International Journal of Engineering & Technology, 7 (4.15) (2018) 447-450

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Research Framework for Supervised Image Classification For Tornado Chaos Phenomena

W Wanayumini*, O S Sitompul, M Zarlis, Saib Suwilo, A M H Pardede

Program Studi Doktor Ilmu Komputer, Fakultas Ilmu Komputer dan Informasi Teknologi
Universitas Sumatera Utara Jl. Universitas
**Corresponding author E-mail:wanayumini@yahoo.co.id*

Abstract

Unattended classification is a classification which is the process of forming classes conducted by computers. The classes formed in this classification are highly dependent on data acquisition. In the process, this classification classifies pixels based on similarity or spectral similarity. While the supervised classification is a classification carried out by the analyst's direction. The purpose of this study is to build a new model of image-based classification based on chaos phenomena through remote sensing to detect the beginning of the emergence of tornadoes. This research optimizes the search for the best value from a data collection of samples of chaos phenomena in tornadoes through a new model called Citra which is supervised by Chaos Discrete Cosine Transform Spectral Angel Mapper Classification (SiChDCosTSamC). The resulting model can then be used as remote sensing to detect the appearance of the initial tornado. Tests will be carried out using the Protected Image Welding on models based on chaotic / chaotic phenomena. Testing will be carried out on a collection of sample image data sourced from SIO, NOAA, US data. Navy, NGA, GEBCO U.S. PGA / NASA Google IBCAO Geological Geological Survey / Copernicus.

Keywords: *Chaos Phenomena, Classification, Supervised Image, Tornado.*

1. Introduction

The segmentation process is a quantitative classification in digital image processing which can also be interpreted as a process of grouping pixels into classes that are determined based on the variables used. The classification can be in the form of something related to similar features based on statistics by a computer[1]. To evaluate the image, classification is needed because in the classification of objects or phenomena on the surface of the earth from a very large number it is simplified in number into several classes to be easily analyzed [2].

Unattended classification is a classification that the process of forming classes is done by a computer. The classes formed in this classification are highly dependent on data acquisition. In the process, this classification classifies pixels based on their similarity or spectral similarity. While the classification that is supervised is the classification that is carried out by the analyst's direction. Classification criteria are determined based on the class identifier obtained by the analyst through the creation of a sample area [3]. The supervised classification technique can be interpreted as a supervised classification technique in remote sensing and can be done in several steps, including select training areas, generate signature files, and classify. Some algorithms that are often used are Maximum Likelihood Classifier (MLC), Minimum Euclidean Distance Classifier (MEDC), Extraction and Classification of Homogeneous Objects (ECHO) Classifiers (ECHOC), Spectral Angel Mapper Classifiers (SAMC), and Parallel Piped Classifiers (PPC) [4].

The Minimum Euclidean Distance Classifier (MEDC) is the average vector for each class and calculates the Euclidean distance from each unknown pixel to the average vector for each class. Pixels are classified to the closest class [5]. The Parallel Piped Classifier (PPC) uses a simple decision rule to classify hyperspectral data. Decision boundaries form dimensions in image data space. The dimensions of Parallel Piped Classifier are determined based on the threshold, standard deviation, mean of the selected class.

The optimal selection of classification in supervised classification, image-based remote sensing has different performance [5]. In remote sensing it is difficult to get high quality field samples in image analysis. Because the sample has a big influence on the results and accuracy of image or image-based classification. The method that has been carried out by researchers is to use various methods including using the RBG & L method. Researchers will use a sample dataset image pickaxe based on chaos phenomena by using the Discrete Cosine Transform equation and the model learning process using the Spectral Angel Mapper Classification (SAMC) method which is expected to produce the most or the best number of pixels based on Numbers Samples and Percentages in the dataset sample process. image Supervised Classification [6].

The purpose of this study was to build a new model of image-based supervised classification based on chaos phenomena through remote sensing to detect early emergence of tornado. This study optimizes looking for the best value of a dataset sample image chaos phenomenon in a tornado through a new model with the name Supervised Image Chaos Discrete Cosine Transform Spectral Angel Mapper Classification (SiChDCosTSamC). The resulting model can later be used as a remote sensing to detect early emergence of a tornado.

The occurrence of tornado winds must also be handled before the disaster occurs, it is necessary to carry out an early warning tool for tornado disasters. the warning device is a smart device that is placed in disaster-prone areas, by synchronizing input data from more than one smart device used, so that the early warning tool

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

can try to submit a decision, so that decisions can be given about what actions will be taken by disaster management [7].

2. State of The Art

Image classification aims to get an image that contains parts that state an object or theme. Each object in the picture has unique symbols or certain colors or patterns. The image-shaped classification, initially begins with visual interpretation to identify the same pixel groups that represent various groups [8]. Digital classification generally can be done in two ways, namely unsupervised classification (Supervised Classification) and Supervised Classification [9]. Unsupervised classification classifies pixels based on their spectral similarities. These classes are not directly related to certain characteristics of features or objects in the image [10].

Supervised classification is a classification whose analysis has a number of pixels representing each desired class or category [8]. Classification criteria are determined based on the class characteristics obtained through the creation of a training area. Determination of the training area is usually carried out based on observations of the earth map. The training area that has been obtained is then used as an input in the classification process for the whole image [8]. The classification is monitored based on the results of the survey the identity and location of the classes are known, analysis of aerial photographs (or previous satellite images), as well as in other ways [11].

Supervised and unsupervised classifications have their own shortcomings. Unsupervised classification has a spectral deficiency which always changes over time, causing the relationship between spectral responses to information classes to be constant. Unattended classification is used in images that have little information. The disadvantage of supervised classification is that the output obtained is not in accordance with the situation in the field, the problem has resulted in the training area that has been created will process data that is not right. The advantage of supervised classification is that it can distinguish classes / clusters well if the training sample obtained is right [12].

Previous research used the supervised classification method and its accuracy was optimized. However, due to the requirements of calculating cluster centers, the supervised classification method traditionally requires the average value of the sample area, and there is no consideration of the entire sample. The study combines the process of dynamic chaotic search with genetic algorithms, to improve search efficiency and improve optimal solutions, and improve the accuracy of classification of remote sensing images using initial sensitivity of conditions and system parameters from chaotic systems and chaotic ergodicity [10]].

3. Supervised Classification Method

There are several methods used in image-based Supervised Classification, including Maximum Likelihood Classification (MLC), Minimum Euclidean Distance Classification (MEDC), Parallel Piped Classification (PPC), ECHO Classification (ECHOC), Spectral Angel Mapper Classification (SAMC) [6]. Spectrum-based physical classification that uses n-dimensional angles to match pixels with training data. This method determines the spectral similarity between two spectra by calculating the angle between spectra and as a vector in space with dimensions equal to the number of bands. Bands in this case are colors, for example using RGB colors (Band1 Red / Red, Bands Green / Green, Band3 Blue / Blue).

This technique, there are two types of general classification procedures and each finds applications in remote sensing processing which is called the supervised classification and unsupervised classification. There are various classification approaches that have been developed and are widely used to produce land cover maps. This is related to logic, from supervised to unsupervised, parametric to non parametric [13]. In this case it can be used as an alternative approach, but often uses a hybrid methodology that uses more than one method [14].

Unsupervised image classification is a method in which image analysis software separates a large number of unknown pixels in an image based on reflectance values into a class or group without direction [15]. There are two grouping methods most often used for unsupervised classification: K-means and Iterative Self-Organizing Data Analysis Technique (ISODATA), both of these methods depend entirely on statistics based on spectral pixels and do not include prior knowledge of the characteristics being studied. On the other hand, supervised classification is a method in which the analyst defines a small area called the training site in the image, which contains predictor variables measured in each sampling unit, and assigns the previous classes to the sampling unit [3].

4. Chaos Basic Concept in the Wind Characteristics of The Tornado

Basically Chaos Theory is a theory that deals with irregular systems. This kind of system can be found in many irregular objects such as waves, clouds, trees, coastlines and so on. At first glance, these systems look random and irregular. But if the division of small parts is done, this large, irregular system is found as a repetition of regular parts. Statistically it can be stated that Chaos is a stochastic behavior of a deterministic system. The system is deterministic (simple, one solution) when stacked will become a stochastic system (complicated, many solutions).

Each periodic movement proposed is usually "unstable", as a result the parameter value will be very much an unstable periodic movement in a system. When this happens, the system dynamics are very complex and chaotic conditions occur again. In dealing with this problem Briggs & Peat put forward three aspects to stop chaos, namely: control, creativity and communication. These three aspects bring and encourage new meaning or purpose to find order in a chaotic state, solve unusual problems, and form links new and able to balance creations and ideas so that they can solve problems properly [16].

A whirlwind that is often also known by the name of the whirlwind is included in one of the most dangerous natural phenomena. A whirlwind or even in the Sumatra region is often referred to as the Bahorok wind, which is a wind that rotates at a speed of up to 63 km per hour. This wind often moves straight, and also usually passes after 5 minutes.

This tornado often occurs during the day and also in the afternoon during the transition season. This wind is often regarded as one of the deadliest types of wind because it can destroy anything that has passed. Until now, there have been many reports of cyclones in many places. A very large whirlwind can even damage people's homes, transportation equipment, trees. So it is no stranger to have passed this wind can make a lot of damage at the same time also cause losses that are not very few. Almost all places in the Indonesian region are prone to the emergence of this one wind disaster. But, even so there are some places that in fact are more often attacked by cyclones when seen from other places.

This often happens in the regions of Nusa Tenggara, Sulawesi and Sumatra. Even on the island of Java there are also some places that are often affected by hurricanes. For example, in the West Java area, whirlwinds usually occur in the areas of Banjar, Garut, Ciamis, and Tasik. In addition, this wind often occurs even in the Sukabumi area and also in the Sumedang area. Because any area can also be affected by this type of wind. A whirlwind often occurs when the transition season is during the day or also in the afternoon. The phase of the occurrence of this tornado has several fairly close links to the phase of cumulonimbus cloud growth. The three phases of the tornado include.

5. Research Methodology

This study uses a model based on chaotic or chaotic phenomena in a dataset sample image of a tornado. Chaos is modeled by the Discrete Cosine Transform equation. The model learning process uses several stages as a comparison, including the Maximum Likelihood Classification, Euclidean Distance Classification, ECHO Classification, Spectral Angle Mapper Classification, and Parralel Piped Classification to get the number of classes, number of pixels, Means value, Standard Deviation value, classification size, Number of Samples and Percentages of sample image datasets. The data source used as a training sample is the image of a tornado taken by using the Google Earth application. Image dataset sample it was obtained from a SIO, NOAA, U.S. Data source. Navy, NGA, GEBCO U.S. PGA Geological Survey / NASA Google IBCAO Landsat / Copernicus. The applications used to obtain sample image datasets include: Photoshop CS3 application, MultiSpecWin32 application, WEKA 3.9 and Matlab application to carry out the supervised classification process before using the proposed model.

The data used is sourced from image datasets taken through the Google Earth application, Data SIO, NOAA, U.S. Navy, NGA, GEBCO U.S. PGA Geological Survey / NASA Google IBCAO Landsat / Copernicus, then the image data is cropped to take samples of distance based datasets with the help of Photoshop CS3 application. The dataset taken as a sample image dataset for the testing process is a sample image dataset or an image of a tornado wind that simulates chaos phenomena. The sample dataset image that will be analyzed is 6 samples images.

The new Supervised Image Classification model that will be carried out by researchers is to take a sample of image datasets through the Google Earth data application SIO, NOAA, U.S. Navy, NGA, GEBCO U.S. PGA Geological Survey / NASA Google IBCAO Landsat / Copernicus. Samples of image datasets are taken based on remote and close sensing from the satellite camera to the object and then do the sampling process of the image dataset sample with the help of Photoshop CS3 application. The results of the extracting are processed with the Matlab 2015 application by adding the dataset sampling results to the image Winds of Puddles based on chaotic phenomena can be seen in Figure 1 as follows :

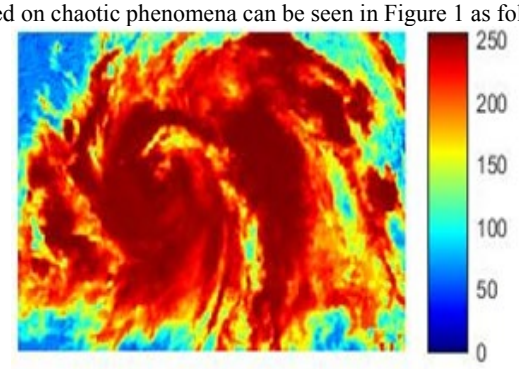


Fig.1. Proposed DCT Model Test Result

The model learning process uses several stages or procedures. The first stage is the clastering process to get the number of classes / clusters, the number of pixels per class / cluster, the mean value (average) and the standard deviation value of C1 (red), C2 (green), C3 (blue). The second stage carries out the classifier process as a comparison, by finding the value of the Maximum Likelihood Classification, Euclidean Distance Classification, ECHO (Extraction and Classification of Homogeneous Object) Classification, Spectral Angle Mapper Classification, and Parralel Piped Classification to get the number of classes, number of pixels, Means value, the value of the standard deviation, the size of the classification, the number of number of samples and the percentage of the sample dataset image whirlwind that simulates chaos phenomenon.

In general, this research process several stages, including:

1. Selection stage and sample preparation of image / image / image datasets,
2. Preprocessing stages
3. Testing Stages
4. Stages of Evauation of Dataset Processing Results

The sample image dataset used is a sample of image datasets taken using the Google Earth application that shows chaotic or chaotic phenomena. The dataset sample is cropped with the help of Photoshop CS3 application to get a sample of image datasets that will be processed using the MultiSpecWin32 application. The MultiSpecWin32 application functions to perform the model testing process using several algorithms as a comparison, including the maximum likelihood classification algorithm, the Euclidean Distance Classification algorithm, the ECHO Classification algorithm, the Spectral Angle mapper Classification algorithm, and the Parralel Piped Classification algorithm to get the number of classes, number of pixels, Means value, the value of the Standard Deviation, the size of the classification, the number of Number Samples and the Percentage of the sample dataset image of a tornado that describes the chaos phenomenon. After testing using several algorithms as a comparison to get the final model proposed by the researcher which algorithm is better. After obtaining a better algorithm based on the model proposed by the researcher, the next step is to evaluate the test results that will be used as a new model.

6. Research Framework

Testing will be carried out by using Unsupervised Image Classification on a model based on chaotic / chaotic phenomena to get the number of classes, number of pixels, mean value, standard deviation and classification size. Testing will be carried out on sample image datasets sourced from SIO, NOAA, U.S. data. Navy, NGA, GEBCO U.S. PGA Geological Survey / NASA Google IBCAO Landsat / Copernicus. The test results can be seen in Table 1. below shows the results of sampling image datasets using the MultiSpecwin32 application. The sample results of the dataset image are obtained from the remote clustering process from satellite cameras to objects.

Table 1. Average Value, Standard Deviation and Classification Size Long Distance from Satellite Cameras to Objects (7161,25 km)

Cluster	Pixels	%	Means			Standard Deviations		Classification	
			C1 (Red)	C2 (Green)	C3 (Blue)	C1 (Red)	C2 (Green)	C3 (Blue)	Size
1	3,5	5.8	152.9	163	186.3	5.9	5.6	6.5	3,6
2	476	0.8	142	152.3	158.1	5.6	4.8	5.3	516
3	397	0.7	117.5	127.7	130.2	4.4	3.4	3.9	420
4	366	0.6	82.4	95.5	93.6	5.8	4.4	5.2	394
5	504	0.8	108.3	118.1	119.1	4.4	3.4	4.1	493
6	**21**	**34.4**	**246.6**	**248.9**	**250.8**	**7.2**	**4.9**	**3.1**	**20**
7	387	0.6	127.9	138.5	142.3	4.4	3.7	5.1	400
8	**233**	**0.4**	**61.5**	**78.4**	**74.9**	**10.3**	**9.2**	**10.5**	**183**
9	437	0.7	96.9	108.1	108.1	4.9	3.7	4.6	433
10	1,1	1.8	81	97	137.5	4.3	3.7	5.9	1,2
11	7,2	11.8	193.8	201.8	220.7	6.5	6.3	6.4	7,4
12	10	17	217.5	224.5	239	7.8	7.4	5.5	10
13	1,7	2.9	112.4	125.1	157.1	5.5	5.1	4.6	1,7
14	5,2	8.6	172.4	181.5	202.8	5.8	5.6	6	5,4
15	2,3	3.9	132.2	143.4	172	6.1	5.7	4.9	2,3
16	1,3	2.2	95.2	109.4	145.3	4.8	4.2	4.4	1,3
17	1,3	2.1	64.8	82	125.7	6.1	5.1	4.8	1,2
18	613	1	36.9	57.7	110.4	5.3	4	5.6	744
19	2,3	3.8	44.7	66.1	121.3	5.2	3.8	5.3	2,2
	60	100	2,3	2,5	2,9	110	94	102	60

Based on table 1. It can be seen that when viewed from the number of clusters, the number of pixels and percentages and the mean value for each rgb color ((red / red, green / green, blue / blue) and the classification size get the highest value among the other clusters. The deviation for each rbg color

((red / red, green / green, blue / blue) has a moderate value, the resulting cluster with the highest value is in cluster 6.

Table 2. Average value, standard distance deviation and classification size nearby from satellite camera to object (2876.77 km) un supervised classification

Cluster	Pixels	%	Channel Means			Channel Standard Deviations			Classification
			1	2	3	1	2	3	Size
1	206	0.3	102.6	109.8	127.8	8.3	8.2	6.7	265
2	738	1.2	125.9	131.9	147	9.1	8.7	7.1	795
3	2,141	3.4	153.8	158.3	169.1	8.3	8.1	7.4	2188
4	3,94	6.3	180	183.7	189.6	8.2	7.8	7.6	3893
5	6,261	10.1	206.3	208.6	212.4	7.4	7.2	6.8	6696
6	19,29	31	232.6	233.6	235.6	6.5	6.3	5.7	19421
7	29,57	47.6	250.8	250.9	251.4	4.4	4.2	3.9	28884
	62,15	100	1,25	1,277	1,333	52	51	45	62,142

Based on table 2. it can be seen that when viewed from the number of clusters, the number of pixels and percentages and the mean value for each RGB color (Red / Red, Green / Green, Blue / Blue) and the classification size get the highest value among the other clusters. While the standard deviation value for each RBG color (Red / Red, Green / Green, Blue / Blue) has a moderate value. The resulting cluster with the highest value is in cluster 7.

7. Test Results Against Remote Revised Image Classification

The test will be carried out by using Supervised Image Classification on a model based on chaotic / chaotic phenomena to get the Number Samples and Percentage of the Maximum Likelihood Classification (MLC) model, Minimum Euclidean Distance Classification (MEDC), ECHO Cassifier (ECHOC), Spectral Angel Mapper Classification (SAMC), and Parallel Piped Classification (PPC). Testing will be carried out on sample image datasets sourced from SIO, NOAA, U.S. data. Navy, NGA, GEBCO U.S. PGA Geological Survey / NASA Google IBCAO Landsat / Copernicus. The test results can be seen in Table 7.1. below shows the results of sampling image datasets using the MultiSpecWin32 application. The results of sample image datasets are obtained from the Classification process based on close range from satellite cameras to objects.

Table 3. Number Samples Value, Percentage of MLC, MEDC, ECHOC, SAMC, and PPC Models Long Distance from Satellite Cameras to Objects (7161.25 km) Supervised Classification.

Cluster	ECHO Classification			Spectral Angle Mapper Classification		Parallel Piped Classification	
	Likelihood Thresholds per class	Number Samples	Percent	Number Samples	Percent	Number Samples	Percent
1	-51.27	2,904	4.81	3,9	6.44	3,79	6.27
2	-49.99	710	1.18	1,5	2.54	503	0.83
3	-49.57	373	0.62	205	0.34	440	0.73
4	-51.95	233	0.39	424	0.7	380	0.63
5	-49.75	465	0.77	435	0.72	455	0.75
6	**-43.29**	**19,37**	**32.07**	**20**	**33.53**	**19,1**	**31.59**
7	-50.55	346	0.57	425	0.7	293	0.49
8	**-56.43**	**379**	**0.63**	**255**	**0.42**	**213**	**0.35**
9	-50.42	480	0.79	380	0.63	354	0.59
10	-49.27	1,013	1.68	1,1	1.89	1,17	1.94
11	-49.76	7,494	12.41	7,6	12.6	7,7	12.74
12	-49.45	11,4	18.87	9,2	15.29	10,3	16.97
13	-47.2	1,868	3.09	1,8	2.91	1,88	3.1
14	-49.93	5,339	8.84	4,8	7.88	4,44	7.34
15	-46.76	2,584	4.28	2,6	4.37	2,13	3.52
16	-48.33	1,177	1.95	1,3	2.23	1,06	1.75
17	-49.84	1,516	2.51	1,4	2.32	1,25	2.07
18	-50.14	623	1.03	1,4	2.4	1,14	1.88
19	-49.7	2,134	3.53	1,3	2.08	1,49	2.46
	-944	60,41	100	60	100	2,4	3.98

8. Future Research Plans

For the perfection of this research, there are several suggestions that must be done in the future, i.e. :

- Build a model and simulation equipped with the implementation of input from a disaster warning device.
- Further discussion about data security to ensure that data from the population is not misused by certain parties.
- Each model and simulation that is designed must be discussed in more detail to find points or things that are impossible to implement [17].

References

[1] T. Dogan and A. K. Uysal, "Intelligent Systems and Applications in Engineering The Impact of Feature Selection on Urban Land Cover Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 6, no. 1, pp. 59–64, 2018.

[2] S. Ge, R. Wang, and D. Yu, "Classification of four-class motor imagery employing single-channel electroencephalography," *PLoS One*, 2014.

[3] J. Al-doski, S. B. Mansor, H. Zulhaidi, and M. Shafri, "Image Classification in Remote Sensing," *J. Environ. Earth Sci.*, vol. 3, no. 10, pp. 141–148, 2013.

[4] D. Camilo, D. O. Duarte, J. Zanetti, and J. G. Junior, "Comparison of supervised classification methods of Maximum Likelihood image , Minimum Distance , Parallelepiped and Neural network in images of Unmanned Air Vehicle (UAV) in," in *Proceedings XVII GEOINFO*, 2016, no. 2004, pp. 12–21.

[5] B. Park, K. C. Lawrence, W. R. Windham, and D. P. Smith, "P s c a h i i f i c," vol. 49, no. 6, pp. 2017–2024, 2006.

[6] Mr. Devi and D. Santhosh Baboo, "Land use and Land Cover Classification using RGB&L Based Supervised Classification Algorithm," *Int. J. Comput. Sci. Eng. Technol.*, vol. 2, no. 10, pp. 2229–3345, 2011.

[7] A. M. H. Pardede *et al.*, "Framework For Patient Service Queue System For Decision Support System on Smart Health Care," *Int. J. Eng. Technol.*, vol. 7, no. 2.13, pp. 337–340, 2018.

[8] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*. 2017.

[9] S. Mishra, T. Yamasaki, and H. Imaizumi, "Supervised classification of Dermatological diseases by Deep learning," no. Dl, 2018.

[10] L. Gao, G. Yiqiang, W. Yanbin, J. Zhengshan, W. Jun, and Z. Luyan, "Remote sensing image classification by the Chaos Genetic Algorithm in monitoring land use changes," *Math. Comput. Model.*, 2010.

[11] M. Chen, Y. Fu, T. Sun, D. Li, and Q. Qin, "Classification of land cover from remote sensing fused image based on ICA-SVM and D-S evidence theory," vol. 7147, no. 40601055, p. 71470E, 2008.

[12] J. St. Peter, J. Hogland, N. Anderson, J. Drake, and P. Medley, "Fine Resolution Probabilistic Land Cover Classification of Landscapes in the Southeastern United States," *ISPRS Int. J. Geo-Information*, vol. 7, no. 3, p. 107, 2018.

[13] K. E. Chlouverakis and J. C. Sprott, "Chaotic hyperjerk systems," vol. 28, pp. 739–746, 2006.

[14] J. a Richards and X. Jia, *Remote Sensing Digital Image Analysis*. 2006.

[15] J. T. Tou and R. C. Gonzalez, "Pattern recognition principles," *Image Rochester NY*, 1974.

[16] J. Briggs and F. D. Peat, "Seven Life Lessons of Chaos : Wisdom from the Science of Change," pp. 1–224, 2000.

[17] A. M. H. Pardede, Y. Maulita, and R. Buaton, "Application modeling ipv6 (internet protocol version 6) on e-id card for identification number for effectiveness and efficiency of registration process identification of population," in *Journal of Physics: Conference Series*, 2018, vol. 978, no. 1.