International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

Linear Norm Tree Based Least Square TSVM Scheme for Wilt Diseased Trees Classification

Mayank Arya Chandra[1*], S S Bedi[2]

Department Of CSIT MJP Rohilkhand University, Bareilly, UP, India

Corresponding Author: machandra100@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i4.1317 | Available online at: www.ijcseonline.org

Accepted: 12/Apr/2019, Published: 30/Apr/2019

Abstract: The latest generation of Least Square Twin Support Vector Machine has achieved amazing outcomes in the field of image classification. This paper presents a new method Linear Norms Tree based Least Square Twin Support Vector Machine for developing a plant disease recognition system, which is based on high-resolution multispectral satellite imagery. This proposed model can identify and classify different kinds of plant diseases. Dataset is a composition of diseased trees and other land cover. That are used to identify whether the tree is diseased or not. Experiments with wilt disease data set carried out indicate that new classifier, Linear Norms Tree based Least Square Twin Support Vector Machine, yields a progressively balanced classification accuracy between classes compared to different classification schemes in resolving the imbalanced classification problem.

Keywords: Twin SVM, SVM, Norm, wilt, diseases, least square, imbalance data

I. INTRODUCTION

The issue of plant disease protection is firmly identified with the issues of climate change [1]. Early identification, correct analysis, and tackle within time play a crucial role in protection plants free from pathogens. Bacteria, viruses and fungi are the main causes of plant diseases [2]. The component and any parts of tree can be infected by diseases. Early plant disease detection improves the productivity in the agricultural sector. It is very hard for the formers to detect the diseases visually because it is very time consuming and regular process to see the every plant in frequently. Sometime such types of observation may leads to inaccurate results. In every situation it is not possible to discuss with specialists for this issues. Thus, Automatic recognition of plant disease is important research area in the agricultural field [3].

Kurvonen et al., proposed a method for classifying the remote sensing SAR images and categories the part of land cover and forest types [4]. Kosaka et al., present an image fusion method to classify the high resolution multispectral remote sensing forest image. These images are panchromatic [5]. Yang et al. categorized Boreal forests trees which are situated in Canada with the help of support vector machine concept. Boreal forest satellite multispectral remote sensing image is taken form LiDAR, RapidEye and the combination of these two [6]. A new scheme has been introduced which is based on genetic programing with SVM. Genetic programing is used to optimize the image features and improve the classification accuracy with reduce the mis-classification. It gives the balanced classification rate for both smaller and majority classes [7].

SVM is pure mathematical foundation due to this it gaining most popularity in engineering science and seems to achieve an excellent performance in different real-world applications [8]. SVM is a supervised learning approach which is worked in the domain of regression and classification [9]. SVM computational complexity is quadratic, whose constraint the size of input values [10]. Over the past several decades, different versions of SVM appeared, such as Lagrangian SVM [11], FSVM, Least square SVM [12], Twin SVM [13], FTSVM.

This paper work on very high spatial resolution multispectral remote sensing japanese oak and pine trees satellite imagery set [14] which is taken from UCI machine learning repository was used to classify the diseased trees in forest area of Japan. Oak and pine are one of the major horticultural crops in Japan The classification was done by using Linear Norms Tree based Least Square Twin Support Vector Machine(LN-T-LSTSVM). To examine the performance of the LN-T-LSTSVM, a number of experiments are conducted on different data set.

Rest of this paper is organized as follows: Reviews of previous research presented in section 2. Proposed methodology described in section 3. Section 4 deals with experiment settings. Finally conclude the paper at the end.

II. LITERATURE REVIEW

Applying the proper management strategies such as vector control through pesticide applications, fungicide applications, and disease-specific chemical applications could lead to early data on plant health and disease identification. It can also help you maintain control of diseases and improve productivity. Several improved approaches were proposed by different researchers for diagnosis of plant disease [15-18]. Rumpf et.al. use the concept of SVM to identification and classification of sugar beet diseases with use of vegetation indices. They identify and differentiate between diseased sugar beet leaf and non-diseased sugar beet leaf [19].

A new scheme is proposed by Johnson for identifying diseased oak and pine trees on very high resolution panchromatic multispectral satellite imagery. In which intensity hue saturation model based pansharpening method are used to achieve the panchromatic images. Apply multiscale OBIA method with SVM for classification of different segmentation levels of land cover [15]. A concept of hyper-spectral diseases indices are used to identify the diseases in crop [16]. A mechanism of near filed acoustical holography is used to identify the diseases tree [17]. Complementary based learning method was applied for improve the classification accuracy. Complement naïve Bayesian (CNB) method consider the minority class during classification process [20]. Unlersen et.al. used the concept of weka tool and MLP, KNN to classify the diseased trees Japanese Oak and Pine Wilt [21]. A concept of support vector based on genetic programming used for the classified the wilt data set [7]. Sheela Jeyarani et.al. used the concept of fuzzy and entropy for selecting the best feature. The value of fuzzy entropy is calculated for each features and filtered by linear search for identifying the best feature and features suitability [22]. Mahdiyah use the concept of ELM with integrating data selection method to classify the wilt data set and other benchmarked data set [23]. Another strategy separation of points by planes process is utilized to classify the wilt data set [24].

SVM is a powerful machine learning tool used in classification problem. It has been commonly applied to numerous real-world pattern recognition difficulties [25]. Computational cost of SVM is high, which is the main shortcoming. So an effective methodology name is twin support vector machine (TWSVM) was proposed to reduce the computational cost of SVM [13]. It is inevitable to solve single large problem to find the hyper-plane that lead to high computational cost as conventional SVM, it is better to solve the two quadratic programming problems (QPPs) of smaller size. Suykens [12] proposed a LS-SVM based on least square linear system as a function rather quadratic, to reduce the computational cost. Some extensions of twin SVM has been made by kumar [26], by utilize the concept of least square in twin SVM and proposed the concept of least square twin support vector machine (LS-TSVM)), where it solve the two linear equation. The tree-based least square twin support vector machine was proposed by Chandra et al. [27] to further overcome the problem of class imbalance problem and reduce the computational cost.

Motivation of their improved computational cost and performance in classification domain, here, this paper utilizes the idea of linear norm tree-based least square twin support vector machine for disease tree classification problem. A Schematic process plan of proposed method is illustrated in Fig.1.

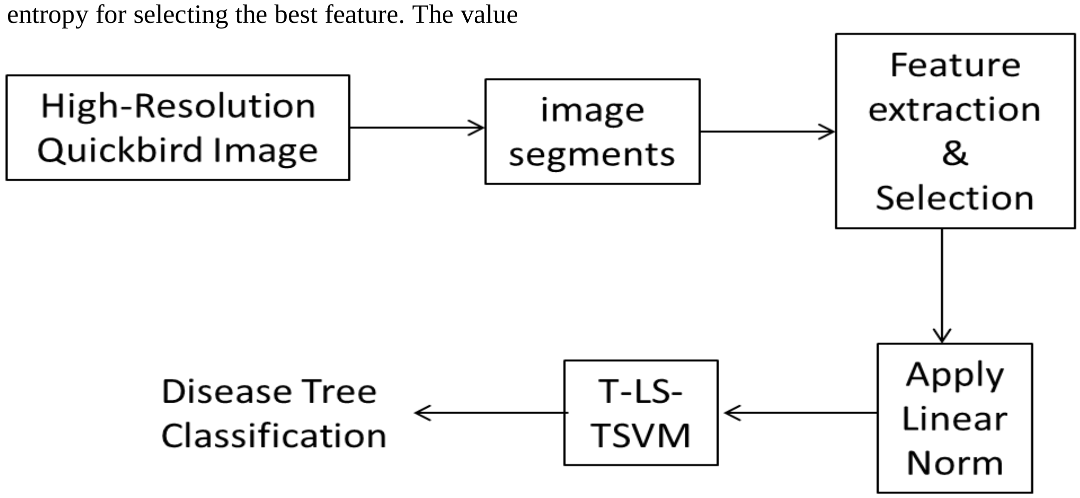

Fig1. Interconnection of the design parts involved.

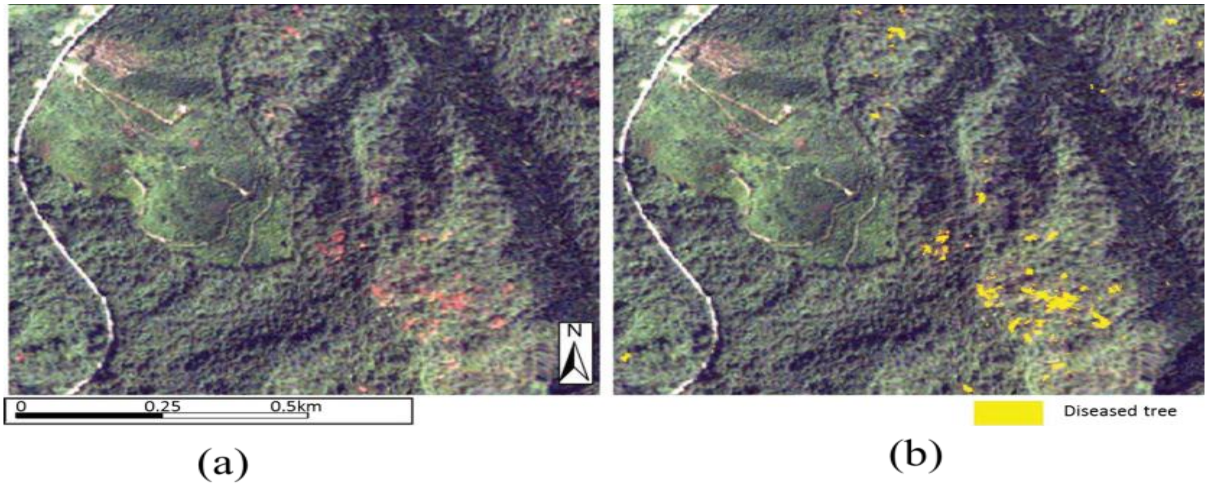

Fig 2. Quickbird satellite images of the study area (a) Classified diseased trees image (b). Light green regions is Diseased oak trees and dark green regions is diseased pine trees [14].

TABLE 1. DATASET DESCRIPTION

Number of Instances	4889
Number of Attributes	6
Missing Values?	NIL
Diseased Trees Set	74
Other Land Cover Set	4265
Number of Class	2

The data values that are used here for identifying the Diseased Trees and non-Diseased Trees sample is taken form UCI Wilt Data Set [14]. The observed region around 3.0 km × 2.5 km, is district of Yonezawa city in Yamagata prefecture, Japan. There is small deciduous broadleaf forest and evergreen needle leaf forest, which consist of brilliant autumn colors of broadleaf trees and the dark green of evergreens, with beautiful residential and agriculture land. The Diseased Trees and non-Diseased Trees are represented in fig.2. It is fine spatial-resolution remote sensing image. This image data set is collection of different segments, which are produced by pansharpened images. Each segment includes the both information first Texture information of panchromatic image band and second spectral information of quickbird multispectral image band. Quickbird images contain multispectral image and pansharpened images. Four bands of multispectral images at 2.4 m resolution with spectral wavelength ranges (blue: 0.450-0.520 µm; Green: 0.520–0.600 µm; Red: 0.630–0.690 µm; NIR: 0.760–0.900 µm). One band of panchromatic images at 0.6 m resolution with wavelength ranges 0.445-0.900 µm. Data set included two target classes of 4889 instances, where each instance consist 6 attributes. Spectral information represented by 4 attributes and texture information represented by 2 attributes. The name of spectral attribute are mean spectral values of Red channel, mean spectral values Green channel, mean spectral values Blue channel, mean spectral values Near Infrared channel values, and texture attributes are standard deviation of pan band, mean texture values of gray level co-occurrence matrix of pan band. This data set is imbalanced data set as it contains 74 instances in diseased trees class and 4265 instances in other land cover class.

III. PROPOSED METHODOLOGY

In this section, we proposed a technique refer as Linear Norms Tree based Least Square Twin Support Vector Machine (LN-T-LSTSVM) carry out experiments on wilt data set (UCI Machine Learning Repository)
These data values normalized at the same time by using linear norm as given formula Eq. (1).

$$\mathrm{la} = (\mathrm{a} - \mathrm{e} * \min)/(\mathrm{e} * (\max - \min)) \tag{1}$$

Where a data is vector, min is minimum values of data vector, max is the maximum values of data vector and e is unit vector.
Following equation are used to find out the LS-T SVM classifiers which represent the two hyperplanes

$$\mathrm{x}^{\mathrm{T}}\mathrm{w}_1 + \mathrm{b}_1 = 0 \tag{2}$$

and

$$\mathrm{x}^{\mathrm{T}}\mathrm{w}_2 + \mathrm{b}_2 = 0 \tag{3}$$

LST-SVM constructs a two hyperplane by using two linear equations as compare to solve a pair of complex QPP's.

$$\min_{\mathrm{w}_1\, \mathrm{b}_1\, \xi_1} \frac{1}{2}(\mathrm{Aw}_1 + \mathrm{eb}_1)^{\mathrm{T}}(\mathrm{Aw}_1 + \mathrm{eb}_1) + \frac{\mathrm{c}_1}{2}\xi_1{}^{T}\xi_1 \tag{4}$$

$$Subject \ -(\mathrm{Bw}_1 + \mathrm{eb}_1) + \xi_1 = \mathrm{e}, \qquad \xi_1 \geq 0 \tag{5}$$

$$\min_{w_2 b_2 \xi_2} \quad \frac{1}{2}(Aw_2 + eb_2)^T(Aw_2 + eb_2) + \frac{c_2}{2}\xi_2{}^T\xi_2 \tag{6}$$

$$Subject\ -(Bw_1 + eb_1) + \xi_2 = e, \qquad \xi_2 \geq 0 \tag{7}$$

The above equation can be resolved by

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -\left(B^T B + \frac{1}{c_1} A^T A\right)^{-1} B^T e \tag{8}$$

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = -\left(A^T A + \frac{1}{c_2} B^T B\right)^{-1} A^T e \tag{9}$$

Calculated decision function from these equations is

$$\text{Class N} = \text{argmin}_{(i-1,2)} \frac{|w_i^T x + b_i|}{\|w_i\|} \tag{10}$$

For binary classification problem LN-T-LSTSVM construct binary classifier and build hyper-plane. In testing phase, it utilizes a binary tree concept that includes root node, non-terminal node, terminal node (leaf node) and it separates one-class from other classes. For given data point tree construction process starts form the root node. The position or movement of given node is specified by the value of decision function. The value of decision function is calculated on the basis of input data points. These output value decide the given data point move either left or right side of binary tree. This process repeated until we achieve the leaf node. The class of leaf is the predicted class of given data sample. Same process is done for multiclass classification problems.

IV. EXPERIMENT RESULT

To check this proposed tree-based LSTWSVM strategy the tests are performed on UCI datasets depicted in Table 1. , we proceed to the parameter values $C = C_1 = C_2$ from the set $\{2^{-8}, \ldots, 2^8\}$ for all cases and δ is 0.5. All computation was carried out on windows 8 Pro, Intel-core- i3 1.7 GHz processor with 4 GB RAM with MATLAB R2013a. Different kernels are used to verify the accuracy of classification depicted in table 2.

TABLE 2. CLASSIFICATION ACCURACY ON DIFFERENT KERNAL

Data Set	**Accuracy (%) Processing Time(s)**		
	Lin	Poly	RBF
Wilt	95.45 0.00027	95.35 0.00035	96.59 0.00026

Table 2 shows the average testing accuracy of proposed method on different kernel function. Equation for calculating the accuracy is defined as

$$\text{Accuracy} = \frac{TP+TN}{FP+FN+TP+TN} \tag{11}$$

Classification accuracy of proposed method is shown with other method on different data set in Table 3. This can be observe from Table 3, that our proposed classifiers perform better and gain highest classification accuracy on all five types of dataset as compared other methods. Proposed method LS-T-LSTSVM classifier took much less time for classification.

TABLE 3. AVERAGE TESTING ACCURACY COMPARISONS FOR EXPERIMENTS WITH OTHER TECHNIQUE

Data Set	**Accuracy (%)**			
	Proposed	**SVM**	**Naïve Bayesian**	**KNN**
Wilt	96.59	80.74	86.37	72.0
Iris	97.56	95.72	96.00	96.66
Skin	94.24	68.51	89.79	88.29
Ecoli	74.58	73.64	60.92	87.5
Shuttle	95.63	72.39	85.00	85.30

V. CONCLUSION

Several studies have been conducted in scientific literature that was focusing on discovering stress in plants using hyper-spectral image processing. Plant disease is also a threat to food security. So detection and classification of a plant disease is a main activity in crop plants in agriculture, horticulture and food security. So in this paper LS-T-LSTSVM method introduced to classify the diseased tree.
LS-T-LSTSVM has been implemented in wide range classification problems. This work proves that LS-T-LSTSVM is suitable candidates for carrying out diseased tree recognition and classification. From table 2, one can observed that RBF kernel function obtains the best performance as compared to linear, polynomial kernel function. The experiments have been conducted on different dataset with other techniques. The result shows in table 3 indicate that LS-T-LSTSVM obtains good results in classification accuracy in comparison.

Even the result of classification is influenced by the value of different parameters to a certain extent. Thus, appropriate parameter selection is an issue and should be addressed in the forthcoming.

REFERENCE

[1] K. A. Garrett, S. P. Dendy, E. E. Frank, M. N. Rouse, and S. E. Travers, "*Climate change effects on plant disease: genomes to ecosystems,"* Annual Review of Phytopathology, Vol. **44**, pp. **489–509**, **2006**.

[2] J. G. Arnal Barbedo, "*Digital image processing techniques for detecting, quantifying and classifying plant diseases,*" SpringerPlus, Vol. **2**, article **660**, pp. **1–12**, **2013**.

[3] Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic, "*Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification*," Computational Intelligence and Neuroscience, Vol. **2016,** pp 1-11, **2016**..

[4] L. Kurvonen, M.T. Hallikainen, "*Textural Information of Multitemporal ERS-1 and JERS-1 SAR Images with Applications to Land and Forest Type Classification in Boreal Zone,"* IEEE Transactions On Geoscience And Remote Sensing, Vol. **37(2)** , pp. **680-689**,**1999**.

[5] N. Kosaka, T. Akiyama, B. Tsai, T. Kojima, "*Forest type classification using data fusion of multispectral and panchromatic high-resolution satellite imageries*," In International Geoscience and Remote Sensing Symposium, Vol. **4**, pp. **2980- 2983**, **2005**.

[6] X.Yang, N. Rochdi, J Zhang, J. Banting, D. Rolfson, C King, B. Purdy, "*Mapping tree species in a boreal forest area using RapidEye and LiDAR data*," In Geoscience and Remote Sensing Symposium (IGARSS), IEEE International, pp. **69-71, 2014**.

[7] M.S.M. Pozi, M. N. Sulaiman, N. Mustapha and T. Perumal, *"A new classification model for a class imbalanced data set using genetic programming and support vector machines: case study for wilt disease classification*," Remote Sensing Letters, **6:7, 568-577, 2015**.

[8] M. A. Chandra & S. S. Bedi, "Survey on *SVM and their application in image classification*," International Journal of Information Technology, pp **1–11, 2018**.

[9] C. Cortes and V. Vapnik, "*Support-vector network*," Machine Learning **20(3):273–297, 1995**.

[10] S. Tao, D Chen, W. Zhao, "*Fast pruning algorithm for multi-output LS-SVM and its application in chemical pattern classification*," Chemometrics and Intelligent Laboratory Systems - hemometrics and Intelligent Laboratory Systems Vol.**96 (1),** pp**. 63-69, 2009**.

[11] O. L. Mangasarian and David R. Musicant, "*Lagrangian Support Vector Machines*" Journal of Machine Learning Research Vol. **1,** pp. **161-177, 2001**.

[12] J. A. K Suykens and J. Vandewalle, *"Least squares support vector machine classifiers*," Neural Process, Letter. **9**, pp**. 293-300, 1999**.

[13] Jayadeva, R. Khemchandani, S. Chandra, "*Twin support vector machines for pattern classification*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. **29**, Issue **5,** pp**. 905–910, 2007**.

[14] B. Johnson, R. Tateishi, N. Hoan, *A hybrid pansharpening approach and multiscale object-based image analysis for mapping diseased pine and oak trees*, International Journal of Remote Sensing, Vol. **34(20)**, pp. **6969-6982, 2013**.

[15] J. G. Arnal Barbedo, *Digital image processing techniques for detecting, quantifying and classifying plant diseases*, Springer-Plus, Vol. **2**, article **660**, pp. **1–12, 2013**.

[16] A. K. Mahlein, T. Rumpf, P. Welke et al., *Development of spectral indices for detecting and identifying plant diseases*, Remote Sensing of Environment, Vol. **128**, pp. **21–30, 2013**.

[17] W. Xiuqing,W. Haiyan, and Y. Shifeng, *Plant disease detection based on near-field acoustic holography*, Transactions of the Chinese Society for Agricultural Machinery, Vol. **2**, article **43,2014**.

[18] J. K. Patil and R. Kumar, *"Feature extraction of diseased leaf images,"* Journal of Signal & Image Processing, Vol. **3**, no. **1**, pp. **60, 2012**.

[19] T. Rumpf, A.-K.Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Pl¨umer, *"Early detection and classification of plant diseases with Support VectorMachines based on hyperspectral reflectance,"* Computers and Electronics in Agriculture, Vol. **74**, no. **1**, pp. **91–99, 2010**.

[20] A. Anagaw, and YL Chang, "A new complement naïve Bayesian approach for biomedical data classification," Journal of Ambient Intelligence and Humanized Computing pp 1–9 (2018).

[21] M F Unlersen and K. Sabanci, "The Classification of Diseased Trees by Using kNN and MLP Classification Models According to the Satellite Imagery," International Journal of Intelligent Systems and Applications in Engineering, Vol **4**, No **2 (2016).**

[22] D. Sheela Jeyarani, A. Pethalakshmi, "*An Efficient Fuzzy Entropy Based Feature Selection Algorithm for High Dimensional Data,"* International Journal of Advanced Research in Computer Science, Vol. **6**, No. **6**, June - July **2015**.

[23] U Mahdiyah, M. I. Irawan, E. Imah, *"Integrating Data Selection and Extreme Learning Machine for Imbalanced Data,"* International Conference on Computer Science and Computational Intelligence (ICCSCI 2015), Procedia Computer Science **59, 221 – 229, 2015**.

[24] C. K. Reddy, D. Hiranmayi, K. Eswaran, G. G. Likhita, B.Saritha, *"Classification of Diseased Plants using Separation of Points by Planes,"* International Journal of Engineering Technology Science and Research, Vol. **4**, Issue **9** September **2017**.

[25] D. Tomar, S. Agarwal, *"Twin Support Vector Machine: A review from 2007 to 2014,"* Egyptian Informatics Journal Volume 16, Issue 1, March 2015, Pages 55-69.

[26] M A. Kumar, M, Gopal, *"Least squares twin support vector machines for pattern classification,"* Expert System Appl **36(3):7535– 7543, 2009**

[27] M.A. Chandra and S.S. Bedi, *"Benchmarking tree-based least squares twin support vector machine classifiers,"* International Journal of Business Intelligence and Data Mining.10.1504/IJBIDM.2018.10009883.

Authors Profile

Mr. Mayank Arya Chandra (Mayank Arya) pursed Bachelor of Technology in CSIT from MJP Rohilkhand University, Bareilly, in 2008, Master of Technology in CSE from GGS IP University, Delhi, in 2012. He is currently pursuing Ph.D. and currently working as Lecturer in Governement Girls Polytehnic Moradabad, BTE Lucknow. . He has published more than 10 research papers in reputed international journals and conferences. His main research work focuses on Information Security, Machine Learning, and Neural Networks.

S.S. Bedi is an Associate Professor at the Department of Computer Science and Information Technology, Institute of Engineering and Technology, MJP Rohilkhand University, Bareilly, India. He is a member of International Association of Engineers (IAENG), IEEE, MIR Labs, Computer Society of India. His areas of interests are information security, network management, digital image processing, quality education through internet. He produced numerous publications in digital image processing and information security. He has published more than 50 research papers in reputed international journals and conferences