Feature selection with Lasso for classification of ischemic strokes based on EEG signals

Hendra Angga Yuwono, Sastra Kusuma Wijaya*, and Prawito Prajitno
Departmen Fisika, Fakultas Matematika & Ilmu Pengetahuan Alam, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia

*skwijaya@sci.ui.ac.id

Abstract. Electroencephalography (EEG) is an electrical signal data that can describe brain activity in which the signal contains important information that can be used to detect several diseases. One of the diseases that can be detected by EEG signals is stroke. The most common type of stroke is the acute ischemic stroke (AIS) due to blockage of blood supply to the brain which can generate the tissue damage in the brain EEG signal recording uses several electrodes where the more electrodes used in the recording, the greater the number of EEG features produced (high dimensional data). This can make it difficult for models of machine learning to have optimal performance on high-dimensional data. In this study, for optimizing the performance of the machine learning model by selecting features with the Least Absolute Shrinkage and Selection Operator (Lasso) method, where this method can select the relevant features by shrinking some coefficients to zero. The type of classification used in this study is random forest with features used for classification are Brain Symmetry Index (BSI), Delta-Alpha Ratio (DAR), Delta-Theta-Alpha-Beta Ratio (DTABR). The results showed that the Lasso method can optimize the performance of learning machines with an accuracy value of 75% with 24 features out of 45 features.

1. Introduction
Stroke is one of the diseases with the highest number that can cause paralysis and death. Stroke is a disease that can cause damage to the brain that appears suddenly which can result in impaired blood circulation in the brain. One of the highest incidence of stroke is the Acute Ischemic Stroke (AIS). AIS is a stroke caused by a blockage of blood that enters the brain causing tissue death in the brain. In general, to diagnose someone with ischemic stroke in general, using Magnetic Resonance Imaging (MRI) and Computerized Tomography Scan (CT-Scan). However, not all hospitals have these tools, such as Type-C hospitals in Indonesia. Because of the limitations of these tools the Electroencephalogram (EEG) can be used as an alternative tool. EEG is a device that can record brain activity with electrodes mounted on a person's head. The electrodes will produce brain wave signals in the form of electrical signals. From the electrical signals will be processed into features that will be input for machine learning. However, the more electrode usage the greater the amount of feature data possessed (high dimensional data).

The large amount of feature data, it can cause the machine learning model to not have optimal performance. In addition, the many features that can cause large memory usage and high computational costs [1]. That is because all features are not necessarily useful or important in classification therefore a feature selection method is needed. This study aims to find the best features
by means of the Least Absolute Shrinkage and Selection Operator (Lasso) feature selection method to improve the performance of the learning machine in classifying. Lasso is a regularization method that can be used to select high-dimensional features. In this paper, there are two types of classification classes, namely ischemic stroke and non-stroke. Some studies have classified ischemic stroke patients using features of EEG signals, such as Rahma using the Brain Symmetry Index (BSI), Delta-Alphato Ratio (DAR) and Delta-Theta-Alphato-Beta Ratio (DTABR) features with classifiers Extreme Learning Machine (ELM) [2]. Then with the same feature, Fitriah has added the Principal Component Analysis method with the XG Boost classifier [3]. Several studies related to feature selection methods using Lasso for EEG data [4]- [6].

2. Materials and methods
This section explains the flow of this research, which consists of signal acquisition to classification. We start from data collection, feature extraction, feature selection and classification process. In this research the signal acquisition and feature extraction process uses the library from MATLAB, while the feature selection and classification process uses the library of Python scikit-learn. The general structure of the proposed method can be seen in the following Figure 1.

![Figure 1. Research flow chart](image)

2.1. Data Collection
This EEG data was obtained from the Rumah Sakit Pusat Otak Nasional (RSPON) located in Jakarta and this data is the same data used in previous studies [2], [3], [7]. This data amounted to 60 patients consisting of 31 healthy patients and 29 acute ischemic stroke patients. The sampling frequency recording used is 512 Hz, which uses two different tools namely Xltex and Biologic. The duration of recording for 30 minutes for each patient with electrode placement in accordance with the international system 10-20. After that, the EEG data will be saved in European data format (.edf file). The number of electrodes used in recording is 32 channels but with reference to the research of Rahma (2017) and Fitriah (2017) this study only uses 18 electrode channels.

2.2. Extraction Features
In this study the features used consisted of BSI, DAR and DTABR. These three features were chosen because they have a strong correlation with ischemic stroke so they can be used for input machine learning [2], [3], [7]. The BSI feature was calculated using the Welch method while the DAR and DTABR features were calculated using the db4 Wavelet method [2][3]. The BSI feature itself is a feature that measures the difference between the right and left spectral power densities [3][8] with the following equation
\[ r_{BSI} = \frac{1}{M} \sum_{j=1}^{M} \frac{R_j^* - L_j^*}{R_j^* + L_j^*} \]  

\( L \) is the square value of the Fourier coefficient of the left hemisphere and \( R \) is for the right hemisphere. \( M \) is the frequency used, for BSI the distance frequency used is \( 1 - 25 \text{Hz} \). BSI is one way to determine the ischemic of the brain by calculating the symmetry of right and left brain waves, which can also be used for neurological deficits in stroke patients [10]. DTABR calculates the ratio between the absolute power difference of the delta and theta waves and the difference in the absolute power of alpha and beta and for DAR is a feature that calculates the ratio between the absolute power of the delta wave and the alpha wave [3], as shown in the following equation

\[ DAR = \frac{y_P(f_D)}{y_P(f_A)} \]  

\[ DTABR = \frac{y_P(f_D) + y_P(f_T)}{y_P(f_A) + y_P(f_B)} \]  

where \( f_A, f_B, f_D, f_T \) denotes alpha, beta, delta, theta and \( y_p \) denotes the absolute power of \( f \), with delta and theta present the brain's slow wave activity, alpha and beta represent the brain's fast wave activity [9].

2.3. Lasso

Least Absolute Shrinkage and Selection Operator (Lasso) is one type of embedded method that has been used in feature selection. Embedded feature selection method by embedding in a classifier construction called penalized feature selection [6]. It is a very useful method for regularization and feature selection [10]. Lasso is a linear regression with L1 regularization, which is shown in the equation [4]

\[ \frac{1}{2N} \sum_{i=1}^{N} (w_i - \alpha_0 - x_i^T \alpha)^2 + \lambda \sum_{j=1}^{n} \left| \alpha_j \right| \]  

where \( N \) is the number of samples, \( w_i \) is the target of the \( i \) sample, \( x_i \) is the \( n \)-dimensional input vector for the \( i \) sample, whereas for \( \lambda \) is a regularization parameter, \( \alpha_0 \) and \( \alpha \) are regression parameters where \( \alpha_0 \) is a scalar and \( \alpha \) is an \( n \)-dimensional vector. During the feature selection process, the non-zero coefficient of features after the shrinkage process is selected as part of the model. The purpose of this process is to minimize errors in predictions. Parameter \( \lambda \) that regulate the strength of the penalty, if the value of \( \lambda \) is large enough then there are no features used as input classification, whereas if the value of \( \lambda \) is close to zero or small then many features are selected [11]. In this study, the value of \( \lambda \) used is 0.0001, 0.001, and 0.01.

2.4. Random Forest (RF)

Random Forest (RF) is one of the best algorithms in terms of classification. Random forest was introduced in 2001 by Leo Breiman [12]. This algorithm uses the ensemble learning method in classification, which uses many decision trees at the training stage and uses the average of each tree at the testing stage. Random forest itself has parameters that can be optimized is the number of trees \( T \) and at each node a random number of features is chosen \( (K) \). The randomization process in the random forest was carried out on the sample data and on the predictor variable to the collection of trees [13]. In the study, the \( T \) value used was 100 with the value \( K = \sqrt{q} \) where \( q \) is the number of features used.
3. Results and Discussions

The features used in this study were 45 features consisting of 9 BSI, 18 DTABR and 18 DAR for each one patient. Of all the features used then selected using the Lasso method. The Lasso method looks for any features that are useful for the classification process. By adjusting the penalty value ($\lambda$), we get the number and types of features that can improve the performance of machine learning as shown in Table 1.

| $\lambda$  | DTABR | DAR  | BSI |
|------------|-------|------|-----|
| 0.0001     | 18    | 18   | 9   |
| 0.001      | 10    | 8    | 6   |
| 0.01       | 0     | 1    | 6   |

In Table 1, when the value of $\lambda = 0.0001$ the number of features selected totals 45 features, the number is equal to the total number of features of this study. The value of $\lambda = 0.001$ the number of features selected was 24 features consisting of 10 DTABR, 8 DAR and 6 BSI, and for $\lambda = 0.01$ the number of features selected was 7 features consisting of 1 DAR and 6 BSI. These results indicate that the smaller the value of lambda the more the number of features selected, conversely if the value of lambda increases, the fewer the number of features selected. That is because the weak features are forced to zero. After these features are selected, these features become an input for the classification of stroke and non-stroke. In this study the data sharing used is 20% for testing data and 80% for training data. The performance of the classification can be seen from the parameters used, in this study the parameters used are accuracy, recall and specificity as shown in Table 2 and Table 3.

| $\lambda$  | Accuracy (%) | Recall (%) | Specificity (%) |
|------------|--------------|------------|-----------------|
| 0.0001     | 100          | 100        | 100             |
| 0.001      | 100          | 100        | 100             |
| 0.01       | 98           | 96         | 100             |

| $\lambda$  | Accuracy (%) | Recall (%) | Specificity (%) |
|------------|--------------|------------|-----------------|
| 0.0001     | 58           | 57         | 60              |
| 0.001      | 75           | 71         | 80              |
| 0.01       | 67           | 62         | 75              |

Accuracy scores are obtained from calculating correctly predicted data against the overall data. Recall is obtained from counting people who actually had a stroke in the population compared to people who actually had a stroke in the classification while for specificity by counting people who were actually non-strokes compared to people who were not a stroke in the classification. In Table 2, shows the results of the performance of the training classification shows satisfactory results in classifying for all $\lambda$ values with values of all parameters above 95%. In Table 3, shows the performance of the predicted results that when $\lambda = 0.001$ get the best performance with the value of all parameters above 70%. While for the worst performance results are generated when $\lambda = 0.0001$ with the value of all parameters below 70%. From these results it shows that there is very overfitting.
when the value of $\lambda = 0.0001$ and $\lambda = 0.01$, which means that during training has a good performance while when given new data can not predict it well. Compared with the previous work, the classification accuracy of Ischemic Strokes with the Lasso feature selection method reaches 75% which is 8% higher than the accuracy done by Fitriah [3], 3% higher than the work done by Rahma [2]. From these results, it shows that the feature selection using the Lasso method can improve the performance of the Ischemic Strokes classification.

4. Conclusion
The results of feature selection using the Lasso method show that the Lasso method can improve the performance of the classification. The best results are when using 24 features consisting of 10 DTABR, 8 DAR and 6 BSI. Our work in the future will add variations to the penalty value of the Lasso method or choose other feature selection methods that have the potential to get the best results.

Acknowledgement
This work is supported by research grant of Indexed International Publication of Student Final Project (Hibah Publikasi Internasional Terindeks untuk Tugas Akhir (PITTA) Mahasiswa), Universitas Indonesia, Grant No. NKB-0667/UN2.R3.1/HKP.05.00/2019.

References
[1] Janecek A Gansterer, W Demel M, and Ecker G 2008 On the relationship between feature selection and classification accuracy Workshop and Conference Proceedings 4 pp 90-105.
[2] Rahma O N, Wijaya S K, Prawito, and Badri C 2017 Electroencephalogram analysis with extreme learning machine as a supporting tool for classifying acute ischemic stroke severity 2017 International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM) doi:10.1109/issimm.2017.8124287
[3] Fitriah N, Wijaya S K, Fanany M I, Badri C, and Rezal M 2017 EEG channels reduction using PCA to increase XGBoost’s accuracy for stroke detection AIP Conf. Proc. 1862 doi:10.1063/1.4991232
[4] Feng J K, Jin J, Daly I, Zhou J, Niu Y, Wang X, and Cichocki A 2019 An Optimized Channel Selection Method Based on Multifrequency CSP-Rank for Motor Imagery-Based BCI System Computational Intelligence and Neuroscience 1–10 doi:10.1155/2019/8068357
[5] Wang J-J, Xue F, and Li H 2015 Simultaneous Channel and Feature Selection of Fused EEG Features Based on Sparse Group Lasso BioMed Research International 1–13 doi:10.1155/2015/703768
[6] Wenting Tu and Shiliang Sun 2010 Spatial Filter Selection with LASSO for EEG Classification (China: East China Normal University)
[7] Chan J Y, Wijaya S K, Prawito and Rahma O N 2019 ELM (extreme learning machine) method for detecting acute ischemic stroke using conventional and specific asymmetry BSI (brain symmetry index) features based on EEG signals AIP Conf. Proc. 2092 doi:10.1063/1.5096691
[8] M J A M Putten, J M Peters, S M Mulder, J A M de Haas, C M A Bruijninckx and D L J Tavy 2004 A brain symmetry index (BSI) for online EEG monitoring in carotid endarterectomy Journal of Clinical Neurophysiology 115 1189-1194
[9] S P Finnigan, M Walsh, S E Rose and J B Chalk 2007 Quantitative EEG indices of sub-acute ischaemic stroke correlate with clinical outcomes Clinical Neurophysiology 118(11) 2525-2532
[10] Fonti V and Belitser E 2017 Feature selection using lasso (VU Amsterdam Research Paper in Business Analytics)
[11] Kaushik Shruti & Choudhury, Abhinav & Jatav, Ashutosh & Dasgupta, Nataraj & Natarajan, Sayee & Pickett, Larry & Dutt, Varun, 2019 Comparative Analysis of Features Selection Techniques for Classification in Healthcare Lecture Notes in Computer Science In press
[12] Geurts P, Ernst D, and Wehenkel L 2006 Extremely randomized trees Machine Learning 63(1) 3–42 doi:10.1007/s10994-006-6226-1

[13] Kuswanto H, Salamah M and Fachruddin M I 2017 Random Forest Classification and Support Vector Machine for Detecting Epilepsy using Electroencephalograph Records American Journal of Applied Sciences 14(5) 533–539 doi:10.3844/ajassp.2017.533.539