Classifying Cultural Heritage Images by Using Decision Tree Classifiers in WEKA

Radmila Janković ✉[0000−0003−3424−134X]

Mathematical Institute of the Serbian Academy of Sciences and Arts, Serbia
rjankovic@mi.sanu.ac.rs

Abstract. This paper presents the first step toward looking for an advanced solution of image classification using decision trees in the Weka software. The aim of the paper is to evaluate the ability of different decision tree classifiers for cultural heritage image classification involving a small sample, based on three types of extracted image features: (1) Fuzzy and texture histogram, (2) edge histogram, and (3) DCT coefficients. The used decision tree algorithms involve J48, Hoeffding Tree, Random Tree, and Random Forest. The results indicate that the Random Forest algorithm performs best in classifying a small sample of cultural heritage images, while the Random Tree performs worst with the lowest classification accuracy.

Keywords: Classification · Images · Heritage · Weka.

1 Introduction

Image classification is the task of classifying images into different groups based on their feature descriptors. Image classification is an extremely important task within the digitalisation of the cultural heritage, as it allows a better database management, more correct search and interpretation. It is important to develop and use automatic classification techniques as these tasks usually involve a high number of images, may be prone to errors, and can take a lot of time to compute [8]. One way to perform image classification is by using decision tree-based algorithms.

Decision tree-based algorithms are an important part of the classification methodology. Their main advantage is that there is no assumption about data distribution, and they are usually very fast to compute [11]. In image classification, the decision trees are mostly reliable and easy to interpret, as their structure consists of a tree with leaves which represent class labels, and branches that use logical conjunction to produce a value based on an "if-then" rule. These values produce a set of rules that can be used to interpret the instances in a given class.

Many algorithms that are utilized for image classification require additional settings adjustments and parameter tuning, which makes the classification task time and effort consuming. Moreover, an open source software used for image classification lowers the costs of such a task, which is another important aspect that needs to be considered when dealing with classification problems. Hence, Weka, as a data mining and machine learning open source software, presents an adequate choice.

The aim of this research is to evaluate how well classification algorithms work when fed with image descriptor data from a small sample in the cultural heritage context. In particular, three types of features are extracted using the Weka software. Then, based on their values, decision tree-based algorithms are applied in order to predict the image classes.

2 Related Work

In [8], the authors used deep learning techniques to classify images of architectural heritage. Two convolutional neural networks were applied for image classification tasks: (1) AlexNet, and (2) Inception V3. Accuracy results obtained from this study were very good, with a mean value over 0.93 [8]. The authors concluded that the deep learning methods are very suitable for classifying heritage images. Moreover, the authors suggested using fine-tuning methods, if there are limited computational resources, or if the dataset is smaller [8].

A research by [11] investigated the efficiency of three data mining technique: (i) decision trees, (ii) support vector machines, and (iii) maximum likelihood algorithms to asses land cover changes from remotely sensed data. In this study, it was noted that the decision tree algorithms performed better than the other two classification methods.

In [16], a building facade classification based on an architectural style of the building was performed, considering only the windows. The approach was based on clustering and learning of local features [16]. A high classification rate was obtained, proving the efficiency of the proposed method.

The authors in [12] applied an alternative dimension of CBIR in order to organize heritage images in two classes: one class involving human activities, and the other involving non-human activities such as images of antique objects. The authors also applied a Naive Bayes algorithm in order to classify images based on the features from the edge histogram [12]. A good accuracy was obtained by this application of image classification, and the approach is considered suitable for the automatic classification of heritage images.

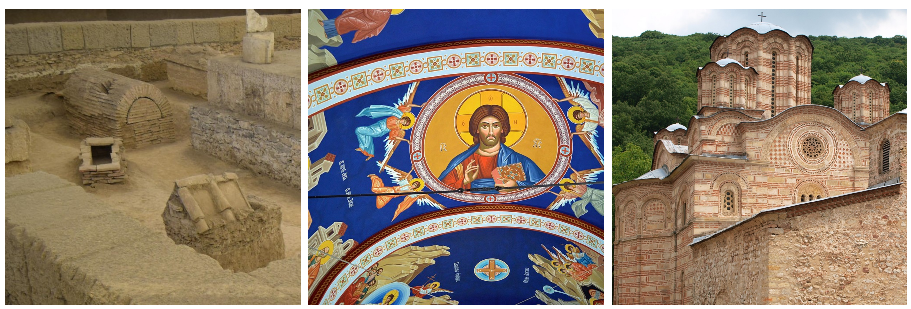
The proposed approach uses Weka, an open source data mining and machine learning software, to classify a small sample of cultural heritage images using decision tree-based algorithms. This approach is based on the fact that image classification needs to be more simple and human-independent, i.e. it needs less manual input, and smaller sample size. This can be accomplished by using Weka's ImageFilters for feature extraction, and then applying classification algorithms on extracted features. Moreover, this research considers only Serbian cultural heritage and is mainly focused on Eastern Orthodox cultural heritage, such as monasteries and frescoes.

The paper is organised as follows. In Section 3 the data and methodology used in this research are described. Then, in Section 4, the analysis is presented followed by a description of the results. Lastly, Section 5 draws conclusions from the study and proposes recommendations.

3 Data and Methodology

3.1 Data

The images used in this experiment have been downloaded from Google images and Flickr, and include 150 colour images in total (50 for each of three classes). The dataset included three classes: (1) Archaeological sites, (2) Frescoes, and (3) Monasteries. Before feature extraction and analysis, all images were firstly cropped and resized to the size of 150×150. A few images from the sample are shown in Figure 1.



(a) (b) (c)

Fig. 1. (a) Archaeological site, (b) Fresco, (c) Monastery.

3.2 Methodology

The main task of this paper is to classify cultural heritage images using Weka software, based on extracted features. The extraction of the features was made in Weka [19], using the ImageFilter package which is based on LIRE [9], a Java library for image retrieval, by applying (1) Fuzzy Color and Texture Histogram (FCTH) filter, (2) Edge Histogram filter, and (3) JPEG coefficients filter. After the feature extraction, four decision tree/based algorithms were applied on the dataset in order to classify images based on their type. The applied decision trees include: (1) J48, (2) Hoeffding Tree, (3) Random Tree, and (4) Random Forest. A 10-fold cross validation was used for testing the model. The results were then compared in order to evaluate which algorithm performs best for image classification problems with the small sample size.

Feature Extraction Methods. FCTH is a descriptor that presents a histogram which includes color and texture information from the image [3]. The FCTH is a combination of three fuzzy units, where in the first unit twenty fuzzy rules are applied so that a 10-bin histogram is produced [3]. This histogram is based on the Hue Saturation Value (HSV) color space. Hence, each of 10 bins matches a preset color [3]. In the second unit, the information about hue of each color is imported, so the histogram expands from 10-bin to 24-bins. In the last unit, a set of texture elements is exported, which are then used as inputs to the third fuzzy system, and the histogram is further expanded to a 192-bins histogram, containing texture information as well [3]. In this unit, a Gustafson Kessel fuzzy classifier is applied and 8 regions are formed, which are then used to map the values of the 192-bins histogram in the interval 0-7 [3].

The Edge Histogram filter is used to extract MPEG7 edge histogram features from the images. It represents the relative frequency and direction of edges in an image block, which are grouped into five categories: (1) vertical, (2) horizontal, (3) 45-degree diagonal, (4) 135-degree diagonal, and (5) non-directional edges [13]. The image block is created by dividing the image space into 4×4 non-overlapping blocks, irrespective of the size of the image [13].

The last feature extraction consists of an image conversion to a JPEG file format based on the Discrete Cosine Transform (DCT) which splits images into parts of different frequencies [10]. The frequencies that are less important are discarded from the image, which affects the quality of an image [10]. Therefore, this process is called "lossy".

Classification Methods. In this paper, four decision tree classification algorithms were used and compared in terms of their classification accuracy and errors: (1) J48 (C4.5), (2) Hoeffding Tree, (3) Random Tree, and (4) Random Forest. The classification is a supervised machine learning methodology, as it classifies instances into one of predefined sets of classes [4].

The J48 algorithm builds a decision tree which classifies the class attribute based on the input attributes. The algorithm is based on the C4.5 algorithm which was developed by Ross Quinlan [14]. The algorithm uses a greedy search method to create decision trees, and allows changing different parameters in order to obtain a better classification accuracy [18].

The Hoeffding Tree algorithm is a decision tree algorithm that can learn from massive data streams [5]. The assumption presumed by this algorithm is that the distribution does not change over time. This algorithm works well with small samples, as it uses the Hoeffding bound which computes the number of observations that are necessary to estimate statistics values within a prescribed precision [5].

A Random Tree algorithm draws a random tree from a set of possible trees [20]. The distribution of trees is considered uniform, as all trees from the set have the same chance of being sampled [20].

A Random Forest algorithm is an ensemble classifier that draws multiple decision trees using a bagging approach, hence the same sample can be selected multiple times, while some other sample may not be selected at all [1]. In-bag samples are used to train the trees, while out-of-the-bag samples are used for internal cross-validation [2]. The algorithm usually yields a very good classification accuracy, and it can handle multicolinearity well [1].

4 Analysis and Results

4.1 Performance Measures

In order to evaluate the classification parameters, the following performance measures were adopted: (1) Percent of correctly classified instances, (2) Kappa statistics, (3) MAE, (4) Precision, (5) Recall, (6) F-Measure, and (7) Time taken to build the model in seconds.

The percentage of correctly classified instances represents a total number of instances that were classified into correct classes. The higher the percentage, the better the model.

Kappa statistics is a measure of true agreement, only taking into account the only agreement that is not purely derived by chance [17]. It is calculated as follows:

$$k = \frac{P_a - P_b}{1 - P_b} \tag{1}$$

where P_a stands for the proportion of observed agreements, and P_b is the proportion of agreements by chance [17].

The strength of the agreement can be interpreted as follows:
≤ 0 = poor, $0.01 - 0.20$ = slight; $0.21 - 0.40$ = fair; $0.41 - 0.60$ = moderate; $0.61 - 0.80$ = substantial, and $0.81 - 1$ = almost perfect agreement [7].

Mean Absolute Error represents a sum of absolute errors between predicted and actual values [15]. Small values of MAE indicate a better model. MAE is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{2}$$

Where n stands for the number of measurements, and e_i represents the absolute error.

Precision and recall are important evaluation factors. Precision or confidence is the proportion of retrieved relevant instances, while recall or sensitivity presents the true positive rate of prediction [6]. Both values can be in an interval 0-1, where values close to 1 mean better prediction relevance.

F-measure is a measure of classification accuracy. It considers both the precision and recall values, and hence presents the harmonic average of these two values. F-measure can be computed as follows:

$$F = 2 * \frac{precision * recall}{precision + recall} \tag{3}$$

The values of F-measure can be in an interval 0-1, with values close to 1 indicating a better accuracy.

4.2 Obtained results

The results of the analysis are shown in Table 1. It is observed that the Random Forest algorithm correctly classified 92.7% of instances, followed by the Hoeffding Tree with 85.3% of correctly classified instances. Also, J48 correctly classified 76% of instances, while the Random Tree achieved the lowest classification rate of 74.7%.

Table 1. Comparison of decision tree classifiers for image classification

Algorithms	J48	Hoeffding Tree	Random Tree	Random Forest
Correctly classified instances	76%	85.33%	74.67%	92.68%
kappa statistics	0.64	0.78	0.62	0.89
MAE	0.16	0.1	0.17	0.22
Precision	0.76	0.85	0.75	0.93
Recall	0.76	0.85	0.75	0.93
F-Measure	0.76	0.85	0.75	0.93
Time taken to build the model (s)	0.22	0.39	0.02	0.28

Moreover, the results indicate substantial (Random Tree) to almost perfect (Random Forest) agreement, based on the values of Kappa statistics that range from 0.62 for Random Tree, to 0.89 for Random Forest.

The values of MAE indicate that Hoeffding Tree performes the best with MAE=0.1, followed by the J48 algorithm (MAE=0.16), Random Tree (MAE=0.17), and Random Forest (MAE=0.22).

This analysis also shows that the highest Precision and Recall are achieved for the Random Forest algorithm, with both values of 0.93. Random Tree, on the other hand, performed the worst in this case with precision and recall values of 0.75.

Moreover, the analysis shows that the Random Forest algorithm achieved the highest f-measure value of 0.93, while the Random Tree achieved the lowest f-measure value of 0.75.

Lastly, the time taken to build the model varies from 0.02s for the Random Tree algorithm, to 0.39s for the Hoeffding tree.

Figure 2 presents the confusion matrices generated from all four algorithms. It can be noted that J48 and the Random Tree algorithms most correctly classified the images belonging to the class of archaeological sites, while the Hoeffding Tree more correctly classified the images of frescoes. The Random Forest algorithm, which achieved the best overall accuracy, performed almost equally well on all three classes.

	Archaeological site	Frescoes	Monasteries
Archaeological site	41	5	4
Frescoes	7	37	6
Monasteries	8	6	36

(a)

	Archaeological site	Frescoes	Monasteries
Archaeological site	43	2	5
Frescoes	3	46	1
Monasteries	5	6	39

(b)

	Archaeological site	Frescoes	Monasteries
Archaeological site	42	5	3
Frescoes	4	34	12
Monasteries	9	5	36

(c)

	Archaeological site	Frescoes	Monasteries
Archaeological site	45	3	2
Frescoes	1	48	1
Monasteries	0	4	46

(d)

Fig. 2. (a) J48 confusion matrix, (b) Hoeffding Tree confusion matrix, (c) Random Tree confusion matrix, (d) Random Forest confusion matrix.

5 Conclusions

This paper analysed the application of methods in Weka for image classification using decision tree-based algorithms. The dataset included a small sample of cultural heritage images, divided in three classes. Three features were extracted from the images using Weka's image filters: (1) fuzzy color and texture histogram, (2) edge histogram, and (3) DCT coefficients. The extracted values were then fed into machine learning algorithms in order to evaluate their performance. Four decision tree-based classifiers were used: (1) J48, (2) Hoeffding Tree, (3) Random Tree, and (4) Random Forest. The obtained results indicate that, in terms of classification accuracy, the best one to use is the Random Forest algorithm. In terms of errors produced by the algorithm, the lowest MAE is found for the Hoeffding Tree, with good classification accuracy of 85%. The Random Tree classifier achieved the lowest performance with 75% of classification accuracy.

This research is an ongoing work only involving a small sample of images (50 per class). Future work will increase the sample size and create a much larger training database. Future work will also include the comparison of clustering, classification and neural network algorithms for cultural heritage images.

Acknowledgments. This work was supported by the Mathematical Institute of the Serbian Academy of Sciences and Arts (Project III44006).

References

1. Belgiu, M., Drgu, L.: Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing **114**, 24–31 (2016)
2. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
3. Chatzichristofis, S. A., Boutalis, Y. S.: Fcth: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. In: Ninth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 191–196. IEEE (2008).
4. Goyal, A., Mehta, R.: (2012). Performance comparison of Nave Bayes and J48 classification algorithms. International Journal of Applied Engineering Research **7**(11), (2012)
5. Hamoud, A. K.: Selection of Best Decision Tree Algorithm for Prediction and Classification of Students Action. American International Journal of Research in Science, Technology, Engineering and Mathematics **16**(1), 26–32 (2016).
6. Kumari, M., Godara, S.: Comparative study of data mining classification methods in cardiovascular disease prediction 1. International Journal of Computer Science and Technology **2**(2), 304–308 (2011)
7. Landis, J. R., Koch, G. G.: The measurement of observer agreement for categorical data. Biometrics **33**(1), 159–174 (1977)
8. Llamas, J., M Lerones, P., Medina, R., Zalama, E., Gmez–Garca–Bermejo, J.: Classification of architectural heritage images using deep learning techniques. Applied Sciences **7**(10), 992 (2017)
9. Lux, M., Marques, O.: Visual information retrieval using java and lire. Synthesis Lectures on Information Concepts, Retrieval, and Services **5**(1), 1–112 (2013)
10. More, N. K., Dubey, S.: JPEG Picture Compression Using Discrete Cosine Transform. International Journal of Science and Research (IJSR) India **2**(1) 134–138 (2013)
11. Otukei, J. R., Blaschke, T.: Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. International Journal of Applied Earth Observation and Geoinformation **12**, S27–S31 (2010)
12. Polpinij, J., Sibunruang, C.: Thai heritage images classification by Nave Bayes image classifier. In: 6th International Conference on Digital Content, Multimedia Technology and its Applications (IDC), pp. 221–224. IEEE (2010)
13. Prajapati, N., Nandanwar, A. K., Prajapati, G. S.: Edge histogram descriptor, geometric moment and Sobel edge detector combined features based object recognition and retrieval system. International Journal of Computer Science and Information Technologies **7**(1), 407–412 (2016)
14. Quinlan, J. R. : C4. 5: programs for machine learning. Elsevier (2014)
15. Sahoo, G., Kumar, Y.: Analysis of parametric and non parametric classifiers for classification technique using WEKA. International Journal of Information Technology and Computer Science (IJITCS) **4**(7), 43–49 (2012).
16. Shalunts, G., Haxhimusa, Y., Sablatnig, R.: Architectural style classification of building facade windows. In: International Symposium on Visual Computing, pp. 280–289. Springer, Berlin, Heidelberg (2011).
17. Sim, J., Wright, C. C.: The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Physical therapy **85**(3), 257–268 (2005).
18. Stojanova, D., Panov, P., Kobler, A., Dzeroski, S., Taskova, K.: Learning to predict forest fires with different data mining techniques. In: Conference on Data Mining and Data Warehouses (SiKDD 2006), pp. 255–258. Ljubljana, Slovenia (2006)

19. Witten, I. H., Frank, E., Hall, M. A., Pal, C. J.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2016).
20. Zhao, Y., Zhang, Y.: Comparison of decision tree methods for finding active objects. Advances in Space Research **41**(12), 1955-1959 (2008)