



STUDIES ON CLASSIFICATION OF fMRI DATA USING DEEP LEARNING APPROACH

Nur Farahana Mohd Suhaimi, Zaw Zaw Htike and Nahrul Khair Alang Md Rashid

Department of Mechatronics Engineering, IIUM, Kuala Lumpur, Malaysia

E-Mail: farahana.ms@live.iium.edu.my

ABSTRACT

Brain as main server for entire human body is a complex composition. It is a challenging task to read and interpret the brain. Functional magnetic resonance imaging (fMRI) has become one of the means to do the task. fMRI is a non-invasive technique to measure brain activity of a human subject according to various stimuli. However, the fMRI datasets for each subject is huge and high-dimensional. For instance, the dataset has four dimensions for 3D images time series. Pre-processing and analysing using pattern recognition are insignificant for datasets with varied anatomical structures and dimensions. On the other hand, supervised learning or biomarker is employed to reduce the curse-of-dimensionality of fMRI datasets. Yet, the process is difficult and subjective to the labeled datasets. Therefore, a well-versed approach in signal processing, natural language processing (NLP) and object recognition, known as deep learning is seen to have higher standard than usual classification approach. Deep learning is the improved version of neural network with higher capability and accuracy. This paper aims to review the deep learning approach in fMRI classifications based on three studies on fMRI data classifications.

Keywords: fMRI, deep learning, classification and neural network.

INTRODUCTION

The brain is a complex composition. It is the main server for entire system of body. Thus, reading and interpreting what is inside the brain becomes a very challenging task. However, studying the brain has proved to be beneficial for disease prevention and detection, psychological treatment and behavioral understanding [1-4].

The functional magnetic resonance imaging abbreviated as fMRI is a non-invasive technique to measure brain activity [5]. It is routinely used to localize various brain states according to various stimuli such as finger tapping, image recognitions and reading. Countless investigations had been done to localize neural actions such as blood percentage in the neurons correlates to various stimuli as mentioned. On the other hand, the magnetic resonance imaging (MRI) is more to mapping the brain structure while fMRI is brain function mapping.

Specifically, fMRI measures haemodynamic changes which is the changes of blood flow. The measurement is known as blood-oxygen-level dependent (BOLD) due to haemodynamic response. It is detected by fMRI scanner that intrinsically has low signal to noise ratio (SNR) output. This is due to various activities influence including the main stimuli. For instance, while recognizing faces in images as main stimuli, brain would also process colors and shapes in the images. Thus, many brain activities will be detected by the BOLD measurement. In addition to low SNR, low temporal resolution of fMRI as compared to EEG requires good design of experiment. On the important note, fMRI is recorded by slices of brain map. Thus, large size of dataset is recorded for an individual subject such as three orientations (axial, sagittal and coronal) [6].

Recently, for analyzing the datasets of fMRI, classification technique has become the main approach

[7]. The technique is one of the machine learning components. This technique can be used to extract exciting new information from neuroimaging data, especially the fMRI datasets. In words, classification is the analogue of regression when the variable being predicted is discrete, rather than continuous. For instance, man and woman classification, Alzheimer's patient or control patient classification and moving and reading activity classification. With multiple subjects, classification technique is expected to give higher and more accurate result.

The fMRI dataset is very huge for each of individual subject. Three orientations and countable slices of brain images resulted in big data. Whole brain pre-processing would incur difficulties and harder to be analyzed compared to normal data capacity. Not only that, data acquisition is crucial process, as different orientation would produce dissimilar dimension for different subjects. This is due to anatomical structure of every subject's brain. In addition, acquired data shown a lot of variation for different subjects although with similar stimuli or activity. The huge data and dissimilar orientations of the brain incur problems in data classification.

One of dominant approaches in brain imaging is multi-voxel pattern analysis (MVPA) approach. This approach must be preceded by feature selection. It is difficult and subjective process for big data classification [8]. Supervised learning is seen to be one of the problem in fMRI analysis as researcher has to identify the correct input and output beforehand. Furthermore, the parameter optimization step might reduce the accuracy of the approach.

In this study, we propose to review the previous researches on classification of fMRI datasets using various types of deep learning. The methods are studied to give particular ideas for practical usage and improvements.



CONVENTIONAL METHODS

Brain mapping is the association of perceptual or cognitive states with specific patterns of brain activity and it is central to human neuroscience [9]. There are single subject, multi subjects, single stimulus and multi stimulus approaches in brain mapping. Conventional brain mapping is univariate that treat each voxel as an independent measurement. General linear model (GLM) is one of conventional methods that utilizes parametric statistical analysis. The analysis aims to categorize brain regions that demonstrate significant signal changes in response to the experimental conditions [10]. The GLM is a massive univariate testing, i.e. one statistic per voxel. The method introduces multiple comparison for same test that relatively increases the processing time.

Other conventional technique is event-related fMRI responses [11]. This technique measures the temporal brain response on related event or stimulus. It introduces experimental design to relate the timing of events to the acquisition of data. As a result, a temporal resolution is acquired. The temporal resolution is important in that approach for dimension reduction step.

CURRENT DEVELOPMENTS

Later on, researchers are using many stimuli approach for an individual subject. The term that is normally used is multivariate and interchangeably with multi-voxel. The stimuli can be distinct or similar to each other. For instance, hand tapping and speaking is distinct to each other, but both activities need muscle movements. Thus, brain mapping classification is essential and introduced in this field.

One that is particularly famous and widely used method for brain mapping classification is MVPA. It employs simple classifier such as linear regression method. And many other researchers are using pattern analysis for mapping the brain, human's or animal's [12-14]. However, high inter-subject variability in brain function becomes a big challenge for researchers. Furthermore, the problem lies in choosing the best classifier.

Later, new approach to encounter the inter-subject variability by supervised learning and classification is tackled. Under the same multi-voxel pattern analysis method, the new approach is to train by feeding samples of input-output pairs. Thus, when tested, the characteristics consistency existence could be classified and labeled. However, varied orientations and dimensions incur another problem, due to difficult acquisition of input-output samples.

There are wide range of developed classifications algorithm. Varied performance has given a platform for practical testing and improvement. Naive Bayes is one of outstanding and robust classifier. With the probability as backbone, the discrete identification is adapted to the Naive Bayes. However, it is only applicable for linear problem. The fMRI datasets are susceptible for non-linearity classification.

Currently developing algorithm, support vector machine (SVM) constructs a separating hyperplane such that the distance from the hyperplane to the nearest data point is maximized [15]. This method adopts Naive Bayes approach with some alteration for non-linearity data adaption [16]. In early development, only two-class is expected to be classified by conventional SVM.

However, there are many approaches for improvement of SVM technique to suite multi-class classification. The primary approach is by multiplying binary classification to be multi-class. All in all, this classification does not cater high dimensionality datasets of fMRI. Dimensionality reduction is introduced but suffers generalization in fMRI classification. Figure-1 shows the general idea of current approach development.

DEEP LEARNING

To address the high dimensional and non-linearity of fMRI datasets, deep learning approach is introduced. Deep learning is a branch of machine learning. It is a wider and longer hidden layers of neural network (NN). In contrast to ANN, deep learning uses low-level abstraction and gradually increases to higher level until the feature is classified.



Figure-1. Current development approach such as Naive Bayes and SVM. The dataset is divided for training and testing data. Then, features extraction is done on the training data. Next, the feature is fed back to training set when not selected. While; if features selection is completed, chosen classification approach is employed, such as Naive Bayes or SVM. Later, the performance is evaluated.

For instance, the first low-level abstraction are all the pixels available in the images. And the next abstraction would be edges. Further down local shapes would be next abstraction in images. However, in practice, the right level is unknown due to automated discovery of abstraction [17]. It has adopted unsupervised learning architecture. This is because; the approach is free of input-output samples. Many researches are tweaking the initialization and architecture of deep learning. Figure-2 shows the simple explanation of deep learning for image recognition.

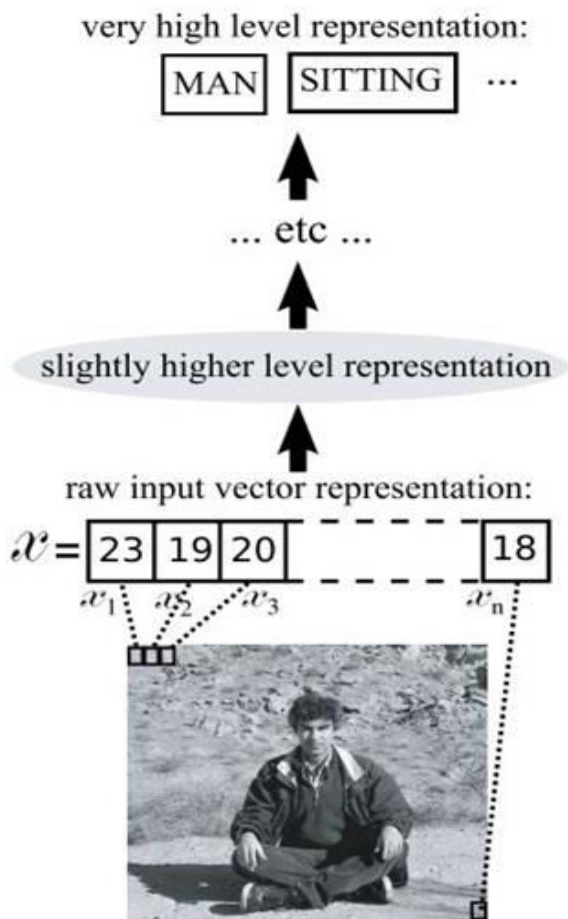


Figure-2. Example of abstraction in deep learning. Extracted from [18].

Common deep learning network architectures include deep belief network (DBN) [19], autoencoders [20], and convolutional neural network (ConvNet) [21]. DBN consists of restricted Boltzmann Machine (RBM). It is a Markov random field that models data distribution and parameterizing it with the Gibbs distribution. While auto-encoder seems easier to train rather than RBM, as each level can be trained separately. DBN and autoencoder adopt bi-directional neural network architecture. On the other hand, ConvNet as introduced in [22] is feed-forward deep neural network that can be trained purely supervised.

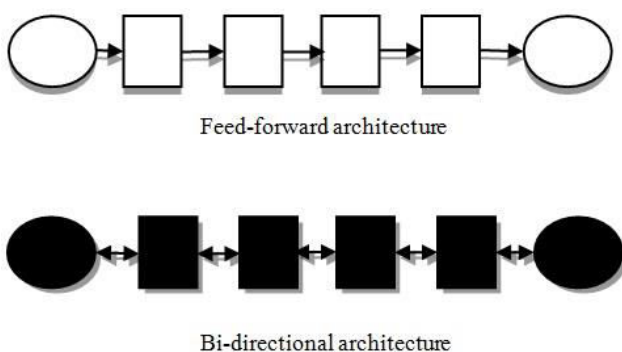


Figure-3. Common deep learning architectures, feed-forward and bi-directional.

For instance, the autoencoder main objective is to encode the input.

$$X_{1:n} = \{x_1, x_2, \dots, x_n\} \quad (1)$$

into some representation of features, $c(X)$ [18]. Using negative log-likelihood minimization, the mean square error criterion is

$$RE = -\log P(X|c(X)). \quad (2)$$

However, the normal squared error is employed if $c(X)$ is Gaussian. The loss function is

$$-\log P(X|c(X)) = -\sum_i^n x_i \log f_i(c(X)) + (1 - x_i) \log (1 - f_i(c(X))) \quad (3)$$

when the inputs, $X_{i:n}$ are either binary or considered to be binomial probabilities. The $f(\cdot)$ is the decoder of the algorithm while $f(c(X))$ is the reconstruction produced by the network.

In adaptation to fMRI data, each visible variable data represents a voxel in fMRI with a real-valued and approximate of Gaussian distribution [8]. This coding and encoding is one of the main approaches in many other areas including neuroscience. Research in this area of imaging attempts to make better representations with the foundation of neural networks and computational capabilities nowadays. Furthermore, many attempting to create models that learn these representations from large-scale unlabeled data.

CASE STUDIES DISCUSSION

Case study 1: Classification on ADHD using deep learning [23]

ADHD is attention deficit hyperactivity disorder which is one of the common mental disorder among children. Poor concentration and excessive activity are among the main symptoms of children with ADHD. Development in neuroscience has helped research in mental disorder department. For ADHD, there exist abnormality in fMRI compared to the healthy. This research has employed deep learning to classify the four distinct subjects; control, combined, inattentive and hyperactive.

Authors employed three important steps; preprocessing, dimension reduction and deep belief network (DBN) classification approach on the datasets. The fundamental step when dealing with fMRI datasets is preprocessing where the main processes are alignment, slice-time correction, normalization, and spatial smoothing. Later on, the authors applied brodmann template to reduce the dimension and convert the data to frequency domain. The max-pooling frequencies is done in assumption that active voxel has higher frequency. The DBN with three hidden layers is chosen for training and testing the frequency domain-datasets. Every two training



neighboring layers is used for backpropagation for tuning the model.

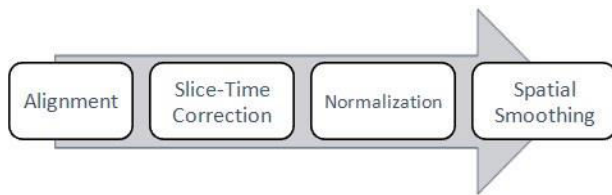


Figure-4. fMRI datasets preprocessing using Statistical parametric mapping software (SPM8) [24].

This classification technique of deep learning approach uses ADHD-200 datasets for three categories. The authors claim to have higher accuracy of 2.2% as compared to ADHD-200 researchers. However, it is observed that the higher dataset category has higher accuracy compared to less datasets. The hypothesis is; the more data trained the higher accuracy and effectiveness of deep learning approach.

Case study 2: Memory encoding and decoding with deep learning [25]

The study aims to obtain fMRI modelling using unsupervised learning representation for memory encoding and decoding based on ten semantic categories. Small subset of measurements, which is 100 and 1024 voxels, are studied. Dimensions of the feature and accuracy of the method become the benchmark for this study.

The method includes autoencoder, one type of deep learning architecture that stacked together and expected to be more efficient compared to MVPA technique. As one of the neural network type, the autoencoder targets to learn two model parameters, $\Theta = \{\theta_1, \theta_2\}$ and two non-linear consecutive feature mapping functions. The functions are used to classify the tasks or feed the input layer to another autoencoder. Authors applied gradient descent to optimize the model parameters.

Results shown that the method known as non-linear feature mapping (NLFM) has higher accuracy when more hidden layers and numbers of voxels employed. Though the baseline MVPA technique has the same accuracy for single hidden layer of NLFM, the feature dimensions are lower than MVPA. In addition, this type of deep learning method gives higher accuracy when higher voxels are used that conveniently proves the previous hypothesis.

Case study 3: Brain parcellation using deep learning [26]

Brain parcellation is the approach of defining the brain based on its functional connectivity. The gap in this area is the lack of standard model protocol due to assumption of similarity in subjects' anatomy. The author aims to develop the ideal protocol of brain parcellation using deep learning. The approach benefited the stacked autoencoder type of deep learning architecture.

Author mentioned the need to pre-training each layer to reduce computation cost despite the conventional approach of deep learning where random initialization is employed. Similar to [25], author used Limited memory BFGS (L-BFGS) optimization algorithm to optimize the parameters. Accordingly, fine-tuning is applied on entire network to improve the performance of stacked autoencoder.

Study shown that four hidden layers has performed better than five hidden layers. Overfitting and less accuracy might occur for five hidden layers. Though the achievement of parcellation is less accurate, nevertheless, the approach enhances the quality of experimental investigations due to no subjects' anatomical dependencies. Thus, the assumption for conservative protocol is avoided. The author suggested more fMRI datasets for higher accuracy.

CONCLUSIONS

There are many types of deep learning architectures. Those reviewed study cases are published in the same year, which is 2014, despite their different background and fMRI datasets. These studies are few of major breakthroughs in fMRI data classification using deep learning approach.

In spite of that, the stacked autoencoder deep learning is proposed in the two out of three studies. The approach is reliable specifically for fMRI datasets due to very few labelled data in neuroimaging field. Thus, an unsupervised learning is favored in this field. Those studies also suggested that more datasets would suggest better classifications and accuracy. Thus far, deep learning approach is developing in many ways and promise more in the future.

REFERENCES

- [1] Dawson G. 2008. Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, Vol. 20, No. 3, pp. 775-803.
- [2] Michel V., Gramfort A., Varoquaux G., Eger E., Keribin C. and Thirion B. 2012. A supervised clustering approach for fMRI-based inference of brain states. *Pattern Recognition*, Vol. 45, No. 6, pp. 2041-2049.
- [3] Kandel E. R. 2014. A new intellectual framework for psychiatry.
- [4] Martin S., Brunner P., Holdgraf C., Heinze H.-J., Crone N. E., Rieger J. and Pasley B. N. 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuro Engineering*, 7: 14.
- [5] Ashby F. G. 2011. *Statistical analysis of fMRI data*. MIT press.



- [6] Suma H. N. and Murali S. 2010. Pattern Classification and Analysis of Brain Maps through fMRI data with Multiple Methods. *International Journal of Computer Applications*, Vol. 1, No. 27, p. 117.
- [7] Pereira F., Mitchell T. and Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, Vol. 45, No. 1, pp. S199-S209.
- [8] Plis S. M., Hjelm D. R., Salakhutdinov R., Allen E. A., Bockholt H. J., Long J. D. and Calhoun V. D. 2014. Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience*, p.8.
- [9] Cabral C., Silveira M. and Figueiredo P. 2012. Decoding visual brain states from fMRI using an ensemble of classifiers. *Pattern Recognition*, Vol. 45, No. 6, p. 2064-2074.
- [10] Henson R. N. A. 2003. Analysis of fMRI time series: Linear time-invariant models, event-related fMRI and optimal experimental design.
- [11] Josephs O., Turner R. and Friston K. 1997. Event-related fMRI. *Human brain mapping*, Vol. 5, No. 4, pp. 243-248.
- [12] De Martino F., Valente G., Staeren N., Ashburner J., Goebel R. and Formisano E. 2008. Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, Vol. 43, No. 1, pp. 44-58.
- [13] Carp J., Park J., Polk T. A. and Park D. C. 2011. Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. *Neuroimage*, Vol. 56, No. 2, pp. 736-743.
- [14] Coutanche M. N., Thompson-Schill S. L. and Schultz R. T. 2011. Multi-voxel pattern analysis of fMRI data predicts clinical symptom severity. *Neuroimage*, Vol. 57, No. 1, pp. 113-123.
- [15] Cortes C. and Vapnik V. 1995. Support-vector networks. *Machine learning*, Vol. 20, No. 3, pp. 273-297.
- [16] LaConte S., Strother S., Cherkassky V., Anderson J., and Hu X. 2005. Support vector machines for temporal classification of block design fMRI data. *NeuroImage*, Vol. 26, No. 2, pp. 317-329.
- [17] Bengio Y. 2013. Deep learning of representations: Looking forward. In *Statistical Language and Speech Processing*. Springer Berlin Heidelberg. pp. 1-37.
- [18] Bengio Y. 2009. Learning deep architectures for AI. *Foundations and trends in Machine Learning*, Vol. 2, No. 1, pp. 1-127.
- [19] Hinton G. E. and Salakhutdinov R. R. 2006. Reducing the dimensionality of data with neural networks. *Science*, vol. 313, No. 5786, pp. 504-507.
- [20] Bengio Y., Lamblin P., Popovici D. and Larochelle H. 2007. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19, 153.
- [21] LeCun Y., Kavukcuoglu K. and Farabet C. 2010. Convolutional networks and applications in vision. In *Circuits and Systems (ISCAS)*, Proceedings of 2010 IEEE International Symposium on. IEEE. pp. 253-256.
- [22] LeCun, Y., & Ranzato, M. 2013. Deep learning tutorial. In *Tutorials in International Conference on Machine Learning (ICML'13)*.
- [23] Kuang, D., & He, L. 2014. Classification on ADHD with Deep Learning. In *Cloud Computing and Big Data (CCBD)*, 2014 International Conference on. IEEE. pp. 27-32.
- [24] Statistical parametric mapping, version 8. <http://www.fil.ion.ucl.ac.uk/spm/>, 2009.
- [25] Firat, O., Oztekin, L., & Yarman Vural, F. T. 2014. Deep learning for brain decoding. In *Image Processing (ICIP)*, 2014 IEEE International Conference on. IEEE. pp. 2784-2788.
- [26] Gravelines, C. 2014. Deep Learning via Stacked Sparse Autoencoders for Automated Voxel-Wise Brain Parcellation Based on Functional Connectivity.