Classification of Baby Cry Sound Using Higuchi’s Fractal Dimension with K-Nearest Neighbor and Support Vector Machine

D Widhyanti¹ and D Juniati²
¹Undergraduate Mathematics Student, Universitas Negeri Surabaya
²Department of Mathematics, Universitas Negeri Surabaya
dyahwidhyanti16030214012@mhs.unesa.ac.id, dwijuniati@unesa.ac.id

Abstract. The crying baby sound is a way to express their physical and psychological conditions. For most people, it is difficult to distinguish the meaning of the sound of a baby's cry because they are almost all similar. In fact, sometimes parents still lack knowledge of understanding their baby's condition from the sound of crying. Therefore, research to classify the types of baby crying sounds mathematically becomes very interesting. This study aimed to classify the types of crying sounds for babies using fractal dimensions, particularly the Higuchi Fractal dimension. The data used in this study were 80 data consisting of 4 types of cries, namely hunger cries, tired cries, stomach ache cries, and uncomfortable cries. In this study k-max values of 10, 16, and 50 were selected as experiments. The results of the fractal dimension value of each sound signal were carried out in the data grouping process. The k-fold cross validation data distribution method with values of k = 5 and 10 were used. The classification process was carried out using the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) methods. Based on the research results, the best accuracy was 78.75% at level 5 decomposition with 5-fold cross validation, k-max value = 10, and K = 9 value on the KNN method. Whereas with the SVM method, the best accuracy was 80% at level 5 decomposition with 10-fold cross validation, k-max value = 10 and c = 10 and using RBF kernel with γ = 10. So in this study, the Support Vector Machine (SVM) method was slightly better than the K-Nearest Neighbor (KNN) method.

1. Introduction
The sound of a baby crying is one sign that the baby is having a problem. Babies cry to express their physical and psychological conditions, such as hunger, pain, fatigue, discomfort, and so on. From the sound of crying babies, it is hoped that parents can understand the purpose of crying because babies cannot communicate directly in language that can be understood by adults. But most people, even parents, still lack the knowledge to understand the baby's condition from the sound of crying. This can confuse and misinterpret the baby's needs and conditions.
Several studies related to the detection of baby sounds have been conducted, for example, automatic detection of crying babies in audio recordings using deep learning method. In the study of Torres [1] and Lavner [2], they used the method of extracting baby crying sound features with the Mel-Frequency Cepstrum Coefficient (MFCC) algorithm and the convolutional neural network (CNN)
algorithm and produced the best results with an accuracy of 82.5%. While in other studies, the application of baby crying sound detection used the method of extracting baby cry sound features using the Mel-Frequency Cepstrum Coefficient (MFCC) and K-Nearest Neighbor (K-NN) algorithm as a classification method. In the study, the best accuracy rate was 79.95% in the classification with k=1 [3]. The research on the recognition of hunger through the sound of crying babies in infants aged 0-9 months was also conducted using Continuous Wavelet Transformation and Neural Network with a percentage of accuracy in recognizing speech signal of 89.7% [4]. In another study, the introduction and classification of infant crying using the Support Vector Machine (SVM) algorithm yields an accuracy value of 76.56% [5].

Fractal geometry is a branch of geometry that studies irregular shapes that are often found in nature that are not found in classical geometry. One of the characteristics of fractal geometry is the fractal dimension which can be valued for non-integer fractions. The concept of the Fractal dimension with the value of a fraction has opened new insights and has been applied in various fields. Safitri [6] classified diabetic retinopathy using fractal dimension analysis of eye fundus image and obtained the accuracy of 89.17%. Khotimah [7] studied iris recognition using feature extraction of box counting Fractal Dimension with the accuracy obtained was 92.63%. Juniati [8] used Fractal Dimension to classify the heart sound recording and the accuracy obtained was 86.17%, and many more. The Higuchi Fractal Dimension is one of several methods to calculate the fractal dimension in the form of time series data. In many cases, the use of the Higuchi fractal dimension is more appropriate than the spectral exponent for analyzing irregular time series. Therefore, research to classify the types of baby crying sounds mathematically becomes very interesting, specially using fractal dimensions. In this study, the Higuchi Fractal dimension was used.

1.1. Baby Language

Newborns are communicating by crying. The baby's cry is a response to internal or external stimuli. Babies cry as a basic instinctive form of communication. Neonatologists can distinguish between different types of cry, although the interpretation of the cry can be subjective and depend on the experience of the listener [9].

Dunstan Baby Language or DBL was first introduced by Priscilla Dunstan, a musician from Australia who has the ability to remember all types of sounds called sound photographs. According to Priscilla, there are 5 types of universal baby crying (applicable to all babies in the world) which have their respective meanings, namely "neh" means hungry, "owh" means tired, "heh" indicates discomfort, "earh" means stomach ache, and "eh" is burp.

In this study, the sound of crying babies was classified into four groups, namely conditions of hunger, fatigue, stomach ache and discomfort based on the fractal dimensions of the baby crying sound signal using the Higuchi method. The baby crying sound signal in different conditions is given in Figure 1 below.

![Baby crying sound signal](image)

**Figure 1.** Baby crying sound signal

2. Classification process of baby crying sound recording

2.1. Feature Extraction

Feature extraction is the process of taking the features of an object that can describe the characteristics of the object. Wavelets are used for the frequency-time analysis of signals, signal compression, signal denoising, and feature extraction [10]. Wavelet transform is a transformation used to analyze moving
signals. Wavelet transform provides a time-frequency representation of a signal. Wavelet transform is better to show the change in frequency with time than the Fourier transform [11]. Discrete Wavelet Transform (DWT) is the transformation of a discrete signal into a wavelet coefficient which is obtained by filtering the signal using two opposite filters, namely the low pass filter and the high pass filter. Fast Fourier Transform (FFT) is a method for transforming sound signals from the time domain into the frequency domain to see the frequency range.

2.2. Higuchi Fractal Dimension
The fractal dimension of an object shows the characteristics of the object in the form of a measure of the complexity of the object as the ratio of its detailed change to scale.[12]. One of the algorithms for calculating fractal dimensions is the Higuchi algorithm. The fractal dimension can show how complex an object is, both real and non-real objects. The fractal dimensions can indicate how complex the leaf margins of a plant are, how dense the tofu is, or how curvy the edges of skin cancer lumps are from MRI images [13]. The Higuchi method is one of the methods used to calculate the fractal dimension value of a waveform. This algorithm is an algorithm for calculating the fractal dimensions of a curve and is suitable for time series data. Suppose given a time series \( X[k] \) where \( k = 1, 2, \ldots, N \). The algorithm for calculating the dimensions of the Higuchi fractal in the time series is as follows [8]:

1. The \( k \) form of the new time series \( X_n^k \) is as follows:
   \[
   X_n^k = \{ X[n], X[n+k], \ldots, X\left[ n + \text{int}\left( \frac{N-n}{k} \right) \cdot k \right] \}
   \] (1)
   Where \( n \) and \( k \) are integers, \( k \) indicates discrete time intervals, while \( n = 1, 2, 3, \ldots, k \) indicates the value of the initial time.
2. The length of each new time series can be defined as follows:
   \[
   L(n, k) = \left\lfloor \left( \sum_{i=1}^{\text{int}\left( \frac{N-n}{k} \right)} |X[n+i.k] - X[n+(i-1).k]| \right)^{\frac{N-1}{\text{int}\left( \frac{N-n}{k} \right) . k}} \right\rfloor
   \] (2)
   Where \( N \) is the length of the original time series, \( \text{int}\left( \frac{N-n}{k} \right) \cdot k \) denotes the normalization factor and \( |X[n+i.k] - X[n+(i+1).k]| = h_i \). Thus \( L(n, k) \) is the normalized number of new segment lengths \( h_i \). Each \( h_i \) denotes a different distance value on the coordinates of the n-distance pair of points, starting at the \( n \) sample, \( X[n], X[n] \) where \( n = 1, 2, \ldots, k \).
3. The length of the curve for the time interval \( k \) is obtained by dividing all of the sub-series \( L(n, k) \) by \( k \). For \( n = 1, 2, \ldots, k \) the equation is:
   \[
   L(k) = \sum_{n=1}^{k} L(n,k)
   \] (3)
4. So that the dimensions of the Higuchi fractal can be defined as the slope of the line which corresponds to the pair \( \left\{ \ln(L(k)), \ln\left( \frac{1}{k} \right) \right\} \) which is estimated using the most suitable linear squares resulting in the following Higuchi fractal dimensions:
   \[
   L(k) = k^{-D}
   \]
   \[
   L(k) = \frac{1}{k^D}
   \]
   \[
   D = \frac{\ln(L(k))}{\ln\left( \frac{1}{k} \right)}
   \] (4)
   Where \( D \) is the dimension of the Higuchi fractal.

2.3. Classification
Classification is the process of dividing things into groups based on the type or nature of the data. Classification is also used to find a model or function that distinguishes groups or class data that aims to estimate the class of an object whose label is unknown.
2.4 K Nearest Neighbor

K-Nearest Neighbor (KNN) algorithm is a method for classifying objects into K classes in a training data based on the closest distance [8].

In general, to calculate the distance between two objects x and y using the Euclidean distance. The Euclidean distance formula is as follows:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$  \hspace{1cm} (5)

In (4) $$x_i$$ is training data and $$y_i$$ is testing data.

The steps to calculate the K-Nearest Neighbor algorithm are as follows:

1. Determine the K parameter (number of nearest distance neighbors).
2. Calculate the Euclidean distance between data to be evaluated on all training data.
3. Sort the distance and determine the nearest neighbor based on the minimum distance to the K-sequence.
4. Collect appropriate classes.
5. Search for the number of classes from the nearest neighbor and set the class as the data class to be evaluated.

2.5 Support Vector Machine (SVM)

Classification is the process of dividing things into groups based on the type or nature of the data. One classification method is the Support Vector Machine (SVM) method. Support Vector Machine is a classification method that was first introduced by Vapnik in 1998. The Support Vector Machine is a novel small-sample learning method, because it is based on the principle of structural risk minimization, rather than the traditional empirical risk minimization principle [14]. This method works by defining the boundary between two classes with the maximum distance from the closest data. SVM is included in the supervised learning class. The SVM concept stems from the problem of classifying two classes so that it requires positive and negative training sets. SVM tries to find the best hyperplane (separator) to separate into two classes and maximize the margin between the two classes.

In general, data in the real world domain are rarely linear in nature, most are non-linear. To solve non-linear problems, SVM is modified by including kernel functions [15]. Hyperplane search depends on the dot product of the data mapped in the high dimensional space, which is $$\Phi(u).\Phi(v)$$. In this new vector space, the hyperplane that separates the two classes can be constructed. Calculation of transformation $$\Phi(u)$$ is very complicated, but can be made easy by using a function called the kernel as follows:

$$K(u, v) = \Phi(u).\Phi(v)$$  \hspace{1cm} (6)

$$f(x_d) = \sum_{i=1}^{n} \alpha_i y_i K(u, v) + b$$  \hspace{1cm} (7)

Kernel function:

1. Linear kernel
   $$K(u, v) = (u \cdot v)$$  \hspace{1cm} (8)

2. Polynomial kernel
   $$K(u, v) = (u \cdot v + c)^d$$  \hspace{1cm} (9)

3. Radial basis function (RBF) kernel
   $$K(u, v) = \exp(-\gamma |u - v|^2)$$  \hspace{1cm} (10)

With $$u$$ : training data, $$v$$ : class of training data, and $$d, \gamma$$ : parameter of the kernel.

3. Method

This section will briefly described the dataset, signal pre-processing, feature extraction process, Higuchi fractal dimensions, Cross Validation, K-Nearest Neighbor, Support Vector Machine and data accuracy testing. The procedure to classify the babies cry sound can be seen in the following flow chart.
3.1 Dataset
The research data used for this study was secondary data. The data used in this study were 80 data consisting of 4 types of cries, namely hunger cries, tired cries, stomach ache cries, and uncomfortable cries. The data was obtained from the website https://github.com/gveres/donateacry-corpus/. The database contains the sound of crying babies released from newborns to 2 years old with male and female sex.

3.2 Signal Pre-Processing
Two processes must be carried out, namely the screening and normalization process. All baby crying sound data has a frequency of 44,100 Hz. In the recording, the baby's crying sound frequency has a different level of noise. There are some recordings where the original signal frequency is higher than the noise frequency. But there are also some recordings with higher noise frequencies than the original sound. To eliminate or reduce the noise, the filtering process was carried out using Audacity software. Then normalization process used to equalize the maximum amplitude interval of each baby's crying sound signal by changing the amplitude interval to -1 to 1, so that the change in amplitude does not affect the process of extracting the baby's cry sound characteristics. Then the data was processed using MATLAB R2015b software.

3.3 Feature Extraction Process
The methods used in the feature extraction process was Discrete Wavelet Transform (DWT) and Fast Fourier Transform (FFT) process. The DWT process is a process of taking the characteristics of the baby's cry sound signal. An important step in the DWT process that aims to take the characteristics of the baby's crying sound signal was the selection of the type of mother wavelet. The level of classification accuracy depends on the type of mother wavelet chosen. In this study, the type of mother wavelet chosen was a member of the Daubechies family, Daubechies 4. After selecting the next type of mother wavelet, the decomposition process. Signals that have been processed with DWT produce decomposition signals at each level of approximation components A1, A2, A3, A4, and A5 with the time domain being converted into the frequency domain with FFT.

3.4 Calculating Fractal Dimension
The approximation signals A1, A2, A3, A4, and A5 resulting from the Fast Fourier Transform (FFT) process were calculated in fractal dimension values using the Higuchi method. The Higuchi method was chosen because this method is very efficient for determining the fractal dimensions of a curve and suitable for time series data. In this study k-max values of 10, 16, and 50 were selected as experiments. The experiment with the k-max value is to find out the k-max value that is suitable for the baby crying sound data.

Figure 2. Flowchart of classification process of babes' crying sound
3.5 Classification
In the classification process requires the process of sharing data that aims to get testing data and training data. In this study the k-fold cross validation data sharing method was used. In this study, the value of k = 5 and 10 was chosen as an experiment. The results of the data sharing process using 5-fold cross validation are then used for the classification process. The KNN and SVM process is based on the Higuchi fractal dimension value of each baby's crying sound signal. In this study, the value of K = 1,2,3,5,9 and 10 was chosen as an experiment in the KNN method. Whereas the SVM method was chosen c = 1,10,20,40,70 and 100 with the kernel functions namely linear kernel, polynomial kernel, and RBF kernel as experiments. The value of c between 1 and 100 is chosen to find out the c value that is suitable for the baby crying sound data. Parameter c states the amount of the penalty due to misclassification (threshold). In the polynomial kernel, degrees (d) = 1 and 10 are used, where degree 1 is the linear kernel. Whereas in the RBF kernel gamma (γ) = 1 and 10 are used as experiments.

3.6 Accuracy
The accuracy of an algorithm will be measured using the Confusion Matrix. The Confusion Matrix is a way to evaluate the classification method in terms of the accuracy of the classification results. To perform the analysis, confusion matrix can be used, which is a matrix of predictions that will be compared with the original class from the input data. The accuracy of the system was calculated by the following equation:

\[
\text{Accuracy} = \frac{\text{number of correct data}}{\text{number of overall data}} \times 100\% \quad (9)
\]

4. Result and Discussion
In this study, several experiments were carried out with k-max values = 10, 16, and 50 for the calculation of fractal dimension values using the Higuchi method on sound signals. Furthermore, from the results of the calculation of the fractal dimensions, data is divided using k-fold cross validation with values of k = 5 and 10. So that all data is partitioned into 5 partitions and 10 partitions, each of which has the same amount of data. Furthermore, the testing and training process is carried out in 5 iterations and 10 iterations. In iteration-i, partition Ui becomes testing data and the rest becomes training data. In the next stage, the data is classified using the K-Nearest Neighbor (KNN) method and the Support Vector Machine (SVM) with the help of WEKA software. In this study, the values of K = 1,2,3,5,9 and 10 were selected as experiments on the KNN method. Whereas in the SVM method, the values of c = 1,10,20,40,70 and 100 were selected with the kernel function, namely the linear kernel, the polynomial kernel, and the RBF kernel as experiments. The c value between 1 and 100 was chosen to determine the value of c which is suitable for the baby crying voice data.

The following is the classification result without using Fast Fourier Transform (FFT) and using Fast Fourier Transform (FFT):

| k-max | k-fold | KNN | Time (second) | k-fold | SVM | Time (second) |
|-------|--------|-----|--------------|--------|-----|---------------|
| 10    | 5      | 40% | 0            | 10     | 40% | 0.08          |
| 16    | 5      | 38.75% | 0        | 10     | 36.25% | 0.06  |
| 50    | 10     | 53.75% | 0        | 10     | 37.50% | 0.06  |

Based on Table 1 it showed the best accuracy results in the infant crying sound classification experiment without using Fast Fourier Transform (FFT) for each k-max value of the Higuchi dimension used. For the k-max value = 50, the best accuracy was 53.75%, which using the 10-fold cross validation data sharing on the KNN algorithm with the time needed for 0 seconds.
Based on Table 2 it showed the best accuracy results in the infant crying sound classification experiment using Fast Fourier Transform (FFT) for each k-max value of the Higuchi dimension used. For the value of k-max = 10, the best accuracy was 80%, which using the 10-fold cross validation data division on the SVM algorithm with the time needed for 0.03 seconds.

5. Conclusion

Based on the analysis of the results of research on the classification of infant crying sounds using Higuchi’s fractal dimension with the K-Nearest Neighbor and Support Vector Machine, it can be concluded that:

1. The results of the study without using the Fast Fourier Transform (FFT) obtained the best accuracy of 53.75% at level 5 decomposition with 10-fold cross validation, k-max = 50 and K = 10 values in the KNN method. Whereas the SVM method obtained the best accuracy of 40% at level 4 decomposition with 10-fold cross validation, k-max = 10 in the Higuchi method and c = 10 and using the RBF kernel with γ = 10.

2. The results of research using the Fast Fourier Transform (FFT) obtained the best accuracy of 78.75% at level 5 decomposition with 5-fold cross validation, k-max = 10 and K = 9 values in the KNN method. Whereas the SVM method obtained the best accuracy of 80% at level 5 decomposition with 10-fold cross validation, k-max = 10 and c = 10 values and using the RBF kernel with γ = 10. So that in this study the Support Vector Machine (SVM) method is better than the K-Nearest Neighbor (KNN) method.

3. In this study, the best k-max value is k-max = 50 for experiments without using Fast Fourier Transform (FFT) and k-max = 10 for experiments using Fast Fourier Transform (FFT). Where the k-max value shows the maximum limit of the discrete time interval when calculating the Higuchi fractal dimension. While the best level of decomposition is level 5 because it can take a clearer feature vector in each dimension of the baby's crying sound.

6. References

[1] Torres R, Battaglino D, and Lepauloux L 2017 Baby Cry Sound Detection: A Comparison of Hand-Crafted Features and Deep Learning Approach

[2] Lavner Y, Cohen R, Ruinskiy D, and Ijzerman H 2016 Baby Cry Detection in Domestic Environment using Deep Learning IEEE Int. Conf. on the Sci. of Elect. Engine. (ICSEE) 1

[3] Limantoro W S, Fatichah C, and Yuhana L 2016 J. Tek. ITS 5 2

[4] Ramadijanti N, B. Sena, and B. Dewantara 2016 Pengenalan Rasa Lapar Melalui Suara Tangis Bayi Umur 0-9 Bulan Dengan Menggunakan Neural Network (Sub Judul: Penapisan Dengan Transformasi Wavelet Kontinyu)

[5] Srijiranon K and Eiamkanitchat N 2014 Application of neuro-fuzzy approaches to recognition and classification of infant cry TENCON 2014 - 2014 IEEE Region 10 Conference, Bangkok, 1

[6] Safitri D W and Juniati D 2017 Classification of diabetic retinopathy using fractal dimension analysis of eye fundus image AIP Conf. Proc. 1867 020011

[7] C Khotimah and D Juniati 2018 J. Phys.: Conf. Ser. 947 012004
[8] D Juniati et al 2018 J. Phys.: Conf. Ser. 953 012202
[9] Bğnicğ I, Cucu H, Buzo A, Burileanu D and Burileanu C 2016 Baby cry recognition in realworld conditions 39th Int. Conf. on Telecomm. and Sign. Processing (TSP), Vienna 315
[10] Keng N W, Leong M S, and Abdelrhman A M 2013 Appl. Mech. Mater 393 953.
[11] Bruns A 2004 J. of Neurosci. Methods 137 321
[12] Cannon JW 2018 The American Math. Monthly 91 594
[13] Juniati D and Budayasa I K 2016 Geometri Fractal dan Aplikasinya (Surabaya: University Press)
[14] Zhang Y 2012 Comm. in Comp. and Inform. Sci. 308
[15] Johra M B 2018 J. Mat. 4 137