INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Classification of Attention Deficit Hyperactivity Disorder using Variational Autoencoder

Azurah A Samah[1]*, Siti Nurul Aqilah Ahmad[2], Hairudin Abdul Majid[3], Zuraini Ali Shah[4], Haslina Hashim[5], Nuraina Syaza Azman[6], Nur Sabrina Azmi[7], Dewi Nasien[8]
Artificial Intelligence and Bioinformatics Group (AIBIG)[1-7], Department of Informatic Engineering[8],
School of Computing[1-7],
Faculty of Engineering[1-7], Faculty of Science Computer[8],
Universiti Teknologi Malaysia[1-7], Institut Bisnis dan Teknologi Pelita Pekanbaru[8]
81310 UTM Johor Bahru, Johor, Malaysia[1-7], Pekanbaru, Indonesia[8]
Email: azurah@utm.my[1]*, aqilahahmad308@gmail.com[2], hairudin@utm.my[3], aszuraini@utm.my[4], haslinah@utm.my[5], nsyaza7@graduate.utm.my[6], nsabrina36@graduate.utm.my[7], dewinasien@lecturer.pelitaindonesia.ac.id[8]

Submitted: 16/9/2021. Revised edition: 25/10/2021. Accepted: 26/10/2021. Published online: 15/11/2021
DOI: https://doi.org/10.11113/ijic.v11n2.352

***Abstract*—Attention Deficit Hyperactivity Disorder (ADHD) categorize as one of the typical neurodevelopmental and mental disorders. Over the years, researchers have identified ADHD as a complicated disorder since it is not directly tested with a standard medical test such as a blood or urine test on the early-stage diagnosis. Apart from the physical symptoms of ADHD, clinical data of ADHD patients show that most of them have learning problems. Therefore, functional Magnetic Resonance Imaging (fMRI) is considered the most suitable method to determine functional activity in the brain region to understand brain disorders of ADHD. One of the ways to diagnose ADHD is by using deep learning techniques, which can increase the accuracy of predicting ADHD using the fMRI dataset. Past attempts of classifying ADHD based on functional connectivity coefficient using the Deep Neural Network (DNN) result in 95% accuracy. As Variational Autoencoder (VAE) is the most popular in extracting high-level data, this model is applied in this study. This study aims to enhance the performance of VAE to increase the accuracy in classifying ADHD using fMRI data based on functional connectivity analysis. The preprocessed fMRI dataset is used for decomposition to find the region of interest (ROI), followed by Independent Component Analysis (ICA) that calculates the correlation between brain regions and creates functional connectivity matrices for each subject. As a result, the VAE model achieved an accuracy of 75% on classifying ADHD.**

***Keywords*—Functional Magnetic Resonance Imaging (fMRI), Variational Autoencoder (VAE), Attention Deficit Hyperactivity Disorder (ADHD), Independent Component Analysis (ICA), Nilearn**

I. INTRODUCTION

The human brain is the direct center of the human nervous system. It can be viewed as a large and complex network for monitoring and controlling all systems in the body, including supervising and executing different bodily functions. All brain regions are interlinked and form one complex integrative system. The brain network consists of multiple interconnected brain regions, and each of the brain regions coordinates simultaneously, leading to a complex brain connectivity pattern. Many imaging methods have been proposed to explore the functionality of the brain, including Magnetoencephalography (MEG), Electroencephalography (EEG), Positron Emission Tomography (PET) and fMRI. The fMRI is considered the most suitable method to determine the functional activity of brain regions because it is non-invasive and displays remarkable resolution.

ADHD could be a common psychological state disorder and noticeable by an eternal pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, as stated by the National Institute of Mental Health (NIH). Scientists have shown that the brains of children with ADHD differ from normal children and some of those differences vary from childhood ages with matures. The brain is categorized into lobes: the frontal lobe, parietal lobe, temporal lobe, and occipital lobe. The most focused lobe for ADHD disorder is the frontal lobe and it is located at the front of the brain, behind the forehead, which functions in helping people to prepare, plan, listen and make decisions.

II. Problem Background

The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) of classification system and criteria were identified through the advancements of neuroscience and clinical health needs. Issues have been recognized by ongoing research identified with the indicative clinical standards delineated in the DSM-IV that, at present, stay uncertain. However, the intimation of a diagnostic threshold for the findings is not compatible such that it appears to delegate "disordered" from "non disordered" individuals. Accordingly, when it is reliant on categorical diagnose, the realities that among disordered people frequently have clear varieties in seriousness are disregarded and people that beneath the limit for a disorder have clinically impeding degrees of side effects (Frick *et al*., 2015). Thus, there is a need for general psychiatric classification and the DSM to begin moving towards incorporating dimensional approaches to the classification (Frick *et al*., 2015).

Past attempts to classify ADHD based on functional connectivity coefficient using the DNN model and the experiment resulted in 95% accuracy. In this study, a VAE model is proposed to measure the accuracy of classification of ADHD from the results of neuroimaging of ADHD-200 dataset from the patients using VAE. This method does not directly use the neuroimaging scans but focuses on the vectorized indicators of functional connections. Next, ICA is used to extract functional connectivity coefficients, train and test on VAE.

Clinical results of diagnosed patients with ADHD are often made subjectively based on doctors' knowledge and skills. This practice applies when doctors tend to predict the result based on the behavior of the patients. This practice ends up in unwanted biases, errors and excessive medical prices that affect the standard of service provided to patients. Therefore, an automatic diagnosis system is designed to take advantage of the collected database in fMRI neuroimaging. An objective method that will take out the image of the patient's brain can help diagnose ADHD. Deep learning such as VAE can develop a knowledge-rich environment on image input, improving the standard of other clinical choices.

The rest of the paper is divided into several sections. Section 3 provides a brief overview of the previous works. Section 4 explains the materials and methods used in the proposed classification model. Section 5 contains results and discussion. The final remarks can be found in section 6.

III. Previous Works

ADHD is categorized as a common neurodevelopmental and mental disorder in children, affecting 5-10% of children, life impairment contribution, low quality of life, and continuity of burden to the affected families (Riaz *et al.*, 2018). ADHD can be seen as a complex disorder that has many symptoms that develop from childhood to adulthood. Therefore, there is no simple way to find a correct solution for this disorder (Sigh *et al.*, 2018). The symptoms of ADHD can be either excessive impulsive, hyperactivity or inattention behaviors which can develop at early childhood age and may carry on until infancy age. Those symptoms bring severe defects and activate financial burdens to society and families (Zou *et al.*, 2017).

fMRI is viewed as generally appropriate towards deciding utilitarian action of the cerebrum areas as it is non-intrusive and shows striking spatial goals (Riaz *et al.*, 2018). Lately, useful availability has been demonstrated to be a significant biomarker towards the segregation of various cerebrum issues. The examination considers that mental issues, for instance, ADHD, can change the useful availability of the cerebrum arrange. The precise ID of the modified practical network instigated by a specific problem is viewed as a significant undertaking that may feature the hidden components of the turmoil. Resting-state fMRI was developed as a promising neuroimaging apparatus to examine the practical action of mind districts. Specifically, fMRI has been utilized to recognize the network changes instigated by issues for ADHD (Riaz *et al.*, 2018).

In addition, fMRI has been widely used in many recent kinds of research to explore brain functionality and connectivity for the classification of neurological disorders. Based on the past studies, there are several neural network method that has been used in classifying ADHD includes Convolutional Neural Network (CNN), Deep Belief Network (DBN) and Recurrent Neural Network (RNN), DNN and VAE. Table I shows the comparison among various artificial neural network (ANN) methods that have been done in previous studies. The VAE is chosen to classify ADHD based on this comparison by considering their learning capability and the amount of data executed. This paper used a standardized ICA analysis on fMRI data, and the functional connectivity coefficients between the regions were extracted based on their connectivity parameters and implemented into VAE classification (Chauhan and Choi, 2020).

TABLE I. Comparison of different ANN model

Method	Author	Advantages	Disadvantages
CNN	(Munoz-Organero, Powell, Heller, Harpin, & Parker, 2018)	The target function can be easily predicted to get a better representation by increasing the depth of this CNN network.	Easy to get overfitting and hard to optimize when the complexity of the network increases.
DBN	(Farzi, Kianian, & Rastkhadive, 2017)	The DBN is learned by the greedy training algorithm on the training data, evaluated on the test data.	Imbalance of data.
RNN	(Munoz-Organero, Powell, Heller, Harpin, & Parker, 2019)	Minimized the reconstruction error when predicting the following 2-second acceleration data for the 4 sensors based on the past 6 seconds of data.	Segments of data will not be characterized adequately by the RNN.
DNN	(Chauhan & Choi, 2020)	Reach 95% accuracy with smaller data and due to the four-layer of sequence model.	The low number of layers was not suggested to use in this module because it deals with functions that are not necessarily linearly separable.

Method	Author	Advantages	Disadvantages
VAE	(Hu, Pei, Jia, & Zhao, 2019)	The VAE model can find nonlinear descriptive features through nonlinear functions by implementing the variational approach for latent representation learning, which uses the probability distribution technique.	Rarely used in training small sample size in the dataset, which can affect the results.

IV. Materials and Method

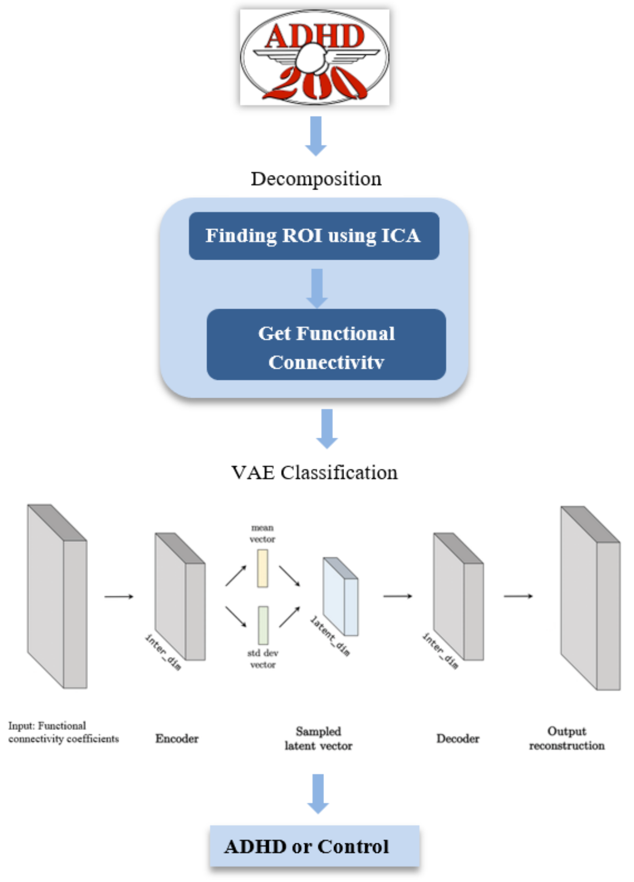

Fig. 1. ADHD classification framework

This study proposed a framework for classifying ADHD patients and control groups that have healthy brains. The structure of this study is illustrated in Fig. 1. Firstly, all the preprocessed fMRI data are decomposed by ICA, and this provides a unique representation of the data and activity maps. Then, a connection matrix of ADHD and controls was created to visualize the functional connectivity. Lastly, the feature vectors are gathered from the extraction of functional connectivity coefficients to train and test the VAE model for final classification.

A. Dataset

TABLE II. ADHD-200 DATA FEATURES

Type	Sample Size
ADHD	20
Typically Developing (TD)	20

In this research, the publicly accessible data of ADHD-200 was used, which tranquil resting-state fMRI (rs-fMRI) and anatomical data from 8 independent institutions. The data were preprocessed by several neuroimaging tools and grouped into three. The first group is called Athena pipeline which uses tools from AFNI and FSL packages. The second group is the NIAK pipeline which uses a neuroimaging analysis kit from CBRAIN software. The third group was the Burner pipeline, which carried out vowel-based morphometry processing using the SPM8 tool. 40 fMRI data are stored in the Python module for neuroimaging data, where the dataset includes 20 typically developed people ("controls") and 20 people diagnosed with ADHD ("treatments"), as stated in Table II (Chauhan and Choi, 2020). There are four types of information included; paths to rs-fMRI data images (fumc), CSV files containing nuisance variables (confounds), preprocessing steps details (phenotypic), and data description (description).

B. Decomposition

Decomposition is the first step in extracting functional connection coefficients, which ICA is widely used to analyze fMRI data. It is vital to determine an appropriate ROI to assess connectivity. Nilearn has a group-level ICA feature called CanICA, which allows a single entity to control the variability associated with the functional network. According to the recommendations of the Nilean document, 20 components were selected to decompose and evaluate the independent components (Hu *et al*., 2019). These independent components use statistical atlas to represent the default mode network (DMN) in the standard anatomical image of the brain. The common components generated by the decomposition are displayed by the probability atlas on the anatomical brain image, as shown in Fig. 2.

However, decomposition is still challenging to understand because different areas of the brain are observed. We summarized the brain signals received from ICA and converted the extracted data into time series to solve this problem. Next, correlation is used to determine the coefficient of functional connectivity, which can be classified as more accurate than other connectivity parameters (tangent, partial correlation) (Hu *et al*., 2019).

Correlation is simply the calculation of the marginal connectivity between pairs of ROIs. Nilearn provides a built-in connectivity measures function to calculate the correlation matrix, which requires two parameters: function connection type data (correlation) and the time series obtained in the previous step. This makes it possible to calculate the relationship matrix for all topics. Fig. 3 shows the average connection matrix of all participants. The diagonal lines in Fig. 3 can be ignored because they show the correlation between each other. We can represent the functional connection coefficient matrix on a predefined brain image (Fig. 4).

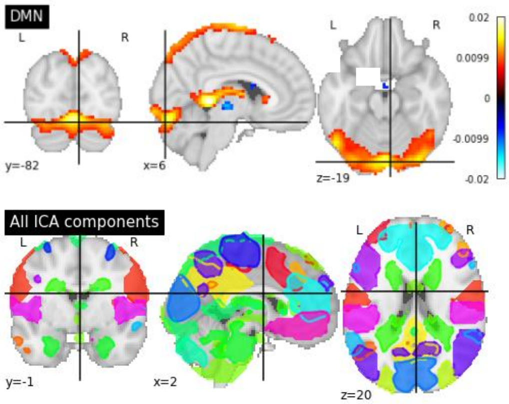

Fig. 2. ICA decomposition

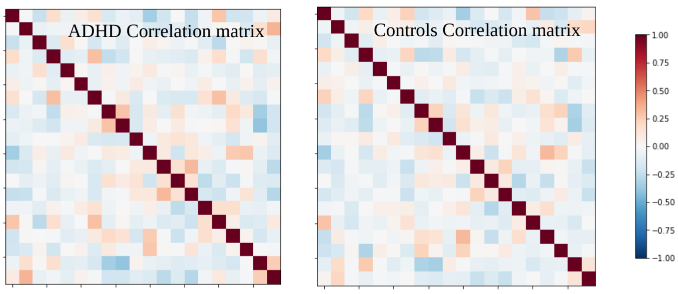

Fig. 3. Functional connectivity coefficient matrices based on correlation

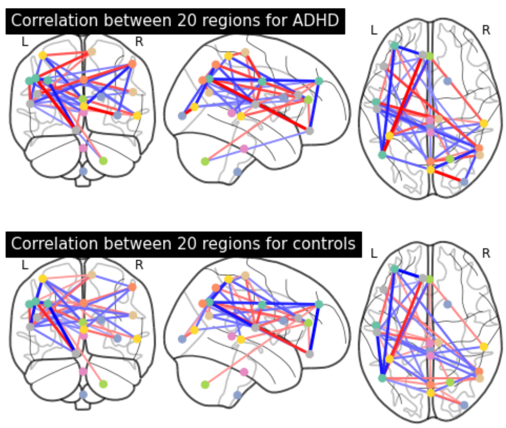

Fig. 4. Visualization of functional connectivity coefficients

Compared with the control group, ADHD connections seem to be less tight, indicating that ADHD-induced functional connections are reduced, while there are fewer connections in the upper parietal cortex (circles) responsible for attention, which appear to be damaged by ADHD (Fig. 4). The collected correlation matrix provides a vectorized measure of functional connectivity and provides a classification input for the VAE model.

C. Variational Autoencoder Classification

The 3 convolutional levels in the encoder layers with 3 × 3 kernels, and the stride is fixed at 2 to achieve spatial subsampling instead of deterministic spatial functions like max-pooling. Each convolutional network is followed by a batch normalization level and an activation level. Then the two fully connected output layers (for mean and variance) are added to the encoder and used to calculate the Kullback-Leibler (KL) divergence loss and the sample latent variable.

A three-layer autoencoder was used for this classification with a specific number of nodes for encoder (32,16,16) and decoder (16,1). The output node in the decoder was 1 because it is a binary classification, and it is three layers due to deal with functions that were not necessarily linearly separable (Chauhan and Choi, 2020). Then, the length of batch size was set to 32 because of the 40 subjects of ADHD-200 data that has been set up. Fig. 5 above shows the VAE model architecture. As demonstrated in Fig. 6, the suggested deep learning method in classifying ADHD can be promoted into 3 phases: data splitting, training and testing of the model, and validation of informative, functional connectivity through VAE classification.

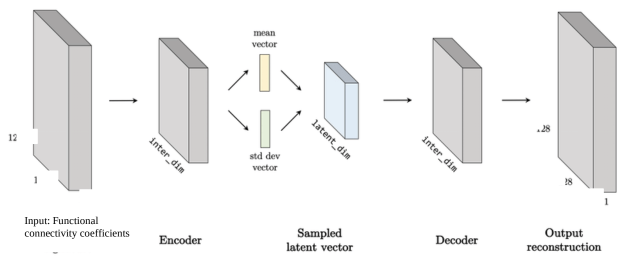

Fig. 5. VAE model architecture

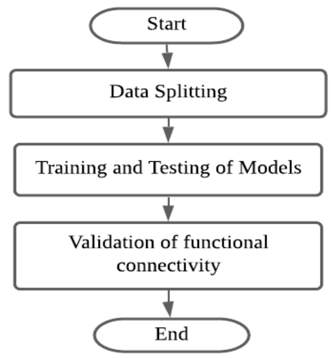

Fig. 6. VAE implementation

The implementation of the VAE model starts with data splitting. The ADHD-200 dataset (Chauhan and Choi, 2020) is split into training and testing sets, which are 32 images and 8 images, respectively. First, training datasets are fit into the VAE model to train the model, followed by testing the model using testing datasets. Then, the model is saved as an R object for future use. The trained model accommodated information such as the accuracy of the model and the loss of the model used. Afterward, the corresponding test set was applied to the trained model to return the predicted label of each data item in the test set. The confusion matrix function provided detailed information about model performance for a better evaluation, such as confusion matrix and accuracy values. Finally, to validate the selected informative, functional connectivity, the confusion matrix is shown to view the VAE performance as the informative, functional connectivity selected by the cluster. The reading of accuracy from the confusion matrix was taken.

D. Performance Measurement

The performance of the method proposed in this study was analyzed using confusion matrix. To measure the result's accuracy, this measurement was used to justify the connection between the performance of the proposed cluster and the selected informative, functional connectivity. The accuracy is the measurement of the proportion of the total number of correct classifications, which includes the parameters True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Table III visualizes the confusion matrix based on binary classification of ADHD and TD subjects. TP refers to the number of correct ADHD, TN refers to the number of right TD, FP refers to the number of incorrect ADHD, and FN refers to the number of incorrect TD.

TABLE III. CONFUSION MATRIX

		Predicted Class	
		ADHD	**TD**
Actual Class	**ADHD**	True Positive (TP)	False Negative (FN)
	TD	False Positive (FP)	True Negative (TN)

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

To analyze the performance of the VAE model, the terms used were accuracy. This term was calculated based on the information from the confusion matrices. The formula can measure accuracy as in Eq. 1. To validate the performance measurement as the comparison with the previous deep learning model, which is DNN, the evaluation method used is identical for both models to show the effectiveness of the proposed method in covering functional information connectivity in the input matrix of the classifier. The DNN and VAE model classifies ADHD using the same training and testing set. The results are compared with existing ANN methods such as DNN.

V. RESULTS AND DISCUSSIONS

This study used the brain functional connectivity derived from 40 subjects of the ADHD-200 fMRI dataset provided from NITRC. Then, the functional connectivity matrix is used to train and test after splitting the data. The VAE classification models were implemented through Google Colaboratory, which includes packages of Python 3.8 and Tensorflow 2.5 libraries with CUDA 10.1 enabler were utilized to implement these models. In addition, some Python libraries which also used in this study, such as Keras, Numpy, Pandas, Matplotlib, and Nilearn.

The neural network's activation function determines how the input's weighted sum is transformed into the output of one or more nodes in the network layer. The choice of activation function has a significant impact on the throughput and performance of the neural network. In addition, different activation functions can be used in other parts of the model. Generally, a differentiable nonlinear activation function is used in the hidden layer of a neural network. Thus, the model learns more complex features than training the network through a linear activation function. There were three activation functions used for hidden layers in this study which are Hyperbolic Tangent (Tanh), Logistic (Sigmoid), and Rectified Linear Activation (ReLu).

Tanh and Sigmoid activation functions were used for the first and last layer, respectively, for the model. It is due to the vanishing gradient problem that was pointed for both activation functions. Besides, the ReLu activation function is also used in hidden layers to overcome the issue of the hidden layer. Lastly, Adam optimization and binary-cross entropy functions were used for training and testing the VAE model. The good model successfully works with functional connectivity matrices of ADHD-200 data. On the other hand, the test loss performance on noisy data was the reason Adam optimizer was used in this study. Also, binary-cross entropy is used in this classification because it secures the output vector so that it is independent of other vectors and is independent for each class. The learning rate was set to 0.0001 with epochs 100 and a batch size of 32.

The VAE model was trained by optimizing reconstruction loss and KL divergence loss by gradient descent. The pre-trained loss is fixed for high-level feature extraction, and KL divergence loss is used to upgrade the encoder, while perceptual feature loss is responsible for updating the encoder and decoder parameters (Hou *et al*., 2017).

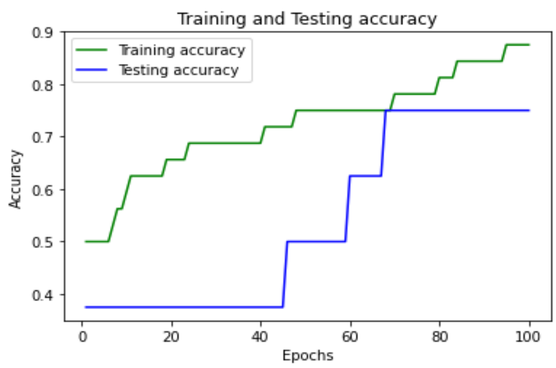

Fig. 7(a). VAE model accuracy

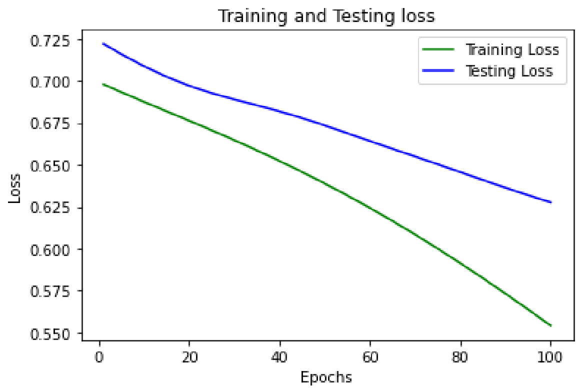

Figure 7(b) VAE model loss

TABLE IV. ACCURACY COMPARISON BETWEEN DNN AND VAE MODEL

Model	Accuracy	Loss
DNN (Chauhan and Choi, 2020)	95%	10%
VAE	75%	62%

The VAE model accuracy graph was shown in Fig. 7(a), while the model loss was shown in Fig. 7(b). According to both figures, the training and testing accuracy generally increases along with epochs and decreases along with loss. The VAE model achieved 75% accuracy, which indicates this model successfully works with functional connectivity matrices of ADHD-200 data. On the other hand, the test loss reached 62% at the last epochs, slightly greater than the train loss at the same epoch, which is 55%. This shows that the model is overfitting where there were too many parameters capable of memorizing the limited amount of training data. The problem of overfitting is also related to the optimal size of the ANN, the existence of outliers in the input set, complex algorithm resolution, and too high of number of training data (Bilbao and Bilbao, 2017).

These findings reveal that fMRI does not always work on imaging data to go through for a classification model, hence the functional connectivity coefficients also can be used as an input for the classification and detection of brain disorders such as ADHD. The accuracy result for this VAE model was compared to existing neural network model research which is DNN (Chauhan and Choi, 2020).

The DNN model achieved an accuracy of 95%, which is greater than the VAE model accuracy as simplified in Table IV. These may cause by the neural network layer, in which the DNN model used 4 layers of sequential model that indicates the model that they are dealing with functions that were not necessarily linearly separable (Chauhan and Choi, 2020). On the other hand, this VAE model used only 3 layers of nodes which are lesser than the other model. Supposedly, different numbers of neurons in the hidden layer will lead to different prediction accuracy (Xing and Li, 2020). For neural networks of other structures, the optimal number of neurons in the hidden layer is different.

The number of hidden layers in the neural network decreases and the error obtained when using the model to predict the test data set increases even if the model predicts the training set correctly due to overfitting (Raut and Dani, 2020). This depends on the performance of the network architecture and the algorithm used in the test dataset. If the network tries to fit the data closely, it will produce significant generalization errors and very high variance due to overfitting. The variance at the output of the network must be smoothened, but if the variance decreases, the deviation will increase to a considerable value and the generalization error will increase again.

VI. CONCLUDING REMARKS

This study proved that classification of fMRI data for ADHD could use the aid of functional. ICA was used to extract functional connectivity correlation coefficients for multiple ROIs. Then, the VAE model was used for the classification. The VAE model had a classification accuracy of 75% lower than past research using DNN, which is 95%. According to the findings of this study, the VAE model has the potential to recognize individuals with ADHD by using functional connectivity coefficients of brain fMRI.

Thus, more future work with larger datasets of ADHD-200 and manual preprocessing steps using Nilearn built-in functions in Python would be captivating to classify ADHD using the VAE model. Many past researchers have been using the public dataset of ADHD-200 Consortium to train and validate their neural network models and the results for the models were outperformed. For example, it is proven by a recent study that used the 4-D CNN method and resulted in 71.3% accuracy (Mao *et al*., 2019). The more the number of samples, the more processing time, including the functional connectivity extraction time and the model training time (Wen *et al*., 2018).

ACKNOWLEDGMENT

The authors wish to thank Universiti Teknologi Malaysia (UTM) and the Malaysian Ministry of Education (MOE) for financial sponsorship of this work under the Fundamental Research Grant Scheme (FRGS/1/2018/ICT02/UTM/02/11).

REFERENCES

[1] Bilbao, I., & Bilbao, J. (2017). Overfitting Problem and the Over-training in the Era of Data: Particularly for Artificial Neural Networks. *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*.

[2] Chauhan, N., & Choi, B.-j. (2020). DNN based Classification of ADHD fMRI Data using Functional Connectivity Coefficient. *International Journal of Fuzzy Logic and Intelligent System, 20*(4), 255-260.

[3] Farzi, S., Kianian, S., & Rastkhadive, I. (2017). Diagnosis of Attention Deficit Hyperactivity Disorder Using Deep Belief Network Based on Greedy Approach. *IEEE*.

[4] Frick, P. J., & Nigg, J. T. (2015). Current Issues in the Diagnosis of Attention Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, and Conduct Disorder. *NIH Public Access*.

[5] Hu, R., Pei, G., Jia, P., & Zhao, Z. (2019). Decoding Regulatory Structures and Features from Epigenomics Profiles: A Road-

map-ENCODE Variational Auto-Encoder (RE-VAE) Model. *Methods*.

[6] Hou, X., Shen, L., Sun, K., & Qiu, G. (2017). Deep Feature Consistent Variational Autoencoder. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*.

[7] Mao, Z., Su, Y., Xu, G., Wang, X., Huang, Y., Yue, W., Sun, L., & Xiong, N. (2019). Spatio-temporal Deep Learning Method for ADHD fmri Classification. *Information Sciences*, 499, 1-11.

[8] Munoz-Organero, M., Powell, L., Heller, B., Harpin, V., & Parker, J. (2018). Automatic Extraction and Detection of Characteristic Movement Patterns in Children with ADHD Based on a Convolutional Neural Network (CNN) and Acceleration Images. *Sensors*.

[9] Munoz-Organero, M., Powell, L., Heller, B., Harpin, V., & Parker, J. (2019). Using Recurrent Neural Networks to Compare Movement Patterns in ADHD and Normally Developing Children based on Acceleration Signals from the Wrist and Ankle. *Sensors*.

[10] Raut P., Dani A. (2020) Correlation Between Number of Hidden Layers and Accuracy of Artificial Neural Network. Vasudevan H., Michalas A., Shekokar N., Narvekar M. (eds). Advanced Computing Technologies and Applications. Algorithms for Intelligent Systems. Springer, Singapore.

[11] Riaz, A., Asad, M., Alonso, E., & Slabaugh, G. (2018). Fusion of fMRI and Non-imaging Data for ADHD Classification. *Computerized Medical Imaging and Graphics*.

[12] Singh, A., Yeh, C. J., Verma, N., & Das, A. K. (2015). Overview of Attention Deficit Hyperactivity Disorder in Young Children. *Health Psychology Research*, 2115.

[13] Wen, D., Wei, Z., Zhou, Y., Li, G., Zhang, X., & Han, W. (2018). Deep Learning Methods to Process fmri Data and Their Application in the Diagnosis of Cognitive Impairment: A Brief Overview and Our Opinion. *Frontiers in Neuroinformatics*, 12.

[14] Xing, Y., & Li, F. (2020). Research on the Influence of Hidden Layers on the Prediction Accuracy of GA-BP Neural Network. *Journal of Physics: Conference Series*, 1486, 022010.

[15] Zou, L., Zheng, J., Miao, C., Mckeown, M. J., & Wang, Z. J. (2017). 3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI.