

Neural Network Input Features Analysis and Selection on EEG Signals by Functional Measurement combined with Genetic Algorithm

Xuguang Song

Research school of computer science
Australia National University,
ACT Australia
u6250082@anu.edu.au

Abstract. EEG signal, which is used widely nowadays in diseases diagnosis areas could be learned by Neural Network and help to mark out potential patients and support doctors to make decisions. However, EEG signal's high complexity and large number of noise data is a big problem for Neural network to learn well and learn quickly. In this research, we are experimenting on 4 different functional measurement models I and W, and their aggregated form C and U are used on a channel wise encoded data set for alcoholism diagnosis to analysis inputs from different frequency bands of EEG signal and their contribution to alcoholism diagnosis. To extend the functional measurement process and improve model I and W's aggregation method, genetic algorithm is combined with model I and W to find out the list of less significant inputs to delete quickly. Result in this paper shows that by removing 2 functional similar or complementing inputs using functional measurement, prediction accuracy is hardly affected due to the large size of inputs but it has been found that among all three frequency bands in encoded EEG signals, frequency band theta and beta's inputs are less significant than frequency band alpha's inputs. Another finding is that Genetic algorithm could find out a number of inputs to delete which have higher average similar or complementing functionality score than the simple functional measurement aggregation method so by using GA-functional measurement to delete multiple inputs could still make the Neural network perform consistently on a good level, thus by applying GA-functional measurement Neural Network's computational complexity could be reduced.

Keywords: Feature Selection, EEG signal, Input analysis, Genetic Algorithm, Neural Network

1 Introduction

EEG signal is proved efficient and useful in diseases diagnosis, for example it has a good performance in diagnosing seizure disorder [1], and according to Tilley et al.'s research [2] we are able to do mood and mental health study based on EEG signal. However, EEG signal has its drawback: large number of inputs and large number of noise data are existing inside [3]. To solve these issues and help Neural Network to study EEG signals efficiently we would like to experiment on inputs functional measurement techniques model I and model W and their aggregated form: model C and model U introduced by Gedeon and Harris [5] on a alcoholism diagnosis data set based on EEG signals [6, 7]. To improve the aggregation process and find out the best list of least significant inputs to delete, we combined Genetic Algorithm with model I and model W to get GA-I model and GA-W model to compare with the performance of aggregated model C and U. By measuring each input's functionality with other inputs, we are able to find out which inputs have less significant functionalities in predicting alcoholism. So that by removing inputs with least significant functionalities we are able to help neural network improve classification accuracy, at the same time reduce computational complexity of the neural network. We also want to analysis different input frequency bands' contribution to alcoholism diagnosis in order to find out which frequency band of EEG signals contributes more in predicting alcoholism.

2 Method

Process diagram attached below shows the main method. As shown in the Figure 1, we first use the preprocessed EEG signal data to train a baseline neural network. Then we use the neural network's weight output [4] and the EEG signal data to apply functional measurement models I and W to get each pair of inputs' significance in prediction. Then, based on the results from models I and W, we tried two approaches to get the list of inputs to delete. First approach is the aggregation method [4] to aggregate model I and W's pairs results into each input's own significance results, by doing

this to get model C and model U's values to rank each input's functional significance then the list of inputs with less significance are to be deleted. Second approach is to use Genetic Algorithm to select a list of inputs which have the biggest average model I scores or model W scores with each other to ensure that all inputs deleted have similar or complementing functionalities with each other. Then according to two different approaches we remove a number of inputs which are least significant and train the neural network with reduced inputs and compare performance with the accuracy before removing the inputs and accuracy after removing the inputs.

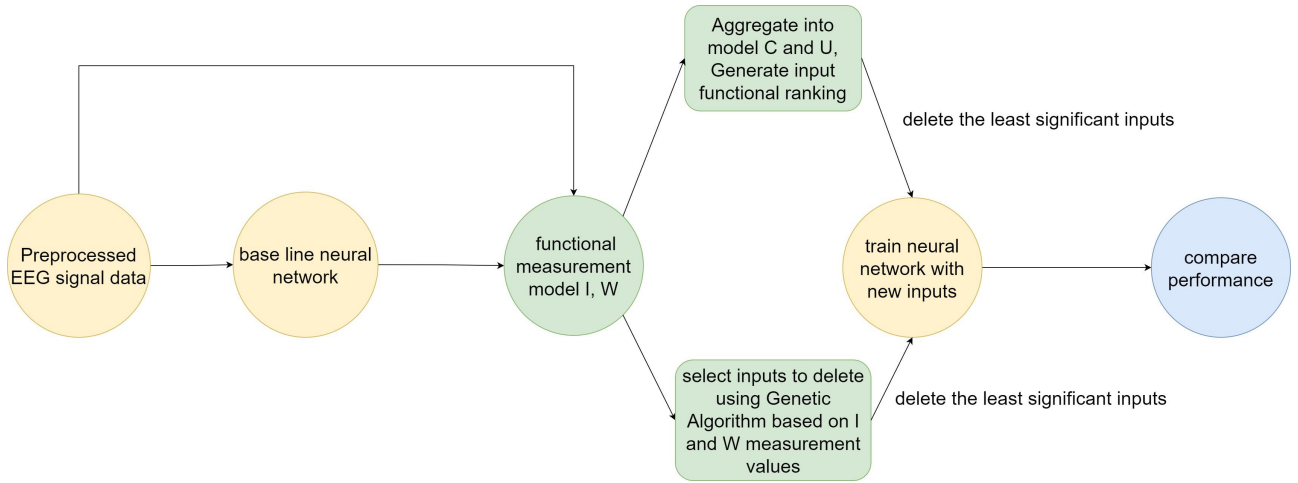


Figure 1. process of inputs' functional measurement and genetic algorithm-functional measurement

2.1 Data set

Data set we were experimenting on is a preprocessed EEG signals data from ANU, Yao et al. [3]'s research. Among all attributes inside, we choose the 'data' attribute as input features and the 'y_alcoholism' as our target. Basically we are using encoded EEG signals to predict whether a subject is diagnosed as alcoholism or not.

'data' attribute is a 11057×192 matrix, it is extracted from the raw data of dimension $64 \times 256 \times 11057$ from University of New York EEG data where 64 is the channel numbers and 256 is the signal extracted [6, 7], the signal then is encoded into 3 different frequency bands, 4 to 7 Hz signals are put into theta band, 8 to 13 Hz signals are put into alpha band and 13 to 30 Hz signals are put into beta band, the preprocessed data then is dimension 64×3 , and then the 64×3 data is flattened into 192×1 to form a vector. The new data is this 192×1 vector of total trial 11057 times so the data matrix is of dimension 11057×192 [3]. To help understand this new data set, a visualization of the 64 channel and 3 frequency bands are shown in the Figure 2 below, in one trial's 64 channels, every frequency band's signal values are shown as bars with different colors.

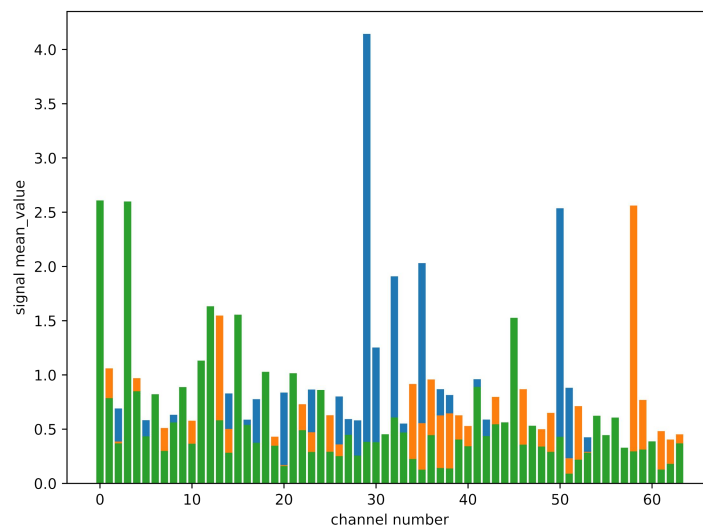


Figure 2. different frequency bands' value in each channel

Another visualization way is provided below as shown in Figure 3, it shows how each frequency band's value is changing among the trials from 0 to 250 in a single EEG channel from all the 64 channels

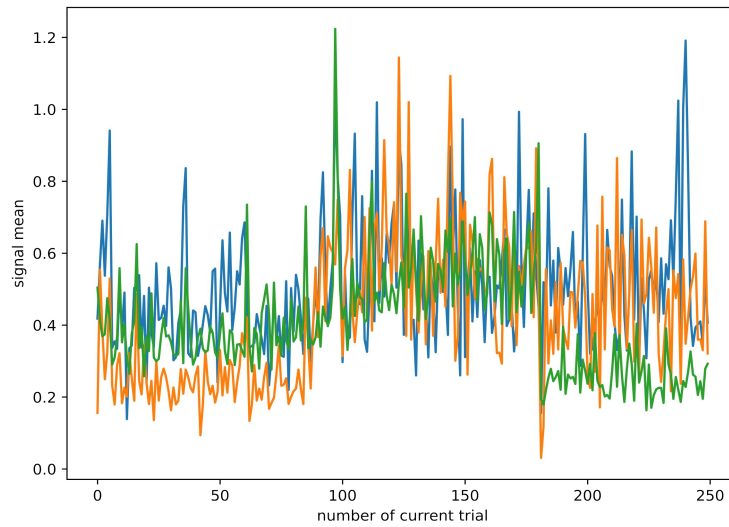


Figure 3. in a single channel how different frequency bands' value changes along time

And then the last attribute is 'y_alcoholism' which contains 0 and 1 inside, 0 represents a subject does not have alcoholism and 1 represents a subject has alcoholism.

2.2 Preprocessing

The features data are divided as 70% training and 10% validation set for hyper-parameter tuning, 20% testing. Normalizing method was tried on the raw data to preprocess but it lowered the neural network's performance, it turns out that normalizing changes the frequency bands values' characteristic and thus affect the neural network's performance. Another preprocess method is to round the signal values to 2 decimal places. This turns small values to 0, thus we are able to reduce the network's calculation complexity.

In order to experiment on neural network's performance within subject and cross subject [3, 8], the original input features 'data' and target value 'y_alcoholism' are also processed into 2 different data sets. The within subject data set is splitting all data from all subjects as training and validating, testing set while the cross subject data set is splitting all subject into different groups for training and validating, testing. One group of subjects' data is used to train models and the other two are used to validate the hyper parameter tuning and test models.

2.3 Task and the baseline neural network

Basically, as mentioned in introduction our task is to use the EEG signal data to train a baseline neural network first for predicting whether a subject is alcoholism (1) or not (0) then applying input functional analysis model I and W based on that to improve the performance of the neural network and analysis different frequency bands of EEG signal's contribution to neural network's prediction. Also a genetic algorithm-functional measurement method is suggested to compare with the aggregation method for model I (which is model C) and aggregation method for model W (which is model U).

The baseline neural network used here has 192 input neurons, 15 hidden neurons and 2 output neurons. It uses sigmoid function as activation. Because of the large number of inputs, we decided to use batch method with size 125 to train the network and a total epoch of 200 is set for the network and Adam optimiser with learning rate 0.0075.

2.4 Technique used

2.4.1 Functional measurement

Technique used to compare two inputs is functional measurement. Two functional measurement models I and W and their aggregated form model C and U are used [4, 5].

Model I calculates two inputs' angle by combining all samples of the two inputs to two high-dimension vectors and calculating the angle based on that, according to Gedeon and Harris's distinctiveness analysis [4, 5]

$$\text{Model_I_angle}(\text{input1}, \text{input2}) = \tan^{-1} \left(\frac{\sqrt{\sum_{i1}^{\text{input_samples}} (\text{sample}(i1) - 0.5)^2 * \sum_{i2}^{\text{input_samples}} (\text{sample}(i2) - 0.5)^2}}{\sum_i^{\text{input_samples}} ((\text{sample}(i1) - 0.5) * (\text{sample}(i2) - 0.5))^2} - 1 \right) \quad (1)$$

Model W calculates two input vectors' angle by their weights to all the hidden neurons. Each input and its weights to all hidden neurons are combined to a high-dimension vector which in our case is a 15x1 vector for calculating angles, according to Gedeon's hidden neuron weights analysis [9]

$$\text{Model_W_angle}(\text{input1}, \text{input2}) = \tan^{-1} \left(\frac{\sqrt{\sum_{i1}^{\text{hidden_neurons}} (\text{weight}(i1, h) - 0.5)^2 * \sum_{i2}^{\text{hidden_neurons}} (\text{weight}(i2, h) - 0.5)^2}}{\sum_i^{\text{hidden_neurons}} ((\text{weight}(i1, h) - 0.5) * (\text{weight}(i2, h) - 0.5))^2} - 1 \right) \quad (2)$$

Note that in equation 2, because the input should be in range 0 to 1 so the weight matrix needs to be normalized first. Result angle between two inputs should be in range 0 degree to 180 degree [5], in distinctiveness analysis if angle is less than 15 degree then the inputs are considered having similar functionality so they can be removed, if angle is more than 165 degree then the inputs are complementing so they can be removed as well [5]. However, here we are using this technique to analyze input contribution so we just need to find out the distance between the angle and 90 degree. That means model_I's values are calculated as shown in equation 3 and model_W's values are calculated as shown in equation 4

$$\text{Model_I}(\text{input1}, \text{input2}) = \text{Abs}(\text{Model_I_angle}(\text{input1}, \text{input2}) - \frac{\pi}{2}) \times \frac{180}{\pi} \quad (3)$$

$$\text{Model_W}(\text{input1}, \text{input2}) = \text{Abs}(\text{Model_W_angle}(\text{input1}, \text{input2}) - \frac{\pi}{2}) \times \frac{180}{\pi} \quad (4)$$

and the smaller means inputs are less similar or complementing.

The way to get aggregated form model C and model U are not mentioned by Gedeon's paper [4], so in this experiment we are going to get the aggregated result by adding up all the angular distance for each input among all the combination pairs including that input:

the way to get model C is shown in equation 5

$$\text{Model_C}(\text{input}) = \sum_j^{\text{pairs}(\text{input}, j)} \text{Model_I}(\text{input}, j) \quad (5)$$

the way to get model U is shown in equation 6

$$\text{Model_U(input)} = \sum_j^{\text{pairs(input, j)}} \text{Model_W(input, j)} \quad (6)$$

2.4.2 Genetic Algorithm - functional measurement

To further support better inputs selections and analysis. We want to suggest another way to select list of inputs to delete instead of using aggregation method of C and U. Current aggregation method is by ranking all the inputs with their sum of distances to all the other inputs. But this method could not make sure the deleted inputs indeed have similar or complementing functionalities. For example, input1 and input2 may have the biggest aggregated distance value but the angle between the two inputs are close to 90 degrees means they have different functionalities, then by removing the two inputs may potentially reduce the accuracy.

To solve this problem we decide to introduce a new inputs deleting method: Instead of delete inputs according to the ranking of aggregated sum of their functional measurement score, we want to change the problem to delete x inputs from 192 inputs so that the average functional measurement score (two inputs' angular difference to 90 degrees) between all the combinations of the x inputs are maximized. That is to maximize:

$$\text{Average_score(delete_inputs)} = \frac{\sum_i^{\text{delete_inputs}} \sum_j^{\text{delete_inputs}} \text{Model_I_or_W}(i, j)}{C_{\text{number_of_delete_inputs}}^2} \quad (7)$$

By doing this we are able to make sure deleted inputs have similar or complementing functionalities. But by using this approach the computational cost is very high. For example if we want to delete 10 inputs from 192 inputs and make sure the average functional measurement score between the 10 inputs are maximized, then we need to search through:

$$C_{192}^{10} = \frac{192!}{10!(192-10)!} = 1.48 \times 10^{16} \text{ combinations} \quad (8)$$

Hence, to solve this high computational cost task we need to introduce the Genetic Algorithm to speed it up and find an optimal solution.

Genetic algorithm-functional measurement:

Here is the settings if we want to remove x inputs.

Chromosome: a list of length x, where each gene in the chromosome is an integer between 0 to 191, indicate which input indices are selected for example [1, 5, 78, 98, 125, 152,, 17, 18].

Fitness Value:

The average value of functional measurement score between all inputs, as explained in equation 7, GA-I uses model I's functional measurement functions between inputs and GA-W uses model W's functions between inputs.

Crossover:

Using one point crossover, because we want to keep most of the inputs in the best individuals as they have higher functional measurement score with each other. Using other crossover methods may break this relation.

Selection:

Using tournament selection to select the best top 4

Mutation:

Mutate each integer in the list to another integer between 0 to 191 with a probability of 20%, this is to create randomness and avoid the genetic algorithm to stuck at local optimum

Initial population is set to 50 and Max generation is set to 225 after some tuning because according to the graph:

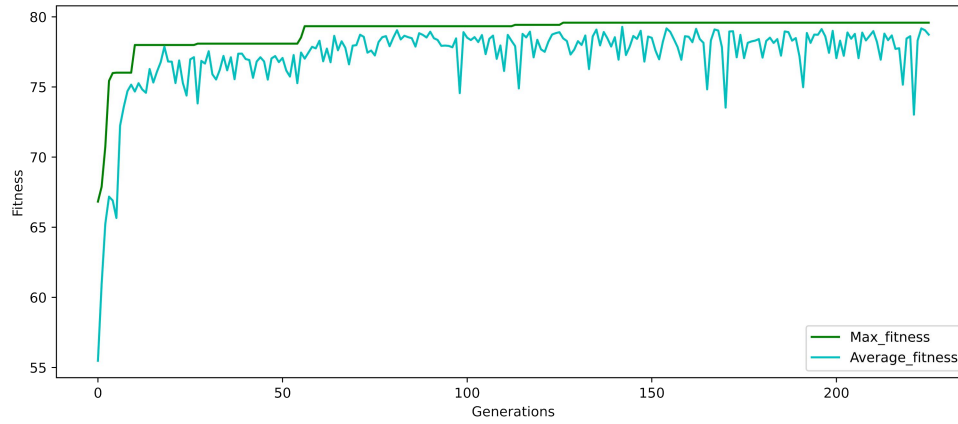


Figure 4. example of GA-functional measurement I model's fitness curve

As shown in the graph, by setting initial population to 50 and max generation to 225, the algorithm could achieve a relatively stable result and at the same time spend only around 15 seconds. Here the max fitness after 225 generations is around 80 with means the average angle between all selected inputs are average 80 degrees away from 90 degrees, that is they have an average angle between each other of 170 degrees or 10 degrees. Which shows that those inputs have similar or complementing functionalities thus by removing this could potentially improve both the predicting accuracy and efficiency.

3. Results and findings

3.1 Result by applying functional measurement in frequency band analysis

Because one of the purposes of our task is to analysis different frequency bands' behavior in predicting EEG signals, by applying basic functional measurement method, we measured pair combinations within inputs of frequency band theta (0-64), frequency band alpha (64-128), and frequency band beta (128-192).

After applying the technique we find the least and most significant inputs among all the inputs as shown in Table 1 below.

model	Most significant inputs index						Least significant inputs index					
I	93	125	113	110	124	64	16	13	14	17	1	4
W	92	93	97	115	128	185	19	13	20	14	128	176
C (aggregate I)	93	0	156	3	114	96	173	155	179	152	182	149
U (aggregate W)	93	101	154	135	45	95	77	83	177	150	90	91

Table 1. 4 model's choice of most and least significant inputs

3.2 Result of GA-functional measurement

Then we compare the choice between functional measurement I, W, C, U and GA-I, GA-W. Average functional measurement scores between the listed inputs are shown.

model	Least significant inputs index						Average functional measurement scores
I	16	13	14	17	1	4	-
W	19	13	20	14	128	176	-
C(aggregate I)	173	155	179	152	182	149	78.00
U(aggregate W)	77	83	177	150	90	91	51.17
GA-I	182	83	173	86	74	68	79.11
GA-W	83	150	31	119	141	75	56.13

Table 2. GA-functional measurement's performance on least significant inputs compare to models C and U

3.3 Result by removing inputs using functional measurement and GA-functional measurement

The accuracy change after removing the 2 least significant inputs is shown in Table 2 and Table 3. In Table 3 is the comparison of our final output with Yao et al.'s Channel-Wise auto-encoder performance and Image-Wise auto-encode performance [3]

	Accuracy within-subject	Accuracy cross-subject
Baseline (before)	87.10%	67.27%
I	88.86%	67.48%
C	86.42%	60.89%
W	85.99%	67.75%
U	87.67%	68.80%

Table 3. Accuracy before and after removing the 2 least contributed inputs

We further tested the model U and W compare to GA-I and GA-W by using it to remove 0, 6, 9, 12 numbers of least significant inputs and the accuracy is shown in Table 4

Number removed	Accuracy of C(aggregate I)	Accuracy GA-I	Accuracy of U(aggregate W)	Accuracy GA-W
0 (baseline)	87.10%	87.10%	87.10%	87.10%
6	85.30%	87.76%	87.72%	86.66%
9	83.30%	87.21%	86.33%	84.67%
12	85.26%	87.25%	85.97%	86.59%

Table 4. accuracy of removing multiple inputs using model U

3.4 Result compare to other methods

	Accuracy within-subject	Accuracy cross-subject
Base line neural network	87.1%	67.3%
Neural network with functional measurement	88.9%	67.8%
Neural network with GA-functional measurement	87.8%	68.1%
Normal Channel-wise autoencoder [3]	86.4%	73.1%
Image-wise autoencoder[3]	91.7%	75.6%

Table 5. Comparison with other methods

3.5 Discussion

One valuable finding shown in Table 1 above is among different model's measurement, frequency bands theta (0-64) and beta (128-192) are not very significant as their inputs appear more on the least significant inputs table and appear less on the most significant inputs. Meanwhile frequency band alpha (64-128) appears the most of the times in most significant inputs so alpha frequency band seems more significant than the other two bands. This indicates that frequency band theta and frequency band beta tend to have similar or complementing inputs inside. So, this tells us that when analyzing EEG signal in diagnosing alcoholism we should focus more on the alpha frequency band.

By looks at Table 2's results, clearly we could see that GA-I achieved average functional measurement score 79.11, higher than functional measurement C model which is aggregation form of model I, it has 78.00 score. GA-W achieve average functional measurement score 56.13 higher than functional measurement U model which is aggregation form of model W, it has 51.17 score. This shows GA-functional measurement is good at finding inputs that has more similar or complementing functionalities than the simple aggregation methods.

Refer to Table 3, we can see that removing two least significant inputs hardly have an effect on improving the neural network's accuracy. This is because the data set introduced in Gedeon 1997's research [10] is a GIS data with 16 inputs [10], and according to the brute force analysis in his paper, removing some certain features can have significant effect on neural network's performance [4]. While it is a different case for EEG signals, EEG signals are more complicated and the similarity or complementary between inputs are very low. This makes input analysis increasingly difficult. Removing a pair of inputs is not going to provide much help in improving network's accuracy.

According to Table 4, using C, U and GA-I, GA-W, we tried to remove 6, 9 and 12 least significant inputs. GA with functional measurement performs more stable when removing more inputs because it is able to keep the deleted inputs to have more similar or complementing functionalities. Another point is we removed in total 12 least significant inputs while the accuracy is still stable. This indicates that we could use the GA-functional measurement method to help reduce computational cost, at the same time keep the neural network's performance.

According to Table 5, although accuracy is slightly improved by removing least significant inputs, the performance is still worse than Yao et al.'s research [3], I think one reason is we are using a simple Neural Network here. It may not be able to outperform the CNN which could learn in a deeper way in Image-wise autoencoder [3]. Another reason is because the functional measurement method has limitation, it is designed to find functional similar and complementing inputs to help the Neural Network to learn more efficiently so not necessarily could improve EEG signals' prediction accuracy as it has large input sized and big amount of noises are included in EEG signals.

4. Conclusion

This paper experiments on 2 different functional measurement models I and W and compared 2 aggregated functional measurement models C and U with GA-model I and GA-model W. It turns out for large input size EEG signals, removing least significant inputs have limited effect on classification accuracy, however the functional measurement is still able to show each input's contribution, according to our analysis, frequency band theta and beta are the bands that have less significant inputs inside which means inputs inside these two frequency bands tend to have similar or complementing functionalities. While frequency band alpha has inputs with higher contribution to alcoholism classification. So when encoding EEG signal channel-wise, more operations could be taken for frequency band theta and frequency band beta in order to reduce the noise and the least significant inputs. Another finding is that GA-model I and GA-model W are better at finding least significant inputs to delete than aggregation model C and U, and using it to remove up to 12 inputs could still get consistent performance compare to model C and U. Thus by removing big number of inputs using Genetic Algorithm-functional measurement, we could reduce the neural network's computational complexity and still be able to have consistent neural network accuracy in prediction.

5. Limitation and future work

Although we could get the inputs with similar and complementing functionality quickly with GA-functional measurement method, due to the large size of input features in EEG signals and the simple structure of current Neural Network, removing those inputs hardly improve the prediction accuracy on the preprocessed EEG signal data set [3]. As we already known that different frequency band could have different contribution in prediction, in the future we could try to use RNN with channel and self attention as introduced by Tao [11], to see if we could achieve a better prediction accuracy. To further verifying and looking into the conclusion about different frequency bands' contribution to alcoholism diagnosis, we could also use GA-SVM method to select inputs that helps improve classification accuracy as Sharma and Gedeon's research [12] showed it has a good performance on EEG signals.

References

1. Alsaadi, T.M., Thieman, C., Shatzel, A., Farias, S.: Video-EEG telemetry can be a crucial tool for neurologists experienced in epilepsy when diagnosing seizure disorders pp. 32-34. *Seizure* (2004),
2. Tilley, S., Neale, C., Patuano, A., Cinderby, S.: Older people's experiences of mobility and mood in an urban environment: a mixed methods approach using electroencephalography (EEG) and interviews. 14(2):151. doi: 10.3390/ijerph14020151 *Int J Environ Res Public Health*. (2017)
3. Yao, Y., Plested, J., Gedeon, T.: Deep Feature Learning and Visualization for EEG Recording Using Autoencoders: 25th International Conference, ICONIP 2018, Proceedings, Part VII, Siem Reap, Cambodia (2018)
4. Gedeon, TD.: Data mining of inputs: analysing magnitude and functional measures, 8(2):209-18. doi: 10.1142/s0129065797000227 *Int J Neural Syst* (1997)

5. Gedeon, TD and Harris, D: Network Reduction Techniques, Proceedings International Conference on Neural Networks Methodologies and Applications, AMSE, vol. 1, pp. 119-126, San Diego (1991)
6. Li, Y., Dzirasa, K., Carin, L., Carlson, D.E.: Targeting EEG/LFP synchrony with neural nets. In: Advances in Neural Information Processing Systems, pp. 4623–4633 (2017)
7. Sykacek, P., Roberts, S.J.: Adaptive classification by variational Kalman filtering. In: Advances in Neural Information Processing Systems, pp. 753–760 (2003)
8. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: EEGnet: A compact convolutional network for eeg-based brain-computer interfaces, arXiv:1611.08024, arXiv (2016)
9. Gedeon, TD.: Indicators of Input Contributions: Analysing the Weight matrix, Proceedings ANZIS'96 International Conference, 4 pages, Adelaide (1996)
10. Bustos, RA and Gedeon, TD. :Decrypting Neural Network Data: A GIS Case Study, Proceedings International Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA), 4 pages, Alès (1995)
11. Tao et al. : EEG-based Emotion Recognition via Channel-wise Attention and Self Attention, IEEE Transactions on Affective Computing PP(99) DOI:10.1109/TAFFC.2020.3025777 (2020)
12. Sharma, N. Gedeon, TD.: Modeling observer stress for typical real environments, Expert Systems with Applications Volume 41, Issue 5, pp. 2231-2238 (2014)