

PAPER • OPEN ACCESS

An Optimal Wavelet Type to Increase Classification Accuracy for Forest Softwood Species

To cite this article: Mohammed Abdulrazaq Kahya *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **571** 012005

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the **collection** - download the first chapter of every title for free.

An Optimal Wavelet Type to Increase Classification Accuracy for Forest Softwood Species

Mohammed Abdulrazaq Kahya, Azzam S.Y. Aladool and Suhaib A. Altamir
College of Education for Pure Sciences, University of Mosul

*¹mohammedkahya@uomosul.edu.iq, ²math.azzam@yahoo.com,
³suhaib.altamir@uomosul.edu.iq

Abstract. Computer vision systems are a very interesting alternative approach for the forest species classification, which is a challenging task that generally requires well-trained human specialists. Hence this work proposed a method concerned with computer vision system to classify 37 different forest softwood species. The images are being decomposed to red, green and blue channels to extract features, each channel is decomposed using a different discrete wavelet transform (DWT) family. The image features are prepared to be classified. Texture images of softwood species are classified using neural network (NN) and support vector machine (SVM) classifiers. Classification accuracy (CA) values for SVM were 94.7297% and 97.2973% for 3rd and 5th level decompositions “coif2” DWT respectively which are better than NN results.

Keywords: Discrete Wavelet Transform; Features Extraction; ordinary wavelet analysis; wavelet packet analysis

1. Introduction

Presently, identification of forest species is a very important process for several reasons, such as the prevention of commercial fraud, the protection of threatened species from extinction, and the promotion of the shipbuilding and construction industry. Therefore, computer vision systems become a very interesting field for the identification.

In the past decades, computer vision applications studies of the wood industry did not take types classification into account [1], [2]. Recently, computer vision has been successfully used to classify forest species. Twenty different Malaysian forest species can be recognized by a system offered by Khalid, Lee [3]. Their experiments depended on database which includes (1,753) images used for training and only (196) for testing, and the classification rate was (95%). Tou, Lau [4] stated two methods to classify forest species using Gray Level Co-occurrence Matrix and neural network. The researcher presents classification rates ranging from (60% to 72%) for five different forest species. de Souza Britto Jr and Sabourin [5] presented a system to recognize 22 different species of the Brazilian flora using GLCM and color-based features. In this paper, database was built of (1270) images of (22) different forest species. It was splitted into 402, 231, and 619 images for training, validation, and testing, respectively. They report a recognition rate in the range of (66.3-80.8) %.

Yusof, Rosli [6] offered a recognition system approach for the species tropical wood using Gabor filter, GLCM with multilayer perceptron back-propagation they report (95.44%) training and (90.33%)



accuracy test. Wang, Qi [7] presented a wood recognition approach using Gabor wavelet coefficients. Their experiment accuracy was 94.58% using the nearest neighbor classifier technique.

Cavalin, Kapp [8] proposed a system to classify 112 Brazilian forest species using multiple feature vectors based on Local binary patterns, GLCM and Linear phase quantization. The recognition rates of the system were around 93.2%.

Yadav, Anand [9] introduced Coiflet discrete wavelet transform family (DWT) for features of microscopic images feature extraction of hardware spaces. They recorded a classification accuracy of about 92.20%. Maruyama, Oliveira [10] presented a method based on local binary patterns with convolutional neural networks for automatic identification depending of wood charcoal species. They obtain a range between 93.9% to 95.7% of recognition rate.

We performed experiments depending on a database which consisted of (2,240) microscopic images which were taken from (112 different Brazilian forest species and divided into two parts hardwood (1500) and softwood (740) (softwood is only considered here) [11]. it is clear that the classification algorithm results are significantly affected by the feature equality input (extracted from the softwood species image). In this paper, an approach is presented for extracting discriminative features of (37) different softwood species of microscopic images using some of the discrete wavelet family. SVM [12]–[14] and NN [14], [15] tested to classify the softwood species into (37) different categories.

2. Overview

2.1. The Discrete Wavelet Transform

The Wavelet Series can be described as a sample version of Continuous Wavelet Transform (CWT) and it is computationally expensive relay for a required resolution. Therefore, The Discrete Wavelet Transform (DWT) based on sub-band coding, is developed for fast computation Wavelet Transform. Hence, less computation time and resources are required [16]. Furthermore, Digital filters and down-samplers $2\downarrow$ can be used to implement two-dimensional DWT. Figure-1 shows the procedures of selecting the one-dimensional FWT of the rows of $f(x, y)$ and the subsequent one-dimensional FWT of the resulting columns. Three sets of detail coefficients are produced including the horizontal, vertical, and diagonal details [17], [18].

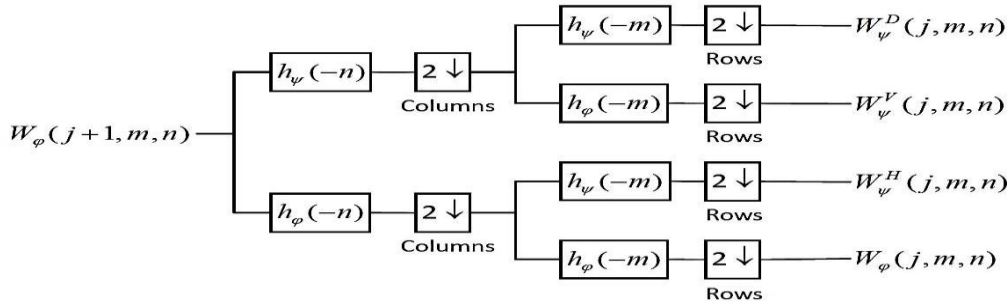


Figure-1: The analysis filter bank of the two-dimensional FWT

The ordinary wavelet analysis can be described as follows: an image can be divided into an approximation and a detail. Then the approximation can also be divided into a second-level approximation and a detail, and then the process is repeated. There are $(n+1)$ possible ways to decompose or encode the image for n -level decomposition (see Figure-2). In wavelet packet analysis the details and the approximations can be divided (see Figure-3). This obtains $(2n)$ different ways to encode the image. This is the wavelet packet decomposition tree [17], [19].

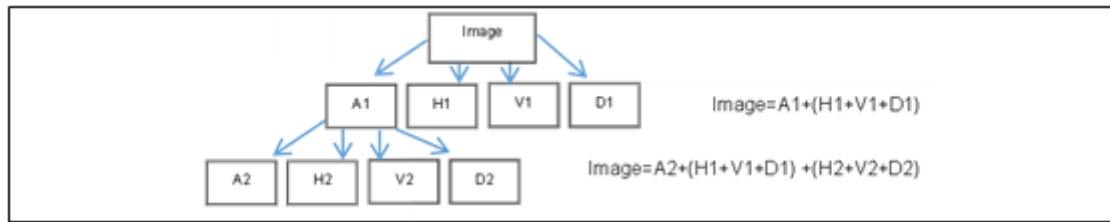


Figure-2: 2nd Level ordinary wavelet analysis

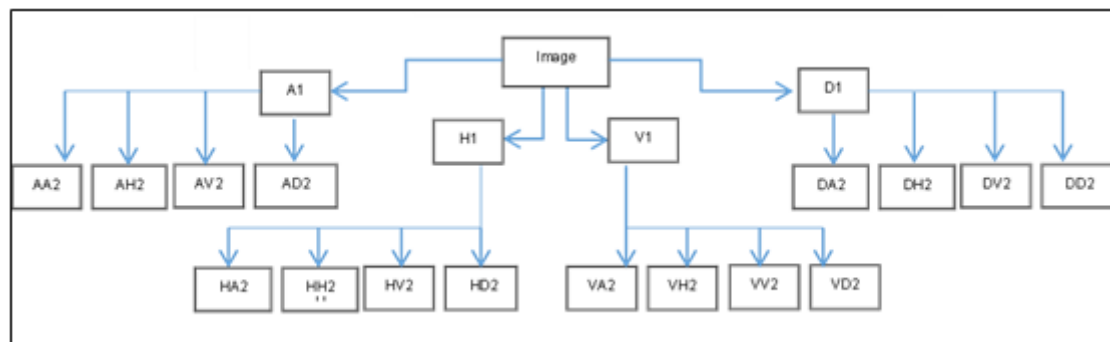


Figure-3: 2nd Level wavelet packet

2.2. Database

In this paper, database contains (37) different forest softwood species which were cataloged by the Laboratory of Wood Anatomy at the Federal University of Parana in Curitiba, Brazil [11]. The resolution of the available images species in PNG (Portable Network Graphics) format is 1024768 pixels. Each species has 20 image samples (see Figure-4).

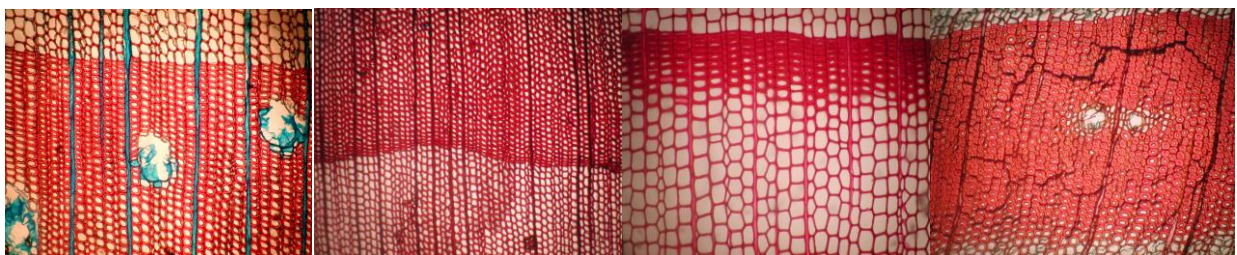


Figure-4: Samples of the database used

3. Method

The diagrams of the presented method for texture classification are as shown in Figure 5 & 6 and Error! Reference source not found.. The following steps describe the proposed method:

Step1: The system applies a wavelet transform of 3rd Level ordinary wavelet analysis in the first experiment (see Figure 5), and 2nd Level wavelet packet analysis on the second experiment (see Figure 6), Error! Reference source not found.. These were applied to the entire textured images and it extracts features from each image detail.

Step2: Feature extraction performed to efficiently represent interesting parts of an image as a feature vector. This approach is more applicable for large size images with reduced feature representation. This is to quickly complete tasks such as image matching and retrieval or classification [20], [21]. Conventional features which was used in this work, which can be described as following:

$$1. \text{ Mean} \quad \mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N P(i, j) \quad (1)$$

$$2. \text{ Standard deviation} \quad \sigma = \text{Sqrt}\left(\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (P(i, j) - \mu)^2\right) \quad (2)$$

$$3. \text{ Skewness} \quad S = \left[\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (P(i, j) - \mu)^3 / \sigma^3 \right] \quad (3)$$

$$4. \text{ Kurtosis} \quad K = \left[\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (P(i, j) - \mu)^4 / \sigma^4 \right] \quad (4)$$

$$5. \text{ Entropy} \quad H = - \sum_{k=1}^L Pr_k (\log_2 Pr_k) \quad (5)$$

$$6. \text{ Energy} \quad E = \sum_{i=1}^M \sum_{j=1}^N P(i, j)^2 \quad (6)$$

$$7. \text{ Root mean square} \quad R = \text{Sqrt}\left(\sum_{i=1}^M \sum_{j=1}^N P(i, j)^2 / MN\right) \quad (7)$$

$$8. \text{ Mean absolute deviation} \quad M1 = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |P(i, j) - \mu| \quad (8)$$

$$9. \text{ Median absolute deviation} \quad M2 = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |P(i, j) - \text{Median}|, \quad (9)$$

Where $P(i, j)$ is the pixel value at point (i, j) of an image of size $(M \times N)$ and Pr_k is the probability of the k -th grey level.

Step3: Standardization of features matrix to standardize data, typically this means fitting data within unity (1), in other words, all data values will be of 0 and 1. The data should be arranged in a form that the value of standard deviation produced from the average is computable. This is followed by the curve of normal distribution showing the each standard deviation probability for a normal distribution. Then to standardize the data, the Z-Score will be calculated. This shows the number of standard deviations from the average which data point falls. The following equation should be used for determining the Z-score of each data point. To determine the Z-Score of each data point, the following equation should be used:

$$F_{\text{standardize}} = \frac{F - \mu_F}{\sigma_F} \quad (10)$$

Where F represents the features matrix

Step4: Stratified using 10-fold cross validation method as it is shown below [14]:

- Dividing data randomly into $(k=10)$ folds (subsets) of equal size.
- Training the model on $(k-1)$ folds using one fold for testing
- Repeating this process k - times so that all folds are used for testing.
- Computing the average performance on the k - test sets.

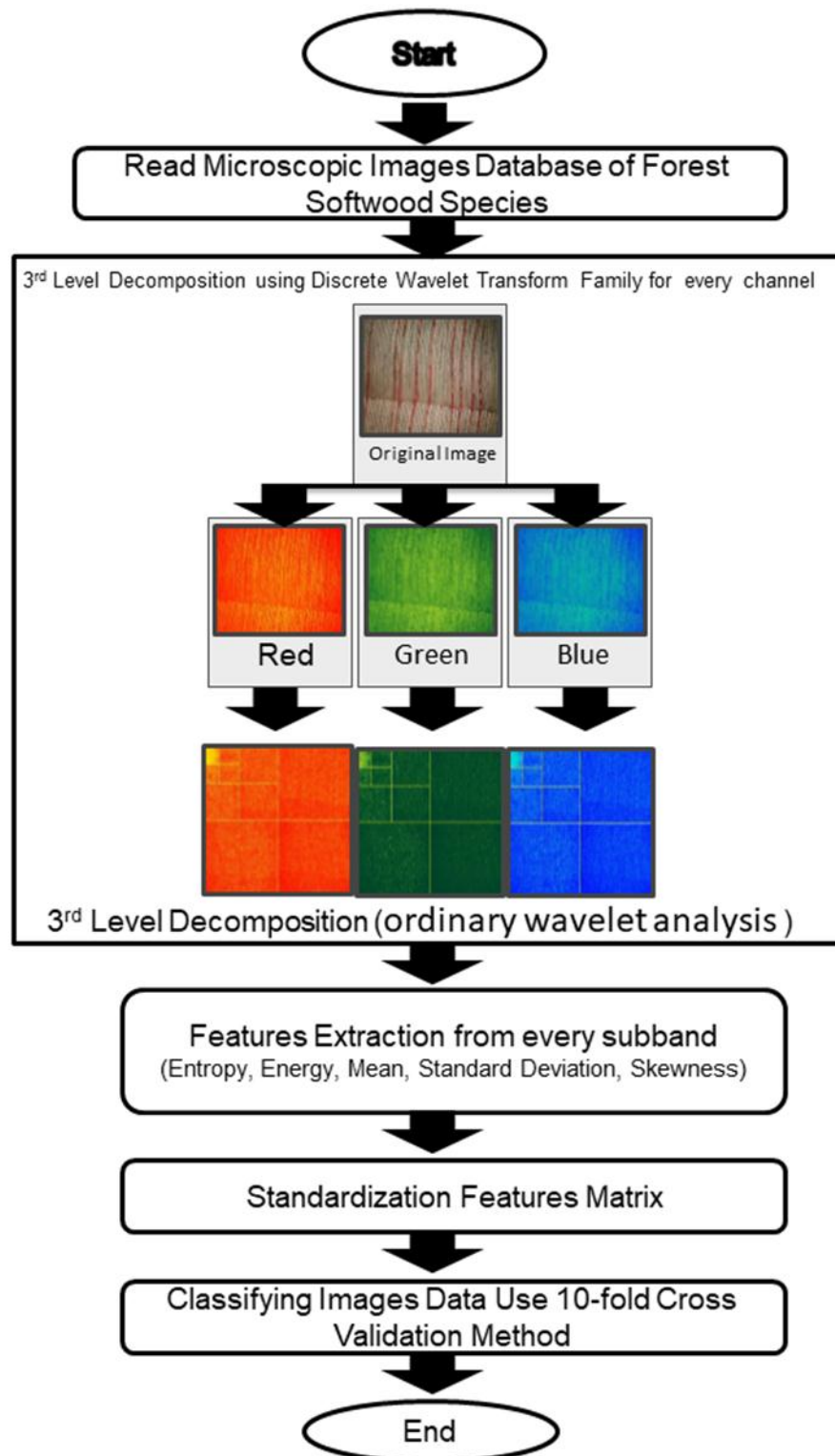


Figure - 5: The diagram of the presented methodology (First experiment)

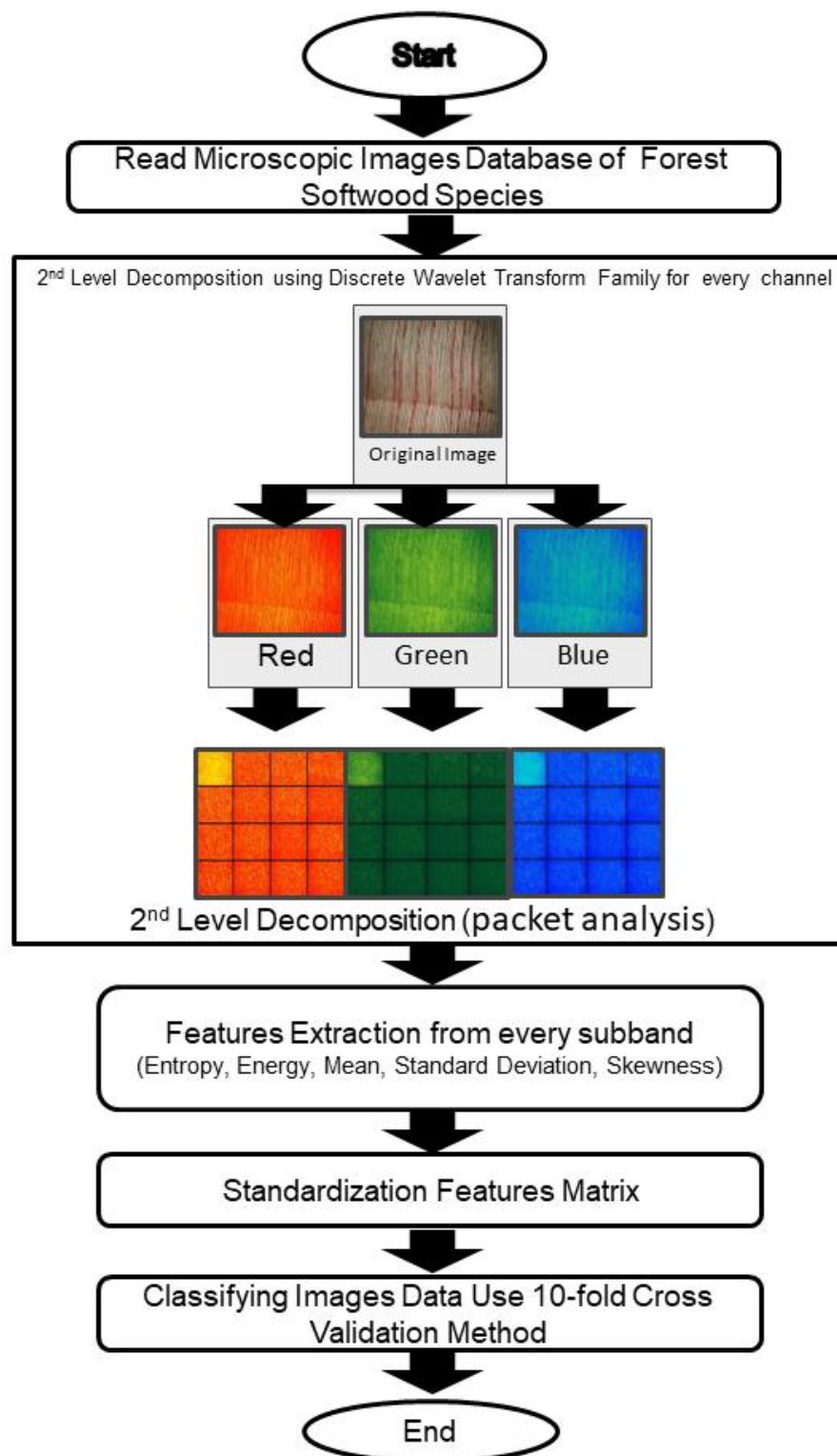


Figure - 6: The diagram of the presented methodology (Second experiment)

4. Results and Discussion

The proposed method performance evaluated by using official classification metrics such as Classification Accuracy, Precision, Recall and MCC [14], [22]. Classification metrics of the SVM and NN listed in Table-1 based on first experiment (use 3rd Level ordinary wavelet analysis, see Figure 5) applied using extracted features of texture images to classify them into (37) kinds of Brazilian Forest Softwood Species.

Table-1: Accuracy of Classification Models Based on First Experiment

Classifier wavelet family	SVM				NN			
	Accuracy %	Precision	Recall	MCC	Accuracy %	Precision	Recall	MCC
haar	92.9730	0.933	0.930	0.928	91.8919	0.921	0.919	0.915
dmey	93.5135	0.938	0.935	0.934	91.7568	0.923	0.918	0.915
db1	92.9730	0.933	0.930	0.928	91.8919	0.921	0.919	0.915
db2	92.9730	0.934	0.930	0.929	93.2432	0.937	0.932	0.930
db3	93.5135	0.938	0.935	0.934	93.3784	0.936	0.934	0.932
bior1.1	92.9730	0.933	0.930	0.928	91.8919	0.921	0.919	0.915
bior1.3	92.9730	0.932	0.930	0.928	91.8919	0.922	0.919	0.916
bior1.5	93.1081	0.934	0.931	0.930	91.6216	0.918	0.916	0.913
rbio1.1	92.9730	0.933	0.930	0.928	91.8919	0.921	0.919	0.915
rbio1.3	93.9189	0.943	0.939	0.938	91.7568	0.923	0.918	0.916
rbio1.5	92.8378	0.932	0.928	0.927	91.7568	0.923	0.918	0.915
Sym2	92.9730	0.934	0.930	0.929	93.2432	0.937	0.932	0.930
Sym3	93.5135	0.938	0.935	0.934	93.3784	0.936	0.934	0.932
Sym4	93.1081	0.935	0.931	0.930	92.7027	0.931	0.927	0.925
Coif1	93.7838	0.941	0.938	0.936	93.2432	0.936	0.932	0.930
Coif2	94.7297	0.950	0.947	0.946	92.8378	0.933	0.928	0.926
Coif3	94.5946	0.950	0.946	0.945	92.4324	0.929	0.924	0.922

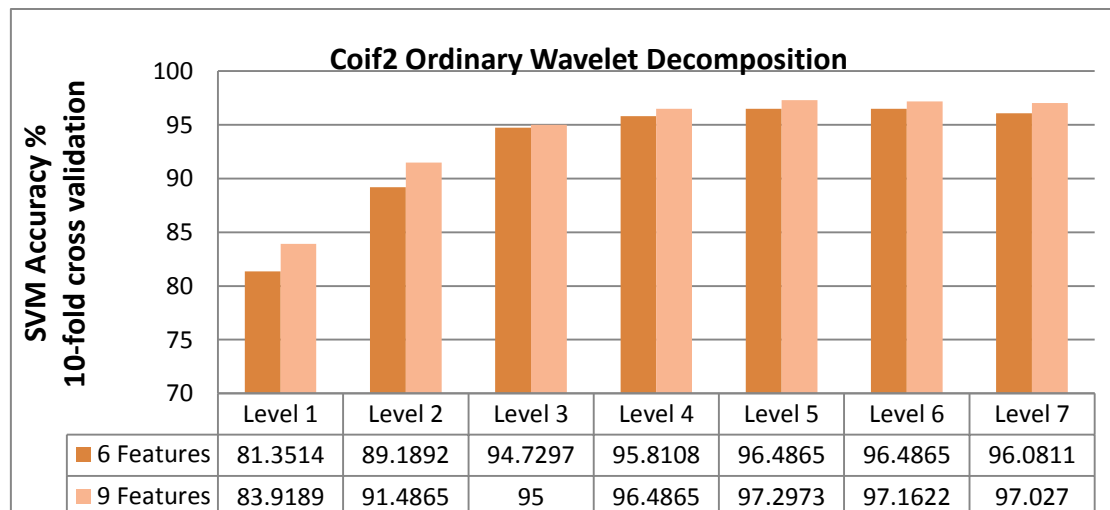
The best wavelet kinds bases can be the chosen depends on the image features as different images have different characteristics. A single mother wavelet produces the wavelet basis function using translating and scaling. Nevertheless, mother wavelet function is neither single nor universal. Therefore, a small set of conditions must be satisfied by the mother wavelet. This set of conditions is selected depends on the image processing problem domain. A small number of convenient computable function with typical signals represented, as the choice of wavelet bases not clear for microscopic image database of softwood. An investigation carried out on softwood species image in order to choose the best wavelet for softwood species images database. The majority of the MATLAB wavelet bases were examined. As it is shown in Table-1, the Coiflet2 wavelet with SVM was chosen for the decomposition of Brazilian forest softwood species Higher levels of (ordinary wavelet analysis) decomposition. This showed promising diagnostic features of the forest softwood species image.

Now the second experiment (use 2nd Level wavelet packet analysis, see Figure-5) shown in Table-2 explained that Haar, Daubechie1, Biorthogonal1.1 and Reverse biorthogonal1.1 wavelets with SVM chosen for the best wavelet packet decomposition of Brazilian forest softwood species.

Table-2: Accuracy of Classification Models Based on Second Experiment

Classifier wavelet family	SVM				NN			
	Accuracy %	Precision	Recall	MCC	Accuracy %	Precision	Recall	MCC
haar	92.027	0.923	0.920	0.919	91.3514	0.919	0.914	0.911
dmey	87.7027	0.882	0.877	0.874	89.5946	0.901	0.896	0.892
db1	92.027	0.923	0.920	0.919	91.3514	0.919	0.914	0.911
db2	89.7297	0.898	0.897	0.894	90.4054	0.910	0.904	0.901
db3	89.7297	0.902	0.897	0.895	90.6757	0.915	0.904	0.905
bior1.1	92.027	0.923	0.920	0.919	91.3514	0.919	0.914	0.911
bior1.3	91.6216	0.920	0.916	0.915	90.8108	0.913	0.908	0.905
bior1.5	91.6216	0.919	0.916	0.914	91.3514	0.919	0.914	0.911
rbio1.1	92.027	0.923	0.920	0.919	91.3514	0.919	0.914	0.911
rbio1.3	87.027	0.874	0.870	0.867	88.6486	0.891	0.886	0.883
rbio1.5	87.8378	0.882	0.878	0.876	88.9189	0.893	0.889	0.885
Sym2	89.7297	0.898	0.897	0.894	90.4054	0.910	0.904	0.901
Sym3	89.7297	0.902	0.897	0.895	90.6757	0.915	0.907	0.905
Sym4	88.7838	0.892	0.888	0.885	89.5946	0.901	0.896	0.892
Coif1	89.5946	0.900	0.896	0.894	90.0000	0.907	0.900	0.896
Coif2	87.973	0.884	0.880	0.877	89.3243	0.900	0.893	0.889
Coif3	88.2432	0.885	0.882	0.879	90.0000	0.903	0.900	0.896

Finally, another type of comparisons between levels of wavelet decomposition effective to classification accuracy is presented. It can be seen from the comparison shown in Figure- that the 5th level is the best level of ordinary decomposition, this based on first experiment (shown in Figure) if nine features Equation (1) to (9) is used.

**Figure-7: Comparison classification accuracy based on levels of ordinary wavelet decomposition**

On the other hand, the graphic in Figure- shown below explained that the 3rd level is the best level of wavelet packet decomposition based on second experiment (shown in Figure-6) if six features Equation (1) to (6) or nine features Equation (1) to (9) are used.

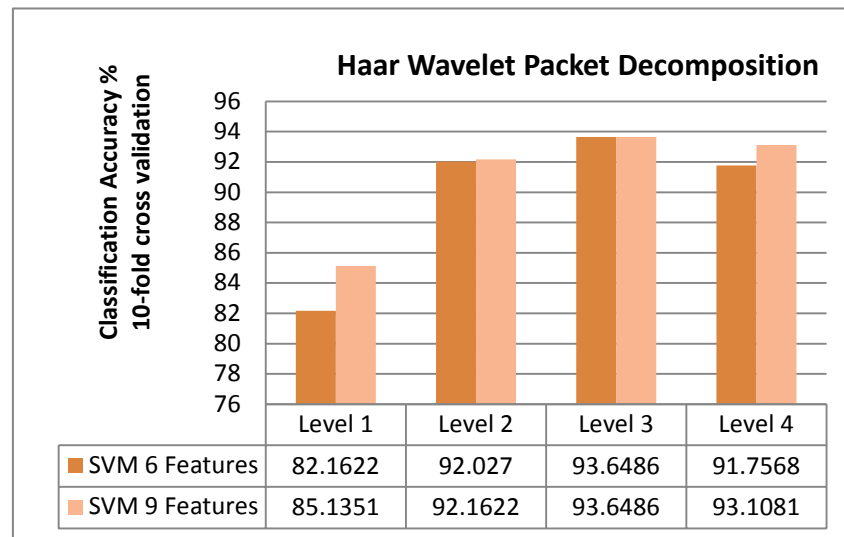


Figure-8: Comparison classification accuracy based on levels of wavelet packet decomposition

5. Conclusion

Automatic classification of Brazilian forest softwood species based on microscopic images of 37 different classes is presented. Two experiments for classification based on ordinary wavelet analysis and wavelet packet analysis are introduced. It is also noted that ordinary wavelet analysis is the best and fastest images wavelet analysis for the classification purpose.

The classification processes with the use of SVM and NN is presented by using the stratified (10-fold) cross-validation. The highest accuracy of classification of (94.7297 %) is achieved by SVM classification algorithms for the features extracted by “coif2” 3rd level DWT (ordinary wavelet analysis). Furthermore, in this work the pre-processed images are evaluated at many levels of image decomposition and show that 5th level with a classification accuracy of 97.2973 % is the best level of ordinary wavelet decomposition. The results highlighted that the proposed method of texture feature extraction based on DWT family by determining mother wavelet function (5th level of ordinary wavelet by choose coiflet2 with SVM), is a practical approach for the forest softwood species classification.

6. References

- [1] L. Thomas and L. Mili, “A robust GM-estimator for the automated detection of external defects on barked hardwood logs and stems,” *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3568–3576, 2007.
- [2] D. N. Buechler and D. K. Misra, “Subsurface detection of small voids in low-loss solids,” in *Sensors for Industry, 2001. Proceedings of the First ISA/IEEE Conference*, pp. 281–284.
- [3] M. Khalid, E. L. Y. Lee, R. Yusof, and M. Nadaraj, “Design of an intelligent wood species recognition system,” *Int. J. Simul. Syst. Sci. Technol.*, vol. 9, no. 3, pp. 9–19, 2008.
- [4] J. Y. Tou, P. Y. Lau, and Y. H. Tay, “Computer vision-based wood recognition system,” in *Proceedings of International workshop on advanced image technology*.
- [5] P. L. Paula Filho, L. S. Oliveira, A. de S. Britto, and R. Sabourin, “Forest species recognition using color-based features,” in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 4178–4181.
- [6] R. Yusof, N. R. Rosli, and M. Khalid, “Using gabor filters as image multiplier for tropical wood species recognition system,” in *Computer Modelling and Simulation (UKSim), 2010 12th International Conference on*, pp. 289–294.

- [7] H. Wang, H. Qi, and X.-F. Wang, "A new wood recognition method based on gabor entropy," in *International Conference on Intelligent Computing*, pp. 435–440.
- [8] P. R. Cavalin, M. N. Kapp, J. Martins, and L. E. S. Oliveira, "A multiple feature vector framework for forest species recognition," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pp. 16–20.
- [9] A. R. Yadav, R. S. Anand, M. L. Dewal, and S. Gupta, "Analysis and classification of hardwood species based on Coiflet DWT feature extraction and WEKA workbench," in *Signal Processing and Integrated Networks (SPIN), 2014 International Conference on*, pp. 9–13.
- [10] T. M. Maruyama, L. S. Oliveira, A. S. Britto, and S. Nisgoski, "Automatic classification of native wood charcoal," *Ecol. Inform.*, 2018.
- [11] J. Martins, L. S. Oliveira, S. Nisgoski, and R. Sabourin, "A database for automatic classification of forest species," *Mach. Vis. Appl.*, vol. 24, no. 3, pp. 567–578, 2013.
- [12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [13] V. Vapnik and V. N. Vapnik, *Statistical Learning Theory*. New York, USA: John Wiley and Sons Ltd, 1998.
- [14] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [15] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
- [16] B. Banoth, "Rotation and Scale Invariant Texture Classification using Log-Polar Wavelet Energy Signatures," 2007.
- [17] S. A. Broughton and K. Bryan, *Discrete Fourier analysis and wavelets: applications to signal and image processing*. John Wiley & Sons, 2018.
- [18] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, New Jersey, USA: Pearson Prentice Hall, 2008.
- [19] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelet toolbox*. Natick, Massachusetts, USA: The MathWorks Inc., 1997.
- [20] M. A. Kahya, W. Al-Hayani, and Z. Y. Algamal, "Classification of Breast Cancer Histopathology Images based on Adaptive Sparse Support Vector Machine," online) Scienpress Ltd, 2017.
- [21] M. A. Kahya, "Classification enhancement of breast cancer histopathological image using penalized logistic regression," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 1, 2019.
- [22] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.