Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 16 Issue 3 Version 1.0 Year 2016
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Crop Coverage data Classification using Support Vector Machine

By Tarun Rao, N. Rajasekhar & N C Naveen
Dayananda Sagar College of Engineering

Abstract- A statistical tool which can be used in various applications ranging from medical science to agricultural science is support vector machines. The proposed methodology used is support vector machine and it isused to classify a raster map. The dataset used herein is of Gujarat state agriculture map. The proposed approach is used to classify raster map into groups based on crop coverage of various crops. One group represents rice crop coverageand the othermillets crop coverage and yet another that of cotton crop coverage.Various statistical parameters are used to measure the efficacy of the proposed methodology employed.

Keywords: mining, SVM, supervised classification.

GJCST-C Classification: C.1.2

CROPCOVERAGEDATACLASSIFICATIONUSINGSUPPORTVECTORMACHINE

Strictly as per the compliance and regulations of:

Crop Coverage data Classification using Support Vector Machine

Tarun Rao [α], N. Rajasekhar [σ] & N C Naveen [ρ]

Abstract- A statistical tool which can be used in various applications ranging from medical science to agricultural science is support vector machines. The proposed methodology used is support vector machine and it isused to classify a raster map. The dataset used herein is of Gujarat state agriculture map. The proposed approach is used to classify raster map into groups based on crop coverage of various crops. One group represents rice crop coverageand the othermillets crop coverage and yet another that of cotton crop coverage.Various statistical parameters are used to measure the efficacy of the proposed methodology employed.

Keywords: Mining, SVM, supervised classification.

I. Introduction

Crop mapping is widely used in agriculture andremote sensing science. Crop mapping using classification methodologies serves various applications in agricultural sciencelike gauging water and soil demand etc.. For such applications information on the spatial distribution of classification error is of particular interest [1].Recent progresses in Information Technology systems, lead to dynamic communication among people of every profession. Information technology systems have changed the way people meet and communicate. There is an increasing tendency of professionals and experts in the agriculture sector to communication best practices in the field of agriculture via the medium of internet. Farmers who use the medium of internet get benefited from the various forums used therein to communicate advanced crop yield technologies. Crop mapping can also facilitate the farmers in planning their crop management in advance and they do not see internet and modern technologies has a hurdle [2].

Data is everywhere, abundant, continuous, increasing and heterogeneous. Extracting meaningful information from that data is useful but very difficult: rich data but poor information is a common phenomenon in the world. Data mining (DM) refer to extracting or mining useful knowledge from large amounts of data. One of the various phases of data mining is classification.

Classification is the process in which available data items are categorized into two or more categories depending on the various criterions. Methodologies in which the class label is known apriori is called supervised classification and those in which class labels are not known apriori are called unsupervised classificationor clustering [3]. Supervised classification can be further categorized as parametric and non-parametric categories. Based on whether or not the approach is based on probability distribution or density functions [4].

A well-known statistical method that can be used to solve optimization problems is Support Vector machines (SVM). The proposed methodology used here is SVM. The data items can be represented as feature vectors in a hyper plane and a line passing through the hyper plane can be used to categorize the data items into two different categories. The line can be considered a naïve form of SVM [6] [7]. The An advantage of SVM as a classification method is that in has feature extraction method in-built in its architecture. SVM is better compared to other existing classification methodologies like Naïve bayes, Artificial neural network, decision tree based classification etc. depending on previous research[8][9].

SVM which is inherently linear in nature. However by using kernel function it can be extended to non-linear space as well. In either of the approaches SVM takes a lot of time to classify the data items. SVM approach is used to solve a multi-class classification problem in this research work Its finds a suitable line which is far off from all equidistant points in the hyper plane [10-16].

SVM has numerous applicationsas inland analysis[10], species mapping[11], medicine[9], error identification[12], text and speech analysis[5,13], signal analysis[14etc... SVM is used in this research to classify raster TIFF datasets. Subsequent section explains about Literature Survey on SVM. Later Proposed methodology is explained followed by result analysis. The final section deals with conclusion followed by references.

II. Literature Survey

a) Introduction to SVM

SVM isa promising methodology which is used in various applications. It solves both two class and multi-class problems[15][16]. Problems in which input data items need to categorized into two categories are called two class problems and the ones in which data items need to be categorized into multiple classes are called multi-class problems[17]. The multi-class classification problem can be solved using divide and

Author α σ ρ: Dayananda Sagar College of Engineering, India. His current research interests include Data mining.
e-mails: tarun636@gmail.com, rajasekhar531@gmail.com, ncnaveen@gmail.com

Global Journal of Computer Science and Technology (C) Volume XVI Issue III Version I Year 2016
1

conquer approach. In this approach the problem can be divided into many two-class problems and in the future the results can be merged to arrive at the final solution to the problem.

b) SVM has two major features

Margin maximization: The classification boundary functions of SVMs maximize the margins, which leads to maximizing generalization performance [18].

SVM can be categorized as linear and non-linear SVM as in Fig 1. In linear SVM the hyper plane categorized under two different class labels by a line passing through the hyperplane[18][19][20].

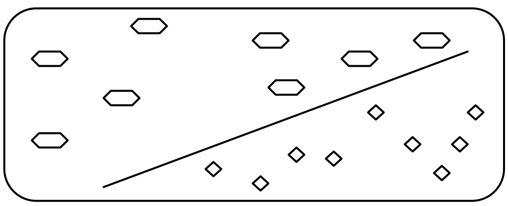


Fig.1: Linear SVM

The line representing the SVM can be denoted by (1)[21]:

$$m\theta i+ \ c>= +1$$
$$m\theta i+ \ c<= -1 \qquad (1)$$

Data items can be represented by (2)[22]:

$$f(x)= \mathrm{sign}(mc+ b) \qquad (2)$$

Where sign() represents sign function, m denotes slope and θ happens to be the angle. Sign function is denoted by:

$$\mathrm{sign}(c)=\begin{cases}1 & \text{if } c>0\\ 0 & \text{if } c=0\\ -1 & \text{if } c<0\end{cases} \qquad (3)$$

Numerous lines might be able to split the planeas different categories but the one that maximizes the distance between itself and the data items in the two categories is known as the support vector as denoted in Figure 2.

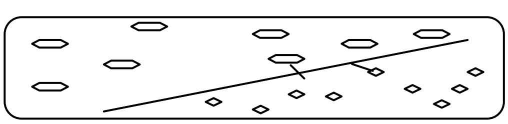


Fig. 2: Distance amid data items in a feature space

The above distance cam be denoted as:

$$M=\frac{(\theta^{+}-\theta^{-}).m}{|m|} \quad = \frac{2}{|m|} \qquad (4)$$

$$h(m)=\frac{1}{2}m^{t}m \qquad (5)$$

subject to $y_i(m\theta_i+ b) >=1, \forall i$

The solution can be denoted with the help of a Lagrange multiplier $\alpha_i$as:

$$m=\sum \alpha_i y_i \theta_i \qquad (6)$$

$b=y_k$ - $m^t x_k$ for any x_k such that Lagrange multiplier $\alpha_k \# 0$

Classifier representation[16]:

$$f(\theta)=\sum \alpha_i y_i \theta_i \, x + b \qquad (7)$$

Systematic nonlinear classification via kernel tricks: SVMs effectively handle non-linear classifications using kernel tricks.

To improve the efficiency of the solution the input data item space can be mapped to a higher dimensional feature space denoted by [18]:

$$K(\theta i,\theta j)=f(\theta_i)^t.f(\theta_j) \qquad (8)$$

The above representation is also known as a kernel function and can be denoted as [23]:

Linear Kernel function $=\theta_i^t\theta_j$

Polynomial kernel function $= (1 + \theta_i^t\theta_j)^p$

Gaussian kernel $= \exp(-\frac{|\theta i-\theta j|^2}{2\sigma^2})$

Sigmoid kernel $= \tanh(\omega_0\theta i\theta j+\omega_1)$ (9)

c) Multi-class SVM

Multi class SVM can be categorized as one-versus-all, one-versus-one, and k-class SVM's[18].

i. *One-versus-all support vector machines*

In this approach SVM classifiers are constructed which separate one class from remaining patterns[18].

ii. *One-versus-one support vector machines*

In this approach k different SVM classifiers are constructed for every pair of classes [18].

iii. *k-Class support vector machines*

In this approachK binary classifiersare built concurrently [18].

III. Proposed Methodology

a) Datasets used

A TIFF data set is used in this research and SVM is used to classify the said data set[24].The TIFF data set is a Gujarat map which has crop coverage data across the state for rice, cotton and millet.

b) Proposed Approach

The TIFF dataset is initially pre-processed. [25]. Later Region Of Interest (ROI) is created from the image. In the next stage training set samples are selected from the ROI. Each of these training set samples correspond to a particular crop coverage in Gujarat map data set used. Three crop coverage's are used for performing the said classification. They are rice, cotton and millet crop coverage's. After the training data sample are collected the SVM classification methodology is used as explained[26]:

Begin
Step 1: Extract features from the data sets
Step 2: Select feature vectors and form the input data set
Step 2: Start dividing the input data set into two sets of data corresponding to two different categories
Step 3: If a data item is not assigned any of theregions mentioned then add it to set of support vectors V
Step 4: end loop
End

Finally the built model is validated against the test data set. Herein the test data set under consideration is the crop coverage area that is not covered as part of the selected training data set sample. One of the key steps involved in the classification process is feature extraction as mentioned below:

Energy (E): It facilitates in computing homogeneousness in the data set and is denoted by:

$$E = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(p(i,j))^2 \tag{9}$$

Contrast(C): Contrast helps identify local data set variation and is denoted as:

$$C = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(i-j)^2 p(i,j) \tag{10}$$

Inverse difference moment (IDM): Local texture alterations can be located using:

$$IDM = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}\frac{1}{1+(j-2)^2}p(i,j) \tag{11}$$

Entropy (S): The data set complexity can be computed by:

$$S = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}p(i,j)\log(p(i,j)) \tag{12}$$

Where μ_k and mxn are the mean and size of the blockB_k

Spatial Frequency (SF): Frequency changes in the data set can be computed using:

$$SF = (RF)^2 + (CF)^2$$

Where

$$RF = \sqrt{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=2}^{n}[I(i,j)-I(i,j-1)]^2} \quad \text{and}$$

$$CF = \sqrt{\frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=2}^{n}[I(i,j)-I(i-1,j)]^2}$$

Variance (V): Level of focus in a data set can be computed using:

$$V = \frac{1}{m\times n}\sum_{i=1}^{m}\sum_{j=1}^{n}(I(i,j)-\mu)^2 \tag{16}$$

Where μ is the mean value of the block image and $m \times n$ is the image size Energy of Gradient (EOG): Measure of focus can also be computed using:

$$EOG = \sum_{i=1}^{m-1}\sum_{j=1}^{n-1}(f_i^2 + f_j^2) \tag{17}$$

Where, f_i = f (i+1, j) - f (i, j)
f_j = f (i, j+1) - f (i, j)

IV. Result Analysis

a) Environment Setting

Agricultural map of Gujarat was used as a dataset to perform the said classification. A region of interest (ROI) was extracted from the map that acted as a training data and it was validated against the complete data segment pertaining to a particular crop in the map. Environment in which the research was undertaken is shown in Table 1[27].

Table.1 : Environment Setting

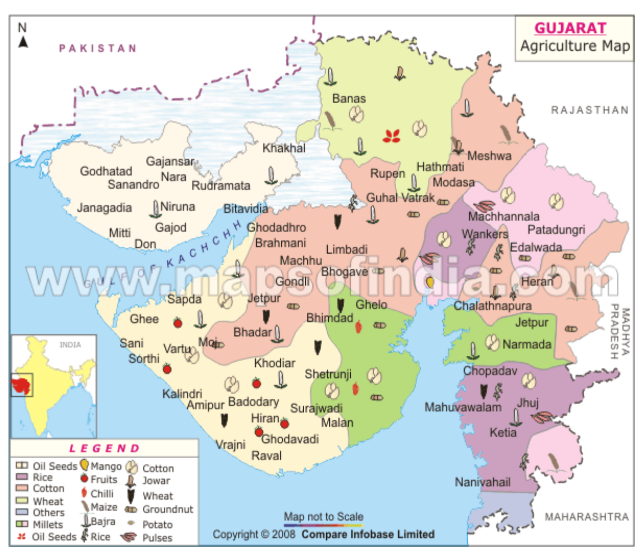
Item	Capacity
CPU	Intel CPU @2 GHz processor
Memory/OS	4GB /WIN 7
Applications	Monteverdi

b) Result Analysis

The ratio of correctly classified and uncorrectly classified data items can be represented using confusion matrix view as mentioned in Table 2. It helps measure the efficacy of the performed classification. Classification results is given in Figure 4.

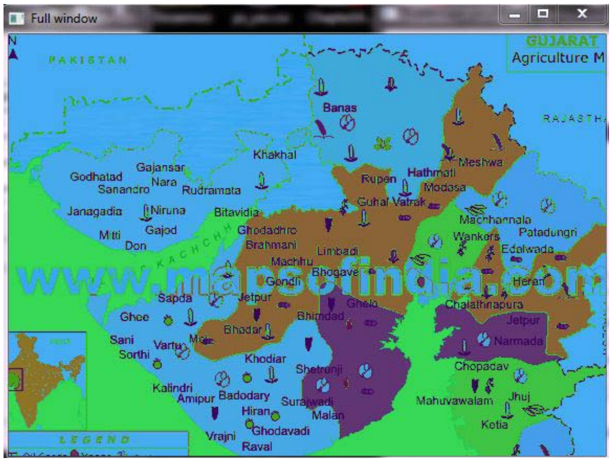
Table. 2 : Confusion Matrix

	Classification result	
	No Event	Event
No Event	True Negative(TN)	False Positive(FP)
Event	FalseNegative(FN)	True Positive(TP)



(a)

Global Journal of Computer Science and Technology (C) Volume XVI Issue III Version I Year 2016



(b)

(c)

Fig. 3 : (a) ROI from the TIFF data set. (b) Classified image with various crop coverage in the state of Gujarat displayed in various colors(Rice-Brown, Millets-Violet, Cotton-Brown). (c) Edge Feature extracted image of the crop data set

Accuracy and kappa statistics are used to measure the efficacy of the classification methodology used.Thes parameters are denoted by equations (18) and (19)[28][29][30]:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{(\text{TP}+\text{FN}+\text{FP}+\text{TN})} \times 100 \tag{18}$$

Kappa statistics=Sensitivity + Specificity - 1 (19)

Confusion matrix in research is mentioned in Table 3.

Table. 3 : Confusion Matrix

Prediction	Reference		
	Rice	Millets	Cotton
Rice	14	0	0
Millets	0	16	0
Cotton	0	0	11

Accuracy and kappa statistics obtained while classifying TIFF data set are mentioned in Table 4.

Table. 4 : Performance measures for TIFF dataset

Data set type	Accuracy	Kappa Statistics
Raster TIFF datasets	100	100

V. Conclusion

SVM classification methodology is used to classify the Gujarat map TIFF data set.Accuracy and kappa statistics parameters are used to measure the efficacy of the said method and the values obtained for the said evaluation parameters prove beyond doubt that the method used classifies the data set with better accuracy.

References Références Referencias

1. Kenichi Tatsumi, Yosuke Yamashiki, Miguel Angel Canales Torres, Cayo Leonidas Ramos Taipe, Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data, Computers and Electronics in Agriculture, Volume 115, July 2015, Pages 171-179, ISSN 0168-1699,http://dx.doi.org/10. 1016/j.compag.2015.05.001.
2. Henrik Skov Midtiby, Björn Åstrand, Ole Jørgensen, Rasmus Nyholm Jørgensen, Upper limit for context–based crop classification in robotic weeding applications, Biosystems Engineering, Available online 15 February 2016, ISSN 1537-5110, http://dx.doi.org/10.1016/j.biosystemseng.2016.01.012.
3. M. Pérez-Ortiz, J.M. Peña, P.A. Gutiérrez, J. Torres-Sánchez, C. Hervás-Martínez, F. López-Granados, A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method, Applied Soft Computing, Volume 37, December 2015, Pages 533-544, ISSN 1568-4946, http://dx.doi.org/10.1016 / j.asoc.2015.08.027.
4. Sybrand Jacobus Muller, Adriaan van Niekerk, An evaluation of supervised classifiers for indirectly detecting salt-affected areas at irrigation scheme level, International Journal of Applied Earth Observation and Geoinformation, Volume 49, July 2016, Pages 138-150, ISSN0303-2434,http://dx.doi.org/10.1016/j.jag 2016.0 2005.
5. Murat Olgun, Ahmet Okan Onarcan, Kemal Özkan, Şahinlşik, Okan Sezer, Kurtuluş Özgişi, Nazife Gözde Ayter, Zekiye Budak Başçiftçi, Murat Ardiç, Onur Koyuncu, Wheat grain classification by using dense SIFT features with SVM classifier, Computers and Electronics in Agriculture, Volume 122, March 2016, Pages 185-190, ISSN 0168-1699, http://dx.doi.org/10.1016/j.compag.2016.01.033.
6. P.J. García Nieto, E. García-Gonzalo, J.R. Alonso Fernández, C. Díaz Muñiz, A hybrid PSO optimized

SVM-based model for predicting a successful growth cycle of the Spirulinaplatensis from raceway experiments data, Journal of Computational and Applied Mathematics, Volume 291, 1 January 2016, Pages 293-303, ISSN 0377-0427, http://dx.doi.org/10.1016/j.cam.2015.01.009.

7. Najafi, B. Ghobadian, A. Moosavian, T. Yusaf, R. Mamat, M. Kettner, W.H. Azmi, SVM and ANFIS for prediction of performance and exhaust emissions of a SI engine with gasoline–ethanol blended fuels, Applied Thermal Engineering, Volume 95, 25 February 2016, Pages 186-203, ISSN 1359-4311, http://dx.doi.org/10.1016/j.applthermaleng.2015.11.009.
8. Yang Long, Fan Zhu, Ling Shao, Recognising occluded multi-view actions using local nearest neighbour embedding, Computer Vision and Image Understanding, Volume 144, March 2016, Pages 36-45, ISSN 1077-3142, http://dx.doi.org/10.1016 / j.cviu.2015.06.003.
9. Sunil S. Morade, Suprava Patnaik, Comparison of classifiers for lip reading with CUAVE and TULIPS database, Optik - International Journal for Light and Electron Optics, Volume 126, Issue 24, December 2015, Pages 5753-5761, ISSN 0030-4026, http://dx.doi.org/10.1016/j.ijleo.2015.08.192.
10. Yang Shao, Ross S. Lunetta, Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 70, June 2012, Pages 78-87, ISSN 0924-2716, http://dx.doi.org /10.1016/j. isprsjprs. 2012.04. 001.
11. Hongji Lin, Han Lin, Weibin Chen, Study on Recognition of Bird Species in Minjiang River Estuary Wetland, Procedia Environmental Sciences, Volume 10, Part C, 2011, Pages 2478-2483, ISSN 1878-0296, http://dx.doi.org/10.1016/j.proenv.2011.09.386.
12. Baoping Tang, Tao Song, Feng Li, Lei Deng, Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine, Renewable Energy, Volume 62, February 2014, Pages 1-9, ISSN 0960-1481, http://dx.doi.org/10.1016/j.renene.2013.06.025.
13. A.D. Dileep, C. Chandra Sekhar, Class-specific GMM based intermediate matching kernel for classification of varying length patterns of long duration speech using support vector machines, Speech Communication, Volume 57, February 2014, Pages 126-143, ISSN 0167-6393, http://dx.doi.org/10.1016/ j.specom.2013.09.010.
14. Rao, T.; Rajasekhar, N.; Rajinikanth, T.V., "Drainage water level classification using support vector machines," Engineering (NUiCONE), 2013 Nirma University International Conference on, vol., no., pp.1, 6, 28-30 Nov. 2013doi: 10.1109/NUiCONE.2013.6780068.
15. Lam Hong Lee, Rajprasad Rajkumar, Lai Hung Lo, Chin Heng Wan, Dino Isa, Oil and gas pipeline failure prediction system using long range ultrasonic transducers and Euclidean-Support Vector Machines classification approach, Expert Systems with Applications, Volume 40,Issue 6,May 2013, Pages 1925-1934,ISSN 0957-4174, http:// dx.doi.org/10.1016 /j.eswa.2012.10.006.
16. Elias Zintzaras, Axel Kowald, Forest classification trees and forest support vector machines algorithms: Demonstration using microarray data, Computers in Biology and Medicine, Volume 40, Issue5, May 2010, Pages 519-524, ISSN0010-4825.
17. Omar Abdel Wahab, Azzam Mourad, Hadi Otrok, Jamal Bentahar, CEAP: SVM-based intelligent detection model for clustered vehicular ad hoc networks, Expert Systems with Applications, Volume 50, 15 May 2016, Pages 40-54, ISSN 0957-4174, http://dx.doi.org/10.1016/j.eswa.2015.12.006.
18. Sarah M. Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, Christopher Leckie, High-Dimensional and Large-Scale Anomaly Detection using a Linear One-Class SVM with Deep Learning, Pattern Recognition, Available online 14 April 2016, ISSN 0031-3203, http://dx.doi.org/10.1016/j.patcog.2016.03.028.
19. Yi Yan, Wei Kai Zheng, JianBao, Ran Liu, An enhanced M-ary SVM algorithm for multi-category classification and its application, Neurocomputing, Volume 187, 26 April 2016, Pages 119-125, ISSN 0925-2312, http://dx.doi.org/10.1016/j.neucom.2015.08.101.
20. Azadeh Sadat Mozafari, Mansour Jamzad, A SVM-based model-transferring method for heterogeneous domain adaptation, Pattern Recognition, Volume 56, August 2016, Pages 142-158, ISSN 0031-3203, http://dx.doi.org/10.1016/j. patcog.2016.03.009.
21. Muhammad Tahir, Asifullah Khan, Protein subcellular localization of fluorescence microscopy images: Employing new statistical and Texton based image features and SVM based ensemble classification, Information Sciences, Volume 345, 1 June 2016, Pages 65-80, ISSN 0020-0255, http://dx.doi.org/10.1016/j.ins.2016.01.064.
22. Yan Leng, Chengli Sun, Xinyan Xu, Qi Yuan, Shuning Xing, Honglin Wan, Jingjing Wang, Dengwang Li, Employing unlabeled data to improve the classification performance of SVM, and its application in audio event classification, Knowledge-Based Systems, Volume 98, 15 April 2016, Pages 117-129, ISSN 0950-7051, http://dx.doi.org/10.1016/j.knosys.2016.01.029.

23. Chao Zhou, Kunlong Yin, Ying Cao, Bayes Ahmed, Application of time series analysis and PSO–SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China, Engineering Geology, Volume 204, 8 April 2016, Pages 108-120, ISSN 0013-7952, http://dx.doi.org/10.1016/j.enggeo.2016.02.009. http://www.infobase.co.in.
24. D.Lu& Q. Weng (2007):A survey of image classification methods and techniques for improving classification performance, International Journal of Remote Sensing, 28:5,823-870,http://dx.doi.org/10.1080/01431160600746456.
25. S. N. Jeyanthi, Efficient Classification Algorithms using SVMs for Large Datasets, A Project Report Submitted in partial fulfillment of the requirements for the Degree of Master of Technology in Computational Science, Supercomputer Education and Research Center, IISC, BANGALORE, INDIA, June 2007.
26. Jennifer A. Taylor, Alicia V. Lacovara, Gordon S. Smith, Ravi Pandian, Mark Lehto, Near-miss narratives from the fire service: A Bayesian analysis, Accident Analysis & Prevention, Volume 62, January 2014, Pages 119-129, ISSN 0001-4575, http://dx.doi.org/10.1016/ j.aap.2013.09.012.
27. David J. Rogers, Jonathan E. Suk, Jan C. Semenza, Using global maps to predict the risk of dengue in Europe, ActaTropica, Volume 129, January 2014, Pages 1-14, ISSN 0001-706X.
28. Rafael Pino-Mejías, María Dolores Cubiles-de-la-Vega, María Anaya-Romero, Antonio Pascual-Acosta, Antonio Jordán-López, Nicolás Bellinfante-Crocci, Predicting the potential habitat of oaks with data mining models and the R system, Environmental Modelling & Software, Volume 25 ,Issue 7, July 2010, Pages 826-836, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2010.01.004.

Global Journal of Computer Science and Technology (C) Volume XVI Issue III Version I Year 2016