Rainfall Prediction Based on Himawari-8 IR Enhanced Image Using Backpropagation

D C R Novitasari¹, B D Supatmanto², M F Rozi³, Hermansah³, Y Farida¹, Rr D N Setyowati¹, Ilham⁶, R Junaidi⁵, A Z Arifin⁷, A R Fatoni⁸

¹ Departement of Mathematics, UIN Sunan Ampel Surabaya, Surabaya, Indonesia
² Weather Modification Technical Unit, Agency for the Assessment and Application of Technology (BPPT), Jakarta, Indonesia
³ Departement of Mathematics Education, University of Riau Islands, Batam, Indonesia
⁴ Department of Environmental Engineering, UIN Sunan Ampel Surabaya, Surabaya, Indonesia
⁵ Departement of Architecture, UIN Sunan Ampel Surabaya, Surabaya, Indonesia
⁶ Department of Information System, UIN Sunan Ampel Surabaya, Surabaya, Indonesia
⁷ Departement of Mathematics, Universitas PGRI Ronggolawe, Tuban, Indonesia
⁸ Forecaster and Analyst, Meteorological Climatological and Geophysics Agency, Surabaya, Indonesia

Abstract. The wider sea area causes greater evaporation of water in Indonesia. In addition, these conditions have an impact on the season that Indonesia has. Indonesia's high rainfall disrupts human activities. As a result, it is very important to detect cumulonimbus clouds using satellite imagery. The satellite image used is intended to be taken two values of the characteristics possessed. Characteristics taken are average cover and average cloud temperature. Previous studies predicting rain were only done using observational data taken at the height of 10 meters. This research predicts using satellite imagery that represents the cloud peak temperature value. Furthermore, the classification of data is done using backpropagation. The results of the classification process using backpropagation obtained the best results on the distribution of 80% training data and 20% testing data, with the activation function logging in the hidden layer and that the output layer. The results obtained indicate the accuracy rate of 88.283%.

1. Introduction

Indonesia is a maritime country. Evaporation from the wider sea causes Indonesia to experience high rainfall during the rainy season. High rainfall in Indonesia disrupts the activities of the Indonesian people. Among the disturbances caused by heavy rains, namely industrial disturbances that require solar irradiation, interference with high waves due to heavy rains, and so forth. Besides the presence of cumulonimbus clouds also disrupt flight [1].

Rain accompanied by winds cause visibility disorders in reduced flights. Also, the presence of cumulonimbus clouds can disrupt flights if the aircraft passes through it. Among the dangers that can occur if the plane passes through the cumulonimbus clouds is the danger posed by cumulonimbus, also caused by an updraft and downdraft that allows the friction of particles and generate electrical charges inside it [2]. The friction of these particles which can cause lightning on cumulonimbus clouds. The above factors can affect the occurrence of rain. The formation of clouds cumulonimbus from the image of the clouds.
Retrieval of cloud images can take through satellite imagery. One that to detect clouds and the degree of saturation in clouds is the image of the Enhanced IR Himawari-8. The use of the Enhanced IR Himawari-8 image because the EH satellite image is an image that describes the temperature of a cloud. The higher the intensity of the red color shown in the Himawari-8 IR Enhanced image, the lower the temperature the cloud has [3]. Visions of the presence of cumulonimbus clouds are very important because cumulonimbus clouds are triggers of hail [4]. Bony Septian Pandjaitan uses Himawari-8 satellite imagery as an image to see the formation of cumulonimbus clouds in an area and has a very fast growth rate [5]. In another case, Ajis Nur Efendi analyzed the satellite imagery of Himawari-8 IR in the event of rain in the Kalimantan region [6].

The Himawari-8 IR satellite imagery obtained has a colour index that corresponds to the temperature the clouds have. The red image obtained produces an image that depicts clouds with low temperatures. The process of getting cloud images that have very low temperatures requires a segmentation process. One method in the segmentation process is the Adaptive Average Brightness Thresh-holding (AABT) method. The AABT method of segmentation by taking features that have more than average light intensity values [7]. Retrieval of features using the AABT method leaves clouds that have very cold temperatures. The results of segmentation using AABT can indicate the presence of cumulonimbus clouds. The results of the segmentation feature taken in the form of a mean value of closure and on the average temperature of the cloud in the desired area. The value of the feature used to represent the state of the cloud in that place [8].

Statistical values in cloud images are useful for inputting the rain prediction process. The two values are associated with several other parameters, including data on wind speed, air pressure, temperature, solar radiation, and water level. The level of influence of each variable can be determined using factor analysis. The results after factor analysis are variables with a high influence value used as input variables. One method used to make predictions is backpropagation [9]. Based on previous research, using backpropagation to predict the occurrence of rain conducted by Nilay S. Kapadia, and Mislan [7] [10]. Other studies predict the weather in an area that has a very good accuracy value using backpropagation [11]. Calculation of the success or accuracy of the model formed can use the method confusion matrix. The output of this method is a matrix. The matrix confusion is resulting shows the number of classifications of true or false in each class.

2. Literature Review

The process of weather classification using satellite imagery requires some understanding of several terms or methods. Explanation of the terms used in the classification process is expected to help understanding in studying the research net. Other than that the explanation of the method in the literature review, is intended to explain some calculation processes to produce the desired output.

2.1. Image IR Enhanced

Image IR is a satellite image that produces a picture of the temperature at the cloud top. Cloud peak temperature values obtained using radiation at a wavelength of 10.4 micrometers. The value generated by the radiation is classified using colour. Black indicates the absence of clouds, blue indicates there is not too much cloud formation. Green to red indicates temperature. The higher the intensity of the red color, the lower the temperature, and the more potential to become cumulonimbus clouds [12]. The results of the Enhanced IR image shown in Figure 1.
2.2. Median Filter
The median filter is one of the methods used to help improve images. Improvements made with the Median Filter utilize the middle value of neighboring pixels. The neighboring pixels obtained are sorted starting from the smallest and then taking the middle value from the neighboring matrix. The obtained middle value replaces the old value as the pixel value that is corrected [13].

2.3. Adaptive Average Brightness Thresholding (AABT)
Adaptive Average Brightness Thresholding (AABT) is one that can be used for segregation. The segmentation is done using AABT by looking at the brightness level in an image. This method is best used for segmenting cloud images because the presence of clouds is illustrated with a higher brightness value compared to the state without clouds. Intake of cloud features obtained by looking at the average brightness possessed in an image. Through the average deduction function, values below the threshold are omitted values while values that exceed the threshold are cloud read features [9][14]. Threshold values can be obtained using Equation (1). With $g_{max}$ is the maximum pixel in the matrix $f(x, y)$.

$$C = \frac{1}{MN} \left( \text{mean}(f(x,y)) + 22.5 \left( \ln(g_{max}) - \ln \left( \text{mean}(f(x,y)) \right) \right) \right)$$  \hspace{1cm} (1)

Where:
- $f(x,y)$ = matrix image
- $g_{max}$ = maximum pixel
- $M$ = Number of raw on matrix
- $N$ = Number of column on matrix

2.4. Principal Component Analysis (PCA)
Based on statistical science Principal Component Analysis (PCA) is a technique used to simplify data. Dataprocess carried out by linear transformation so that new coordinates are formed with maximum variance [15]. Principal Component Analysis (PCA) methods can be used to reduce data dimensions without significantly reducing the characteristics of the data. PCA changes the original variables that correlate with each other into a new set of variables that are smaller and mutually independent. The principal component analysis is the analysis of the prefix process of the next analysis. For example, PCA is used as a factor analysis input. Mathematically. PCA is formed from variance and covariance through a linear combination of input variables. PCA can be used to check several correlated variables [16]. The value of variance ($s^2$) and covariance can be obtained using Equation (2) and Equation (3). Where $x$ is
the input, \(x\), and \(y\) is the target of the input. This value is used for pattern recognition on each variable using Equation (4).

\[
s^2 = \sum_{i=1}^{n}(\bar{x} - x_i)^2
\]

\[
\text{Cov}(x, y) = \sum_{i=1}^{n} \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1}
\]

\[
|A - \lambda| = 0
\]

2.5. Artificial Neural Network

The artificial neural network is one of the optimization methods which uses a clone of the way nerves work in the brain. Like nerves in the brain, artificial neural networks are also formed using several layers. The information obtained is sent or delivered from layer one to another layer. Artificial neural networks are composed of three layers. An input layer, an output layer, and the hidden layer. Each layer has a different number of neurons. The input layer is the layer that contains the data that is owned. The output layer contains the target value that we expect. While the hidden layer is used to overcome conditions or problems that have not been resolved, or in other words, the hidden layer is used to overcome complex problems from the data held [17]. Each layer in the artificial neural network has a different relationship or relationship. The weight value expresses the degree of interrelation between layers. The more information that can be conveyed to other neurons, the higher the weight value they have. Conversely, the value of the weight gets smaller if the information delivered is not conveyed [18].

2.6. Confusion Matrix and Accuracy

The confusion matrix is one of the methods used to calculate accuracy on the classification results obtained. Matrix produced at Confusion Matrix helps see the classification details. The number of distribution of classification results is displayed in each element in the matrix confusion. The confusion matrix gets the sum for each class. The size of the matrix rows and columns confusion corresponds to the number of classes data. The column in the matrix confusion shows the class that the data belongs to, while the row in the matrix confusion shows the classification class. The matrix form produced by the confusion matrix is shown in Table 1 [19]. In the results, table confusion matrix shows the results of the formation of the matrix confusion. The matrix element of the first row of the first column shows the amount of data properly classified in the positive class. Element contained in the second row, the first column, is the amount of data that is classified incorrectly on the positive class. Whereas in the first row of the second column, the number of classifications is wrong in the negative class. The second row of the second column is an element that shows the amount of data correctly classified in the negative class [20].

| Classification Result | Classified as          |
|-----------------------|------------------------|
| +                     | True Positive          |
| -                     | False Negative         |

| Classified as          |
|------------------------|
| True Positive          |
| False Negative         |

Matrix results formed by using a confusion matrix. It can be used to determine the accuracy of the system. The accuracy value can be obtained from the amount of data properly classified in each class compared to the total amount of data. Calculation of accuracy by using the confusion matrix is calculated using Equation (5).
Where:

\( TP = True \ Positive \)
\( TN = True \ Negative \)
\( FP = False \ Positive \)
\( FN = False \ Negative \)

3. Research Methods

This research is a quantitative study. This type of quantitative research is intended because of the use of quantitative data in the form of nominal or numbers. The data used has two types of data. The first data is data on wind speed, temperature, barometric pressure, solar radiation, and water temperature. The second data is cloud image data. Cloud image data is data taken from the Himawari-8 satellite. The Himawari-8 satellite imagery data used in this study is a type of Enhanced IR Himawari-8 imagery. Satellite image processing is intended to obtain 2 feature extraction values. the feature extraction values include the average temperature, and the average cloud cover

![Flowchart](image)

Figure 2. Flowchart

In this study, the Himawari image was taken with an intensity of red color. taking red intensity due to cold temperatures in the image are marked in red. The next step is to eliminate noise using the Median Filter process. In the next stage, the segmentation process is carried out to obtain cloud images with very cold temperatures. The segmentation process is carried out using the Adaptive Average Brightness Thresh-holding (AABT) method.

Cumulonimbus cloud intake using AABT is obtained by equation (3). After getting the cloud with a temperature below zero degrees, then cut it. The cutting in question is cutting the image from the region image. After getting an image that depicts the cloud conditions in the region, the feature values are taken
from that image. Extraction of the image features is taken two values that represent cloud imagery. that is, the average cloud value and the average cloud temperature can be taken.

Two values obtained from the image data then collected with data on wind speed, temperature, barometric pressure, and water temperature. The combination of all data becomes input data in the process backpropagation. All existing data are analyzed factors to see the correlation of some data collected. Factor analysis using Principal Component Analysis (PCA) using Equation (3), and (4). Factor analysis is used to determine the correlation value of each observed variable. Variables that have high correlation values are used as input variables in the backpropagation process. Before entering into the backpropagation process, normalization is first carried out. Normalization is intended to equalize the values of several variables in the data set.

If the process produces an optimal model, then the testing process uses the obtained model. the optimal model produced by backpropagation is in the form of several weight values. The weight value is used to be the multiplier value in the testing process. The testing process is finished by optimal model. Each data entered in the model produces predictions of rain. The results of the process are testing calculated for accuracy using the confusion matrix. Calculation of accuracy using Equation (5). The accuracy produced by the confusion matrix is the value of success achieved by a model.

4. Results and Discussion

Data in this study consist of wind speed, air temperature, air pressure, solar radiation, water temperature, and water level data. The data is data obtained from the Meteorology Climatology and Geophysics Agency. The data are 15489 and have 5 parameters, resulting in Data Matrix with a size of 15489 x 5.

In addition to these data, in this research the predictions rain also uses satellite imagery. Satellite imagery is used to take the average value of cloud cover and average cloud temperature from the image. The satellite images used in this study are images taken from the Himawari-8 satellite. The Enhanced Himawari-8 image used is an image of Indonesia's clouds in region two. Indonesia region two in the image of Himawari-8 covers the regions of Sumatra to West Nusa Tenggara. Retrieval of Himawari-8 image data is done using the Filezilla application, which is an application used to access data from satellites. In Enhanced Imagery, the value of each pixel is a value that indicates the cloud peak temperature. One example of the data of the satellite image used in this study is shown in Figure 3.

Both of these data are taken from July until January 2019. The data used are data taken every ten minutes. Also, ten minutes of data retrieval is intended to get the learning pattern of which is increasing In this study; the data used were divided into two, namely training data and testing data. Distribution of data is training 80% of the total data, while data testing is 20% of the total data. Tests were also carried out using 60% data training and 40% data testing.

Figure 3 is a picture showing a cloud-covered region with temperatures below -40°C. This value looks at the colour that covers the area. The colour covering the region is orange, which shows temperatures below -40°C. Because the retrieval of the value fixed on the image with a reddish colour, at a later stage we can take the red value on the original image. The results of taking red intensity on an RGB image showed in Figure 3.
Figure 3. Enhanced IR Imagery on November 13, 2018

The results of taking basic red colours change colour to white, which shows the intensity of the colour. The next step is cleaning noise. Cleaning Noise in the image using a median filter. Median Filter takes the median value on neighboring pixels as a fix to the pixel value. This value corrected to all pixel values in the image. After it is fixed, the noise in the image updated with the new value. Next is the segmentation process. The segmentation process is obtained by an image that presents clouds with low temperatures. The segmentation process gets an image that has a very high red intensity. The process is expected to get clouds that are indicated as cumulonimbus clouds. Cloud segmentation using the method Adaptive Average Brightness Thresh-holding (AABT). Segmentation using the AABT method, performed by calculating the cutoff value. The cutoff value is used to benchmark at the segmentation stage. By using this method, the cutoff value is 112.23.

\[
cutoff = 88.40 + 22.5(\ln(255) - \ln(88.4)) = 112.23
\]

The cutoff value that has been obtained is used at the segmentation stage. If a pixel has a value of less than 112.23, then the pixel value is replaced by 0. If the pixel value is more than 112.23, the pixel value is left fixed. Image segmentation results using the Adaptive Average Brightness Thresh-holding method are shown in Figure 4. Furthermore, to obtain the value of cover in the city of the region, image capture is needed in the area. Intake is done by cutting results in the area used to find the average cover value and the average value of the image. Take the average value of cloud cover obtained by the average value of 98.8304%. Whereas for the average image obtained by taking a total pixel value of 81125 with the number of pixels that are read cumulonimbus clouds is 342 pixels, obtained an average cloud temperature of 237.2076 in pixel values.

\[
could\ cover = \frac{338}{342} x 100\% = 98.8304\%
\]

\[
could\ temperature = \frac{81125}{342} = 237.2076
\]

The average value of cloud temperature, and cloud cover are collected with other observational data. The merger of the two data aims to make one with other parameters. The results of the merger used as input to the backpropagation process. Table 2 is the result of the backpropagation process using several experiments. Also, an experiment created by changing the variables used by doing factor analysis. The results of the factor analysis are two PC, PC 1 consisting of wind, temperature, solar radiation, and water temperature variables. Whereas PC 2 is an average variable cover, and average temperature.
The next step is classification using backpropagation. In the process of getting optimal results, the target used is one if the data produces rain and zero if it does not cause rain. Training using backpropagation uses a number of different treatments. Different treatments use to find out the best treatment with the best results. Some treatments used are the number of neurons in the hidden layer, the type of activation function, and the distribution of data training-testing. In addition, there are several constants used in the process backpropagation, namely the maximum number of iterations of 100,000, the minimum error is \(1 \times 10^{-5}\).

![Figure 4. Segmentation Results Using AABT Method](image)

### Table 2. Results of the Backpropagation Model

| Data Sharing | Number of Hidden Layer | Activation Function | Accuracy Results Before Factor Analysis | After Factor Analysis (PC1) | After Factor Analysis (PC2) |
|--------------|------------------------|---------------------|----------------------------------------|-----------------------------|-----------------------------|
| Training     | Testing                |                     |                                        |                             |                             |
| 80%          | 20%                    | 50                  | Tansig Tansig                          | 46.41704                    | 45.15817                    | 87.605                      |
|              |                        | 50                  | Tansig Logsig                          | 40.51001                    | 40.34861                    | 84.797                      |
|              |                        | 50                  | Logsig Tansig                          | 44.51259                    | 45.51323                    | 88.283                      |
|              |                        | 50                  | Logsig Logsig                          | 40.38089                    | 40.18722                    | 85.539                      |
|              |                        | 30                  | Tansig Tansig                          | 46.51388                    | 48.57973                    | 88.089                      |
|              |                        | 30                  | Tansig Logsig                          | 40.96191                    | 40.6714                     | 81.859                      |
|              |                        | 30                  | Logsig Tansig                          | 44.89994                    | 45.44868                    | 88.154                      |
|              |                        | 30                  | Logsig Logsig                          | 40.96191                    | 40.41317                    | 87.863                      |
|              |                        | 60%                 | 50                  | Tansig Tansig                          | 71.46546                    | 72.43383                    | 87.967                      |
|              |                        | 50                  | Tansig Logsig                          | 69.56101                    | 69.30278                    | 87.130                      |
|              |                        | 50                  | Logsig Tansig                          | 71.4816                     | 72.7889                     | 88.144                      |
|              |                        | 50                  | Logsig Logsig                          | 69.36733                    | 71.27179                    | 87.275                      |
|              |                        | 30                  | Tansig Tansig                          | 71.25565                    | 71.77211                    | 87.483                      |
|              |                        | 30                  | Tansig Logsig                          | 69.33505                    | 69.17366                    | 86.873                      |
|              |                        | 30                  | Logsig Tansig                          | 71.4816                     | 72.69206                    | 87.597                      |
|              |                        | 30                  | Logsig Logsig                          | 69.30278                    | 69.28664                    | 87.243                      |
After getting the optimal model, the model is tested using data testing. Table 2 shows that the maximum model is in the third row. The experiment carried out by taking the variable average cover and cloud temperature. This model resulted in 88.238% success. The accuracy value calculated using the matrix confusion method. The matrix confusion table in the experiment show in Table 3. The matrix results obtained calculated in accordance with the accuracy of the value in the confusion matrix

\[
\text{Accuracy} = \frac{2651 + 84}{2651 + 316 + 47 + 84} \times 100\% = 88.283\%
\]

| Classification Result | Classified as Not raining | Raining |
|------------------------|---------------------------|---------|
| Not raining            | 2651                      | 316     |
| Raining                | 47                        | 84      |

5. Conclusion

The results obtained accuracy prior to analysis by the following factors the factor analysis was done to experience significant differences. The accuracy obtained after doing the factor analysis is much higher than before the factor analysis. After the factor analysis, the accuracy value obtained was 88.283%, an increase compared to before the factor analysis i.e. 71.4816%. The most optimal accuracy value when using two variables at the classification stage is the average cover and average cloud temperature.

References

[1] S. Handoyo, Aviapedia: ensiklopedia umum penerbangan, vol. 1. Penerbit Buku Kompas, 2011.
[2] C. Emiliani, Planet earth: cosmology, geology, and the evolution of life and environment. Cambridge University Press, 1992.
[3] C. Andronache, “Characterization of Mixed-Phase Clouds: Contributions From the Field Campaigns and Ground Based Networks,” in Mixed-Phase Clouds, Elsevier, 2018, pp. 97–120.
[4] J. A. I. Paski, D. S. Permana, A. Sepriando, and D. A. S. Pertwi, “Analisis Dinamika Atmosfer Kejadian Hujan Es Memanfaatkan Citra Radar dan Satelit Himawari-8 (Studi Kasus: Tanggal 3 Mei 2017 di Kota Bandung) Atmospheric Dynamics Analysis of Hail Event Utilizing Radar and Himawari-8 Satellite Imagery (Case Study: May 3, 2017 in Bandung City),” 2017.
[5] B. S. Pandjaitan and M. I. Damayanti, “Kajian Mikrofisis Awan Menggunakan Satelit Himawari 8 pada Kejadian Hail (Studi Kasus: Kejadian Hail di Jakarta Tanggal 28 Maret 2017) Study of Cloud Microphysic Using Himawari 8 Satellite on the Hail Stone Event (Case Study: Hail Stone Event in Jakarta on 28 March 2017).”
[6] A. N. Efendi and M. Fachrurrozi, “Pemanfaatan Data Satelit Ir 1 Himawari-8 untuk Mengurangi Dampak Kerugian Materiil dan Jiwa Akibat Bencana Alam di Wilayah Melawi Kalimantan Barat The Utilization of Ir 1 Himawari-8 Satellite Data to Reduce the Impact of Material Loss and Soul Due to Natural Disaster.”
[7] N. S. Kapadia, D. P. Rana, and U. Parikh, “Weather Forecasting using Satellite Image Processing and Artificial Neural Networks,” Int. J. Comput. Sci. Inf. Secur., vol. 14, no. 11, p. 1069, 2016.
[8] P. Roy, S. Dutta, N. Dey, G. Dey, S. Chakraborty, and R. Ray, “Adaptive thresholding: A comparative study,” 2014 Int. Conf. Control. Instrumentation, Commun. Comput. Technol. ICCICCT 2014, no. July, pp. 1182–1186, 2014.
[9] D. C. Rini, “Klasifikasi Sinyal EEG Menggunakan Metode Fuzzy C-Means Clustering (FCM) Dan Adaptive Neighborhood Modified Backpropagation (ANMBP),” *J. Mat. “MANTIK,”* vol. 1, no. 1, p. 31, 2015.

[10] S. Hardwinarto and M. Aipassa, “Rainfall monthly prediction based on artificial neural network: A case study in Tenggarong Station, East Kalimantan-Indonesia,” *Procedia Comput. Sci.*, vol. 59, pp. 142–151, 2015.

[11] M. Narvekar, P. Fargose, and D. Mukhopadhyay, “Weather Forecasting Using ANN with Error Backpropagation Algorithm,” in *Proceedings of the International Conference on Data Engineering and Communication Technology*, 2017, pp. 629–639.

[12] J. A. I. Paski, A. Sepriando, and D. A. S. Pertiwi, “PEMANFAATAN TEKNIK RGB PADA CITRA SATELIT HIMAWARI-8 UNTUK ANALISIS DINAMIKA ATMOSFER,” vol. 4, no. 3, 2017.

[13] G. Gupta, “Algorithm for image processing using improved median filter and comparison of mean, median and improved median filter,” *Int. J. Soft Comput. Eng.*, vol. 1, no. 5, pp. 304–311, 2011.

[14] A. Fanani, A. Yuniarti, and N. Suciati, “Geometric Feature Extraction of Batik Image Using Cardinal Spline Curve Representation,” vol. 12, no. 2, pp. 397–404, 2014.

[15] A. J. A. Arma, “Aplikasi Analisis Faktor dengan Metode Principal Component Analysis dan Maximum Likelihood dalam Faktor-faktor yang Memengaruhi Pemberian Makanan Tambahan pada Bayi Usia 0-6 Bulan di Desa Pematang Panjang Kecamatan Air Putih Kabupaten Batubara Tahu,” *Kebijakan, Promosi Kesehat. dan Biostatistik*, vol. 1, no. 2, 2015.

[16] G. H. Dunteman and M. S. Lewis-Beck, “Principal components analysis. 1989,” *Newbury Park. Calif.*

[17] P. J. Braspennings, F. Thuijsman, and A. J. M. M. Weijters, *Artificial neural networks: an introduction to ANN theory and practice*, vol. 931. Springer Science & Business Media, 1995.

[18] D. C. R. Novitasari, A. Lubab, A. Sawiji, and A. H. Asyhar, “Application of Feature Extraction for Breast Cancer using One Order Statistic, GLCM, GLRLM, and GLDM,” *Adv. Sci. Technol. Eng. Syst. J.*, vol. 4, no. 4, 2019.

[19] M. Bramer, *Principles of data mining*, vol. 180. Springer, 2007.

[20] S. Visa, B. Ramsay, A. L. Ralescu, and E. Van Der Knaap, “Confusion Matrix-based Feature Selection.,” *MAICS*, vol. 710, pp. 120–127, 2011.