Weed Classification Analysis Using Localized Multiple Kernel Learning (LMKL)

Shahrani Shahbudin, Noor Fithratul Ain Ishak, Hasmaini Mohamad, Murizah Kassim
Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor
shahranis@salam.uitm.edu.my

***Abstract*—Weed classification is a need in the agricultural research to improve the weed control system. There are many kernel-based learning algorithms to identify weed images proposed in the literature; however, most of the weed classification technique proposed a single kernel-based algorithm. Recently, the Localized Multiple Kernel Learning (LMKL) instead of using a single kernel has been proposed for the classification technique that can enhance the interpretability of the decision function and improve performances. LMKL is composed of a kernel-based learning algorithm and a parametric gating model to assign local weights to kernel functions. These two components are trained in a coupled manner using a two-step alternating optimization algorithm. The learning algorithm is derived from three different gating models (softmax, sigmoid, and Gaussian), which applies the LMKL framework on the machine learning problems of binary classification. Therefore, in this work, feature vectors of weed images extracted using the Gabor Wavelet and the Fast Fourier Transform (FFT) were employed to analyze weed pattern images using LMKL algorithms. The result with the aid of gating model are visualized and discussed to prove the performance of LMKL classifier. The results showed the visualization using six types of combinations kernels for all set feature vectors are different for each weed dataset.**

***Index Terms*—Fast Fourier Transform; Gabor Wavelet; Localized Multiple Kernel Learning; Weed Classification**.

I. Introduction

Weed is commonly known as the unwanted plant in human-controlled settings, such as farm fields, gardens, lawns, and parks. Weed is considered as the unwanted plant because they interfere with food and fiber production in agriculture by competing with the desired plants for their food resources, providing hosts for plant pathogens and shelter for animal pests.[1] Therefore, it is compulsory to add an automatic weed detection to the selective patch spraying as a practical solution to reduce the amount of chemical herbicide used in the agricultural practices. [2]

There are many kernel-based learning algorithms to identify weed images proposed in the literature, such as kernel perceptron, support vector machines (SVM), Gaussian processes, principal components analysis (PCA), canonical correlation analysis, ridge regression, spectral clustering, linear adaptive filters and many others. For example, in [3], the Sequential Support Vector Machine Classification is proposed for Small-grain Weed Species Discrimination with Special Regard to *Circium* and *Galium Aparine,* this method obtained an overall classification accuracy of 97.7%. Another research proposed by Francois et al. was on the Bayesian Classification and Unsupervised Learning for isolating weeds in Row Crops, which gives an average of 85% for classification of multiple weeds. [4]

Researchers argued that most of the proposed weed classification techniques were a single-kernel-based algorithm [13-15]. Recently, the Multiple Kernel Learning (MKL) has been proposed for the classification technique as an alternative for the use of a single kernel. It is argued that all of the single-kernel-based learning algorithm can be transformed into a Multiple Kernel Learning algorithm [5, 6].

According to Mehmet Gonen [7], even though MKL classifier combines two types of kernels, it cannot capture the localities exist in the data. The Localized Multiple Kernel Learning (LMKL) gives a more accurate decision boundary by using the gating model that divides the input spaces into regions, according to its kernel function. [8]

Therefore, this paper proposed a binary classification of weed between the *Narrow* and the *Broad* species using LMKL with an addition of gating model rather than using only the decision boundary for the classification visualization.

The rest of this paper is organized as follows: the proposed method is described in Section II. Section III presents the setup of the experiment. The results obtained are presented and discussed in Section IV and finally, Section V provides the conclusion of the paper.

II. Methodology

A. Extracted Data

In this study, 200 data of weed images (100 data from *Narrow* species and 100 data from *Broad* species) were used as the classification input. According to [9], all the weed images were resized to 100 by 100 pixels images using modified Excess Green (MExG) method to separate between the plant and the soil. The original images are then converted into a grey scale image.

(a) (b)

Figure 1: Images of weed species (a) Broad (b) Narrow [2]

Two methods, which are the Gabor Wavelet and the FFT algorithm were applied to find the extracted feature vector of image. The Gabor Wavelet method represents the images that were locally normalized in intensity and decomposed in spatial frequency and orientation. The fast Fourier Transform (FFT) decomposed a discrete signal into its frequency components and shuffles the low frequency components to the corners. Four sets of different orientation, θ were used in this study, which are 0° , 45°, 90° and 134°. Then, the FFT of the extracted Gabor features were computed and produced six feature vectors of "difFFTgabor", which are [0° & 45°], [0° & 90°], [0° & 135°], [45° & 90°], [45° & 135°] and [90° & 135°] [2]. The detailed of feature extraction process can be found in [9].

The analysis of LMKL classification is made based on the two combinations from the six feature vectors extracted from the weed dataset. Figure 2 shows an overall view of weed classification using LMKL.

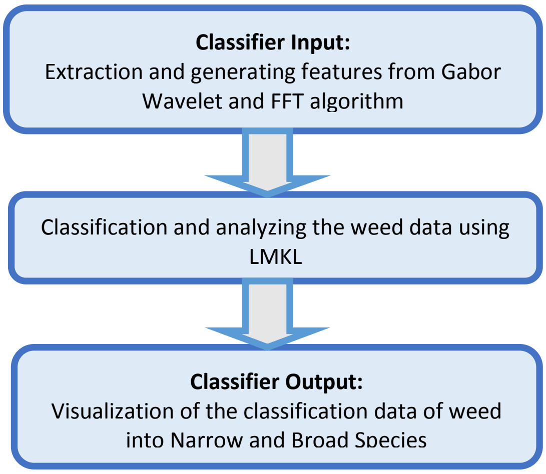


Figure 2: Weed classification system using LMKL

B. Overview of Localized Multiple Kernel Learning

In this study, the derivation of LMKL framework is extended from SVM to the other kernel-based learning algorithms using gating model for selecting the appropriate kernel function locally. The original SVM discriminant function is:

$$f(x) = (w, \Phi_m(x)) + b \tag{1}$$

where $\boldsymbol{w}$ is the weight coefficient, $\boldsymbol{b}$ is the threshold, and Φm is the mapping function for the feature space. [5] In MKL, the rewrite discriminant function is proposed in order to allow local combination of kernels:

$$f(x) = \sum_{m=1}^{p}((w_m, \Phi_m(x)) + b) \tag{2}$$

where $\boldsymbol{w}$m is the weight coefficients and $\boldsymbol{p}$ is the number of kernels. [5] In LMKL, the original SVM formulation is derived with a new MKL discriminant function that allows local combinations of kernels:

$$f(x) = \eta_m(x)(w, \Phi_m(x)) + b \tag{3}$$

where ηm (x) is the gating function that choose feature space m as a function of input x [10]. Assuming that the regions use of kernels are linearly separable, the gating model can be expressed as:

$$\eta_m(x) = \frac{\exp((\eta_m, x) + v_{m0})}{\sum_{k=1}^{p}\exp((\eta_k, x) + v_{k0})} \tag{4}$$

where **vm, vm0** are the parameters for the gating model and the softmax guarantees nonnegativity. Originally, the softmax gating model investigated is:

$$\eta_m(x|V) = \frac{\exp(\langle v_m, x^0\rangle + v_{m0})}{\sum_{h=1}^{P}\exp(\langle v_h, x^0\rangle + v_{h0})} \qquad \forall m \tag{5}$$

where $\boldsymbol{x^0 \epsilon R^0}$ is the input instance in the feature space of gating model. $\mathbf{V} \in \boldsymbol{R^{P\times(D_g+1)}}$ contains the gating model parameters $\{\boldsymbol{v_m, v_{m0}}\}_{\boldsymbol{m=1}}^{\boldsymbol{p}}$.[11] The last discriminant function after determining the *gating function* is:

$$f(x) = \sum_{i=1}^{n}\sum_{m=1}^{p}\alpha_i y_i \eta_m(x) K_m(x, x_i)\eta_m(x_i) + b \tag{6}$$

III. Experimental Setup

From the weed dataset, a random half is reserved as the test set and the remaining half is resampled using 5x2 cross-validation to generate ten training and validation sets. The validation sets of all folds are used to optimized C=1 and 10.

The main algorithm for LMKL is implemented in MATLAB software downloaded from [12]. This proposed method allows the combination of kernels either from the same type of kernels or even with the different type of kernels. [7] There are three common kernels used to perform the simulations. The kernels used are linear kernel (KL), polynomial kernel (KP), and Gaussian kernel (KG) in the following equation 7, 8 and 9 respectively.

$$K_L(x_i, x_j) = (x_i, x_j) \tag{7}$$

$$K_P(x_i, x_j) = \left((x_i, x_j) + 1\right)^q \tag{8}$$

$$K_P(x_i, x_j) = \exp\left(-\left\|x_i - x_j\right\|^2 / s^2\right) \tag{9}$$

The second degree (q=2) is used for the polynomial kernel while in the Gaussian kernel, s is estimated as the average nearest neighbor distance between the instances in the training set. The validation from all the folds are used to optimize the regularization parameter, C with values 1 and 10. [7]

There are six combinations of kernel used in this study. The combination of kernels that have been used are; KL– KL, KP–KP, KG–KG, KL–KP, KL–KG and KP–KG. This paper will only discuss the result from the combination of kernels, which are KL–KP, KL–KG, and KP–KG.

IV. Results and Discussion

The detailed of classification rate of each combination set are shown in term of confusion matrix. Table 1, 2 and 3 show the confusion matrix for the combination of kernel, namely the KL-KG, KL-KP and KP-KG with three set of pairs of feature vector, namely Set A as combination of feature vector [0° & 45°] and [0° & 90], Set B for pairs of [0° & 90°] and

[0° & 135], and finally, Set C for the combination of [0° & 45°] and [0° & 135].

With reference to all tables and based on the three combinations of kernels; KL-KG, KL-KP and KP-KG with the value of C; C=10, the highest classification rate is obtained by Set A with 100% and Set B with the rate of 99%. However, for Set C, the lowest rate recorded is 89% for the combination of kernel of KL-KP.

Table1
Confusion Matrix for LMKL with KL-KG

Predicted Output	Actual Output Value of C = 10 Set A		Set B		Set C	
	N	B	N	B	N	B
Narrow(N)	50	0	49	0	44	4
Broad(B)	0	50	1	50	6	46
Classification Rate (%)	100		99		90	

Table 2
Confusion Matrix for LMKL with KL-KP

Predicted Output	Actual Output Value of C = 10 Set A		Set B		Set C	
	N	B	N	B	N	B
Narrow(N)	50	0	49	0	44	5
Broad(B)	0	50	1	50	6	45
Classification Rate (%)	100		99		89	

Table 3
Confusion Matrix for LMKL with KP-KG

Predicted Output	Actual Output Value of C = 10 Set A		Set B		Set C	
	N	B	N	B	N	B
Narrow(N)	50	0	49	0	44	4
Broad(B)	0	50	1	50	6	46
Classification Rate (%)	100		99		90	

The visualization of the weed classification for the three set feature vectors are shown in Figure 3. It shows that the visualization is different for each data set that uses three types of combinations of kernels. As we can see from the figure, there are two solid lines, which are green and purple. The green solid line is the decision boundary that generates to split broad and narrow weed. The purple solid line shows the boundaries calculated from the gating models that classify the kernel function.

The gating model divides the input space into two regions accordingly to the kernel function used and the decision boundary is induced in each region. From the visualization of gating model of LMKL, we can see that the gating boundary produced when training Set C gives nearly equal combination weights for each kernel, while the crisp output produced when training Set B. However, the classification rate on Set B is higher than Set C due to the simpler datasets, whilst the classification rate for Set C is low due to complicated datasets.

The detailed results of Weed Classification using LMKL, including the number of Support Vector (#SV) and Percentages of Classification Rate (ACC) using C=1 and C=10 are shown in Table 4.

Table 4
Detailed result of Weed Classification using LMKL including Number of Support Vector (#SV) and Percentages of Classification Rate (ACC) using C =1 and C=10.

Set of Data	Value of C	LMKL with KL-KG #SV	ACC (%)	LMKL with KL-KP #SV	ACC (%)	LMKL with KP-KG #SV	ACC (%)
Set A	10	10	100	7	100	9	100
	1	10	98	10	97	15	99
Set B	10	5	99	7	99	9	99
	1	12	99	12	99	29	99
Set C	10	29	88	34	89	33	90
	1	30	89	30	88	34	89

According to Table 4, Set A gives superb classification rate using combination kernel of KL-KP with the smallest number of support vectors, which is 7 using C=10. Meanwhile, Set B recorded the highest value of accuracy rate (99%) using the combination kernel of KL-KG with the smallest number of support vectors, 5 and C=10. Different from Set C, it recorded the best accuracy rate and the smallest number of support vector using combination kernel of KP-KG with 90% and 33 respectively using C=10. It can be concluded that LMKL algorithm can find more reasonable decision boundary and fewer support vector with optimal C value, C=10 for different type of combination kernels.

Table 5
Comparison the result of performance between LMKL and SVM with RBF

Set of Data	Optimal Value of C	LMKL with KL-KG, KL-KP or KP-KG #SV	ACC(%)	SVM with RBF [2] #SV	ACC(%)
Set A	10	7 KL-KP	100	7	99
Set B	1	12 KL-KG	99	22	97
Set C	10	33 KP-KG	90	28	83

To validate and verify the results, the performance of LMKL with three different combination kernels is compared to the conventional SVM [2]. Table 5 shows the comparison results between LMKL with three combination kernels and SVM with single kernel, radial basis function, and RBF. The result of weed using conventional SVM is obtained from the previous work [3]. Based on Table 5, it depicts that the classification rate using LMKL algorithm gives the highest accuracy rate compared to the conventional SVM. As we can see from the table, for LMKL, all sets recorded superb classification rate with a range of 90% to 100% compared to SVM with a range 83% to 99%. For comparison purpose, the optimal value of C for set B, C=1 is used.

In terms of support vector, only set B reduced the number of support vector. However, set A remains the same for both classifiers. Meanwhile, only set C increases the number of support vector compared to the conventional SVM. It can be concluded that by applying LMKL algorithm, it improved the accuracy rate and the number of support vector in weed classification.

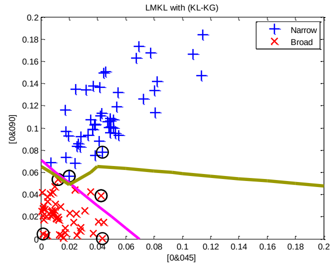
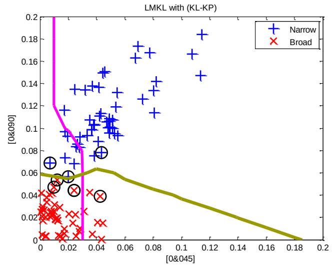
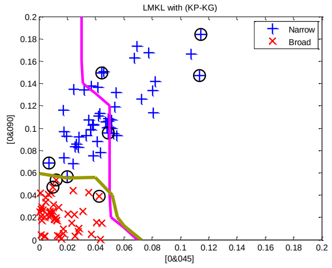
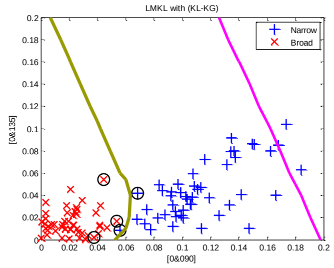
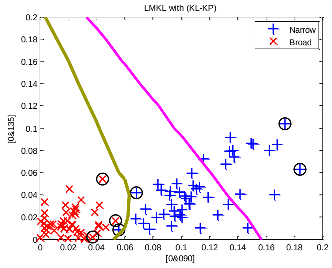
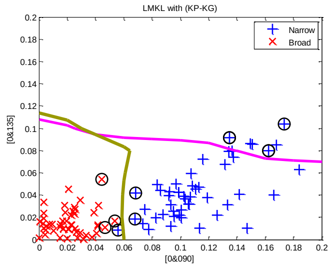
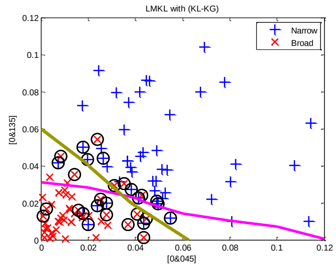
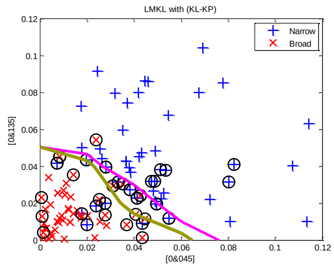
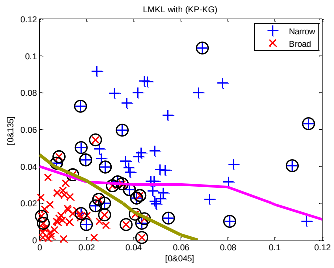
Pair of Set Feature Vectors	LMKL with KL-KG	LMKL with KL-KP	LMKL with KP-KG
Set A [0° & 45°] and [0° & 90°]			
Set B [0° & 90°] and [0° & 135°]			
Set C [0° & 45°] and [0° & 135°]			

Figure 3. Comparison Results of LMKL between 3 different of combination kernels (a) KL-KG, (b) KL-KP and (c) KP-KG with optimal value C = 10

V. CONCLUSION

This paper highlighted the used of gating model in LMKL classifier using three different sets of combination kernels linear, polynomial and Gaussian kernel to analyze weed recognition task to identify weed type as either *Broad* or *Narrow*. Overall, set A gives the best classification rate for LMKL classifier compared to the conventional SVM. The results revealed that the optimal feature vectors to represent the weed images using the ‘‘*difFFTgabor*” feature vectors is from set A with feature vector [0° & 45°] and [0° & 90°]. Furthermore, LMKL identifies the relevant parts of each input image separately using the gating model as a saliency detector on the kernels on the image patches, and confirm that LMKL obtains better classification results than the conventional SVM for weed recognition task.

ACKNOWLEDGMENT

This work was supported by Universiti Teknologi MARA (UiTM), Selangor, Malaysia under LESTARI grant project (Grant Code 600-IRMI/DANA 5/3/LESTARI (95/2015)).

REFERENCES

[1] Sorlini, F.C.a.C., *Applied and Environmental Microbiology*. 2008.

[2] S.Shahbudin, A.H., S. A. Samad, M.M.Mustafa, A.J. Ishak, *Optimal Feature Selection for SVM based weed Classification via Visual Analysis*. 2010.

[3] Rumpf, T.R., Christoph Weis, Martin Sökefeld, Markus Gerhards, Roland Plümer, Lutz, *Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsium arvense and Galium aparine*. Computers and Electronics in Agriculture, 2012. 80(0): p. 89-96.

[4] De Rainville, F.-M., et al., *Bayesian classification and unsupervised learning for isolating weeds in row crops*. Pattern Analysis and Applications, 2014. 17(2): p. 401-414.

[5] Alain Rakotomamonjy, F.B., Stephane Canu, Yves Grandvalet., *SimpleMKL*. Journal of Machine Learning Research, Microtome Publishing, 2008, 9,, 8 Sep 2008.

[6] G¨onen, M., *Multiple Kernel Learning Algorithms*. Journal of Machine Learning Research 12 2011.

[7] G¨onen, M., *Localized Multiple Kernel Learning Algorithms* 2008.

[8] Yina Han, K.Y., Yuanliang Ma, Guizhong Liu *Localized Multiple Kernel Learning via Sample-wise Alternating Optimization*. IEEE Transactions On Cybernetics, 2014. 44.

[9] Asnor Juraiza Ishak, M.M.M.a.A.h., *Weed Feature Extraction Using Gabor Wavelet and fast Fourier Transform* International Symposium on information & Communication Technologies, Nov 2006.

[10] Alain Rakotomamonjy, F.B., Stephane Canu, Yves Grandvalet., *More Efficiency in Multiple Kernel Learning*. 2004: p. 8.

[11] Gönen, M.a.E.A., *Localized algorithms for multiple kernel learning*. Pattern Recognition, 2013: p. 46(3): 795-807.

[12] Gonen, M., *Localized Multiple Kernel Learning*. 2010. URL : *http://users.ics.aalto.fi/gonen/icml08.php*

[13] Razavi-Far, R., Palade, V. & Zio, E. Invasive weed classification.Neural Comput & Applic , 2015, 26: 525. doi:10.1007/s00521-014-1656-3

[14] A Tannouche, K Sbai, Y Ounejjar, M Rahmoune, A Rahmani, A Fast and Efficient Approach for Weeds Identification Using Haar-Like Features, American-Eurasian Journal of Sustainable Agriculture, 2015, 9 (4), 44-48

[15] MH Siddiqi, SW Lee, AM Khan, Weed Image Classification using Wavelet Transform, Stepwise Linear Discriminant Analysis, and Support Vector Machines for an Automatic Spray Control System.J. Inf. Sci. Eng. 2014, 30 (4), 1227-1244