

DISCRIMINATION BETWEEN HEALTHY AND DAMAGED SESAME LEAVES USING HYPERSPECTRAL DATA

Anjana N. Ghule

Assistant Professor, Information Technology Department,
Government Engineering College Aurangabad
anghule@geca.ac.in

Prof. Dr. R. R. Deshmukh

Professor, Computer Science and Information Technology Department,
Dr. Babasaheb Ambedkar Marathwada University Aurangabad
rrdeshmukh.csit@bamu.ac.in

Chitra M. Gaikwad

Assistant Professor, Information Technology Department,
Government Engineering College Aurangabad
cmgaikwad@geca.ac.in

Abstract - Non-imagery field spectroscopic data is very high dimensional and thereby provides appropriate information about the object. Conversely, the high dimension of field spectroscopic data causes a very strong correlation between adjacent wavelengths and heavy redundancy in the information. This experimentation was performed to study the spectral behavior of healthy and unhealthy Sesame leaves on the electromagnetic spectrum (EMS). Three types of sesame leaves viz. healthy(H), damaged level1(L1), and damaged level2(L2) were selected and Non-imagery hyperspectral measurements of healthy and infected leaves were collected in the Visible and Near-infrared region (Vis-NIR), in the spectral range of 400nm-1050nm. The VNIR region was divided into three sub-regions i) Visible(Vis):400nm-680nm ii) Red Edge(RE):681nm-781nm and iii) Near-Infrared(NIR):782nm-1050nm. ReliefF supervised attribute selection with ranker method was used as filters for dimensionality reduction, further Random Forest and J48(C 4.5) machine learning classifiers were applied on all the wavelengths in the regions as well as the first 20 wavelengths. The maximum classification accuracy was obtained in the visible region and it is revealed that distinct wavelengths affecting the health of the crop were found in the Visible region than RE and NIR regions. The classification accuracy of the Random Forest classifier was found better than the J48 classifier.

Keywords: Hyperspectral leaf reflectance; wavelength selection; ReliefF; J48 ; Random Forest.

1. Introduction

Til is a common name of Sesame seed in India. These nutrient-rich oil seeds are used in various Indian cuisines and also used in cosmetics, pharmaceutical products, and lubricants. India shares approximately 22% of the world's sesame production. Not much attention was paid to this crop so far by the researchers. This experimentation was aimed to study the spectral behavior of healthy and damaged white sesame leaves.

Crop and vegetation health monitoring, yield prediction, crop type classification, identification, and crop species discrimination are some of the applications of remote Sensing. Crops have different biophysical and biochemical characteristics like chlorophyll a and b, total chlorophyll, nitrogen content, a carotenoid pigment, anthocyanin, plant stress, plant moisture, etc. [1]. Reflectance at specific wavelengths in Visible, Red Edge, NIR and SWIR regions shows the change in the spectral pattern if there is any change in the biophysical or biochemical property of the crop[2]. Hyperspectral non-imagery spectroscopic data provides spectral measurements over a large number of wavelengths. Analytical Spectral Devices (ASD) spectroradiometer records spectral reflectance over the range of 350nm -2500nm. This covers the Visible Region, Near Infrared-NIR Region and Shortwave Infrared-SWIR Region of the electromagnetic EMS spectrum which helps in the study and analysis of the crops. High dimensions of the hyperspectral(HS) datasets are used for detailed study of crop characteristics and their spectral behavior but also suffer from multicollinearity and high correlation along many adjacent wavelengths. Therefore, the selection of the most significant wavelengths is one of the crucial

parts which further helps in finding the impact of the disease on the EMS and improves classification tasks.[4][5][6]

Diago M.P.et.al. experimented to discriminate four grape wine varieties by taking leaf reflectance measurements, which were recorded using a camera of 1040 wavelengths from HS images. Partial Least Square (PLS) based classifier obtained 92% classification accuracy. Anhelina Zapolska et.al. carried out research to discriminate between healthy and diseased *Olea europaea* L.. Linear Discriminant Analysis (LDA), PLS Regression and Principal Component Regression (PCR) methods were used to find significant and optimal wavelengths [7][8]. Shafri, Helmi Z. M., et al used Analysis of Variance(ANOVA) for finding the set of optimal spectral bands from spectral reflectance(SR) and the first derivatives(FD). This subset of wavebands was further used for the classification of healthy and unhealthy oil palms, and also for the segregation of Papyrus plants from other species[9][10]. Lei Wang et.al. also used ANOVA for selecting 37 most significant wavelengths out of 288 wavelengths, from SR of *Lycium barbarum* medicinal plant to find its origin[11].

Kumar et al. collected in-situ tea plantation HS data to analyze the spectral behavior for different treatments and its effect on the yield. PCA and SDA methods were used for finding optimum and significant wavelengths [12]. SDA was also used by Manjunath et al. for selecting optimum bands from field HS data to discriminate, cole crops, ornamental plants and pulses [13].

This research paper focuses on discriminating between healthy and unhealthy crops using a feature selection algorithm. Healthy leaves and leaves infected at two damage levels are used for the experimentation. ReliefF supervised attribute selection method along with Ranker algorithm was used for ranking and feature wavelength selection. Feature selection was applied for three regions independently and for the complete VNIR region as well. After feature selection, two classification techniques J 48 and Random Forest were applied with 10-fold cross-validation.

2. Material and Methods

2.1. Sample collection and spectral measurements

Field leaf samples were collected from Krishi Vidnyan Kendra, Aurangabad (19.851128, 75.298525) Maharashtra, India on 7th September 2019. Healthy and infected plants of Sesame were identified by visual judgment. Leaves were categorized into three-level 1) Healthy(H) 2) Damaged Level1 (L1) and Damaged Level2 (L2) clipped and kept in ziplock polythene bags. Leaf spectral reflectance was measured in a lab using ASD FieldSpecPro spectroradiometer and RS3 software [14][15].

Raw .asd spectral measurements taken were converted to data files by ViewSpec Pro 6.2, software. Spectral Reflectance (SR) were further divided into three regions, Visible Region (350nm-670nm), Red Edge Region(671nm-780nm) and Near Infrared(NIR) region(781nm -1000nm) and three separate datasets were created. Figure 1 shows Spectral Reflectance-SR curves of Sesame leaves in the range of 350nm-2500nm.

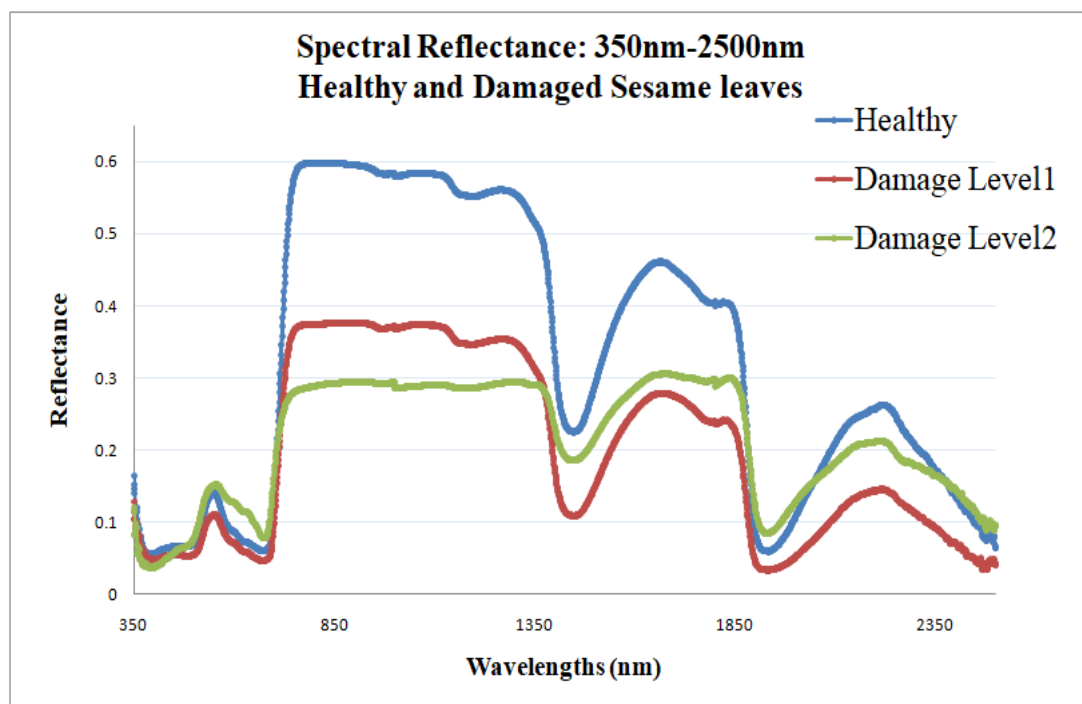


Figure 1: Spectral Reflectance of healthy and damaged Sesame leaves in the range of 350nm-2500nm

2.2. Spectral feature selection and classification

The spectral reflectance was divided into three regions, Visible Region (350nm -670nm), Red Edge Region(671nm-780nm) and Near Infrared(NIR) region(781nm -1000nm) and three separate datasets were created. Figure 1 shows Spectral Reflectance-SR curves of Sesame leaves.

ReliefF is a supervised attribute selection method, which is a sixth variation(A-F) from the family of Relief algorithms. ReliefF evaluates significant features by sampling an instance repeatedly. It uses Manhattan's Distance measure for calculating the distance between the neighbors.

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \quad (1)$$

Distance $d(i, j)$ satisfies :

- (1) $d(i, j) \geq 0$: distance is Non-negativity
- (2) $d(i, i) = 0$: the distance of an object to itself is 0

Weight of attribute A, $W[A]$ is calculated by probabilities: $W[A] = P(\text{different value of A} | \text{nearest instance from different class}) - P(\text{different value of A} | \text{nearest instance of the same class})$

k nearest from the same class are counted as hits and k nearest from the other classes are counted as misses scoring is updated for the target instance. The ranker Algorithm is used for ranking the significant wavelengths.

J48 is a decision tree classifier. In this method, a binary tree is constructed to model the classification process. Following are the steps to construct the tree

- (1) Check whether all instances belong to the same class, then the tree represents a leaf and the leaf is labeled with that class.
- (2) For each attribute, information gain is calculated.
- (3) Depending upon the gain, the best splitting attribute is selected [16][17]

Random Forest classifier creates a set of multiple decision trees. It randomly selects samples and a tree is generated for each sample, this avoids overfitting problems. Each tree participates in voting for selecting the best classification. Prediction result with the majority of votes is selected [18][19].

Classification of three types of leaves was performed for all the wavelengths of the sub-regions viz. Vis-region, Red Edge Region, Near Infrared region for 20 ranked wavelengths and all the wavelengths separately, subsequently complete 652 ranked wavelengths and 20 ranked wavelengths were also used for classification from Vis-NIR(400nm-1050nm) region.

3. Results and Analysis

Feature wavelength selection, ranking and classification were performed in WEKA 3.6.9, an open-source and free software for machine learning. The high dimensionality of the dataset, in comparison with the size of the dataset sometimes leads to overfitting. For selecting and ranking significant wavelengths ReliefF supervised attribute selection filter with ranker algorithm was used.

The classification was performed using J48(C4.5) and Random Forest tree-based Classifiers. To find a minimum number of wavelengths for classification, classification was carried out on 20 ranked wavelengths and results were compared with all ranked wavelengths in the range of Vis(400nm-680nm), Red-Edge(681nm-781nm), NIR(782-1050) and Vis-NIR(400nm-1050nm) separately.

After the feature wavelength selection by ReliefF and rankers method, the accuracy results of the two classifiers are shown in the Table1: 1) The maximum accuracy of 93.75 % was attained in the Visible Region (400nm-681nm) by all ranked wavelengths using J48 and the same accuracy with 20 ranked wavelengths by Random Forest. When the Visible region of the spectrum with 281 wavelengths was used by both classifiers results are almost the same. 2) Using the spectrum of total 101 wavelengths from the RE region from 681nm-781nm, the accuracy of the J48 classifier was 91.67% and 89.58 % by Random Forest classifier for all the wavelengths and first 20 wavelengths. 3) Conversely the accuracy of the J48 classifier was 89.58% and 91.67 % by RF classifier for all the wavelengths and first 20 wavelengths in the NIR region (781nm-1000nm). 4) For the complete Vis-NIR region 88.46% accuracy was obtained by the J48 classifier when reflectance values for all 651 wavelengths were used, but the accuracy of 90.38% was observed when the first 20 ranked wavelengths were used for classification. The same accuracy of 90.38% was calculated by Random Forest classifier for all 651 wavelengths 82.69% for 20 wavelengths.

Table1: Comparison of classification accuracy

		Vis-281 (400nm-680nm)		Red Edge-101 (681nm-781nm)		Near Infrared-268 (782nm-1050nm)		Vis-Near Infrared-651 (782nm-1050nm)	
		C4.5	Random Forest	C4.5	Random Forest	C4.5	Random Forest	C4.5	Random Forest
Accuracy	20	91.67	93.75	91.67	89.58	89.58	91.67	90.38	82.69
	All	93.75	89.58	91.67	89.58	89.58	91.67	88.46	90.38

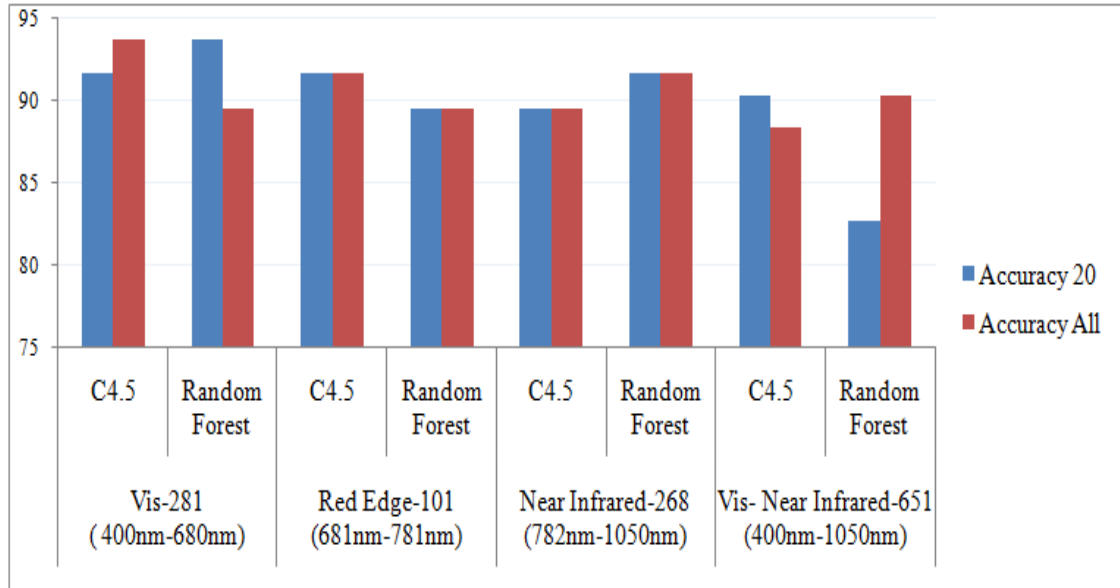


Figure2: Classification Accuracy graph

Table 1 displays classification accuracy obtained using J48 and Random Forest classifiers for the first 20 ranked wavelengths SR and all 651 wavelengths SR. It is also observed that both J48 and Random Forest classifiers demonstrated the same results in the Red Edge region and NIR region for 20 ranked wavelengths and all wavelengths. Figure 2 depicts the result of classification accuracy.

Table 2 shows the classification results with the weighted average of True Positive(TP), False Positive(FP) and F-score along with Kappa values. Using ReliefF attribute selection, the highest TP rate of .938 was observed in the Vis-region with 20 wavelengths by Random Forest and the same TP rate for all wavelengths using all wavelengths by J48.

Table2: Comparison of Classification measures

		Vis-272 (400nm-671nm)		Red Edge-110 (672nm-781nm)		Near Infrared-268 (782nm-1050nm)		Vis-Near Infrared-651 (400nm-1050nm)	
		C4.5	Random Forest	C4.5	Random Forest	C4.5	Random Forest	C4.5	Random Forest
TP Rate	20	0.917	0.938	0.917	0.896	0.896	0.917	0.904	0.827
	All	0.938	0.896	0.917	0.896	0.896	0.917	0.885	0.904
FP Rate	20	0.044	0.031	0.046	0.053	0.056	0.044	0.051	0.087
	All	0.035	0.053	0.046	0.053	0.056	0.044	0.061	0.049
F-score	20	0.917	0.937	0.917	0.896	0.896	0.917	0.904	0.827
	All	0.938	0.896	0.917	0.896	0.896	0.917	0.885	0.903
Kappa	20	0.8742	0.9059	0.8739	0.8431	0.8421	0.8743	0.8551	0.74
	All	0.9055	0.8433	0.8739	0.8431	0.8421	0.8743	0.8261	0.8552

In all three regions, the Visible region was dominantly affected. It is even observed that classification accuracy and other classification measures were almost the same for all wavelengths and 20 wavelengths.

4. Conclusions

Non-imagery spectral measurements obtained by ASD fieldSpec4 provide detailed information about the crop characteristics over EMS. The above experimentation was performed to find the affected region and minimum subset of wavelengths from Spectral Reflectance, which can be used for discrimination between healthy and unhealthy vegetable leaves at two damage levels. To study the impact of the disease, separating the SR into different regions EMS was found more functional. After attribute selection and ranking, classification results were similar for the first twenty wavelengths and all wavelengths, so it can be concluded that to reduce the classification complexity first 20 SR at the first 20 wavelengths can be used. Performance of tree-based classifiers J48 and Random Forest were also similar, but Random Forest performed better than J48.

Further experimentation can be done on more levels of damage and impact on other sub-regions EMS like Short-Wave-IR can also be observed. First Derivative Transform and Continuum Removal data may be used for more comprehensible results.

5. Acknowledgement

Department of Science and Technology - Funds for Infrastructure under Science and Technology (DST-FIST) has supported the Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India, under section No. SR/ FST/ ET1340. This research work has been done under the same support. Authors wish to extend sincere gratitude towards the Department and University authorities for providing required resources.

References

- [1] Prasad S. Thenkabail, M. K. (2014). Hyperspectral Hyperion Images and Spectral Libraries of Agricultural Crops, *Hyperspectral Remote Sensing of Vegetation and Agricultural Crops. Photogrammetric Engineering & Remote Sensing*, 80 (8), 694-723.12.
- [2] R. N. Sahoo, S. S. (2015). Hyperspectral remote sensing of agriculture. *CURRENT SCIENCE, Special Section Hyperspectral Remote Sensing*, 108 (5), 848-859.
- [3] Ray, J. P., Use of Hyperspectral Remote Sensing Data for Crop Stress Detection: Ground-Based Studies. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science*, 38 (8), 2010, pp.562-570
- [4] Mozghan Abbasi, J. (2019). Optimal Spectral Wavelengths for Discriminating Orchard Species Using Multivariate Statistical Techniques. *Remote Sensing*, 12 (63), 1-16.
- [5] Prospere K., M. K. (2014). Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. *Remote Sensing*, 6 (9), 8494-8523.
- [6] Mirzaei M., M. S. (2019). Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. *International Journal of Applied Earth Observation and Geoinformation*, 80, 26-37.
- [7] Diago, Maria P., et al., Identification of Grapevine Varieties Using Leaf Spectroscopy and Partial Least Squares. *Computers and Electronics in Agriculture*, vol. 99, 2013, pp. 7-13. Crossref, doi:10.1016/j.compag.2013.08.021.
- [8] Zapolska, Anhelina, et al., Linear Discriminant Analysis of Spectral Measurements for Discrimination between Healthy and Diseased Trees of *Olea Europaea* L. Artificially Infected by *Fomitiporia mediterranea*. *International Journal of Remote Sensing*, vol. 41, no. 14, 2020, pp. 5388-98. Crossref, doi:10.1080/01431161.2020.1731931.
- [9] Shafri, Helmi Z. M., et al., Spectral Discrimination of Healthy and Ganoderma-Infected Oil Palms from Hyperspectral Data. *International Journal of Remote Sensing*, vol. 32, no. 22, 2011, pp. 7111-29. Crossref, doi:10.1080/01431161.2010.519003.
- [10] Adam, Elhadi, et al., Multispectral and Hyperspectral Remote Sensing for Identification and Mapping of Wetland Vegetation: A Review. *Wetlands Ecology and Management*, vol. 18, no. 3, 2009, pp. 281-96. Crossref, doi:10.1007/s11273-009-9169-z.
- [11] Wang, L., Li, J., Qin, H., Xu, J., Zhang, X., & Huang, L. (2019). Selecting Near-infrared Hyperspectral Wavelengths Based on One-way ANOVA to Identify the Origin of *Lycium Barbarum*. 2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS), 122-125.
- [12] Kumar, Amit, et al. "Field Hyperspectral Data Analysis for Discriminating Spectral Behavior of Tea Plantations under Various Management Practices." *International Journal of Applied Earth Observation and Geoinformation*, vol. 23, 2013, pp. 352-59. Crossref, doi:10.1016/j.jag.2012.10.006.
- [13] Manjunath, K. R., et al., Discrimination of Spectrally-Close Crops Using Ground-Based Hyperspectral Data. *Journal of the Indian Society of Remote Sensing*, vol. 39, no. 4, 2011, pp. 599-602. Crossref, doi:10.1007/s12524-011-0099-x.
- [14] <http://www.asdi.com/applications/remote-sensing/spectral-remote-sensing>. (n.d.). Retrieved January 21, 2016, from <http://www.asdi.com/applications/remote-sensing/spectral-remote-sensing>.
- [15] Ghule A., Deshmukh R.R. (2020) Improved Linear Extrapolation Technique for Crop Health Monitoring Using Hyperspectral Data. In: Bansal J., Gupta M., Sharma H., Agarwal B. (eds) *Communication and Intelligent Systems. ICCIS 2019. Lecture Notes in Networks and Systems*, vol 120. Springer, Singapore. https://doi.org/10.1007/978-981-15-3325-9_8
- [16] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (The Morgan Kaufmann Series in Data Management Systems) (3rd ed.). Morgan Kaufmann.
- [17] Huang, W., Guan, Q., Luo, J., Zhang, J., Zhao, J., Liang, D., Huang, L., & Zhang, D. (2014). New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2516-2524. <https://doi.org/10.1109/jstars.2013.2294961>
- [18] Quinlan, J. R. (1996). Improved Use of Continuous Attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, 77-90. <https://doi.org/10.1613/jair.279>
- [19] Xuqing Li, Xiangnan Liu *, Zhihong Du, Cuicui Wang, A Random Forest Model for Estimating Canopy Chlorophyll Content in Rice Using Hyperspectral Measurements, 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)

Authors Profile



Mrs. Anjana N. Ghule, M.E. Computer Science and Engineering from Dr. B. A. M University Aurangabad (MS) India. Currently working as an Assistant Professor, Department of Information Technology, Government College of Engineering, Aurangabad, (MS), India. Life member CSI, ISTE, ISRS and IETE. EC member, IETE Aurangabad Centre Published research papers in International conferences and reputed journals. Her research interests include Machine Learning and Remote Sensing. She can be reached at anghule@geca.ac.in.



Dr. R. R. Deshmukh, M.E., M.Sc. (CSE) Ph.D. FIETE, PEIN Fellow, Working as Professor, Department of CSIT, Dr. B.A.M. University, Aurangabad, (MS), India. Coordinator of DST-FIST program, University Coordinator, MHRD GIAN program, Chairman, IETE Aurangabad Centre 2014–2018, Organized Zonal ISF-2016, Sectional President, ICT Section ISCA-2019, Coordinator, महाराष्ट्र राज्य मराठी विश्वकोश निर्मिती मंडळ, महाराष्ट्र शासन, Life member ISCA, CSI, ISTE, IEEE, IAEng, CSTA, IDES, ACEE. Management Council, Senate, Academic Council Member at University, Edited Twelve books, Published more than 225 research papers in reputed Journals, Editor-In-Chief, CSI Journal of Computing. Visited Russia, USA, China, Spain, Philippines, Uzbekistan, Thailand for academic work. He can be reached at rrdeshmukh.csit@bamu.ac.in



Mrs. Chitra M. Gaikwad has received a degree in M.E. Computer Science and Engineering from Dr. B. A. M University Aurangabad (MS) India. Currently working as Head and Assistant Professor, Department of Information Technology, Government College of Engineering, Aurangabad, (MS), India. Life member ISTE and IETE. Published research papers in International conferences and reputed journals. Her research interests include Image Processing and Remote Sensing. She can be reached at cmgaikwad@geca.ac.in.