Coupling ground and airborne-based hyperspectral (HyMap) detection over rice canopy to predict leaf area index (LAI) and SPAD value using support vector machine (SVM) technique in irrigated wetland rice, west Java, Indonesia

Muhammad Evri[1], Anto Satrio Nugroho[1], Sidik Mulyono[1], Arif Darmawan[1], Muhamad Sadly[1],Nadirah[1], Nani Hendiarti[1], Osamu Kashimura[2], Chiaki Kobayashi[2], Atsushi Uchida[3], Aya Uraguchi[3], Hozuma Sekine[3]

[1]*Agency for the Assessment and Application of Technology (BPPT), Jl. MH. Thamrin no. 8, Jakarta-10340, Indonesia, phone number: +62-21-3169706, e-mail: evri@tisda.org*

[2]*Earth Remote Sensing Data Analysis Center, FOREFRONT TOWER 3-12-1 Kachidoki, huo-ku, Tokyo 104-0054, Japan, phone number:+81-3-3533-9310, e-mail : kobayashi@ersdac.or.jp*

[3]*Mitsubishi Research Institute, Inc., 3-6, Otemachi 2-chome Chiyoda-ku, Tokyo 100-8141, Japan, phone number: +81-3-3277-0551, e-mail: a-uchida@mri.co.jp*

Key words: Hymap, Hyperspectral, LAI, PLS, Rice, Support Vector Machine, SPAD.

ABSTRACT: Canopy spectral of paddy rice with various growth stages coincidently measured by using field spectroradiometer and HyMap sensors (350 - 2500 nm) that onboard an airplane in tropical irrigated wetland paddy rice covered Indramayu district in West Java province, Indonesia. Meanwhile, destructive sampling undertaken in the same time as well to obtain biophysical parameters such as leaf area index (LAI) and SPAD value during the airplane passed over the area of interest. Sensitive waveband that obtained by exploring overall wavebands toward biophysical parameters to predict biophysical parameters such as LAI and SPAD value. Reflectance value act as predictors to predict LAI and SPAD which is analyzed using support vector machine (SVM) method with Kernel-based Machine Learning methods. SVM currently aroused many attentions due to its robustness in dealing with data with high dimensionality, including hyperspectral data or images. In principle, SVM works as a binary classifier to find the best separating hyperplane in the feature space. By modifying its loss function, SVM can be applied to regression problem in hyperspectral and HyMap hyperspectral remotely sensed data. The study by using SVM method is directed to obtain better performance than other methods such as artificial neural networks (ANN). The result of this paper demonstrated that SVM yield better outcomes in predicting rice crop variables than other method such as ANN. SVM successfully predicted the LAI at correlation coefficient higher than 0.9, while for SPAD the score is 0.77. Artificial Neural Network (ANN) showed slightly lower score for SPAD, while for LAI (10 cm) the CC differs 0.2 lower than SVM. In the case of Subang experiments, SVM outperforms ANN.

1. INTRODUCTION

An airborne campaign and field data campaign held by Agency for The Assessment and Application of technology (BPPT) in cooperation with Earth Remote Sensing Data Analysis Center (ERSDAC) to measure the hyperspectral data and crop variables over rice canopy. During the campaign this mission has recorded huge of hyperspectral data (high data dimension)

High data dimension of hyperspectral data often lead to constraint in data processing and analysis as found in traditional statistics. In this paper, a support vector machine (SVM) method is applied to hyperspectral data analysis. This method does not suffer the limitations of data dimensionality and limited samples.

The foundations of the SVM have been developed by Vapnik (1995) and are gaining popularity in field of machine learning due to many attractive features and promising empirical performance. The Support Vector Machines (SVM) method have been recently introduced in the statistical learning theory domain for regression and classification problems, and applied to the regression of hyperspectral data. The technique directed to finding the optimal separation surface between data classes in order to achieve identification of the most representative training samples of the side of the class. These samples are called *support vectors*. If the training data set is not linearly separable, a kernel method is used to simulate a non-linear projection of the data in a higher dimension space, where the classes are linearly separable. Besides that, unless statistical estimations, a small number of training samples (understood that those are representative) is enough to find the support vectors. Then, this kind of classifier reveals very interesting properties for hyperspectral data processing, because it does not suffer from the particular phenomenon (limited number of training samples, the analysis rate decreases as the dimension increases) and it may perform class separation even with means very closed to each other with a small number of training samples.

SVM are benchmarked to well--known neural networks such as multilayer perceptrons (MLP), Radial Basis Functions (RBF) and Co-Active Neural Fuzzy Inference Systems (CANFIS). Models are analyzed in terms of efficiency and robustness, which is tested according to their suitability to real-time working conditions whenever a preprocessing stage is not possible.

Owing to the development of computer's technology, variable selection can be easily performed and can provide more robust models, where models can transfer more readily and allow non expert users to build reliable models only with limited expert intervention. Further, computer aided selection of variables may be the only approach for some models.

The objective of this paper is to obtain better model performance by using SVM method than other methods such as neural networks to predict crop variables. In addition, this approach is compared to other approaches for obtaining comparison values for getting knowledge upon better, robust and stable models to precisely estimate crop variables.

2. MATERIALS AND METHOD

2.1 Experimental site and data measurement

The experiment was carried out at tropical irrigated wetland in Indramayu (15 km x 30 km) and Subang (19 km x 22 km) Districts, West Java, Indonesia (Fig.1) from June 25 until July 1, 2008. This area is the rice centre that contains some rice cultivars such as Ciherang (Indramayu), ketan, ciherang and IR-42 (Subang). During the experiment, there are three growth stages recorded, namely vegetative, reproductive and ripening. The spectral value classified according to growth stage.

Crop variables of rice, such as leaf area index (LAI) and SPAD value measured quantitatively using LICOR-LAI 2000 and chlorophyll meter (SPAD-502 Minolta) ,respectively through in situ measurement of some hills.

Canopy spectral irradiance in the range of 350 to 2500 nm was measured using spectroradiometer FieldSpect ASD (1 nm intervals) during the experiment. Meanwhile, airborne hyperspectral using sensor HyMap (Hyperspectral Mapper; contain 126 bands; 450 nm~ 2480 nm) undertaken over passed defined target (Subang and Indramayu Districts) with 2 flights for both areas within the experimental time.

The field hyperspectral and crop variables (LAI and SPAD) measurements were performed in the same time at decided area of interest in Subang (6 sites with @ 500 m x 500 m), and

Indramayu (9 sites with @ 500 m x 500 m). One sampling area (500 m x 500 m) consist of 10 selected “quadrate” with the size of 10 m x 10 m, where the measurements carried out in detail inside quadrate. The measurement performed with five replications to reduce the effect of changing sky conditions. The aperture (FOV 25^{o}) of FieldSpect pointed downward at 10 and 50 cm above the canopy. Spectralon (Labsphere, Inc., USA) reference panel was used to optimize the FieldSpect performance prior to taking canopy reflectance measurements.

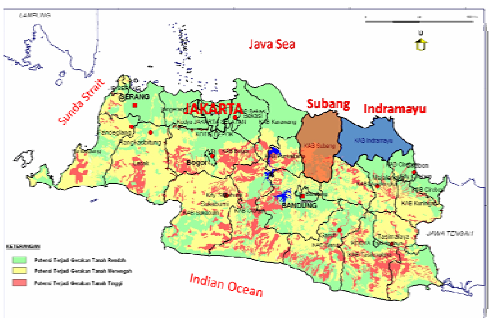

Fig. 1. Research areas (Subang and Indramayu)

2.2 Preprocessing of spectrum data

Spectral domain which displayed low signal-to-noise ratio in both ends of the spectrum (350 ~ 2500) was first omitted from the reflectance data sets. Ground spectral (FieldSpec) wavebands were averaged 15 nm to match with HyMap interval wavebands to reduce the noise and amount of data for analysis. Smoothing data performed using Savitzky-Golay filter (Savitzky and Golay 1964).

2.3 Linear Support Vector Machine

Support Vector Machine (SVM) currently received increasing attentions due to its interesting property and promising performance in many situations. SVM in principal is a linier classifier that is trained to obtain an optimal classification hyperplane on the feature space. The optimal hyperplane is obtained by maximization of the “margin”: a criterion which is defined by the distance between the training samples and the hyperplane [1].

Let us denote each example as $\vec{x}_i \in \Re^d$, $i = 1,2,\cdots,l$. l is the number of examples. Each example is labeled by $y_i \in \{-1,+1\}$, -1 represents the negative class and +1 represents the positive class. Assumed that both positive and negative class is perfectly separated by a hyperplane in *D*-dimensional feature space. The hyperplane could be represented by

$$\vec{w}.\vec{x} + b = 0 \tag{1}$$

Examples $\vec{x}_i$ that belong to the negative class are the ones that satisfy

$$\vec{w}.\vec{x}_i + b \le -1 \tag{2}$$

and, examples $\vec{x}_i$ that belong to the positive class should satisfy

$$\vec{w}.\vec{x}_i + b \ge +1 \tag{3}$$

The optimal margin is calculated by maximizing the distance between the hyperplane and the closest patterns, which is formulated by $1/\|\vec{w}\|$ ($\|\vec{w}\|$ is the norm of vector $\vec{w}$). This problem can be formulated as Quadratic Programming (QP) problem, by minimizing Eq.(4) under constraint (5).

Minimize:

$$\|\vec{w}\|^2 \tag{4}$$

Subject to:

$$y_i(\vec{x}_i.\vec{w} + b) - 1 \ge 0, \quad \forall i \tag{5}$$

The solution to the problem can be obtained by Lagrange Multiplier.

$$L(\vec{w},b,\vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 - \sum_{i=1}^{l} \alpha_i \,(y_i\,(\vec{x}_i.\vec{w}+b)-1) \qquad (6)$$

α_i is the Lagrange multipliers corresponding to example $\vec{x}_i$ that takes zero or positive values. The solution of the optimization can be obtained by minimizing L in respect to $\vec{w}$ and b, and maximizing L in respect to α_i. Equation (6) can be modified to maximization problem that contains only α_i, as shown by Eq. (7)

Maximize:

$$\sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \vec{x}_i.\vec{x}_j \qquad (7)$$

Subject to:

$$\alpha_i \geq 0 \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \qquad (8)$$

The solution of this problem produces many α_i with zero values. The data corresponding to the non-zero α_i are the support vectors, i.e. patterns with smallest distance to the hyperplane.

The previous description is under assumption that both classess are perfectly separable by the hyperplane. However, both classes commonly are not perfectly separable, hence Eq.(5) could not be satisfied. SVM solves this problem by introducing slack variabel ξ_i $(\xi_i \geq 0)$. Equation (5) is modified as follows :

$$y_i(\vec{x}_i.\vec{w}+b) \geq 1-\xi_i, \quad \forall i \qquad (9)$$

Accordingly, Eq. (4) is modified as follows:

$$\text{Minimize} \quad \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{l}\xi_i \qquad (10)$$

Paramater C controls the tradeoff between margin dan classification error. The greater value of C shows the greater is the penalty given to each classification error.

2.4 Non Linear Support Vector Machine

To work with non-linear problem, the example $\vec{x}$ is mapped onto a higher dimensional feature space by a mapping function $\Phi(\vec{x})$. By this transformation, both classes will become linearly separable in the new feature space. Let us denote the mapping as follows

$$\Phi : \Re^d \rightarrow \Re^q \quad d < q \qquad (11)$$

The training phase in SVM is conducted based on the optimization problem as previously stated. As shown in Eq.(7), the optimization needs to calculate the dot product of two samples in the new feature space which is denoted by $\Phi(\vec{x}_i).\Phi(\vec{x}_j)$. This computation could be obtained indirectly, without knowing the tranformation function Φ. This strategy is named as *Kernel Trick*. Instead of computing the dot product in the new feature space, it is possible to use the following kernel function, as given by Eq.(12) :

$$K(\vec{x}_i,\vec{x}_j) = \Phi(\vec{x}_i).\Phi(\vec{x}_j) \qquad (12)$$

Various function can be used as kernel function. In this study we used Polynomial Kernel as defined in Eq.(13) and Radial Basis Function (RBF) Kernel as defined in Eq (14).

$$K(\vec{x}_1,\vec{x}_2) = (\vec{x}_1 \cdot \vec{x}_2)^p \qquad (13)$$

$$K(\vec{x}_1,\vec{x}_2) = \exp\left(-\gamma\|\vec{x}_1-\vec{x}_2\|^2\right) \qquad (14)$$

The decision function of test sample $\vec{x}$ by non-linear SVM is obtained as follows:

$$f(\Phi(\vec{x})) = \vec{w}.\Phi(\vec{x})+b \qquad (15)$$

$$= \sum_{i=1,\vec{x}_i\in SV}^{l} \alpha_i y_i \Phi(\vec{x}).\Phi(\vec{x}_i)+b \qquad (16)$$

$$= \sum_{i=1,\vec{x}_i\in SV}^{l} \alpha_i y_i K(\vec{x},\vec{x}_i)+b \qquad (17)$$

SV (Support Vectors) is the subset of training set $\vec{x}_i$ with corresponding $\alpha_i \neq 0$.

2.5 Support Vector Machine for Regression

Although SVM is basically a binary classifier, it can be modified to solve regression problem by modifying its loss function[2][3]. In simple linear regression, we have to minimize a regularized error function given by

$$\frac{1}{2}\sum_{i=1}^{l}\{f_i - y_i\}^2 + \frac{\lambda}{2}\|\vec{w}\|^2 \quad (18)$$

to obtain sparse solutions, the quadratic error function is replaced by an *ε-insensitive error function*. The function is defined as follows:

$$E_\varepsilon(f(\vec{x}) - y) = \begin{cases} 0, & if\,|f(\vec{x}) - y| < \varepsilon; \\ |y(\vec{x}) - t| - \varepsilon, & otherwise \end{cases} \quad (19)$$

Therefore, we minimize a regularized error function given by

$$C\sum_{i=1}^{l} E_\varepsilon(f_i - y_i)^2 + \frac{1}{2}\|\vec{w}\|^2 \qquad (20)$$

C is the regularization parameter.

By introducing two slcak variables $\xi_i \geq 0$ and $\hat{\xi}_i \geq 0$, the error function for Support Vector Regression is rewritten as follows

$$C\sum_{i=1}^{l} E_\varepsilon(\xi_i + \hat{\xi}_i)^2 + \frac{1}{2}\|\vec{w}\|^2 \qquad (21)$$

which must be minimized subject to constrains

$$\xi_i \geq 0 \qquad (22)$$

$$\hat{\xi}_i \geq 0 \qquad (23)$$

$$y_i \leq f(\vec{x}_i) + \varepsilon + \xi_i \qquad (24)$$

$$y_i \geq f(\vec{x}_i) - \varepsilon - \hat{\xi}_i \qquad (25)$$

The problem can be solved using Lagrange Multipliers, and the prediction for new inputs can be made using the following equation

$$f(\vec{x}) = \sum_{i=1}^{l}(a_i - \hat{a}_i)K(\vec{x}, \vec{x}_i) + b$$

in which a_i and $\hat{a}_i$ are Lagrange Multipliers.

3. DATA PREPARATION

The database is obtained from observation & measurements conducted in Indramayu and Subang, Indonesia. The reflectance data measured at height 10 cm and 50 cm used as training set for SVM, while the HyMap data were used as testing set to evaluate the accuracy of the model. The performance of the model was evaluated based on the correlation coefficient and Root Mean Square Error (RMSE), between the predicted value and the correct values of HyMap data. The performance of SVM was compared to the results obtained by artificial neural network (multilayer perceptron trained by backpropagation algorithm). Detail information of the data is depicted in Tab.1, together with the experimental results.

4. EXPERIMENTAL RESULTS

As widely known in hyperspectral data processing, a large number of waveband combinations cause the reflection coefficient among close wavebands is highly correlated and can lead to overfitting and multicollinearity. Therefore the study continued with applying SVM regression models to predict LAI and SPAD by using all spectrum of hyperspectral to overcome overfitting and multicollinearity. The purpose of using these methods are to gain more precise and stable information of which waveband range were most sensitive to predict LAI and SPAD that away from overfitting and multicollinearity. In addition, the accuracy and the predictability of models can be improved to obtain better, robust and stable predictor to precisely estimate crop variables.

Table 1 demonstrated the results analysis of SVM (regression) attributed to LAI and SPAD. Experiments were conducted using WEKA software [4]. In the case of experiment using Indramayu data, most of the result showed that both SVM and ANN achieved almost similar accuracy. SVM successfully predicted the LAI at correlation coefficient higher than 0.9, while for SPAD the score is 0.77. ANN showed slightly lower score for SPAD, while for LAI (10 cm) the CC differs 0.2 lower than SVM. In the case of Subang experiments, SVM outperforms ANN, however the predicition of SPAD is not satisfied.

In general it seems that result analysis of prediction toward LAI by using ANN demonstrated by coefficient correlation (CC) with the height of 50 cm is more significant than that of the height of 10 cm for both Indramayu and Subang Districts. Meanwhile, in case of using SVM method, the prediction model toward LAI indicated simmilar significance (CC) for both height (10 cm and 50 cm), for both sites (Indramayu and Subang districts) , that achieved more than 0.90. Prediction model toward SPAD using ANN and SVM methods for both height (10 cm and 50 cm) in both sites still represented significant values that achieved higher than 0.70.

In general, predictive model using either ANN and SVM methods over canopy reflectance produced more accurate prediction for LAI than that of SPAD. It can be checked by looking the CC value and RMSE. The higher CC and smaller RMSE represented the fitness of models.

Quantitative value of LAI obtained from *in situ* measurement providing more robust value than that of SPAD which was qualitatively measured at some points to represent the chlorophyll condition in leaves

Table 1 Prediction of LAI, SPAD based on reflectance observed in various wavelengths

Site	**Height**	**Crop variables**	**Height**	**Num. of Quadrates**		**Num. of observed wavelength**	**ANN**		**SVM**	
				Training set	**Testing set (hymap)**		**CC**	**RMSE**	**CC**	**RMSE**
Indramayu	10 cm	LAI	10 cm	64	38	116	0.72	1.45	0.93	0.61
	50 cm	LAI	50 cm	64	38	116	0.96	0.43	0.98	0.97
	10 cm	SPAD	10 cm	65	39	116	0.70	7.93	0.77	10.93
	50 cm	SPAD	50 cm	65	39	116	0.71	4.14	0.77	7.50
Subang	10 cm	LAI	10 cm	40	40	116	0.67	2.53	0.92	0.69
	50 cm	LAI	50 cm	40	40	116	0.82	1.24	0.90	0.80
	10 cm	SPAD	10 cm	40	40	116	0.12	3.07	0.25	16.79
	50 cm	SPAD	50 cm	40	40	116	-0.01	4.49	0.30	3.24

5. CONCLUSION

The result of this paper demonstrated that SVM yield better outcomes in predicting rice crop variables than other method such as ANN. Meanwhile training neural models is unfeasible when working with high dimensional input spaces. For better practical use, the SVM model using hyperspectral data should be further improved for practical crop variables prediction to support robust predicting for management of rice.

6. REFERENCES

[1] Bishop, C.M., 2006. Pattern Recognition and Machine Learning, Springer.
[2] Cristianini, N. and Taylor, J. S., 2000. An Introduction to Support Vector Machine and other kernel-based learning methods, Cambridge University Press.
[3] Savitzky, A., and Golay, E.J.M., 1964, Smoothing and difference of data by simplified least square procedures. Analytical Chemistry, Vol. 36, pp. 1627-1639.
[4]
[5] Smola, A.J and Schoelkopf, B., 1998. A Tutorial on Support Vector Regression. In NeuroCOLT2 Technical Report Series.
[6] Witten, I.H., and Frank, E., 2005. Data Mining: Practical machine learning tools and techniques, 2nd Edition, Morgan Kaufmann, San Francisco.
[7] Vapnik, V., 1995. The Nature of Statistical Learning Theory, chapter 5. Springer-Verlag, New York.