Combination of Coarse-Grained Procedure and Fractal Dimension for Epileptic EEG Classification

Dien Rahmawati*1, Achmad Rizal2, Desri Kristina Silalahi3
1,2,3School of Electrical Engineering, Telkom University, Bandung, Indonesia
e-mail: *1dienrahmawati@telkomuniversity.ac.id, 2achmadrizal@telkomuniversity.ac.id, 3desrikristina@telkomuniversity.ac.id

Abstract
Epilepsy, cured by some offered treatments such as medication, surgery, and dietary plan, is a neurological brain disorder due to disturbed nerve cell activity characterized by repeated seizures. Electroencephalographic (EEG) signal processing detects and classifies these seizures as one of the abnormality types in the brain within temporal and spectral content. The proposed method in this paper employed a combination of two feature extractions, namely coarse-grained and fractal dimension, a challenge to obtain a highly accurate procedure to evaluate and predict the epileptic EEG signal of normal, interictal, and seizure classes. The result of classification accuracy using variance fractal dimension (VFD) and quadratic support machine vector (SVM) with a number scale of 10 is 99%, which is the highest one, excellent performance of the predictive model in terms of the error rate. In addition, a higher scale number does not determine a higher accuracy in this study.

Keywords—epilepsy, EEG classification, coarse-grained, fractal dimension, support vector machine

Received October 19th, 2021; Revised October 30th, 2021; October 31st, 2021
1. INTRODUCTION

Several types of brain disorders can be experienced by humans, such as dementia, stroke, epilepsy, Alzheimer’s, and brain cancer. However, the most common brain disorder that most people suffer from is epilepsy. According to WHO, as many as 50 million people worldwide have epilepsy, so it is not wrong to say it is the most common brain disorder [1].

Epilepsy can affect anyone, from any background, and at any age. One of the characteristics of epilepsy is the presence of seizures that repeatedly occur in certain parts of the body or the whole body, which can also cause fainting and not being able to control bladder function. Several previous studies have classified EEG signals from epilepsy using various feature extraction methods, including the multilevel wavelet-based entropy (MWPE) method with the highest accuracy of 94.3% [2], Katz Fractal with the highest accuracy of 98.7% [3], Higuchi Fractal Dimension (HFD) with the highest accuracy of 97.3% [4], Multidistance Signal Level Difference with the highest accuracy of 97.7% [5]. However, with the complexity of EEG signals from epilepsy, researchers continue to look for the best method for classifying EEG signals from epilepsy, including those carried out in this study.

In addition to feature extraction methods, the classifiers used by previous researchers also vary, including support vector machine (SVM) [3], convolutional neural network [6], and deep learning [7]. Support Vector Machine is a classifier that is often used to detect and predict epilepsy [8][9]. From studies [8][9], it is known that SVM can have a sensitivity of more than 90% for predicting epilepsy.

In this study, two feature extractions were used, namely the coarse-grained procedure and the fractal dimension. The coarse-grained procedure is performed to divide the signal into several scales. After being divided into several scales, the next step is to look for feature extraction, namely fractal dimensions. The fractal dimension is crucial because we need a measure for fractal complexity in the discussion of fractals. The measure of the self-similarity of the signal is called the fractal dimension. By using the fractal dimension, it is expected that the size of the self-similarity level of the EEG signal is assumed to be different for each type of EEG signal. The combination of coarse-grained procedures is expected to make signal differences at different scales more prominent.

2. METHODS

Figure 1 shows the method used in this study. The EEG signal is entered in a coarse-grained procedure to obtain several new signals with different scales. For each signal, the fractal dimension is calculated to characterize the EEG signal. Furthermore, the value of this fractal dimension will be classified by SVM to be assessed as a normal, interictal, or seizure EEG signal. The following subsections will explain the details of each process.

EEG Dataset

At this time, many datasets can be accessed by anyone, one of which is the EEG Dataset from the University of Bonn[10]. Many previous studies[2], [3], [5] have used the EEG dataset. Likewise, in this study, we also used the EEG Dataset from the University of Bonn. The selection of the EEG Dataset includes a large number of data, namely 300 data with a total of 3 classes, namely normal, epilepsy with seizures, and epilepsy without seizures. In addition, the data is free from artifacts noise because it was recorded using a sampling frequency of 173.61 Hz and then filtered using a 40 Hz Low Pass Filter. Epilepsy data were derived from pharmaco-resistant focal-onset epilepsy patients undergoing preoperative evaluation. The recorded EEG data from this epilepsy patient was recorded by the Department of Neurology at the University of Bern. Examples of signal data from normal, interictal, and seizure classes are shown in Figure 2.
Combination of Coarse-Grained Procedure and Fractal Dimension for ... (Dien Rahmawati)
2.1 Coarse-grained Procedure

Physiological signals such as EEG signals have multiple time scales. However, conventional algorithms cannot be used for various time scales. Costa et al. [11] have developed a new algorithm that can calculate multiscale entropy (MSE) for complex time series, which is called the coarse-grained procedure. In general, the concept of the coarse-grained procedure is to use down sampling and smoothing [12]. Equation coarse-grained procedure as in Eq. 1 [13]:

\[ y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=1}^{\tau} y_{j+1}^{(\tau)} , \quad 1 \leq j \leq N \]  

Where \( \{x_1, x_2, ..., x_i, ..., x_n\} \) is a 1-dimensional time series, \( y_j^{(\tau)} \) is a consecutive coarse-grained time series, \( \tau \) is the scale factor, and \( N \) is the length of the original time series. The scale used in this study is 1 to 20. Please note that for a scale factor of 1, it is the same as the raw signal. If \( \tau = 2 \), then \( y_j^{(2)} = \frac{[x(2i-1)+x(2i)]}{2} \), meanwhile if \( \tau = 3 \), then \( y_j^{(3)} = \frac{[x(3i-2)+x(3i-1)+x(3i)]}{3} \) and so on. Thus, the coarse-grained procedure is a process of decomposing signals at different scales or levels. Graphically the coarse-grained procedure is as in Figure 3.

![Figure 3 Coarse-grained procedure](image)

2.2 Fractal Dimension

Fractal dimension is a parameter used to define signal complexity. The value represents the level of self-similarity, which is a repeating pattern that occurs in several different signals [14]. The value of the fractal dimension will increase if the emergence of self-similarity signal patterns is increasing. In addition, the higher the fractal dimension value, the more complex the signal being tested [15]. The ability of fractal dimensions to define signal complexity makes it a great tool for modeling biological signals, such as EEG and ECG, which tend to have high complexity and irregularity. In addition, fractal dimensions are very well used to analyze nonlinear behavior and state in chaotic systems such as EEG [16]. There are several kinds of fractal dimensions used in this study, including Box Counting[14], Higuchi Method[17], Katz Method[18], Sevcik Method[15], Variance Method[19], Petrosian C[20], Petrosian D[20].

**Box-Counting Method**

The box-counting method is one of the methods used to analyze fractals [21]. The technique used by the box-counting way is to use the box as a measure. The box used will cover the figures or curves for which the fractal dimension will be searched. The value of \( r \) determines the box size. Meanwhile, the number of boxes that cover figures or curves is represented by \( N \).
With the representation of \( r \) and \( N \), the total number of boxes used to cover curves is affected by the size of the box. Mathematically, the box-counting method can be written as in Eq. 2

\[
D_B = \lim_{r \to 0} \frac{\log N}{\log(1/r)}
\]  

(2)

From Equation 2, it can be concluded that the box-counting dimension is obtained from the ratio of \( \log N \) to \( \log(1/r) \), where \( r \) approaches 0. With this statement, it can be said that when the box size comes to 0, then the overall figures or curves will be covered by a box. In simple terms, the number of boxes with different sizes to cover the curves is done by the box-counting method. Calculations with various sizes of this box, of course, require a long calculation time. However, even so, the box-counting method has the advantage that it can be implemented for complex or simple fractals on natural and artificial fractals [22].

2.2.1 Katz Method

Katz method is a fractal dimension calculation method using curve length and planar extent of waveforms [23]. Fractal dimension calculation equation using Katz method for a signal \((x_i, y_i)\) of length \(N\) is defined as in Eq. 3 [18].

\[
KFD = \frac{\log_{10}(n_2)}{\log_{10}(d/\text{dist}(1, i))^{\log_{10}(n_2)}}
\]

(3)

where \( Lc \) is total curve length, and \( d \) is the maximum distance of two sequential points between the initial point \((x_1, y_1)\) to the maximum distance or the curve diameter. The value of \( Lc \) can be defined using Eq. 4.

\[
Lc = \sum_{i=1}^{N-1} \text{dist}(i, i + 1)
\]

(4)

where \( \text{dist} (i, i + 1) \) is a distance of two sequential points. The value of \( d \) can be expressed in Eq. 5.

\[
d = \max [\text{dist}(1, i)], \quad i = 2, \ldots, N
\]

(5)

2.2.2 Sevcik Method.

Sevcik Method (SFD) is a method to get the fractal dimension value by using a set of \(N\) sample values from a waveform [24], [25]. This method is derived from the Hausdorff Dimension derivative \((D_n)[25]\). The equation for calculating fractal dimensions using the Sevcik Method is shown in Eq. 6[15].

\[
SFD = 1 + \frac{\ln (Lc)}{\ln(2, (N-1))}
\]

(6)

On Eq. 6, \( Lc \) is the total length as expressed in Eq. 4. Then, \( N \) is the sample value. There is another variation of the Sevcik Method, namely normalization in the x and y-axis before implementing \( Lc \) and SFD calculation. Normalization of the x and y-axis causes the topology of the metric space to not change under linear transformation so that all axes are the same[25]. Eq. 7 displays the normalization equation on the x-axis.

\[
x_i' = \frac{x_i}{x_{max}}, \quad i = 1, \ldots, N
\]

(7)
where \( x_i \) is the initial value in x-axis and \( x_{\text{max}} = \max_{x_i} \). Eq. 8 displays the normalization equation on the y-axis.

\[
y_i^* = \frac{y_i - y_{\text{min}}}{y_{\text{max}} - y_{\text{min}}}, \quad i = 1, ..., N
\]  

where \( y_i \) is the initial value in x-axis, \( y_{\text{min}} = \min_{y_i} \), and \( y_{\text{max}} = \max_{y_i} \).

### 2.2.3 Variance Method

The variance fractal dimension method (VFD) is a method that calculated \( s(t) \) using Hurst Exponent (H), whose derived from the properties of fractions Brownian motion (\( fBM \)) as expressed in Eq. 9.

\[
H = \lim_{\Delta t \to 0} \frac{\log(\exp(\log(\Delta s_{\Delta t})))}{\log(\Delta t)}
\]  

where \( H \) is signal smoothness, \( (\Delta s)_{\Delta t} = s(t2) - s(t1) \) and \( \Delta t = t_2 - t_1 \). The VFD can be determined as Eq. 10.

\[
VFD = E + 1 - H
\]  

where \( E \) is Euclidean dimension. Because of the Euclidean one-dimensional signal values is 1, then VFD equation can be simplified as Eq. 11.

\[
VFD = 2 - H
\]

The value of \( \Delta t \) on the VFD can be different depending on the purpose. If VFD used to separate signal and noise, then the value of \( \Delta t \) is 1. Then, if it is used to separate several data components, then the value of \( \Delta t \) is more than 1.

### 2.3 Support Vector Machine

Support Vector Machine is one of the supervised learning methods in machine learning. The Support Vector Machine works by creating an input-output mapping based on a set of labeled training data. Support Vector Machine can be used for classification and regression. Basically, the Support Vector Machine can separate data by finding the best hyperplane by finding the maximum margin. This method can be used for high dimension space. There are several types of classifiers in the Support Vector Machine, including Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian.
Figure 4 Illustration of class separation using the Support Vector Machine

2.4 K-Fold Cross-Validation

One way to validate machine learning is to use K-Fold Cross-Validation. K-Fold Cross-Validation is a method for validation by dividing the data set into several parts or folds, depending on the K value. Furthermore, each fold is used as a test set at several points [26]. In this study using K = 5, the data set used in this study will be divided into five parts. After being divided into five parts, the first iteration is carried out where the first part is used to test the model, and the remaining four parts are used to train the model. Then, in the second iteration, we use the second part as a test set and parts 1, 3, 4, and 5 as a training set. These processes are repeated until all parts are used as a test set.

3. RESULTS AND DISCUSSION

Epilepsy EEG signals in this study were divided into three classes, namely normal (O), interictal (N), and seizure (S). The result of the coarse-grained process is changing the signal variance, not changing the signal's shape. This change is caused by differences in the scale of the coarse-grained process. The relationship between signal variance and scale (τ) is inversely related, i.e., the larger the scale (τ), the smaller the signal variance. Vice versa, the smaller the scale (τ), the greater the signal variance. After the coarse-grained procedure is performed, the fractal dimension calculation process is then carried out. Figure 1-7 shows a graph of the fractal dimension values of 5 kinds of fractal dimensions with a scale of τ = 1-20. The fractal dimension values in the BCFD and HFD methods range from 1-2, but it is more than that for KFD, SVD, and VFD. This is partly due to the lack of KFD, SVD, and VFD ability in calculating the fractal dimension. However, these results were not a problem in this study. Due to the calculation of the fractal dimension in this study, it is used for features only.

Based on the fractal dimension calculation using HFD, KFD, SFD, VFD, the graph of the fractal dimension value to the scale has a similar pattern. Namely, the fractal dimension value to the scale has a similar pattern. Namely, the fractal dimension value will increase to a specific scale then the fractal dimension value will stagnate. However, in contrast to the fractal dimension calculation using BCFD, the fractal dimension value tends to be high when the scale is small. Still, the value decreases in comparison to the increasing scale. As the results of BCFD calculations, these results are used to classify lung sounds [27].

The fractal dimension value of the epilepsy class with seizures (S) tends to always be higher than normal class (O) and interictal (N). Meanwhile, the normal class (O) in several fractal dimensions, namely BCFD and KFD, has the lowest value. However, initially, at SVD and VFD, the normal class (O) had the lowest score, but after passing the scale of 10, it began to rise and became higher than the interictal (N). These results show that although seizure (S), normal (O), and interictal (N) have similar patterns, they have different features of fractal dimensions. This is what can be used to separate the three classes using a classifier, namely SVM.
Figure 5 Higuchi Fractal Dimension value with scale τ = 1-20

Figure 6 Box-counting Fractal Dimension values with scale τ = 1-20

Figure 7 Sevcik Fractal Dimension value with scale τ = 1-20

Figure 8 Katz Fractal Dimension value with scale τ = 1-20

Figure 9 Variance Fractal Dimension value with scale τ = 1-20

Accuracy using five kinds of fractal dimensions, namely VFD, SFD, BCFD, KFD, and HFD with six kinds of SVM classifier namely Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian SVM, shown in Table 1 and Table 2. The highest accuracy is 99% by using a VFD with a scale of 10 and a quadratic SVM classifier. In contrast, the lowest accuracy is 60.7% on fractal dimensions BCFD with a scale of 20 and the Fine Gaussian SVM classifier type. When viewed from various classifiers and their scales, the best accuracy is using fractal dimension VFD because almost all accuracy is more than 98%. Only when using the Medium Gaussian SVM and Coarse Gaussian SVM classifiers with a scale of 10 have accuracy below 98%, namely each 94.7% and 87.7%, respectively. Meanwhile, the worst accuracy is using BCFD, and the accuracy is between 60.7% to 74.3%.

When examined for each fractal dimension method, one method with another method does not always have the same accuracy at the same scale and classifier. In BCFD, the best accuracy occurs on a scale of 10 with the Fine Gaussian SVM classifier, with a percentage of 74.30%. Then, on HFD, the best accuracy is on a scale of 10 with the Quadratic SVM and Cubic
SVM classifiers, with a percentage of 95.70%. Then, on KFD, the best accuracy is on a scale of 20 with the Medium Gaussian SVM classifier, with a percentage of 88.30%. In SFD, the best accuracy is 98.30%, on a scale of 20, and the Classifier Linear SVM. Finally, on the VFD, the best accuracy is 99.00%, obtained on a scale of 10, and the Quadratic SVM classifier.

From the calculation of epilepsy EEG accuracy using a combination of coarse-grained and fractal dimensions, it can be seen that the higher the scale size does not always represent high accuracy. It is proven that the highest accuracy is on a scale of 10 (using VFD and Quadratic SVM), not on a scale of 20.

**Table 1 Accuracy of BCFD, HFD, KFD**

| Classifier            | BCFD | HFD | KFD |
|-----------------------|------|-----|-----|
|                       | 10   | 15  | 20  | 10  | 15  | 20  | 15  | 20  |
| Linear SVM            | 68.70| 72.30| 70.70| 92.70| 93.30| 92.30| 92.30| 92.30| 92.00| 92.30| 92.00|
| Quadratic SVM         | 72.70| 70.30| 86.67| 95.70| 92.00| 90.70| 93.30| 93.00| 91.70| 91.70| 91.70|
| Cubic SVM             | 62.30| 65.30| 67.70| 95.70| 95.70| 91.70| 95.70| 95.70| 95.70| 91.70| 91.70|
| Medium Gaussian SVM   | 74.30| 71.00| 60.70| 92.00| 92.30| 91.70| 91.70| 91.70| 91.70| 91.70| 91.70|
| Coarse Gaussian SVM   | 74.30| 72.30| 70.70| 91.30| 92.30| 91.70| 91.70| 91.70| 91.70| 91.70| 91.70|

**Table 2 Accuracy of SFD, VFD**

| Classifier            | SFD | VFD |
|-----------------------|-----|-----|
|                       | 10  | 15  | 20  | 10  | 15  | 20  |
| Linear SVM            | 97.70| 98.00| 98.30| 98.70| 98.70| 98.00|
| Quadratic SVM         | 98.30| 98.00| 98.00| 99.00| 97.00| 98.70|
| Cubic SVM             | 97.00| 98.00| 97.70| 98.70| 98.00| 98.00|
| Fine Gaussian SVM     | 97.30| 97.30| 98.00| 98.00| 98.00| 98.00|
| Medium Gaussian SVM   | 96.70| 98.00| 94.70| 98.00| 98.00| 98.00|
| Coarse Gaussian SVM   | 86.30| 96.70| 97.00| 87.70| 98.00| 98.00|

**Table 3 Comparison with other studies using fractal dimensions and SVM**

| Reference | Data classes                | Method | Number of features | Accuracy (max) |
|-----------|----------------------------|--------|-------------------|----------------|
| [3]       | Normal, interictal, seizure (3 classes) | The Katz fractal dimension (KFD) on EEG sub band | 5 | 98.7% |
| [4]       | Normal, interictal, seizure (3 classes) | Multi time-interval HFD | 7 | 98% |
| [28]      | Normal, epileptic (2 classes) | HFD, KFD, SFD | 25 | 100% |
| [29]      | Normal, interictal, seizure (3 classes) | Multidistance signal level difference (MSLD) and fractal dimension | 10 | 99% |
| Proposed method | Normal, interictal, seizure (3 classes) | Coarse-grained procedure and FD | 10 | 99% |

Table 3 shows a comparison of previous studies that used fractal dimensions and SVM as a classifier. KFD calculated on the Alpha, Beta, Delta, Theta, and Gamma sub bands of EEG signals yielded the highest accuracy of 98.7% [3]. Meanwhile, the calculated HFD at various time-interval resolutions produces an accuracy of 98% [4]. In the study by Silalahi et al., MSLD was used to split the ECG and FD signals as a feature resulting in 99% accuracy. The difference with this research is in the signal splitting process using MSLD with the same FD calculation. The accuracy of both is the same, which is achieved with the same number of
features. The proposed method produces accuracy that is proven to have high accuracy compared to previous studies.

The weakness of the proposed method is the determination of the scale on the coarse-grained procedure is done by trial and error. However, this can be overcome by looking at the resulting characteristics, as shown in Figure 5 - Figure 9. Seen at a scale > 10, the differences between data classes become smaller. The advantages of this method are simple computation and the possibility to perform feature selection to increase accuracy. The use of feature subset selection to select features at a specific scale gives the possibility of increasing accuracy. The use of this method on other biomedical signals is interesting for further research.

4. CONCLUSION

In this study, classifying epileptic EEG signals is proposed using a coarse-grained procedure and fractal dimensions. The coarse-grained procedure process will produce several signals with different scales, then calculate the fractal dimensions. A coarse-grained approach helps break signals in various scales to see the change of signal from its coarse form to a more acceptable form than the process of averaging several signal scales. Meanwhile, the fractal dimension helps calculate the complexity of the EEG signal. Classification using SVM produces the highest accuracy of 99% using VFD. A larger scale that reflects a more significant number of traits does not result in higher accuracy. The proof is the highest accuracy is achieved when using ten scales, not when using 20 scales. The proposed method produces a reasonably high accuracy compared to other methods that use fractal dimensions and SVM. The proposed method has the potential for use in other biomedical signals.

REFERENCES

[1] WHO, “Epilepsy: the disorder,” Epilepsy Atlas, 2005.
[2] I. Wijayanto, A. Rizal, and S. Hadiyoso, “Multilevel Wavelet Packet Entropy and Support Vector Machine for Epileptic EEG Classification,” 2018, doi: 10.1109/ICSTC.2018.8528634.
[3] I. Wijayanto, A. Rizal, and A. Humairani, “Seizure Detection Based on EEG Signals Using Katz Fractal and SVM Classifiers,” 2019, doi: 10.1109/ICSIITech46713.2019.8987487.
[4] A. Rizal, Estananto, and R. D. Atmaja, “Epileptic EEG signal classification using multiresolution higuchi fractal dimension,” Int. J. Eng. Res. Technol., vol. 12, no. 4, 2019.
[5] A. Rizal and S. Hadiyoso, “Sample entropy on multidistance signal level difference for epileptic EEG classification,” Sci. World J., vol. 2018, 2018, doi: 10.1155/2018/8463256.
[6] R. San-Segundo, M. Gil-Martín, L. F. D’Haro-Enríquez, and J. M. Pardo, “Classification of epileptic EEG recordings using signal transforms and convolutional neural networks,” Comput. Biol. Med., vol. 109, 2019, doi: 10.1016/j.compbiomed.2019.04.031.
[7] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, “Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification,” Front. Neurol., vol. 11, 2020, doi: 10.3389/fneur.2020.00375.
[8] H. T. Shiao et al., “SVM-Based System for Prediction of Epileptic Seizures from iEEG Signal,” IEEE Trans. Biomed. Eng., 2017, doi: 10.1109/TBME.2016.2586475.
[9] P. W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, “Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG,” 2008, doi: 10.1109/MLSP.2008.4685487.
Combination of Coarse-Grained Procedure and Fractal Dimension for … (Dien Rahmawati)
[28] M. Schneider, P. N. Mustaro, and C. A. M. Lima, “Automatic recognition of epileptic seizure in EEG via support vector machine and dimension fractal,” in 2009 International Joint Conference on Neural Networks, Jun. 2009, pp. 2841–2845, doi: 10.1109/IJCNN.2009.5179059.

[29] D. K. Silalahi, A. Rizal, D. Rahmawati, and B. Sri, “Epileptic Seizure Detection using Multidistance Signal Level Difference Fractal Dimension and Support Vector Machine,” J. Theor. Appl. Inf. Technol., vol. 99, no. 4, pp. 909–920, 2021.