and apply binary logistic regression (BLR) and a support vector machine (SVM). The output layer of the final RBM as the input to a classifier and k-nearest neighbor (k-NN), support vector machines (SVM), and logistic regression as classifiers are used in a modular fashion in (Turner et al. (2014)). Classification of L/R Hand Movement EEG Signals is performed by (Alomari et al. (2013)), where feature vector to the classifiers (ANN and SVM) included the Event-Related (De) Synchronization (ERD/ERS) and movement-related cortical potentials (MRCP) features in addition to the mean, power and energy of the activations of the resulting Independent Components (ICs) of the epoched feature datasets. It is important to understand that the choices made in implementing baseline correction can influence the results of analyses and that differences in the baseline correction procedure may be one reason for inconsistent results across studies (Roach & Mathalon (2008)). Multiple methods such as simple subtraction of baseline values (Spencer et al. (2004)), percent change from baseline (Hoogenboom et al. (2006)), decibels normalization (Delorme & Makeig (2004)), and baseline-adjusted z scores (Le Van Quyen et al. (2001); Lachaux et al. (1999); Rodriguez et al. (1999)) have been applied to nullify the effect of the baseline from EEG data.

Recent advancements in deep learning shows that multi-layered neural networks are excellent at low-level feature extraction from raw data for automated learning and discriminative tasks without significant manual feature engineering. A deep neural network model extracts hierarchical features from the training data (Hinton & Salakhutdinov (2006)) through the use of multiple layers of latent variables. Deep Learning is an emerging branch of machine learning with a strong emphasis on modeling multiple levels of abstraction (from low-level

features to higher-order representations, i.e., features of features) from data (Deng & Dong (2014); Bengio et al. (2013)). For example, in a typical image processing application while low-level features can be partial edges and corners, high-level features may be combination of edges and corners to form parts of an image. Deep Learning generally revolves around implementing complex model structures composed of nonlinear transformations such as sigmoid function in order to learn the higher representations of data (Bengio et al. (2013)). In the last decade, Deep Belief Network (DBN) has emerged as an attractive option for data dimensionality reduction (Hinton & Salakhutdinov (2006)), feature learning (Coates et al. (2011)), and solving classification problems (Larochelle & Bengio (2008)). Several other deep learning architectures such as Convolutional Neural Networks, Stacked Denoising Autoencoders, and Deep Recurrent Neural Networks have also gained immense traction recently as they have been shown to outperform all other state-of-the-art machine learning tools for handling very large dimensional data spaces for learning features in order to perform detection, classification and prediction. A Convolutional Neural Network (CNN) (LeCun et al. (1998); Kavukcuoglu et al. (2010)) is used in particular to automatically learn the multi-resolution features from images for object classification. Recent studies carried out in 2015 (Giering et al. (2014); Sarkar et al. (2014)) at United Technologies Research Center have shown that Deep Convolutional Neural Networks (DCNN) can be used for multi-modal sensor registration and occlusion detection in vision based tasks.

This paper proposes an architecture based on deep learning which monitors human activity in real time by fusing multiple EEG sensors in an unconstrained environment (without baseline) and also selects a smaller sensor suite (similar performance) for a lean data collection system. The main contributions of this paper are summarized as:

- real-time classification of cognitive activity from spatially non-collocated multi-probe EEG by applying deep learning techniques (DBN and DCNN) without any significant amount of data preprocessing and manual feature engineering,
- sensor selection for designing smaller EEG sensor suite, that is more usable at a typical PHM shop-floor environment, without degrading the performance,
- performance validation on extensive amount of data, which was collected from several subjects while they were performing multiple tasks (listening and watching) in PHM service training session,
- semantic validation of the trained deep network based on domain knowledge.
- comparison with machine learning benchmark techniques for multiple activity classification tasks.

The paper is organized in five sections, including the present

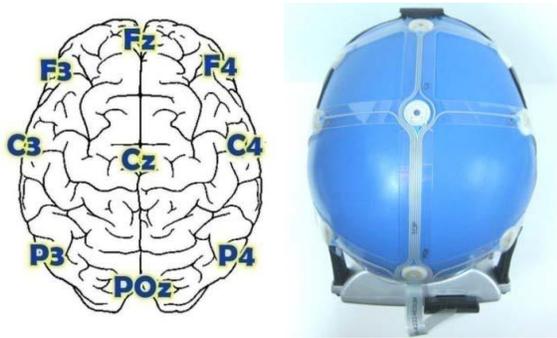


Figure 1. Position of the EEG sensors on the scalp

one. Section 2 describes the data collection process by a multi-probe (nine probes) EEG apparatus, which serves as a test apparatus for experimental validation of the proposed architecture for real-time human activity monitoring. Section 3 describes the proposed framework for task classification along with its building blocks via explaining the concepts of DBN and DCNN. Section 4 presents the capability and advantages of the proposed approach. Finally, the paper is summarized and concluded in Section 5 with selected recommendations for future research.

2. DATA COLLECTION

Extensive data collection was performed mainly in a repair and maintenance training scenario. The EEG sensors collect data generated from brain waves via sensors placed on the scalp. The wireless EEG sensor set used in this study was developed by Advanced Brain Monitoring (ABM). The system combines a 1.5 V battery-powered headset with a sensor placement system, following international standards (Jasper (1958)). The EEG sensor strip has 9 channels of EEG data collected from the following scalp sites, Fz, F3, F4, Cz, C3, C4, POz, P3, P4. Figure 1. below (left) shows the position of the sensor sites on the scalp and pictured (right) shows the sensor strip as it would be positioned on the participants head. A non-toxic electrolyte gel applied to the scalp improves the acquisition of the brain signal. This gel has a similar chemical make-up to sweat and wipes out of the hair with a tissue. The EEG system also has the capability of collecting heart rate data and inter-beat heart rate interval data. The system uses two sensors: 1 sensor placed on the right clavicle or collar bone and the other across the body on the first rib on the left side. The sensor operates using a 9V battery, and the data is transmitted to the main computer wirelessly.

To begin the experimental session, participants were fitted with the 9-channel, wireless EEG system consisting of a soft elastic band, a plastic strip containing the EEG sensors and a

transmitter box to stream the data to the laptop for collection. Before participating in the main tasks, a baseline of each participant's psychophysiological state was taken. An impedance check of all sensors was performed to ensure that proper contact with the scalp was made. When the sensors were settled and read under 20 mV, a set of baseline tasks (a 3-choice vigilance task, an eyes-open rest task, and an eyes-closed task) was performed to characterize the participant's brain patterns before participating in the experimental task. Although the baselines were collected for each participant as a part of the usual experimental procedure, those were not considered in the activity monitoring analysis for this paper. Once the baseline was completed, the participant was ready to enter the scenario. Participants were asked to complete several tasks, each lasting approximately five minutes. The tasks mostly centered on the repair and maintenance of a Roof-Top Unit (RTU) and are as follows: (i) Music task – an aural task during which participants close their eyes and listen to a selection of music, (ii) Audio Task – an aural task during which participants listened to the RTU manual read aloud to them by the experimenter (iii) Reading task – a visual task during which participants are asked to read pages from the RTU manual, and (iv) Video Task – a visual task during which participants view the use of the augmented reality RTU maintenance application on a tablet. Between tasks, participants moved to a different area of the room in order to separate the parts of the experiment. After each task, with the exception of the music task, participants were asked questions to encourage attention and engagement with the material. Completing each task once was considered completing one block. Also the tasks were performed in a random order. Each participant was intended to complete 4 blocks. In this paper, the four tasks are grouped into two major classes which are 'watching' (task (i) and (ii)) and 'listening' (task (iii) and (iv)). Limited by the current Institutional Review Board (IRB), a total of three people participated in the study. Two participants completed four blocks each. At the finish, each participant had generated 15 to 20 minutes of data per task. The sampling frequency of the data was 256Hz.

3. CLASSIFICATION FRAMEWORK

This section describes the proposed framework which analyzes multi-sensor EEG data for classification of human activity. Figure 2 shows the whole architecture in a modular fashion. Synchronized windows are traversed over EEG sensor time series with overlap and windows from each sensor are concatenated. Each of the sensor windows is denoised by a simple low-pass filter and major muscle movements are filtered out. After that, filtered window data are converted to power spectral density (PSD) via fast fourier transform (FFT). A large vector containing all normalized PSDs from each sensor are fed into the deep neural network. Specifically for DCNN, spectrogram is fed into the deep learning mod-

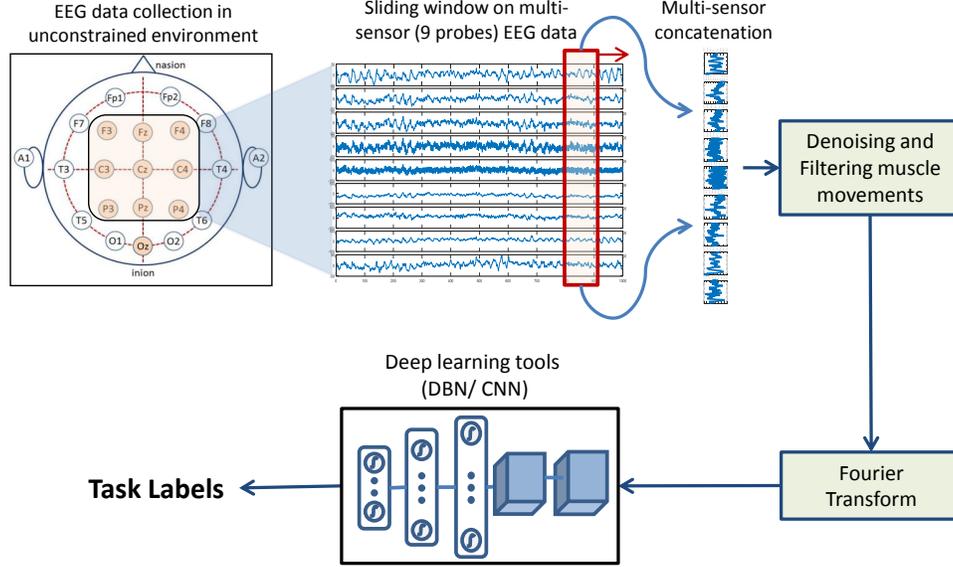


Figure 2. Classification Framework

ule. Deep neural networks (DBN and DCNN) are trained by backpropagation and stochastic gradient descent (Hinton & Salakhutdinov (2006)) based on the output task labels. While testing, the whole tool chain is operated in a feed-forward fashion in real-time and provides an output containing task decision for each window of EEG data. Detailed description of DBN and DCNN are provided as follows.

3.1. Deep Belief Network (DBN)

Deep Belief Networks (DBN) is a type of deep neural networks consisting of multiple layers hidden variables (Hinton & Salakhutdinov (2006)). As shown in figure 3, DBN architecture is built up by putting multiple layers of Restricted Boltzmann Machines (RBM) on top of one another. RBM is a type of energy-based generative probabilistic graphical model that learns a probability distribution over the input space to optimally explain the observed data. Structurally, RBM is a bipartite graph that connects visible units (the inputs) and hidden units except the same type of units. Due to the presence of latent variables, a single layer of RBM is powerful enough to represent complex distributions. The capacity to represent nonlinearity is further increased when multiple hidden layers are stacked on top of each other, with the outputs of lower layer becoming the input of adjacent higher layer.

Multi-layered Deep Belief Networks are notorious for optimization to be caught in poor local minima due to the large number of parameters in the model. However, it has been discovered that DBNs can be trained in an unsupervised manner to help initialize better weights as opposed to using randomized weights, leading to a superior generalization performance. Pretraining is performed in a greedy layer-wise man-

ner. The weights and biases of the first RBM stack is updated iteratively based on an unsupervised training criterion. After a user-defined stopping condition (e.g. maximum number of iterations), the parameters from this layer is fixed and the outputs of the layer (a new representation for the raw input) becomes the input of another layer for pretraining in a similar fashion. Essentially, the objective is to find the hidden unit features that are more common in the training inputs than in the random inputs, such that the pretrained weights may help to guide the parameters of that later towards better regions in the parameter space.

Consider a single RBM stack with hidden units \mathbf{h} . The probability of observing a sample \mathbf{v} is

$$\Pr(\mathbf{v}) = \frac{e^{-F(\mathbf{v})}}{\int e^{-F(\mathbf{v})}} \quad (1)$$

where

$$F(\mathbf{v}) = -\log \sum_h e^{-E(\mathbf{v}, \mathbf{h})} \quad (2)$$

and

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{b}^T \mathbf{v} - \mathbf{c}^T \mathbf{h} - \mathbf{h} \mathbf{W}^T \mathbf{v} \quad (3)$$

Pretraining seeks to find the set of parameters $\{\hat{\mathbf{W}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}\}$ (i.e., layer weights, visible unit biases and the hidden unit biases, respectively) that maximizes the expected log-likelihood of the training data \mathbf{V} . Thus, the optimization problem can be formally represented as:

$$\{\hat{\mathbf{W}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}\} = \arg \max_{\mathbf{W}, \mathbf{b}, \mathbf{c}} \mathbb{E} \left[\sum_{\mathbf{v} \in \mathbf{V}} \log \Pr(\mathbf{v}) \right] \quad (4)$$

and the problem is solved by stochastic gradient descent. Con-

vergence in learning is confirmed by the fact that each newly pretrained layer guarantees an increase on the lower-bound of the log-likelihood of the data, hence improving the model.

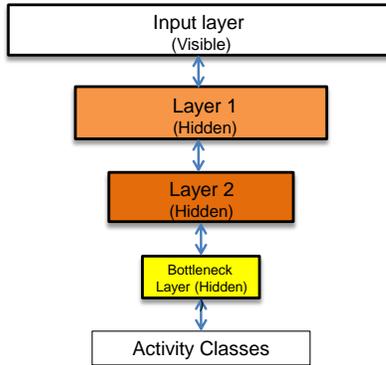


Figure 3. Deep belief network (DBN)

The pretrained network is finetuned using an error backpropagation algorithm. For a classification problem, the class labels are compared against the neural net outputs based on an input vector via an error metric that becomes the cost function of the algorithm (Larochelle & Bengio (2008)). Specifically, the loss function ℓ to be minimized for a dataset \mathbf{V} , parametrized by θ is:

$$\ell(\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{c}\}, \mathbf{V}) = - \sum_{i=0}^{|\mathbf{V}|} \left[\log \left(\Pr(Y = y^{(i)} | v^{(i)}, \mathbf{W}, \mathbf{b}, \mathbf{c}) \right) \right] \quad (5)$$

where $y^{(i)}$ denotes the class index. All weights and biases in the network are then optimized by the algorithm to produce a fully trained model for further classification.

3.2. Deep Convolutional Neural Network (DCNN)

Among various deep learning techniques, Deep Convolutional Neural Network (DCNN) (LeCun et al. (1998); Krizhevsky et al. (2012)) is an attractive option for extracting pertinent features from images in a hierarchical manner for detection, classification, and prediction. Regarding time series analysis, DCNN has been used for automatic speech recognition (ASR) (Abdel-Hamid et al. (2014)) where a spectrogram of the phenoms served as the input. In this paper, the spectrogram of the concatenated EEG sensors are used as input. DCNNs are also easier to train while achieving a comparable (and often better) performance despite the fact that there are fewer parameters relative to other fully connected networks with the same number of hidden layers. DCNN has fewer parameters because the filters share weights for a feature map (LeCun et al. (1998)).

In DCNNs, data is represented by multiple feature maps in each hidden layer as shown in the figure 4. Feature maps

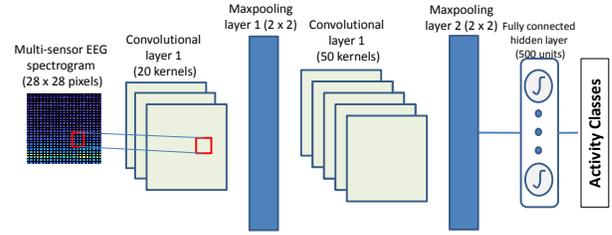


Figure 4. Deep Convolutional Neural Network (DCNN)

are obtained by convolving the input image by multiple filters in the corresponding hidden layer. To further reduce the dimension of the data, these feature maps typically undergo non-linear downsampling with a 2×2 or 3×3 max-pooling. Max-pooling essentially partitions the input image into sets of non-overlapping rectangles and takes the maximum value for each partition as the output. After max-pooling, multiple dimension-reduced vector representations of the input is acquired and the process is repeated in the next layer to learn a higher representation of the data. At the final pooling layer, resultant outputs are connected to the fully connected layer where sigmoid outputs from the hidden units are post-processed by a softmax function in order to predict the class that possesses the highest joint probability given the input data. This way, spectrogram structures of the EEG sensors at different tasks can be learned. For more detailed description on how DCNN works in general, refer to (LeCun et al. (1998); Krizhevsky et al. (2012)).

4. RESULTS AND DISCUSSIONS

This section discusses the training parameters and the performance results that are obtained when the proposed framework is applied on EEG data for activity classification.

4.1. DBN/DCNN Parameters for Training

The EEG data collected from first and third participants are used for constructing training set and remaining data from the second participant is used for testing. Each participant performed around 20 minutes of each of the 4 tasks. This generates around 40 minutes of data for each of the 'listening' and 'watching' activity classes. For training, a window length of 5 seconds with an overlap of 4.5 seconds is chosen to capture adequate slow time-scale dynamics along with fast time scale transience. This parameters are considered after a 2 dimensional grid search over window lengths and overlaps. For example, at smaller window size such as 2 seconds, activity recognition performance drops by more than 10%. Combining participant one and three, around 14000 windows of multi-sensor concatenated time series are produced, that are equally divided among classes of activities. Seventy percent of this data is used for training and remaining thirty percent is

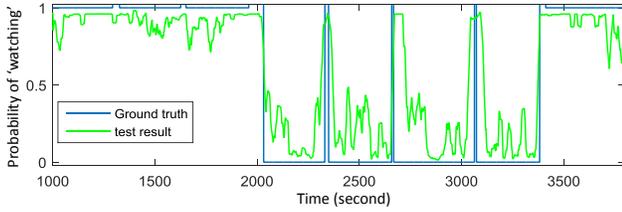


Figure 5. Real-time monitoring performance (ground truth: 0 - listening and 1 - watching)

used as the validation set for the deep learning models. After the training and validation set are constructed, they are denoised by ABM system (see section 2) via eliminating both the frequencies more than $120Hz$ and the frequencies mainly responsible for muscle movements. Each segment of the concatenated denoised signal, representing an EEG sensor, is converted to power spectral density (PSD) and normalized to a range of $[0, 1]$ for DBN. The PSD array for each sensor is 640 unit long because the sampling frequency is $256Hz$ and window length is 5 seconds. The length of the input vector of the DBN is 5760 (640×9) after concatenation of 9 sensors. The DBN is comprised of three layers with 4000, 1000, and 20 hidden units for the first, second, and third hidden layer respectively (see figure 3). A learning rate of 0.01 is used for the stochastic gradient descent algorithm while performing both pre-training and supervised finetuning. Pre-training is performed in batches of 50 samples and each layer undergoes 50 complete iterations of pre-training before moving onto the next layer. During supervised finetuning, classification errors on the validation data is compared against the errors from training set as a measure to prevent overfitting to the available data. For the two-class problem (classifying 'listening' and watching'), the optimized model is obtained around 200th iteration after which the validation error becomes consistently higher than the training error.

For DCNN, each segment of the concatenated denoised signal is converted to spectrogram and normalized to a range of $[0, 1]$. For each training window, the input concatenated spectrogram is resized to a greyscale image of size 28×28 pixels. In the first convolutional layer as shown in figure 4, 20 filters of size 5×5 pixels reduce the input image to 20 feature maps of 24×24 pixels. Next, the feature maps are downsampled by a 2×2 max-pooling layer, resulting in pooled maps of 12×12 pixels. Each of these maps goes through another convolutional layer with 50 filters of 5×5 pixels which produces feature maps of 8×8 pixels, and 4×4 pixels pooled maps after max-pooling. All generated maps are connected to the fully connected layer of 500 hidden units followed by class labels. A learning rate of 0.1 along with a batch size of 500 are used for stochastic gradient descent. For the two-class problem, the optimized model is obtained around 350th epochs under the similar criteria as DBN.

Table 1. Comparison of classification performances with two major classes of activities: 'listening' and 'watching'

Method	Test
	accuracy (%)
k-NN	74.08
SVM	86.61
DBN	91.15
DCNN	91.63

Table 2. Comparison of classification performances with all four tasks

Method	Test
	accuracy (%)
k-NN	38.94
SVM	43.55
DBN	48.76
DCNN	49.50

4.2. Testing and Performance Comparisons

The proposed framework is tested on the EEG data from the second participant. The overlap is reduced to 4 seconds while testing, such that the framework gives a decision regarding the ongoing task every second. A snapshot of the testing phase is shown in the figure 5, which shows the ground truth (0 - listening and 1 - watching) and the probability of 'watching'. If a simple threshold of 0.5 is chosen for the binary classification, a classification accuracy of 91.15% and 91.63% can be achieved by DBN and DCNN respectively.

The performance in binary classification from deep learning are compared to the state-of-the-art techniques such as k nearest neighbor (k-NN) classifier (Bishop (2006)) and support vector machine (SVM) (Bishop (2006)) as shown in the table 1. For fair comparison, k-NN and SVM are optimized for k (optimal $k = 5$) and kernel parameters respectively based on the training and validation set exactly same as the proposed deep learning framework. Table 1 shows that DCNN and DBN perform significantly better than SVM and k-NN. Although the 4-task classification task is a tough problem in this setting due to wide overlap of EEG features among tasks, the proposed framework is tested in classifying all the four tasks for completeness. Table 2 presents that both the DCNN and DBN perform better than k-NN with a margin of around 10%. As the data is limited, a three-way cross validation (different participant for testing) is performed and it is observed that the performance varies within 2% of the reported ones in the previous tables.

4.3. Feature Visualization and Sensor Selection

In the DBN, perfect class representations at the output layer i.e., $[1 \ 0]$ for listening and $[0 \ 1]$ for watching are backpropagated through the optimal network to visualize the representative inputs for two classes. Those representative inputs

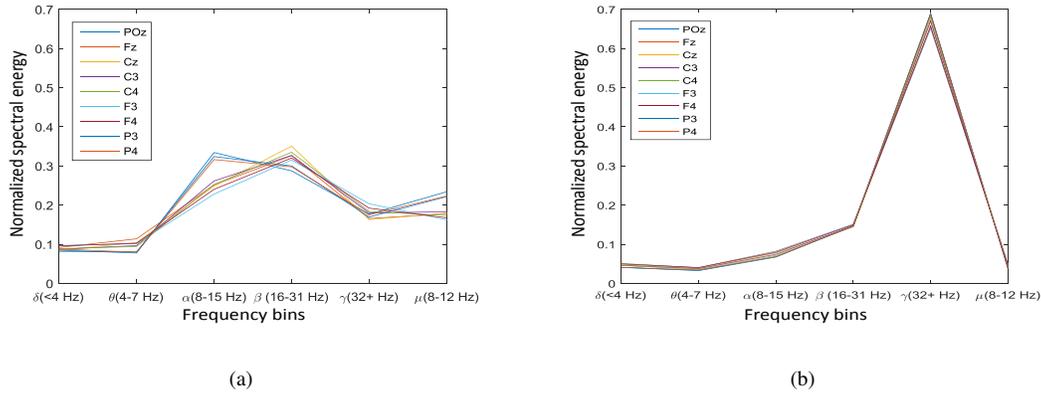


Figure 6. Spectral energy distribution among multiple frequency bins from trained neural network for (a) class 1: Listening, (b) class 2: Watching

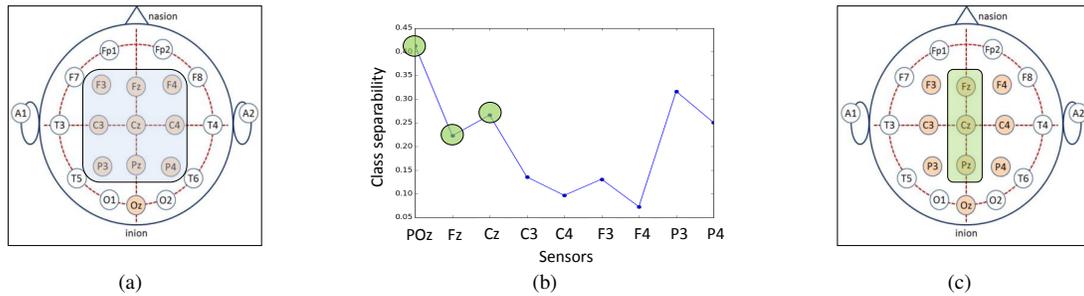


Figure 7. (a) All 9 EEG probes, (b) sensor selection, and (c) Selected EEG probes

are divided among multiple spectral bins, which are important to domain experts, to identify the validity of the network. A spatio-spectral distribution of normalized spectral energy over various frequency bins are presented by figure 6. Figure 6 shows that α dominance of P sensors while listening shifts to β zone during watching. Also the presence of dominant γ activity during watching strongly supports the domain knowledge regarding cross-modal (audio-visual) sensory processing and short-term memory matching of recognized objects (Kisley & Cornwell (2006)). These observations show that the DBN has learned a model without significant preprocessing, which has an adequate semantic meaning according to the domain experts (Kisley & Cornwell (2006)).

A backward feature selection procedure (Bishop (2006)) is carried out based on DBN. Figure 7(b) reveals that the P sensors (POz, P3, P4) in general and Fz and Cz contribute the most towards class separability. If a sensor suite is created containing Pz, Fz and Cz (see 7(c)) based on the class separability criterion, it still classifies two activities with an accuracy as high as 90.1%. The suite of Pz, Fz and Cz is selected instead of just P sensors to capture spatial variability. Hence

the 9-probe EEG sensor suite can be reduced to a smaller 3-probe (Pz, Fz and Cz) sensor suite for this type of activity classification problem while keeping the performance similar. This observation paves the possibility of making EEG sensor suite less clunky and more user-friendly.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a framework consisting of deep learning, that recognizes human activities in real time via the fusion of multiple EEG sensors in an unconstrained environment (without baseline) and selects a smaller sensor suite (similar performance) for a lean data collection system. Classification of human activity from spatially non-collocated multi-probe EEG by applying deep learning techniques is performed without any significant amount of data preprocessing. The ability to label data with the tasks would also enable attention and workload experiments to be performed in unconstrained (natural) and collaborative setting. Two major types of deep neural networks namely, deep belief network (DBN) and deep convolutional neural network (DCNN) are used in this paper at the core of the framework. It is observed that both

DBN and DCNN exhibit more than 91% accuracy at classifying activities such as 'listening' and 'watching'. Comparison with machine learning benchmark techniques reveals superior performance of the deep learning tools. Main advantages of the proposed framework include simple preprocessing, training without baseline and testing on unforeseen subject, which make this framework more generalizable to broader array of applications relevant to EEG data understanding. Sensor selection via backward feature selection shows the possibility of designing smaller EEG sensor suite while keeping the performance equivalent. Feature visualization and validation against domain knowledge regarding spectral energy distribution supports the fact that the deep networks learn semantic and useful features. Future work will attempt to validate this approach on more data collected from larger group of participants for finer classification of tasks along with quantifying the robustness. Future research will also investigate fusion among broader array of heterogeneous sensors such as heart monitors, galvanometers and eye trackers.

REFERENCES

- Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(10), 1533–1545.
- Alomari, M. H., Samaha, A., & AlKamha, K. (2013). Automated classification of l/r hand movement eeg signals using advanced feature extraction and machine learning. *arXiv preprint arXiv:1312.2877*.
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8), 1798–1828.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer, New York, NY, USA.
- Chambayil, B., Singla, R., & Jha, R. (2010). Eeg eye blink classification using neural network. In *Proceedings of the world congress on engineering* (Vol. 1, pp. 2–5).
- Coates, A., Ng, A. Y., & Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. In *International conference on artificial intelligence and statistics* (pp. 215–223).
- Delorme, A., & Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1), 9–21.
- Deng, L., & Dong, Y. (2014). Foundations and trends® in signal processing. *Signal Processing*, 7, 3–4.
- Giering, M., Reddy, K., & Venugopalan, V. (2014). Multi-modal sensor registration for vehicle perception via deep neural networks. Retrieved from <http://arxiv.org/abs/1412.7006>
- Güler, N. F., Übeyli, E. D., & Güler, İ. (2005). Recurrent neural networks employing lyapunov exponents for eeg signals classification. *Expert Systems with Applications*, 29(3), 506–514.
- Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504–507.
- Hoogenboom, N., Schoffelen, J.-M., Oostenveld, R., Parkes, L. M., & Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. *Neuroimage*, 29(3), 764–773.
- Iscan, Z., Dokur, Z., & Demiralp, T. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Systems with Applications*, 38(8), 10499–10505.
- Jasper, H. H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375. Retrieved from <http://ci.nii.ac.jp/naid/10020218106/>
- Kavukcuoglu, K., Sermanet, Y. L., P. Boureau, Gregor, K., Mathieu, M., & LeCun, Y. (2010). Learning convolutional feature hierarchies for visual recognition. In *Nips*.
- Kisley, M. A., & Cornwell, Z. M. (2006). Gamma and beta neural activity evoked during a sensory gating paradigm: Effects of auditory, somatosensory and cross-modal stimulation. *Clinical Neurophysiology*, 117(11), 2549–2563.
- Kottaimalai, R., Rajasekaran, M. P., Selvam, V., & Kannapiran, B. (2013). Eeg signal classification using principal component analysis with neural network in brain computer interface applications. In *Emerging trends in computing, communication and nanotechnology (ice-ccn), 2013 international conference on* (pp. 227–231).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Nips*.
- Lachaux, J.-P., Rodriguez, E., Martinerie, J., Varela, F. J., et al. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4), 194–208.
- Larochelle, H., & Bengio, Y. (2008). Classification using discriminative restricted boltzmann machines. In *Proceedings of the 25th international conference on machine learning* (pp. 536–543).
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Le Van Quyen, M., Foucher, J., Lachaux, J.-P., Rodriguez, E., Lutz, A., Martinerie, J., & Varela, F. J. (2001). Comparison of hilbert transform and wavelet methods for the analysis of neuronal synchrony. *Journal of neuroscience methods*, 111(2), 83–98.
- Morimoto, T., & Sketch, S. (n.d.). Classifying the brain's

motor activity via deep learning.

- Naderi, M. A., & Mahdavi-Nasab, H. (2010). Analysis and classification of eeg signals using spectral analysis and recurrent neural networks. In *Biomedical engineering (icbme), 2010 17th iranian conference of* (pp. 1–4).
- Niedermeyer, E., & da Silva, F. H. L. (2005). *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.
- Omerhodzic, I., Avdakovic, S., Nuhanovic, A., & Dizdarevic, K. (2013). Energy distribution of eeg signals: Eeg signal wavelet-neural network classifier. *arXiv preprint arXiv:1307.7897*.
- Patnaik, L. M., & Manyam, O. K. (2008). Epileptic eeg detection using neural networks and post-classification. *Computer methods and programs in biomedicine*, 91(2), 100–109.
- Peters, B. O., Pfurtscheller, G., & Flyvbjerg, H. (1998). Mining multi-channel eeg for its information content: an annotated method for a brain–computer interface. *Neural Networks*, 11(7), 1429–1433.
- Roach, B. J., & Mathalon, D. H. (2008). Event-related eeg time-frequency analysis: an overview of measures and an analysis of early gamma band phase locking in schizophrenia. *Schizophrenia bulletin*, 34(5), 907–926.
- Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., & Varela, F. J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*, 397(6718), 430–433.
- Sarkar, S., Venugopalan, V., Reddy, K., Giering, M., Ryde, J., & Jaitly, N. (2014). Occlusion edge detection in rgb-d frames using deep convolutional networks. Retrieved from <http://arxiv.org/abs/1412.7007>
- Spencer, K. M., Nestor, P. G., Perlmutter, R., Niznikiewicz, M. A., Klump, M. C., Frumin, M., ... McCarley, R. W. (2004). Neural synchrony indexes disordered perception and cognition in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 101(49), 17288–17293.
- Subasi, A., Alkan, A., Koklukaya, E., & Kiyimik, M. K. (2005). Wavelet neural network classification of eeg signals by using ar model with mle preprocessing. *Neural Networks*, 18(7), 985–997.
- Tsoi, A., So, D., & Sergejew, A. (1993). Classification of electroencephalogram using neural network. In *Neural information processing system* (Vol. 6, pp. 1–7).
- Turner, J., Page, A., Mohsenin, T., & Oates, T. (2014). Deep belief networks used on high resolution multichannel electroencephalography data for seizure detection. In *2014 aaii spring symposium series*.
- Vartak, A. A., Fidopiastis, C. M., Nicholson, D. M., Mikhael, W. B., & Schmorow, D. (2008). Cognitive state estimation

for adaptive learning systems using wearable physiological sensors. In *Biosignals (2)* (pp. 147–152).

BIOGRAPHIES



Soumalya Sarkar is a senior research scientist at United Technology Research Center (UTRC). He has received his B. Eng. Degree in Mechanical Engineering in 2010 from Jadavpur University, Kolkata, India. He received M.S. in Mechanical Engineering in 2013 and M.A. in Mathematics in 2014 from Penn State University. Dr. Sarkar received his Ph.D. in Mechanical Engineering from Penn State University in June, 2015. He has 5+ years of experience on Deep Learning, Anomaly Detection and Estimation, Sensor Fusion, Spatiotemporal Analysis, Situation Awareness, Statistical Signal Processing & Machine Learning and Probabilistic Graphical Models. He co-authored 30 peer-reviewed publications including 13 Journal papers, 2 book chapters and recently filed one US patent. Dr. Sarkar is a member of the American Society of Mechanical Engineers (ASME), Institute of Electrical and Electronics Engineers (IEEE) and PHM Society.



Kishore K. Reddy is a Research Scientist at United Technologies Research Center (UTRC). He is currently leading the Data Analytics initiative at UTRC primarily focusing on PHM applications using Deep Learning. Prior to UTRC, he worked in activity recognition to perform video surveillance from rooftop and aerial cameras. Over the past 10+ years he has worked in the areas of controls, computer vision, big data analytics, Human Machine Interaction (HMI) and machine learning (deep learning). He has co-authored more than 15 research articles with over 600 citations. Dr. Reddy has a Ph.D. in Electrical Engineering from the University of Central Florida, M.S. in Electronic Systems and Engineering Management from Fachhochschule Sdwestfalen, Germany and a bachelors degree in Electrical and Communications Engineering from JNTU, India. He interned at Kitware Inc., NY and Haute cole valaisanne, Switzerland.



Alex Dorgan is a Human Machine Interaction (HMI) Research Analyst at United Technologies Research Center (UTRC) in E. Hartford, CT. She works on multidisciplinary teams focusing on innovation and research in areas ranging from Data Analytics to HMI and designs for the intersection of humans and technology. She has been involved in the interaction design and information architecture of human-machine and human-computer interfaces, as well as with emerging interaction modalities such as virtual

reality, augmented reality, and wearable devices. She uses user interviews, wireframes and quick prototype techniques to as a means to communicate the user experience design process and has been involved in the collection of psychophysiological data for evaluation of those tools. Dorgan studied Art at Skidmore College in Saratoga Springs, NY.



Michael Giering is the Technical Fellow of Data Analytics and Machine Learning at United Technologies Research Center (UTRC) in E. Hartford, CT. In his previous role at Group Leader of the Decision Support & Machine Intelligence group he led the internally funded Service Technologies Initiative from 2013-2015, where multi-modal deep-learning methods have been developed for PHM, navigation and real-time human in the loop systems. He also heads UTRCs machine learning, human-machine interaction (HMI) and cloud analytics portfolio. Giering has more than eight years of experience at UTRC developing analytics systems such as Bayes Net building diagnostics and fleet diagnostic methods for aerospace. Previously, he was a member of the central research organization at Mars Inc. for 10 years working in the areas of machine learning, large scale data mining, currency and commodity hedging, and long-term weather forecasting.