International Journal of Information Processing, 10(3), 24-34, 2016
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

A Comparative Analysis of RF and PSO based Feature Selection Techniques and their Effect on the Plant Leaf Image Classification

A Kumar[a], V Patidar[a], P Saini[a], D Khazanchi[b]

[a]Sir Padampat Singhania University Udaipur, India, Contact:arunkumarsai@gmail.com

[b]University of Nebrasaka, Omaha, USA.

To understand the digital images, it involves understanding the pattern of millions of pixels with respect to their intensity and contrast values, which distinguish them from one digital image from the other. To understand and display these pixel patterns, the concept of feature selection not only reduces the dataset understudy, it helps in understanding the data pattern, but also improves the automatic classification of digital images. In the present study, the plant leaf image texture features have been extracted using Gabor filter and then these feature sets have been subjected to random forest based ensemble technique for feature selection and PSO based feature selection techniques. The classification results for the feature sets prepared using these two techniques have been subjected to Random Forest classification algorithm. This work has utilised both the dorsal and ventral leaf images for discrimination of plants on the basis of digital images. This work has analysed the accuracy results for dorsal and ventral leaf images for plant classification.

Keywords : Dorsal Side, Gabor Filter, Leaf Images, Particle Swarm Optimization, Ventral Side.

1. INTRODUCTION

The plants and animals have lived on this planet earth for centuries and are part of ecological balance of the nature. But due to the rapid development of the human society and also due to the technical advancement and the human need for better roads, bridges and houses, there has been a reckless felling of trees and cutting of vegetation to pave the way for roads and bridges. The development on one end is leading to disappearance of flora and fauna, though essential in maintaining the ecological balance. But, at the same time, the human quest for identifying and scientifically classifying the plants and their sub-species and then devising methods for preserving them for the future before the plant species get extinct, has been going on in scientific world since decades.

The plants have been studied for their flowers, leaves, seeds and fruits. There are millions of different plant species, but many of the sub species are still unknown and would die and become extinct, before their turn comes up to know them. Therefore, there is a need for automatic plant identification and scientific classification methods which could speed up the process of knowing the individual plant species. The biologist and computer scientists have been playing their roles in suggesting newer methods for identifying the plant species. The computer vision methods have revolutionized the work of automatic plant classification and are based on finding suitable characteristic features from the digital images and then suitably classifying them in to various species. As the data collected from the digital images is enormous, there is a need to find subset of the data, which would do the same work as that by the whole dataset.

To reduce the large dataset to a smaller subset, the role of feature selection algorithms is pivotal and is evolving day by day. By using the feature selection methodology, there is a drastic improvement in the average predictive clas-

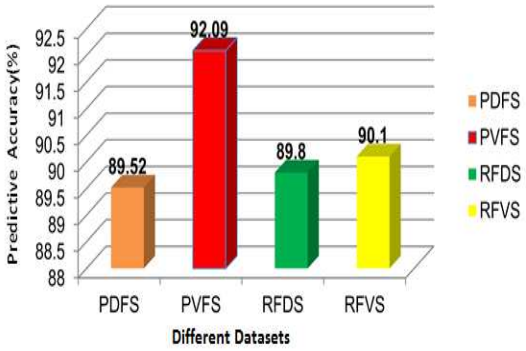

Figure 13. Predictive Classification Accuracy Results for RF and PSO based Feature Subsets

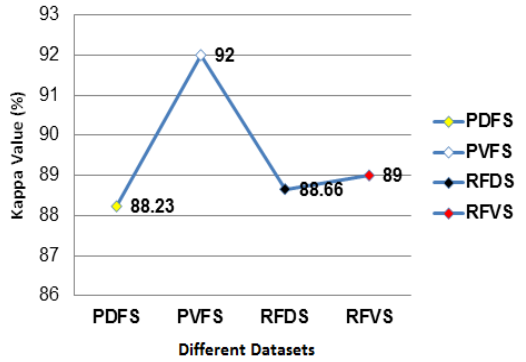

Figure 14. Kappa Accuracy Results for RF and PSO based Feature Subsets

leaves of 30 different plant species to discriminate the different leaf images. This work has extracted shape features like eccentricity, area, perimeter, major axis and minor axis. These features have been passed through probabilistic neural network (PNN) and has obtained a predictive accuracy value as high as 91.41% which is lower than PVFS (92.09%) but higher than the other three datasets of the present study which has been portrayed through Figure 15.

The researcher [14] has worked with 100 different plant species and has extracted shape features using Fourier descriptors and merged them with other shape features like aspect ratio, roundness factor, irregularity, solidity and convexity. The overall predictive accuracy

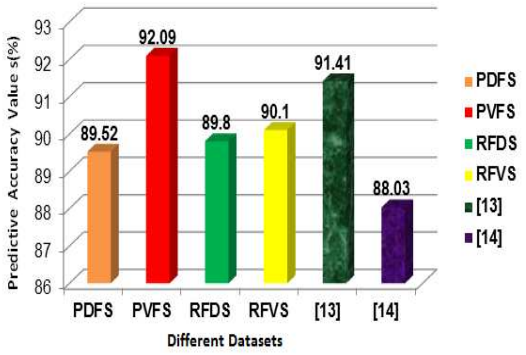

Figure 15. Comparison of the Present Work with the Work of Researchers [13] and [14]

value achieved by this work is 88.03% using Bayes Classifier, which is lower than the results obtained through all the datasets created in the present study and has been portrayed through Figure 15.

6. CONCLUSIONS

By selecting optimized feature subset using RF or PSO-CFS technique, the size of the overall dataset, be it dorsal or ventral has reduced considerably as mentioned in Section 3. On observing the predictive accuracy values obtained for PDFS and PVFS, the PVFS dataset provides better predictive accuracy results as compared to PDFS. On observing the predictive accuracy values obtained for RFDS and RFVS, the RFVS dataset provides better predictive accuracy results as compared to RDFS. Therefore, the objective of this study, to utilize the ventral sides of the leaves has been achieved using both the feature selection techniques. This study shows that the ventral side of the leaf images can be another alternative for the extraction of unique features for leaf image classification and the predictive accuracy results for the ventral side are faring better as compared to the dorsal side results, and this substantiates the proposition of this study.

REFERENCES

1. R M Haralick, K Shanmugam and I Dinstein. Textural Features for Image Classification, *In IEEE Transactions on Systems, Man and Cybernetics*, 6:269-285, 1973.
2. H Tamura, S Mori and T Yamavaki. Textural Features Corresponding to Visual Perception, *In IEEE Transactions on Systems, Man and Cybernetics,* 8:460-472, 1978.
3. D Zhang, A Wong, M Indrawan and G Lu. Content-Based Image Retrieval Using Gabor Texture Features, *In Proceedings of the IEEE Pacific-Rim Conference on Multimedia, University of Sydney, Australia,* pages 91–110, 2000.
4. Gabor Filter. *URL: https://en.wikipedia.org/wiki/Gaborfilter*
5. D Dunn, W Higgins and J Wakeley. Texture Segmentation Using 2-D Gabor Elementary Function, *IEEE Transactions on Pattern Analysis, Machine Intelligence,* 16:130–149, 1994.
6. T Wei. *Corrplot: Visualization of a Correlation Matrix. R package version 0.73, 2013, URL: http://CRAN.R-project.org/package=corrplot.*
7. C Strobl, A L Boulesteix, A Zeileis and T Hothorn. Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution, *BMC Bioinformatics,* pages 8–25, 2007.
8. R Genuer, J M Poggi and C Tuleau-Malot. Variable Selection using Random Forests, *Pattern Recognition Letters,* 31:2225–2236, 2010.
9. M Hall, E Frank, G Holmes, Pfahringer B Reutemann and I H Witten. The WEKA Data Mining Software: An Update, *ACM SIGKDD Explorations,* 11(1):10–18, 2009.
10. R Poli, J Kennedy and T Blackwell. Particle Swarm Optimization: An Overview, *Swarm Intelligence,* 1(1):33–57, 2007.
11. R Development Core Team. R: A Language and Environment for Statistical Computing, *R Foundation for Statistical Computing, Vienna, Austria.* ISBN 3-900051-07-0, 2008.
12. J Sim and C C Wright. The Kappa Statistic in Reliability Studies: Use, Interpretation and Sample Size Requirements, *Physical Therapy,* 85(3):257–268, 2005.
13. J Hossain and M A Amin. Leaf Shape Dentification Based Plant Biometrics, *In Proceedings of* 13^{th} *International Conference on Computer and Information Technology, IEEE Xplore Press, Dhaka, Bangladesh,* DOI: 10.1109/ICCITECHN.2010.5723901, pages 458–463, 2010.
14. A Kadir. Leaf Identification Using Fourier Descriptors and Other Shape Features, *Gate to Computer Vision and Pattern Recognition,* 1(1):3–7, 2015.

Arun Kumar obtained his BE, ME and Ph.D in Image Processing. He is presently working as an Associate Professor in the Department of Computer Science and Engineering and is having Industry and teaching experience of 19 years. He is a member of ISTE, IAENG and EURASIP.

Vinod Patidar is a Professor and Head of the Department of Physics at Sir Padampat Singhania University, Udaipur, India. His research interests include Nonlinear Dynamics & Chaos Theory, application of chaos in Cryptography and Atomic Collision Theory. His research has been published in many international journals of repute with 1600 citations. He is a life member of IPS, IPA, IAPT, ISAMP, ISCA, IAENG and a senior member of IACSIT.

Deepak Khazanchi is a Professor of Information Systems and Quantitative Analysis, Associate Dean for Academic Affairs and Community Engagement and Internationalization Officer in the College of Information Science and Technology at the University of Nebraska at Omaha. His research has been published and presented in national/international peer-reviewed journals and conferences.

Poonam Saini is MCA, M.Phil(C.S.). She is having teaching experience of 14 years and presently working as an Asst. Professor in the Department of Computer Science and Engineering. She is a life member of CSI.