

Determination of the Optimum Feature Extraction Technique Using Wavelet in Classification Phase of Growing Rice

Muhamad Irsyad Habibi*, Sani M. Isa*, Sidik Mulyono**

*Master in Computer Science, Bina Nusantara University, Indonesia

**Agency for The Assessment and Application of Technology (BPPT) Jakarta, Indonesia

irsyad.habibi@gmail.com, sani.m.isa@binus.ac.id, sidik.mulyono@bppt.go.id

Abstract— The purpose of this study is to determine the optimal wavelet-based feature extraction technique based for rice growth stage classification. Data are obtained from the Badan Pengkajian dan Penerapan Teknologi (BPPT). We implemented two decomposition approach i.e. standard wavelet decomposition and wavelet packet analysis with *coif1*, *coif2*, *coif3*, *db2*, *db3* and *haar* as the wavelet basis. The level of decomposition on wavelet decomposition begins from 3 to 11, while on wavelet packet analysis starts from the decomposition level 3 to 6. From each subband we extracted the following features: mean, median, skewness, kurtosis, residual energy, energy, standard deviation, and variance. We used *k*-nearest-neighbor, naive bayes, support vector machine and decision tree as the classifier. The highest accuracy of the wavelet decomposition is 90.41% with *db2* as wavelet basis and 11 level of decomposition using support vector machine as the classifier. Wavelet packet analysis approach gives 80.17% accuracy on Haar wavelet basis and 3 level decomposition using decision tree as the classifier. Based on the experimental results, support vector machine and decision tree have better performance than *k*-nearest-neighbor and Naive Bayes on 77 of total 84 trials.

Keyword— *feature extraction, decision tree, k-nearest-neighbour, naive bayes, support vector machine, wavelet*

I. INTRODUCTION

Indonesia is an agricultural country where most people live by farming, so that agriculture is an important sector that plays an important role in the welfare of the Indonesian population lives [1]. The agricultural sector is the backbone of Indonesia's rice because rice is the main food source of carbohydrates for its residents [2].

According to the [3] the population of Indonesia is increasing every year. One of the consequences is the rice production must also be increased. Efforts are needed to increase rice production with precision farming. Precision farming is a scientific effort to improve the management of the farm with the application of information technology and satellite based technology to identify, analyze, and manage spatial and temporal variables of agronomic parameters [4].

Satellite imagery can be used to predict the total area planted, the age of the rice plant, the area harvested and

productivity estimation [5]. One of the implementation remote sensing technology in the field of agriculture is making hyperspectral image data to determine the condition and age of the rice plant. It is necessary to estimate rice yield in favor of government policies in import rice for food requirements in Indonesia. To get a model in estimation of rice yield with high accuracy must be preceded by the determination of stage rice plant [6]. Wavelet tranform can be use to determine the growth stage of rice, according to [7] wavelet transform has been developed as a powerful analysis tool for signal processing and also successfully applied in applications such as image processing, data compression and pattern recognition. Therefore, in this study we will determine the optimal wavelet-based feature extraction technique for rice growth stage classification.

II. BACKGROUND AND THEORY

A. Rice plants grow stages

The duration of the growth of the rice plant is from 3 to 6 months, depends on the variety and soil environment where the plant grows. Agronomist easier to see in three stages i.e. vegetative, reproductive and maturation. In the vegetative stage, there are three sub-stages i.e. germination, seedling and tillering. There are three reproductive sub-stages i.e. booting, heading and flowering. And for the maturation stage contained a milky, dough and yellow stage [8].

B. Vegetative Stage

In vegetative stage, there are 3 sub-stages: germination, seedling and tillering. The germination stage is marked by the coleoptile emerges from the pack of seeds. Seedling stage, the emerging leaves coleoptile. Then in the tillering stage marked by growing seedling. Vegetative stage characterized by an active tillers, plant height increases gradually and the emergence of leaves regularly [9].

C. Reproductive Stage

Reproductive stage has three stages, namely booting, heading and flowering. Reproductive stage of growth

characterized by stem elongation, reduction in the number of tillers and leaf emergence last. Panicle initiation occurs about 25 days before heading when the panicle has grown by about 1 mm and can be identified by visual or microscope by cutting the stalks. Flowering usually occurs from 1000 hours to 1300 hours in a tropical environment [8].

D. Maturation Stage

The maturation stage divided into three stages: milky, dough and yellowing. This stage can be determined by its texture and color of rice growth. Maturation is also influenced by temperature, time is about 30 days for the area [8].

E. Hyperspectral

Hyperspectral image is a spectral overdetermined, which means they provide spectral information sufficient to identify and distinguish the unique spectral material. Hyperspectral imagery gives information that is more accurate and more detailed than the possibility of other types of remote sensing data [10].

Most of hyperspectral imagery to measure the reflection of radiation with a series of narrow and adjacent wavelengths. Kind of detail pixel spectra can provide more information about the surface than pixel multispectral spectrum [10].

F. Wavelet

Wave (wave) is a function that moves up and down space and time on a periodic basis. While wavelet is a wave that is limited or localized, or it can be said as a short wave. Wavelet focus to energy in space and time [11]. The wavelet transform decomposes signals and shift into a scaled version of the mother wavelet function. Energy variation of hyperspectral different signals on each scale can be detected automatically and provide useful information for hyperspectral image classification [12]. Discrete Wavelet and wavelet transform have proven to be the first step that is suitable for measuring the signal classification [13]. Wavelet analysis procedure is to apply a wavelet prototype function, which is usually called the mother wavelet. [14].

G. Wavelet Decomposition

Wavelet decomposition can decompose the signal into low frequency to represent the optimal approach and high frequencies represent the details [12]. At each parent node is divided into two nodes and only one node that will continue to be divided into two split back.

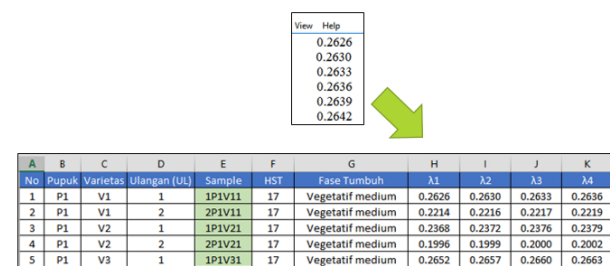
H. Wavelet Packet Analysis

On wavelet packet analysis, every detail coefficients vector also decomposed into two parts as in the wavelet decomposition. However, on wavelet packet analysis is defined as a complete binary tree where each parent node is divided into two nodes again [12].

III. METHODOLOGY

This research is divided into several stages i.e. problem formulation, determine the goals and scope of the research, literature study, software installation, testing, analysis, and the last is the conclusion and suggestions. For the first step is to formulate the problem, then the next step is to determine the purpose of this research based on problem and set the scope so that focus on this research. Further, searching the literature related to this research. Once the literature collected and adopt the part that is needed for this research.

Starting from the screening data that originally contained the data Farmland Response and Farmland Ultra Genjah is then taken only where the data used in this study. Radiographic spectral measurement data in Farmland response originally presented in the form of text files, where each file represents each measurement. The contents of each file in the form of measurement results of spectral reflectance of different wavelengths range from 0.0611 to 1.0625 and has a frequency of up to 826. The contents of each file then transferred in the form of tables and grouped based on the growth stage. Figure 1 shows the examples the data. After data are moved and grouped then stored into one file with .xlsx format. But upon further inspection, there are some empty data. Thus in this research, we eliminates the empty rows of data.



A	B	C	D	E	F	G	H	I	J	K
No	Pupuk	Varietas	Ulangan (UL)	Sample	HST	Fase Tumbuh	λ1	λ2	λ3	λ4
1	P1	V1	1	1P1V11	17	Vegetatif medium	0.2626	0.2630	0.2633	0.2636
2	P1	V1	2	2P1V11	17	Vegetatif medium	0.2214	0.2216	0.2217	0.2219
3	P1	V2	1	1P1V21	17	Vegetatif medium	0.2368	0.2372	0.2376	0.2379
4	P1	V2	2	2P1V21	17	Vegetatif medium	0.1996	0.1999	0.2000	0.2002
5	P1	V3	1	1P1V31	17	Vegetatif medium	0.2652	0.2657	0.2660	0.2663

Figure 1. Sample data transfer process of rows and columns being grouped by growth stage

To more clearly understand the combination in this research can be seen in figure 2 for wavelet decomposition and figure 3 for wavelet packet analysis below, the arrows in the scheme illustrates one process running classification results until it's complete.

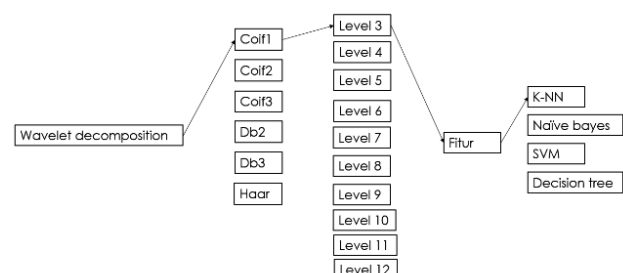


Figure 2. Combination scheme wavelet decomposition

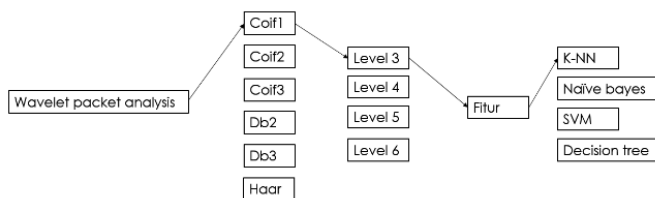


Figure 3. Combination scheme wavelet packet analysis

The filtered and grouped data then imported and stored in the .mat format to make it easier when running the feature extraction process. We developed a feature extraction function as can be seen in figure 4. To illustrate the process of importing the data to the extraction of such features in the image 5. The output of this process in the .xlsx format. The total result of a combination of extraction amounted to 84 obtained from wavelet decomposition with wavelet basis amounted to 6, the decomposition level 3 to level 12, which amounted to 10 and wavelet packet analysis with wavelet basis amounted to 6, the decomposition level 3 to level 6, amounting to 4.

```

for R = 1:439
    [C,L] = wavedec(data(R,:),3,'coif1');
    A3 = C(1:107);
    D3 = C(108:214);
    D2 = C(215:424);
    D1 = C(425:839);

    fitur(R,:)=[mean(A3),std(A3),median(A3),var(A3),skewness(A3)
    Table_decomposition_coif1_level_3(R,:)=table(mean(A3),std(A3)

end

t3=table2array(Table_decomposition_coif1_level_3);
xlswrite('Table_decomposition_coif1_level_3.xlsx',t3);

```

Figure 4. Example function on wavelet decomposition wavelet base coif1 with level 3

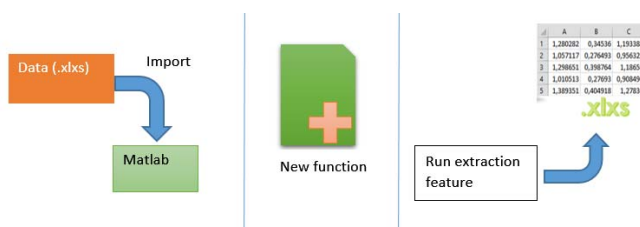


Figure 5. Illustration data import process to feature extraction

The extracted features then imported to Rapid Miner Studio. In Rapid Miner Studio, there is main process window that is used for classification. In the main window process, there is data that has been imported, the type of classification and validation. Total output of all classifications combined with wavelet amounted to 336 results. Where the figure is a combination of the total extraction amounted to 84, multiplied by 4 classifications so get 336.

From the results that have come out, the output will be analyzed the level of accuracy and confusion matrix by comparing each of the results, and then selected the most optimal. Once it is done conclusions and suggestions. For

outlines, the steps of this research can be seen in the figure 6 below.

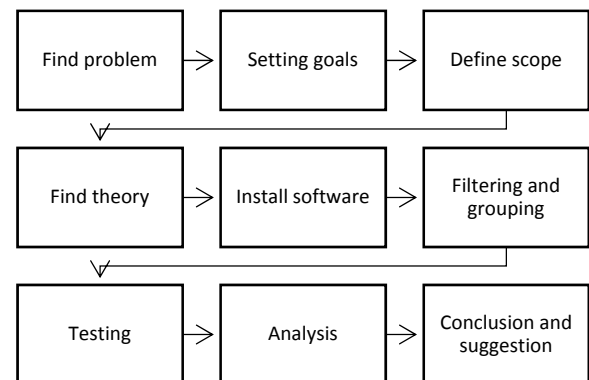


Figure 6. Step - step

IV. EXPERIMENT AND RESULT

In this section, we will analyze the accuracy of the results. Based on the experiment results, we found seven patterns that have certain characteristics that are discussed in this section.

A. Classification Best On Wavelet Decomposition

In the wavelet decomposition using decomposition level that begins from 3 to 12. The highest classification is obtained with a support vector machine on base and db2 wavelet decomposition level 11 with accuracy amounted to 90.41%. While the lowest accuracy obtained on the classification k-nearest-neighbor basis and coif2 wavelet decomposition level 3 with an accuracy of 28.94%. In figure 19 displays confusion matrix of classification support vector machine on base and db2 wavelet decomposition level 11. For the vegetative stage medium has the highest accuracy compared to other stages of 96.88%, of which there is one prediction error at the end of the vegetative stage totaling 2 data. In the stage of seed has the lowest accuracy compared to other stages of 80.95%, of which there are two prediction error at the end of the vegetative stage, amounting to 6 and in stage 6. The prediction error that occurs within the limits of tolerance for the prediction error is more often stage before or afterwards.

accuracy: 90.41% +/- 5.13% (mikroc: 90.43%)

	true Vegetatif m	true Vegetatif al	true Biji	true Malay	true Pra-pati	true Biji matang	true Panen	class precision
pred. Vegetatif r	62	5	0	0	0	0	0	92.54%
pred. Vegetatif r 2	100	6	6	0	0	0	0	87.72%
pred. Biji	0	1	51	4	1	1	0	87.93%
pred. Malay	0	5	6	53	1	0	0	81.54%
pred. Pra-pati	0	0	0	0	60	1	0	98.36%
pred. Biji matang	0	0	0	0	1	61	1	96.83%
pred. Panen	0	0	0	0	0	1	10	90.91%
class recall	96.88%	90.09%	80.95%	84.13%	95.24%	95.31%	90.91%	

Figure 19. The results of the best classification in the wavelet decomposition

B. Classification Best On Wavelet Packet Analysis

In wavelet packet analysis using decomposition level that begins from 3 to 6. The highest classification decision tree obtained at Haar wavelet basis and decomposition level 6 with an accuracy of 80.17%. While the lowest accuracy obtained in

the classification with k-nearest-neighbor basis and coif3 wavelet decomposition level 3 with an accuracy of 24.61%. In Figure 20 displays confusion matrix of the classification decision tree on Haar wavelet basis and decomposition level 6. For the vegetative stage medium has the highest accuracy compared to other stages of 92.19%, of which there are three prediction errors at the end of the vegetative stage, amounting to 3, mature seeds amount stage 1 and stage harvest totaled 1 data. In stage malay had the lowest accuracy compared to other stages of 68.25%, of which there are three prediction error at the end of the vegetative stages totaling 10 seed stage numbered 7 and the pre starch totaling 3 data. In this confusion matrix, the prediction error which happens quite a lot and far apart with the actual stage. Compared with the confusion matrix in figure 19, the confusion matrix in figure 20 is worse.

accuracy: 80.17% +/- 4.94% (mikro: 80.18%)								
	true Vegetatif m	true Vegetatif a	true Biji	true Malay	true Pra-pati	true Biji matang	true Panen	class precision
pred. Vegetatif r	59	5	0	0	0	0	0	92.19%
pred. Vegetatif : 3		82	6	10	2	1	0	78.85%
pred. Biji	0	6	50	7	0	0	0	79.37%
pred. Malay	0	17	7	43	2	0	0	62.32%
pred. Pra-pati	0	1	0	3	58	3	0	89.23%
pred. Biji matang	1	0	0	0	0	51	2	94.44%
pred. Panen	1	0	0	0	1	9	9	45.00%
class recall	92.19%	73.87%	79.37%	68.25%	92.06%	79.69%	81.82%	

Figure 20. The results of the best classification in the wavelet packet analysis

C. The pattern of a specific pattern

In this study, the authors found a 7 pattern as a whole and specifically. Overall pattern means is on the wavelet decomposition and wavelet packet analysis, while the intended pattern in particular is the one type of wavelet. For the overall pattern will be discussed in section D and E, then for the pattern will be specifically discussed in section F to J.

D. Comparison of classification k-nearest-neighbor and Naive Bayes with support vector machine and decision tree

In the wavelet decomposition and wavelet packet analysis, comparison of the accuracy of classification support vector machine and decision tree is almost always higher than the use of classification k-nearest-neighbor and Naive Bayes by comparison truths amounted to 77 from a total of 84 trials. Comparisons are made is the wavelet decomposition level and wavelet same basis. There are seven cases in which the support vector machine classification or decision tree is lower than the k-nearest-neighbor or Naive Bayes. The seventh case is on wavelet decomposition with the Haar wavelet basis decomposition level 3, 4 and 5. Then on the basis of wavelet packet analysis wavelet decomposition level coif2 5, base coif3 wavelet decomposition level 4 and 5, the base db2 wavelet decomposition level 5.

E. Comparison of classification k-nearest-neighbor with Naive Bayes, support vector machine and decision tree

In the wavelet decomposition and wavelet packet analysis, comparison of the accuracy of the classification k-nearest-neighbor tends to be lower compared with Naive Bayes classification, support vector machine and decision tree of truth comparison, amounting to 59 from a total of 84 trials.

Comparisons are made is the wavelet decomposition level and wavelet same basis. There are 25 cases in which the classification k-nearest-neighbor higher than the Naive Bayes classification or a support vector machine or a decision tree. It was found on the basis wavelet decomposition level wavelet decomposition coif1 6 to 12 basis coif3 level wavelet decomposition of 7 to 10 basis db2 wavelet decomposition level 8 basis db3 wavelet decomposition level 9 to 12 basis Haar wavelet decomposition level 3,4,5, 6,7,8,10,11 and on the basis of wavelet packet analysis haar wavelet decomposition level 6.

F. Comparison of the classification decision tree with k-nearest-neighbor, Naive Bayes and support vector machine on wavelet packet analysis

On wavelet packet analysis, decision tree classification produces the best accuracy than the classification k-nearest-neighbor, Naive Bayes and support vector machine by comparison truth amounting to 24 from a total of 24 trials. The comparison is done is on wavelet packet analysis, the level of decomposition and wavelet same basis.

G. Classification of k-nearest-neighbor and Naive Bayes on wavelet packet analysis

Accuracy on wavelet packet analysis using classification k-nearest-neighbor and Naive Bayes tend to increase each level compared classification support vector machine and decision tree of truth comparison numbering 9 of a total of 12 trials. Analysis is conducted by reading charts and accuracy values one by one on the basis of wavelet k-nearest-neighbor and Naive Bayes on wavelet packet analysis. There are three classifications and wavelet basis which has not always increased accuracy values in the wavelet basis coif3 classification decision tree, db3 wavelet basis of classification decision tree and Haar wavelet basis classification k-nearest-neighbor.

H. Comparison of accuracy at the lowest and highest decomposition level

Accuracy on wavelet decomposition at the level of decomposition of 12 tend to be higher than the level of decomposition 3 with a ratio of truth, amounting to 22 from a total of 24 experiments and on wavelet packet analysis at the level of decomposition of 6 is always higher than the level of decomposition of 3 by comparing the truth amounting to 24 from a total of 24 trials. Comparisons are made on wavelet decomposition is to compare the level of decomposition of 3 to 12 on a wavelet basis and the same classification, and for wavelet packet analysis compares the level of decomposition of 3 to 6 on wavelet basis and the same classification. There are two cases where the wavelet decomposition with the Haar wavelet basis, classification k-nearest-neighbor and decision tree decomposition of 3 higher level of accuracy than the decomposition level 12.

I. Classification of support vector machine on wavelet decomposition

In the wavelet decomposition when the level of decomposition level decomposition of 7 to 12, support vector machine classification accuracy is almost always superior to the classification k-nearest-neighbor, Naive Bayes and decision tree of truth comparison, amounting to 35 from a total of 36 trials. The comparison is to compare the classification of support vector machine with k-nearest-neighbor, Naive Bayes, decision tree at each level of decomposition ranging from base wavelet coif1 level of decomposition of 7 to 12 and then a base wavelet coif2 level of decomposition of 7 to 12 and a base wavelet coif3 level of decomposition 7 to 12 and so on until the base wavelet haar. There is one classification of support vector machine is lower than the classification decision tree on the basis wavelet decomposition coif3 on level 8.

J. Classification Naive Bayes on wavelet decomposition

In the wavelet decomposition, Naive Bayes classification accuracy tends to stay constant in the range of 51.00% to 63.55% compared with other classification.

V. CONCLUSION

From the research that has been done we can conclude several things such as feature extraction techniques on the basis db2 wavelet decomposition level 11 with the classification of support vector machine is able to perform the classification stage of growing rice with an accuracy of 90.41%. Classification support vector machine and decision tree is almost always higher than the use of classification k-nearest-neighbor and Naive Bayes by comparison truths amounted to 77 from a total of 84 trials. The future work can added wavelet bases as gaussian, shannon, morlet, meyer and others that can be more valid. Classification used can be added so that the results are valid.

ACKNOWLEDGMENT

Thanks to BINUS University for supporting the students to publish their research. Thanks to Sani M. Isa for being my research supervisor.

REFERENCES

- [1] F. Yani, "Peranan Tanaman Padi Sawah Terhadap Perekonomian Di Kabupaten Kampar Propinsi Riau," 2012.
- [2] A. R. Enhas and others, "Perbedaan Indeks Glikemik Beberapa Menu Makanan Berbahan Dasar Nasi. 2014," 2015.
- [3] B. P. Statistik, "Kepadatan Penduduk Menurut Provinsi 2000-2014." Diakses pada, 2015.
- [4] P. Mondal, M. Basu, and P. B. S. Bhadoria, "Critical Review of Precision Agriculture Technologies and Its Scope of Adoption in India," *Am. J. Exp. Agric.* 1(3) 49-68, 2011, vol. 1, no. 3, pp. 49-68, 2011.
- [5] W. Wahyunto and B. Heryanto, "Pendugaan produktivitas tanaman padi sawah melalui analisis citra satelit," *Inform. Pertan.*, vol. 15, pp. 853-869, 2006.
- [6] F. Maspiyanti, I. Fanany, and A. M. Arymurthy, "Klasifikasi Fase Pertumbuhan Padi Berdasarkan Citra Hiperspektral dengan Modifikasi Logika Fuzzy (Paddy Growth Stages Classification Based On Hyperspectral Image Using Modified Fuzzy Logic)," *J. Penginderaan Jauh dan Pengolah. data Citra Digit.*, vol. 100, no. 1, 2013.
- [7] P.-H. Hsu, "Evaluating the Initialization Methods of Wavelet Networks for Hyperspectral Image Classification," *ISPRS-International Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, pp. 83-89, 2016.

- [8] J. McLean, B. Hardy, and G. Hettel, *Rice Almanac*, 4th edition. 2013.
- [9] P. C. RAM, "Maclean, J.L., Dawe, D.C., Hardy, B. and Hettel, G.P. (eds) Rice almanac. 3rd edn," *Ann. Bot.*, vol. 92, no. 5, p. 739, 2003.
- [10] P. Shippert, "Introduction to hyperspectral image analysis," *Online J. Sp. Commun.*, vol. 3, 2003.
- [11] S. T. Ali Mustofa and others, "Steganografi Pada Citra Digital Menggunakan Metode Discrete Wavelet Transform," *J. Mhs. TEUB*, vol. 1, no. 3, 2013.
- [12] P.-H. Hsu, "Feature extraction of hyperspectral images using wavelet and matching pursuit," *{ISPRS} J. Photogramm. Remote Sens.*, vol. 62, no. 2, pp. 78-92, 2007.
- [13] S. Pittner and S. V. Kamarthi, "Feature extraction from wavelet coefficients for pattern recognition tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 1, pp. 83-88, 1999.
- [14] L. A. Barford, R. S. Fazzio, and D. R. Smith, "An introduction to wavelets. Hewlett-Packard Laboratories." Technical Publications Department, 1992.



Muhamad Irsyad Habibi born in Jakarta, Indonesia and date of birth is August 11, 1992. Educational background is bachelor of information technology. Now study in Binus University Master Information Technology in Jakarta, Indonesia.



Sani M. Isa (M'2015) Finished his Ph.D. in Computer Science, University of Indonesia on 2014. Currently he is a researcher and lecturer at Master in Computer Science, Bina Nusantara University, Indonesia. His research interests are data mining, signal/image processing, remote sensing, and biomedical engineering.