



IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: X Month of publication: October 2017 DOI: http://doi.org/10.22214/ijraset.2017.10172

www.ijraset.com

Call: 🕥 08813907089 🔰 E-mail ID: ijraset@gmail.com



Supervised Classification of Rainfall Coverage Data in Andhra Pradesh Using Support Vector Machine

Kolluru Venkata Nagendra¹, Dr. Maligela Ussenaiah².

¹Research Scholar, Asst. Professor², Department of Computer Science Vikrama Simhapuri University SPSRNellore.

Abstract: Support vector machines have been used as a classification method in various domain's including and not restricted to species distribution and land cover detection. In this paper the support vector machine classification method is applied to geographical information systems data represented in the form of a TIFF map. The dataset used herein is of Andhra Pradesh state rainfall map. Support vector machine classification method is used herein to classify the image into categories based on rainfall coverage in various districts. One category represents excessive rainfall coverageand the othernormal rainfall coverage and yet another that of deficient rainfall coverage. Later the classifier is evaluated using kappa statistics and accuracy parameters.

Keywords: Classification, Data Mining, Support Vector Machine, remote sensed data.

I. INTRODUCTION

Meteorological satellite data have been operationally used in weather services for more than 30 years. During this period, forecasting of severe weather based on satellite remote sensing data has been a challenging task. Early warnings of severe weather, made possible by timely and accurate forecasting will help prevent casualties and damages caused by natural disasters. However, measuring and monitoring rainfall events remains challenging and costly. Indian Summer Monsoon Rainfall (ISMR) is a complex phenomenon involving atmosphere, land, ocean and many other domains. When we speak about a large phenomenon like monsoon, we are aware that there are many variables involved that may very strongly relate to monsoon, variables that may relate weakly, variables that may not be related to monsoon but are present in climate scenario like in theories such as sea breeze1. A question arises here as to what are the variables to be considered for predicting or formulating association rules so as to give better results[1].

Nowadays, the advances in Information and Communication Technologies (ICT), lead to an alternative mode of communication among people of every profession. They have changed the way people meet and communicate. Communication regarding excessive rainfall and floods also can be passed on through ICT systems. Therefore, evidence suggests that communication needs to continue to embrace the use of the Internet[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 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 wherein a class label is assigned to unlabeled data vectors. Classification can be further categorized as supervised and unsupervised classification. In supervised classification the class labels or categories into which the data sets need to be classified into is known in advance. In unsupervised classification the class label is not known in advance[3]. Unsupervised classification can be subdivided into non-parametric and parametric classification. Parametric classifier method is dependent on the probability distribution of each class. Non parametric classifiers are used when the density function is not known[4].

Support Vector machines(SVM) is a classification technique based on the statistical learning theory. It is successfully used to solve optimization problems on large sets. In this paper SVM are used to perform the said classification. Herein the data vectors are represented in a feature space. The purpose of the algorithm is to find a hyperplane that splits optimally the trainingset. It helps in solving equallytwo class and multi class classification problem[6][7]. The aim of the said hyper plane is to maximize its distancefrom the adjoining data points in the two regions. Moreover, SVM's do not have an additional overhead of feature extraction since it is part of its own architecture. Latestresearch hasproved that SVM classifiers providebetter classification results when one uses spatial data setsas compared to other classification algorithms like Bayesian method, neural networks andk-nearest neighbors classification methods[8][9].

Support Vector Machine(SVM) is a well-known machine learning method, which has been used frequently in classification problems because of its strong theoretical foundation and its good performance in practice.[9]. A SVM is a linear classifier, but in most cases, it



is practically restrictive. SVM can be easily extended to a nonlinear classifier by mapping the input space into a high dimensional feature space through a kernel function. However, training times can be prohibitively long for both cases, even with specially tailored quadratic programming solvers, growing exponentially with the number of classes in the problem. In this work we propose an SVM-based linear programming formulationtosolvethemulti-classclassificationproblemefficiently. We based our work on the concept of the center of the configuration [1,16] to obtain a point which is equidistant to all classes, while the classification functions are constructed based on this point.

SVM have been used to classify data in various domains like land cover classification[10], species distribution[11], medical binary classification[9], fault diagnosis[12], character classification[5], speech recognition[13],radar signal processing[14],habitat prediction etc... In this paper SVM is used to classify remote sensed data sets. Two formats of remote sensed data viz. raster formatand comma separated value(CSV) file formats have been used for performing the said classification using SVM.

Our next section describes Background Knowledge about SVM classifiers. In section 3materials and methods viz. data acquired and the proposed methodology have been discussed. Performance analysis is discussed in Section 4. Section 5 concludes this work and later acknowledgement is given to the data source followed by references.

II. BACKGROUND KNOWLEDGE

A. Overview of SVM Classifier

Support vector machine (SVM) is a promising methodology which is used in various applications. They have a strong mathematical foundation. They are used to solveboth two class and multi class classification problem[15][16]. In a two class problem the input data has to be categorized as two diversecategories wherein each category is assigned a unique class label[17]. A multi class classification problem can be divided it into multiple two class classification problems and solved by aggregating the individual results to get the final result of the multi class problem.

Success of SVM depends on their strong mathematical foundation that conveys several significant properties:

Margin maximization: The classification boundary functions of SVMs maximize the margins, which leads to maximizing generalization performance.SVM can be categorized into non-linear and linear SVM. Data can be represented in space as shown in Fig 1. Linear SVM can be geometrically represented by a line which divides the data space into two different regions thus resulting in classifying the said data which can be assigned two class labels corresponding to the two regions[18][19][20].

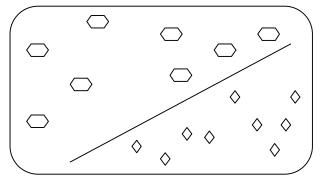


Fig.1.The Hyperplane

The line mentioned herein is called a hyperplane and can be mathematically represented by equation (1)[21]:

$$\begin{split} &m\theta_i + b >= +1 \\ &m\theta_i + b <= -1 \\ & (1) \\ & \text{The data points can be represented by equation (2)[22]:} \\ & f(x) = sgn(m\theta + b) \\ & (2) \end{split}$$

where sgn() is known as a sign function, which is mathematically represented by the following equation:

 $\operatorname{sgn}(\theta) = \begin{cases} 1 & \text{if } \theta > 0 \\ 0 & \text{if } \theta = 0 \\ -1 & \text{if } \theta < 0 \end{cases}$ (3)



There can be many hyperplanes which can divide the data space into two regions but the one that increases the distance amid the bordering data points in the input data space is the result to the two class problem. The adjoining data points close to this hyperplane are called support vectors. This concept can be illustrated geometrically as in Figure 2.

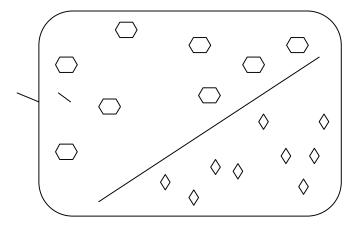


Fig.2.Distance of the nearest data vectors from the Hyperplane

The margin width can be represented mathematically by the equation:

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

This maximization problem viz. maximizing the distance between the hyperplane and the adjoining support vectors can be represented as a Quadratic Optimization Problem as in equation(5)[22][23]:

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

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

The solution for this problem can be provide by a Lagrange multiplier α_i which is associated with every constraint in the main problem. The solution can be represented as:

 $m = \sum \alpha_{i} y_{i} \theta_{i}$ $b = y_{k} - m^{t} x_{k} \text{ for any } x_{k} \text{ such that Lagrange multiplier } \alpha_{k} \# 0 \qquad (6)$ The classifier can be denoted as[16]: $f(\theta) = \sum \alpha_{i} y_{i} \theta_{i} x + b \qquad (7)$

1) Systematic nonlinear classification via kernel tricks: SVMs effectively handle non-linear classifications using kernel tricks. If every data point in the input data space is generalized onto a higher dimensional feature space which can be represented as [18]: $K(\theta_i, \theta_i) = f(\theta_i)^t f(\theta_i)$ (8)

This is also called a kernel function. It is computed using an inner dot product in the feature space. Various kernel functions can be used to do the said mapping as mentioned in the below equations[23]:

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

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

Gaussian radial based kernel function =

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

Sigmoid kernel function= $tanh(\beta_0\theta_i\theta_j+\beta_1)$ (9)

One of the major advantages of SVM is that feature selection is automatically taken care by it and one need not separately derive features.

B. Multi-classSupport Vector Machines

In this section we describe the multi-class Support Vector Machines approach in its three most common forms(one-ver- sus-all SVM, one-versus-one SVM, and k-class SVM).



- One-versus-all support vector machines: One-versus-all support vector machines are the simplest and probably the earliest formulation for multi-class SVM. This approach constructs k binary SVM classifiers, where each one separates one class from the remaining training patterns.
- 2) One-versus-one support vector machines: Another important SVM-based multi-class classification method is known as one-versus-one(OvO) Support Vector Machine. This method constructs K(K-1)/2 binary SVM classifiers, one for every pair of classes. For training data from the k-th and the l-th classes, k≠l(k < l), OvO-SVM solves the following binary classification problem</p>
- *3) k-Class support vector machines*: In Weston and Watkins [36], an all-together approach for multi-class SVM by solving one single optimization problem was proposed. This approach constructs K binary classifiers simultaneously.

III. MATERIALS AND METHODS

A. Data Acquisition

Herein SVM methodology is applied to a TIFF dataset. The TIFF data set used for the said classification is in raster format[25]. Raster image is a collection of pixels represented in a matrix form. Raster images can be stored in varying formats. A map of Andhra Pradesh state that comprises of the rainfall coverage's for various districts are used.

IV. PROPOSED METHOD

The data under consideration is first preprocessed. [26]. Later Region Of Interest is extracted from the dataset. In the next stage training set samples are selected from the ROI. Each of these training set samples correspond to a particular Rain fall in Andhra Pradesh map data set used. After the training data sample are collected the SVM classification methodology is applied. Algorithm that explains implementation of SVM is given below[27]:

Begin

Step 1: Loop the n data items

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 the

egions 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 Rain fall coverage area that is not covered as part of the selected training data set sample.

V. PERFORMANCE ANALYSIS

A. Environment Setting

Rain fall map of Andhra Pradesh 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 Rain fall in the map. The proposed method has been implemented under the environment setting as shown in Table 1[28][29].

Table.1.Environment Setting			
	Capacity		
Item			
CPU	Intel CPU @2 GHz processor		
Memory	4GB RAM		
OS	Windows 7 32-bit		
Tools	Monteverdi tool		

VI. RESULT ANALYSIS

Performance of a Classification method can be measured using parameters of a confusion or error matrix view depending on whether the event is correctly classified or no event is correctly classified as shown in Table 2. And the classified results for the image used sets are demonstrated in Figure 3.



Table.2.Confusion / Error Matrix View				
Real group	Classification result			
	No Event	Event		
No Event	True Negative(TN)	False Positive(FP)		
Event	False Negative(FN)	True Positive(TP)		

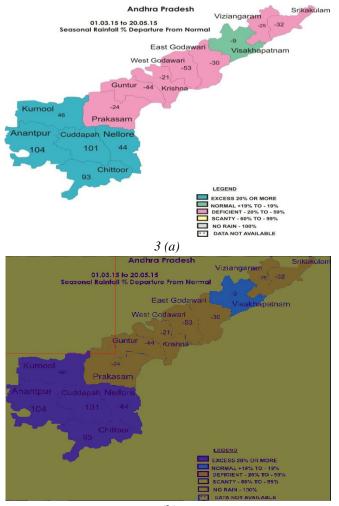




Fig.3. (a) Region of Interest from the input raster data set. (b) Classified image with excess, normal and deficient rainfall data in state of Andhra Pradesh displayed in various colors(Excess-Violet, Deficient-magenta, Normal-Blue)

In this paper the parameters used to evaluate the classification is Accuracy and kappa statistics. The formulae for accuracy, specificity, sensitivity and kappa statistics are provided by equations (10), (11), (12) and (13)[30][31][32]:

Accuracy = $\frac{\text{TP+TN}}{(\text{TP+FN+FP+TN})} \times 100$	(10)
Specificity = $\frac{\text{TN}}{(\text{TN+FP})} \times 100$	(11)
Sensitivity $=\frac{TP}{(TP+FN)}x100$	(12)
Kappa statistics=Sensitivity + Specificity - 1	(13)

The confusion matrix or error matrix view for SVM Classifier while classifying raster TIFF data setis given in Table 4.



Prediction	Reference		
ricultul	Excess	Normal	Deficient
Excess	14	0	0
Normal	0	16	0
Deficient	0	0	11

Performance Measures using evaluation metrics are specified in Table 5 which are calculated using equations (10), (11), (12) and (13).

Table 5 Performance measures for CSV and restor detector

Table.5 Feriorinanc	e measures for CS	v and faster datasets
Data set type	Accuracy	Kappa Statistics
Raster TIFF	100	100
datasets		

VII. CONCLUSION

In this paper SVM classification method is used to build a classification model for a TIFF dataset. The dataset used herein is of Andhra Pradesh rainfall map. The map comprises of Rain fall coverage for various districts. The methodology used classifies the map based on Rain fall coverage. The performance of SVM is calculated using kappa statistics and accuracy parameters and it is established that for the given data setSVM classifies the raster image dataset with great accuracy. The SVM classification methodology discussed herein can help in environment monitoring, land use, mineral resource identification, classification of remote sensed data into roads and land etc.. in the future.

REFERENCES

- Yubin Yang, Hui Lin, Zhongyang Guo, Jixi Jiang, A data mining approach for heavy rainfall forecasting based on satellite image sequence analysis, Computers & Geosciences, Volume 33, Issue 1, January 2007, Pages 20-30, ISSN 0098-3004, <u>http://dx.doi.org/10.1016/j.cageo.2006.05.010</u>.
- [2] V.M. Mantas, Z. Liu, A.J.S.C. Pereira, A web service and android application for the distribution of rainfall estimates and Earth observation data, Computers & Geosciences, Volume 77, April 2015, Pages 66-76, ISSN 0098-3004, http://dx.doi.org/10.1016/j.cageo.2015.01.011.
- [3] Borja Calvo, Pedro Larrañaga, José A. Lozano, Learning Bayesian classifiers from positive and unlabeled examples, Pattern Recognition Letters, Volume 28, Issue 16, 1 December 2007, Pages 2375-2384, ISSN 0167-8655, http://dx.doi.org/10.1016/j.patrec.2007.08.003.
- [4] kumar, G. Sahoo, Analysisof Parametric & Non Parametric Classifiers for Classification Technique using WEKA, I.J. Information Technology and Computer Science, 2012, 7, 43-49 Published Online July 2012 in MECS DOI:10.5815/ijitcs.2012.07.06
- [5] Ismail Hmeidi, Bilal Hawashin, Eyas El-Qawasmeh, Performance of KNN and SVM classifiers on full word Arabic articles, Advanced Engineering Informatics, Volume 22, Issue 1, January 2008, Pages 106-111, ISSN 1474-0346, 10.1016/j.aei.2007.12.00
- [6] Jing Hu, Daoliang Li, QinglingDuan, Yueqi Han, Guifen Chen, Xiuli Si, Fish species classification by color, texture and multi-class support vector machine using computer vision, Computers and Electronics in Agriculture, Volume 88, October 2012, Pages 133-140, ISSN 0168-1699, 10.1016/j.compag.2012.07.00
- [7] Xiaowei Yang, Qiaozhen Yu, Lifang He, Tengjiao Guo, The one-against-all partition based binary tree support vector machine algorithms for multi-class classification, Neuro computing, Volume 113, 3 August 2013, Pages 1-7, ISSN 0925-2312, http://dx.doi.org/10.1016j.neucom.2012.12.048.
- [8] 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, 10.1016/j.isprsjprs.2012.04.001
- [9] Rajasekhar, N.; Babu, S.J.; Rajinikanth, T.V., "Magnetic resonance brain images classification using linear kernel based Support Vector Machine, "Engineering (NUiCONE), 2012 Nirma University International Conference on , vol., no pp.1,5, 6-8 Dec 2012 doi10.1109/NUICONE.2012.6493213.
- [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, <u>http://dx.doi.org/10.1016/j.proenv.2011.09.386</u>
- [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, <u>http://dx.doi.org/10.1016/j.renene.2013.06.025</u>
- [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] Hsun-Jung Cho, Ming-TeTseng, A support vector machine approach to CMOS-based radar signal processing for vehicle classification and speed estimation, Mathematical and Computer Modeling, Volume 58, Issues 1–2, July 2013, Pages 438-448, ISSN0895-7177, http://dx.doi.org/10.1016/j.mcm.2012.11.00



- [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, Volume40, Issue5, May2010, Pages 519-524, ISSN0010-4825, http://dx.doi.org/ 10.1016/j.compbiomed.2010.03.006.
- [17] Xinjun Peng, Yifei Wang, Dong Xu, Structural twin parametric-margin support vector machine for binary classification, Knowledge-Based Systems, Volume 49, September2013, Pages63-72, ISSN0950-7051, http://dx.doi.org/10.1016/j.knosys.2013.04.013
- [18] Maysam Abedi, Gholam-Hossain Norouzi, Abbas Bahroudi, Support vector machine for multi-classification of mineral prospectivity areas, Computers & Geosciences, Volume 46, September 2012, Pages 272-283, ISSN 0098-004, http://dx.doi.org/10.1016/j.cageo.2011.12.014
- [19] Liu ,Mingjun Wang, Jun Wang, Duo Li, Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar, Sensors and Actuators B: Chemical, Volume 177, February 2013, Pages 970-980, ISSN0925-4005, http://dx.doi.org/10.1016/j.snb.2012.11.071.
- [20] classification using Support Vector Machines, ISPRS Journal of Photogrammetry and Remote Sensing, Volume85, November 2013, Pages102-119, ISSN 0924-2716, http://dx.doi.org/10.1016/j.isprsjprs.2013.08.007.
- [21] Steve R. Gunn, Support Vector Machines for Classification and Regression, A Technical Report Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, University Of South Hampton, May 199
- [22] Till Rumpf, Christoph Römer, Martin Weis, Markus Sökefeld, Roland Gerhards, Lutz Plümer, Sequential support vector machine classification for small-grain weed species discrimination with special regard to Cirsiumarvense and Galiumaparine, Computers and Electronics in Agriculture, Volume 80, January 2012, Pages89-96, ISSN0168-1699, http://dx.doi.org/10.1016 /j.compag.2011.10.018.
- [23] Nemmour, Youcef Chibani, Multiple support vector machines for land cover change detection: An application for mapping urban extensions, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 61, Issue 2, November 2006, Pages 125-133, ISSN0924-2716, http://dx.doi.org/10.1016/j.isprsjprs. 2006.09.004.
- [24] Sujit Narwade, Mohit Kalra, Rajkumar Jagdish, Divya Varier, Sagar Satpute, Gautam Talukdar (2011), Collaborative work of Environmental Information System (ENVIS) Centre and Important Bird Areas Programmes-Indian Bird Conservation Network (IBA-IBCN) projects of the BNHS, India, Zookeys, Narwade, S., et al (2011): Literature based species occurrence data of birds of North-east India
- [25] www.maptell.com/
- [26] 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/10.1080/014311606007464
- [27] [27] 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 200
- [28] R Core Team},R:A Language and Environment for Statistical Computing, R Foundation for Statistical Computing,Vienna,Austria,2013,http://www.R-project.org
- [29] www.orfeo-toolbox.org/otb/monteverdi.html
- [30] 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.
- [31] David J. Rogers, Jonathan E. Suk, Jan C. Semenza, Using global maps to predict the risk of dengue in Europe, Acta Tropica, Volume 129, January 2014, Pages 1-14, ISSN 0001-706X,
- [32] 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, Jissue 7, July 2010, Pages 826-836, ISSN 1364-8152, http://dx.doi.org/10.1016/j.envsoft.2010.01.004.





K.Venkata Nagendra, working as Assistant Professor in the Department of Computer Science Engineering at Geetanjali Institute of Science and Technology, Nellore, Andhrapradesh, India. He has 6 years of experience in the field of teaching. He is a research scholar in Vikrama Simhapuri University. He did his M.Tech in ANU, Guntur. His areas of interests are Data warehousing and Data Mining and Cloud Computing.



Dr.Maligela Ussnaiah, working as Assistant Professor in the Department of Computer Science in Vikrama Simhapuri University, Nellore, Andhra Pradesh, India. He is having 8 years of teaching experience. He did his PhD in computer science from SriKrishna Devaraya University, Ananthapur, Andhra Pradesh. His areas of interests are Networks, Mobile Wireless Networks, Data warehousing and Data Mining and Image processing.











45.98



IMPACT FACTOR: 7.129







INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 🕓 (24*7 Support on Whatsapp)