Feature Selection with Genetic Algorithm for Alcoholic detection using Electroencephalogram

Indah Agustien Siradjuddin
Informatics Department
University of Trunojoyo Madura
Telang Raya, Kamal – Madura, Indonesia
Indah.siradjuddin@trunojoyo.ac.id

Rivaldi Septasurya, M. Kautsar Sophan, Noor Ifada, and Arif Muntasa
Informatics Department
University of Trunojoyo Madura
Telang Raya, Kamal – Madura, Indonesia

***Abstract*—Electroencephalography is a technique of recording human brain waves through electrodes mounted on the human scalp. The obtained signal data or Electroencephalogram (EEG) can be used for the alcoholic detection as a substitution of the urine test or breathalyzer test, since in the urine or breathalyzer test, is time limited. Therefore, the usage of EEG for alcoholic detection is proposed in this research. There are four stages are proposed in this research. First, noise removal of the recorded signal by separating the signal using Independent Component Analysis. Second, feature extraction with Discrete Wavelet Transform. Third, extracted features from the previous stage are selected based on the importance of each feature using Genetic Algorithm, and final stage is the alcoholic detection with Backpropagation neural network. Experiments were conducted on 64 channels of EEG data. The feature selection stage using Genetic Algorithm made the accuracy of the detection process is increased, i.e. the maximum average accuracy of the detection is 79.38%. This feature selection stage also reduced the used feature for the detection process, up to 48% of all obtained features, in the extraction feature stage.**

Keywords*— *Electroencephalogram; Feature Selection; Genetic Algorithm; Independent Component Analysis

I. INTRODUCTION

Electroencephalogram (EEG) is the electrical signal data that is obtained and recorded from the human brain waves. This signal data contains valuable information about the brain activities [1], therefore, the valuable information can be used as the data to detect some diseases, especially the neurological diseases, such as the epileptic detection [1]–[6]. Currently, the EEG signal is used not only to detect neurological disease, but other purposes, for instance, to detect the vigilance level of the operator who works on monotonous activity but attention demanding tasks [7], detection of the brain abnormalities [8], alcoholic detection [9]–[11], and etc.

This research used the EEG signal data for the alcoholic detection. According to the data from WHO (World Health Organization), approximately 3.3 million people deaths every year because of the alcohol. Collected data from WHO shows that more than 50% people in the world are alcoholic, and many of them are minors. Consuming too much alcohol will affect the performance of human brain and possible to harm other people, environment, and the society. Therefore the detection of alcoholic is proposed in this research.

There are commonly two ways to detect alcoholic, i.e. breathalyzer test and urine test. The breathalyzer test is easily manipulated by removing the odor from the breath, meanwhile the urine test requires more time, since the alcohol is detected from the urine after 20-24 hours the person consumed alcohol. This research will use the EEG signal to classify whether a person is an alcoholic or a normal people, since the data shows the brain activity and cannot be manipulated.

The EEG signal is obtained by recording the brain wave that requires several electrodes mounted on the top of the human scalp. There are 64 electrodes mounted throughout the head. One electrode represents one channel and there are 64 channels in the used dataset. Huge amount of the extracted data from the EEG signal that requires time computation in the detection proses, in addition, not all the extracted features are considered as important features for the detection process. Therefore this research proposes the feature selection to obtain the important features for the detection process using Genetic Algorithm. The Genetic Algorithm is used for this stage since; the algorithm is a heuristic method to solve the optimization problem, hence the best combination features for the detection process is achieved using the method.

There are four main required stages that we proposed in this research, i.e. noise removal by separating the signal, feature extraction, feature selection, and detection of the alcoholic.

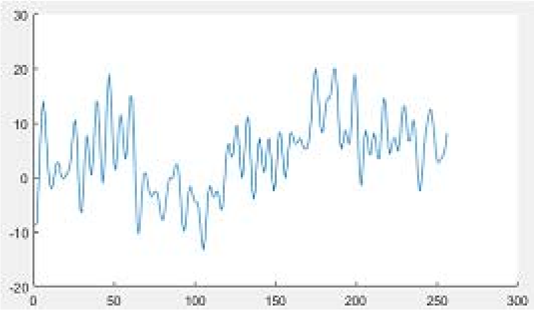
II. SIGNAL SEPARATION USING ICA

Electrodes are placed near to each other to record the brain wave; therefore the noises are obtained from the EEG recording process, since the signal from an electrode will affect the signal from the other electrode. To get the original EEG signal of an electrode, the separation process is required. This research used the Independent Component Analysis (ICA) of the Blind Source Separation method to separate the EEG signals [11]. If there are two signal data ($x_1(t)$ and $x_2(t)$) that are recorded from two electrodes. This recorded data is weighted sum of the original signal from each electrodes (s_1 and s_2), as seen in (1).

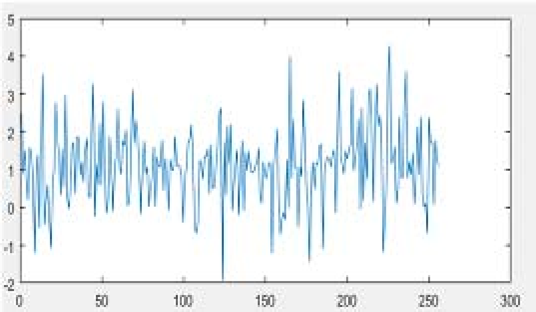
$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$
$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) \quad (1)$$

where the t is the time index, s_1 and s_2 are the original signal that recorded from source 1 and 2, and x_1 and x_2 are recorded signal from source 1 and source 2.

The objective of ICA method is to obtain the original signal by estimating the parameters, i.e. a_{11}, a_{12}, a_{22}, and a_{23}. The example of the EEG signal can be seen in Fig. 1.



(a)



(b)

Fig. 1. Example of EEG signal from a channel ; (a) the recorded signal and (b) the obtained original signal using ICA

III. FEATURE EXTRACTION

Features are extracted from the frequency domain the EEG signal, therefore the EEG is converted into frequency domain using Discrete Wavelet Transform (DWT). The DWT method divides the signal into low frequency signal (or known as approximation) and high frequency signal (or known as detail) using the highpass and lowpass filter [11], [12]. The separated signal EEG is decomposed into low and high frequency using DWT, and for this research features are obtained from the high frequency (detail information). To decompose the signal into high frequency signal, the signal is calculated using (2)

$$A_{high}[N] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \quad (2)$$

Where N is the decomposition level, h is highpass filter, and x is input signal.

There are seven features that are extracted from the high frequency signal, i.e.

1. Maximum, the maximum value of the signal data in each channel of EEG
2. Minimum, the minimum value of the signal data in each channel of EEG
3. Mean, the mean value of EEG in each channel, as in (3)

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \quad (3)$$

where N is the total data and A is the high frequency signal.

4. Standard deviation, spread of the high frequency data as seen in (4)

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N} |A_i - \mu|^2} \quad (4)$$

where μ is the mean of high frequency EEG signal

5. Approximate Entropy (ApEn) is used to measure the regularity of index bias, relative inconsistency, and data dependency, as shown in (5)

$$ApEn(A,m,r) = \ln\left[\frac{C_m(r)}{C_{m+1}(r)}\right] \quad (5)$$

Where Sn is the signal, m is the size of the signal data, r is tolerance value, and Cm is repeated data pattern.

6. Sample Entropy (SpEn) is approximate entropy with non-negative sign as (6)

$$SpEn(A,m,r) = -\ln\left[\frac{C_m(r)}{C_{m+1}(r)}\right] \quad (6)$$

7. Wavelet entropy to measure the data using wavelet theory as in (7)

$$E1(A) = -\sum_{i}^{N} A_i \log_2(A_i)^2 \quad (7)$$

where A is the signal data and i is the index data

IV. FEATURE SELECTION AND DETECTION

There are seven features extracted in each channel of the EEG data. Total features are used for the classification are *7 x numberOfChannel*. In this dataset, there are 64 channels; hence 448 features are extracted from the feature extraction. More number of features makes the required time computation for the detection is increasing. In addition, not all features are considered as important features for the detection process. Therefore, in this research, we proposed feature selection using Genetic algorithm, in order to increase the detection accuracy and lower the computational time since the number of selected features are fewer than the original number of features.

The Genetic Algorithm is a promising method to find the best feature combination by finding the best solution from all the candidate solution (optimization problem) [13]–[17]. The best solution is obtained from the fitness value of each the candidate solution. The higher fitness value then the higher possibility of the candidate solution is being selected.

The first step of the genetic algorithm is chromosome encoding, i.e. determine the chromosome code. Since in the experiment, we used 64 channels, hence, there are 448 features are used in the classification. These 448 features are encoded in binary string with 448 bit length. Bit '0' represents that the feature is not selected, on the contrary, bit '1' represents that the feature is selected feature. The example of the chromosome can be seen in Table 1.

TABLE I. ENCODED CHROMOSOME

Features	1	2	3	4	5	···	150	151	···	448
Chromosome	0	0	1	0	1	···	1	0	···	1

Table 1 shows that feature 1, 2, 4, 151 are not selected for the detection process, meanwhile feature 3, 5, 150, and 448 are selected for the detection process. The chromosome in the population is generated randomly.

Second step is the selection process. This process requires the fitness calculation. This research uses the accuracy of the detection process for the fitness value [18], [19]. For the detection stage, we use Backpropagation neural network. The Chromosomes are selected based on the fitness value using the Roulette wheel. The illustration of selection process is depicted in Fig. 2. In this figure, for the illustration, there are only 28 bits of chromosome is shown, instead of the real bits of chromosome, i.e. 448 bits of chromosome.

No	Chromosome	Fitness Value
1	0111111011100010100010100001	0.6633
2	1010011011000000000100110110	0.6867
3	1011111001100011100001101010	0.6900
4	0100110101001100110100101001	0.6733
5	0001011011110000110011101100	0.6867
6	1101001101110010011110010100	0.6967

Roulette Wheel

No	Chromosome
6	1101001101110010011110010100
3	1011111001100011100001101010
5	0001011011110000110011101100
5	0001011011110000110011101100
6	1101001101110010011110010100
2	1010011011000000000100110110

Fig. 2. Chromosome selection using Roulette Wheel

In the selection process based on roulette wheel method, the chromosome with higher fitness value will have higher probability to get selected; on the contrary, the chromosome with lower fitness value, will have lower probability to get selected. As seen in Fig.2, there are six chromosome are generated in the first step of genetic algorithm, and chromosome 6 has the highest fitness value, therefore, chromosome 6 has the highest probability to get selected, i.e. chromosome 6 is selected twice. Meanwhile, chromosome 4 has the lowest fitness value, and therefore has the lowest probability to get selected.

Third step is the crossover. In this step, bits of the generated position are changed between two chromosomes (offsprings), therefore there will be new chromosomes with new bits are generated from this process, as seen in Fig. 3.

	Bit-1	Bit-2	...	Bit-89	Bit-90	Bit-91	Bit-92	Bit-93	...	Bit-448
Offsprings	1	0	...	0	1	0	0	0	...	1
	0	1	...	1	0	1	1	1	...	0
New Chromosomes	1	0	...	0	0	1	1	0	...	1
	0	1	...	1	1	0	0	1	...	0

Fig. 3. Crossover Process

The last step is the mutation. Bit of the selected position of the chromosomes is changed. Bit '1' will be changed into bit '0', and otherwise, bit '0' is changed into bit'1'. The process of selection, crossover, and mutation are repeated for a number of iteration, or known as generation.

Backpropagation neural network is used as the classifier to classify whether the input signal is alcoholic or a normal person. There are three layers in this network, i.e. input, hidden, and output layer. The number of the neuron in the input layer depends on the number of selected features from the feature selection process. If ten features are selected, then the number of neuron in input layer is ten.

V. RESULT AND DISCUSSION

For the experiments, we used dataset that are taken from UCI Machine learning repository [20]. Each signal data is recorded from 64 electrodes; therefore there are 64 channels in the dataset. The EEG signal was sampled at 256 Hz for one second. In this dataset, there are 10 alcoholic people are recorded 20 times, and 10 control (normal) people are recorded 20 times. Therefore, 400 available data are used in these experiments.

The accuracy of the detection based on all features (448 features) and selected features are compared for the experiments. Three scenarios based on the number of data training are used in the experiment, i.e. 20%, 40%, 60% of dataset is used as data training respectively. The network is trained with 500 epochs, and since the weight of the network initially is generated randomly, therefore the trial is conducted five times in each scenario.

For the selection process using Genetic Algorithm, the number of chromosome in population is ten, and the searching for the best candidate for the feature selection is stop when the number of generation is reached or the minimum accuracy is achieved. In the experiment, the number of generation is set to 10 and the minimum accuracy is 0.95.

The result of the first scenario is shown in Table 2. In the first scenario, there are 80 data are used as the data training, and the remainder of the dataset, i.e. 320 data are used for the data testing. In the first scenario, 232 features are selected using the Genetic algorithm, meanwhile there are 448 features are obtained from the feature extraction process. Hence, 48.12% features are considered as less important features for the alcoholic detection.

TABLE II. RESULT OF THE FIRST SCENARIO (DATA TRAINING=80, DATA TESTING=320)

Trial	Accuracy (%)	
	Selected Features	**All Features**
1	64.38	66.56
2	66.25	65
3	66.87	66.56
4	64.69	66.25
5	65.94	64.69
Average	**65.626**	**65.812**

TABLE III. RESULT OF THE SECOND SCENARIO (DATA TRAINING = 160, DATA TESTING =240)

Trial	Accuracy (%)	
	Selected Features	**All Features**
1	75.42	70.83
2	75.42	74.58
3	75	73.75
4	74.17	74.17
5	75	74.17
Average	**75.002**	**73.5**

TABLE IV. RESULT OF THE THIRD SCENARIO (DATA TRAINING = 240, DATA TESTING = 160)

Trial	Accuracy (%)	
	Selected Features	**All Features**
1	80	75
2	80.63	78.75
3	78.13	75.62
4	80	74.38
5	78.13	75
Average	**79.378**	**75.75**

The average accuracy of the first scenario using feature selection process is 65.626%, meanwhile the 65.812% accuracy is achieved when all features are used for the detection process. In this scenario, the achieved accuracy between the proposed feature selection process is similar with the achieved accuracy with all features are used in the detection process. However the different result is obtained in the second scenario.

In the second scenario, i.e. 40% of dataset is used for data training (data training = 160, data testing = 240), the number of features being selected with Genetic algorithm is 214, i.e. 47.77% of all obtained features from the feature extraction process (448 features). In this scenario, the average accuracy of proposed method is 75.002%, meanwhile the average accuracy using all features (448 features) is 73.5%. In this scenario, the proposed method accuracy is higher with the conventional one.

The result of the third scenario shows the similar trend with the second scenario. In the scenario, 231 features are selected from the genetic algorithm, i.e. only 51.56% of all obtained features are used for the detection process. The achieved average accuracy is 79.378%, meanwhile, using all features for the detection process achieved 75.75% average accuracy. Therefore, the proposed method achieved higher accuracy compare to the conventional method. The comparison of average accuracy between the proposed method and the conventional method is shown in Fig 4.

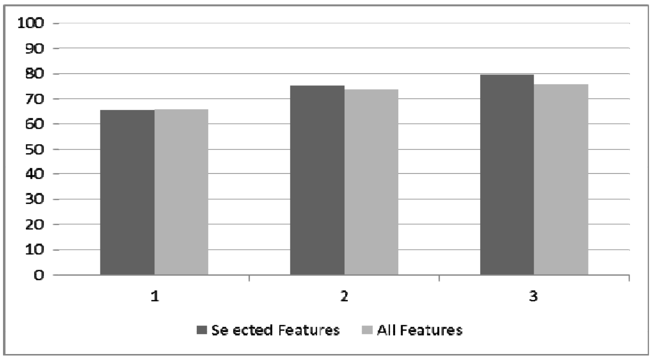


Fig. 4. The comparison of average accuracy

The results of experiments show that the number of selected features is almost 50% of all total number of features. However, the fewer used features does not make the achieved accuracy is lower, compare to the usage of all features for the detection process. Generally, the conducted experiments showed that the fewer used features for the detection process achieves higher accuracy compare to the all features are used for the detection process. The higher accuracy with fewer numbers of features is achieved since not all obtained features are considered as important features for the detection process. Therefore the selection feature process is required to obtain the best combination features for the detection process.

VI. CONCLUSION

Feature selection using Genetic Algorithms is proposed in this research, for the alcoholic detection from Electroencephalogram signal data. The genetic algorithm chooses the best combination features to represent the data, hence it is not necessary to use all extracted features for the alcoholic detection. Experiments showed that the selected features achieved higher accuracy than all features that are used for the detection. However the increasing accuracy is insignificant, approximately 5 % increased. For further research, the improvement is required in the proposed method. The improvement can be made on the selection data of the training process, and also the elitist strategy can be used to select the chromosome with higher fitness, therefore the generated chromosome for the next generation is, only the chromosomes with higher fitness value.

References

[1] O. Salem, A. Naseem, and A. Mehaoua, “Epileptic seizure detection from EEG signal using Discrete Wavelet Transform and Ant Colony classifier,” in *Communications (ICC), 2014 IEEE International Conference on*, 2014, pp. 3529–3534.

[2] D. Gajic, Z. Djurovic, S. Di Gennaro, and F. Gustafsson, “CLASSIFICATION OF EEG SIGNALS FOR DETECTION OF EPILEPTIC SEIZURES BASED ON WAVELETS AND STATISTICAL PATTERN RECOGNITION,” *Biomed. Eng. Appl. Basis Commun.*, vol. 26, no. 2, p. 1450021, Apr. 2014.

[3] R. Ebrahimpour, K. Babakhan, S. A. A. A. Arani, and S. Masoudnia, “Epileptic seizure detection using a neural network ensemble method and wavelet transform,” *Neural Netw. World*, vol. 22, no. 3, p. 291, 2012.

[4] G. Dogali and M. R. Bozkurt, “The detection of normal and epileptic EEG signals using ANN methods with Matlab-based GUI,” *Int. J. Comput. Appl.*, vol. 114, no. 12, 2015.

[5] S. S. P. Kumar and L. Ajitha, “Early detection of epilepsy using EEG signals,” in *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 2014, pp. 1509–1514.

[6] P. Kshirsagar and S. Akojwar, “Classification & Detection of Neurological Disorders using ICA & AR as Feature Extractor,” *Int. J. Ser. Eng. Sci. IJSES*, vol. 1, no. 1, Jan. 2015.

[7] V. Tuulik, A. Rajaa, A. Meister, and A. Lossman, “Neural network method to determine the vigilance levels of the central nervous system, related to occupational chronic chemical stress,” *Technol. Health Care*, vol. 5, no. 3, pp. 243–251, 1997.

[8] M. Kalaivani, V. Kalaivani, and V. A. Devi, “Analysis of EEG Signal for the Detection of Brain Abnormalities,” *Int. J. Comput. Appl. Year*, 2014.

[9] L. Hussain, W. Aziz, S. A. Nadeem, S. A. Shah, and A. Majid, “Electroencephalography (EEG) Analysis of Alcoholic and Control Subjects Using Multiscale Permutation Entropy.”

[10] Z. Ek, A. Akg, and M. R. Bozkurt, “The classificaton of eeg signals recorded in drunk and non-drunk people,” *Int. J. Comput. Appl.*, vol. 68, no. 10, 2013.

[11] N. T. Rachman, H. Tjandrasa, and C. Fatichah, “Alcoholism classification based on EEG data using Independent Component Analysis (ICA), Wavelet de-noising and Probabilistic Neural Network (PNN),” in *Intelligent Technology and Its Applications (ISITIA), 2016 International Seminar on*, 2016, pp. 17–20.

[12] L. C. Cheong, R. Sudirman, and S. S. Hussin, “Feature extraction of EEG signal using wavelet transform for autism classification,” *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 19, pp. 8533–8540, 2015.

[13] Z. Li *et al.*, “A genetic algorithm based wrapper feature selection method for classification of hyper spectral data using support vector maching, A genetic algorithm based wrapper feature selection method for classification of hyper spectral data using support vector maching,” *Geogr. Res.*, vol. 27, no. 3, pp. 493–501, May 2008.

[14] Y. Kaya and H. Pehlivan, “Feature selection using genetic algorithms for premature ventricular contraction classification,” in *2015 9th International Conference on Electrical and Electronics Engineering (ELECO)*, 2015, pp. 1229–1232.

[15] M. Hasanuzzaman, S. Saha, and A. Ekbal, “Feature subset selection using genetic algorithm for named entity recognition,” in *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, 2010.

[16] I. Rejer and K. Lorenz, “Genetic algorithm and forward method for feature selection in EEG feature space,” *J. Theor. Appl. Comput. Sci.*, vol. 7, no. 2, pp. 72–82, 2013.

[17] “Feature selection using genetic algorithm for classification of schizophrenia using fMRI data,” *J. Artif. Intell. Data Min.*, vol. 3, no. 1, 2015.

[18] S. Aalaei, H. Shahraki, A. Rowhanimanesh, and S. Eslami, “Feature selection using genetic algorithm for breast cancer diagnosis: experiment on three different datasets,” *Iran. J. Basic Med. Sci.*, vol. 19, no. 5, pp. 476–482, May 2016.

[19] I. Ahmad, “Feature Selection Using Particle Swarm Optimization in Intrusion Detection,” *Int. J. Distrib. Sens. Netw.*, vol. 11, no. 10, p. 806954, Oct. 2015.

[20] “UCI Machine Learning Repository: EEG Database Data Set: Support.” [Online]. Available: https://archive.ics.uci.edu/ml/support/eeg+database. [Accessed: 01-Oct-2016].