An Accurate HOG based Exemplar Pyramid Method for Image Classification of *Astragalus* L. Taxa

İrfan EMRE[a]* (iD), Turker TUNCER[b] (iD), Sengul DOGAN[b] (iD), Murat KÜRŞAT[c] (iD), Osman GEDIK[d] (iD), Yaşar KIRAN[e] (iD)

[a] Firat University, Faculty of Education, Department of Basic Education, TR-23119, Elazig, Turkey

[b] Firat University, Faculty of Technology, Department of Digital Forensics Engineering, TR-23119, Elazig, Turkey

[c] Bitlis Eren University, Faculty of Science and Arts, Department of Biology, TR-13000, Bitlis, Turkey

[d] Kahramanmaras Sutçu Imam University, Faculty of Agriculture, Department of Field Crops, TR-46000, Kahramanmaras, Turkey

[e] Faculty of Science and Arts, Department of Biology, Firat University TR-23119, Elazig, Turkey

ARTICLE INFO

ABSTRACT

As known from the literature, machine learning (ML) is one of the popular researches have been used variable areas. In this work, a novel exemplar pyramid method is presented to accurately classify *Astragalus* L. taxa by using their chromosome images. To implement ML to biological images, the proposed exemplar pyramid method is used. Histogram of Oriented Gradients (HOG) is utilized as feature generator. The proposed exemplar pyramid method consists of preprocessing, feature generation and concatenation, feature selection and classification phase. 10 classifiers are chosen to train and test the extracted features. According to results, the proposed exemplar pyramid generates discriminative features. because five of the used 10 classifiers achieved 100.0% classification rate.

1. Introduction

The classification of plants not only contribute to the botany but also supports the ecology, conservation, genetics and phytochemical studies (Wang et al., 2017). Traditional plant taxonomy studies use morphological, phytochemical, molecular and anatomical characters but these studies require experts in botany and sometimes these features aren't not enough to solve the systematical problems (Liu and Kan, 2016).

The studies showed that new approaches such as machine learnings are used in many areas such as financial analysis, transportation, health care, clinical diagnosis, image identification and also plant identification studies (Singh and Singh, 2021). The combined use of image processing and machine learning have greatly supplied to the solutions to describe of plant species using images from leaves, fruits or flowers (Guillermo et al., 2016; Naresh and Nagendraswamy, 2016; Yiğit et al., 2019). Machines perceive the objects and evaluate the results with little or human input in the machine

learning is an artificial intelligence approach (Singh and Singh, 2021). The image pattern identification was used to classify an image or a desired object or region into one of categories (Uchida, 2013). Image processing includes a series of transforms to increase the quality of the image and the image characteristics (Siraj et al., 2010). Histogram of Oriented Gradient (HOG) is one of the reliable feature extraction methods for image or pattern detection (Fernandes et al., 2016). It splites the input images into small square cells and later counts the histogram of gradient directions based on the basic differences (Ebrahimzadeh and Jampour, 2014).

The goal of this study is to compare of nine *Astragalus* L. (*Astragalus camptoceras* Bunge*, Astragalus cretaceus* Boiss,, *Astragalus campylorhynchus* Fisch. & C.A.Mey.A., *Astragalus suberosus* Banks & Sol., *Astragalus leporinus* Boiss. var. *hirsutus* (Post) D.F.Chamb, *Astragalus densifolius* Lam. subsp. *densifolius, Astragalus lamarckii* Boiss., *Astragalus anthylloides* Lam., *Astragalus odoratus* Lam.) taxa by using chromosome figures based on HOG feature extraction method. *Astragalus* L., is the biggest genus in the Fabaceae, has over the 3000 species around the world and it is represented by 478 taxa, of which 224 are endemic in Turkey (Çeçen et al., 2016; Albayrak and Kaya, 2019). The basic chromosome number of *Astragalus* L. species varied from x:7,8,11-15 (Sheidai et al., 2009). Also, Gedik et al. (2019) found that somatic chromosome numbers of nine *Astragalus* taxa in this study are 2n:2x:16, 32 and 2n:4x:32 chromosomes. They indicated that *Astragalus* taxa have metacentric and sub-metacentric chromosomes and one of them (*A. densifolius* Lam. subsp. *densifolius*) has sat-chromosome (Gedik et al., 2019). Although the chromosome morphology studies were largely used in classification studies, calculations and the determination of other characters in these studies are done manually using traditional methods that are open to error. Therefore this study aims to prevent such errors by using machine learning.

2. Dataset

Plant materials were collected from natural habitats and samples were stored in Bitlis Eren University Herbarium (BEUH). Karyological analysis of nine *Astragalus* L. taxa was conducted based on the method proposed by Elci (1982). Microphotographs of 387 cells in the metaphase stage of mitosis belong to each species were taken using Olympus BX51 light microscope after staining process completed. The chromosome images of nine *Astragalus* species were utilized as input for the proposed method. Attributes of the used dataset were given as follows. The used dataset consists of nine classes. There are19-57 images in the 1st-9th class (1st class is *Astragalus camptoceras* Bunge; 2nd class is *Astragalus cretaceus* Boiss., 3rd class is *Astragalus campylorhynchus* Fisch. & C.A.Mey.A., 4th class is *Astragalus suberosus* Banks & Sol., 5th class is *Astragalus leporinus* Boiss. var. *hirsutus* (Post) D.F.Chamb, 6th class is *Astragalus densifolius* Lam. subsp. *densifolius*, 7th class is *Astragalus lamarckii* Boiss., 8th class is *Astragalus anthylloides* Lam., 9th class is *Astragalus odoratus* Lam.). These images were RGB and were JPG. Size of these images were 3072 x 2304.

3. A Histogram of Oriented Gradients based Exemplar Pyramid Method

The main purpose of this study is to automatically classify *Astragalus* specifies by proposing a new ML method. To achieve this goal, a new hand-crafted feature generation and concatenation method is presented. The presented method is named as HOG based exemplar pyramid method. The proposed HOG based exemplar pyramid method is inspired by deep learning methods. Deep learning methods can extract low level, medium level and high level features but they have long execution times because need millions parameters set. Therefore, we presented a hand-crafted feature extraction method to obtain low level, medium level and high level features. Schematical demonstration of the proposed HOG based exemplar pyramid method was shown in Fig. 1.
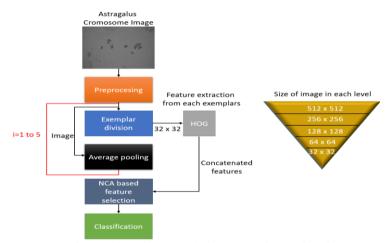

Figure 1. Schematical explanation of the proposed HOG based exemplar pyramid image classification method

Psuedo code of the proposed exemplar pyramid method was also shown in Algorithm 1 for better understanding. Algorithm 1 summarized the proposed HOG based exemplar pyramid image classification method.

Algorithm 1. Psuedo code of the proposed HOG based exemplar pyramid method.

Input: Microscopic Astragalus image (Im) with size of 3072 x 2304 x 3.
Output: Predicted values (PV) with length of 387.

```
0: Load microscopic Astragalus image.
1: Apply RGB2Gray conversion to Im.
2: Resize gray level Im to 512 x 512.
3: for i=1 to 5 do // Create levels
4:     counter = 1;
5:     [m, n] = size(Im);
6:     for k=1 to m step by 32 do // Divide 32 x 32 sized exemplars into Im.
7:         for l=1 to n step by 32 do
8:             Exm = Im(k: k + 31: l: l + 31);
9:             feature((counter − 1) ∗ 324 + 1: counter ∗ 324) = HOG(Exm);
10:            counter = counter + 1;
11:        end for l
12:    end for k
13:    Im = averagePool(Im); // Apply average pooling to Im.
14: end for i
15: Select 128 most discriminative features.
16: Classify selected 128 features.
```

Phases of the proposed HOG based exemplar pyramid image classification method were explained in sub-sections for better understanding.

3.1. Preprocessing

Size of the used *Astragalus* Chromosome images is 3072 x 2304 x 3 and these microscopic images were color (RGB images). In the used preprocessing phase, RGB2Gray conversion and image resizing were used and steps of this phase are given as below.

Step 1: Apply RGB2Gray transformation to Raw *Astragus* image. Mathematical notation of the RGB2Gray transformation was given in Eq. 1.

$$gray = Im(:,:,1)*0.299 + Im(:,:,1)*0.587 + Im(:,:,1)*0.114 \quad (1)$$

where $gray$ was grayscale image with size of 3072 x 2304.

Step 2: Resize grayscale image to 512 x 512 sized image.

$$gray = imresize(gray, 512 \text{ x } 512) \quad (2)$$

where $imresize$ defines resizing function.

3.2. Exemplar division

The proposed *Astragalus* chrosome image classification method was an exemplar classification method. To extract features in detail, exemplar feature extraction with levels were used. Therefore, the second phase of the proposed *Astragalus* chrosome image classification method were exemplar division. The generated 512 x 512 sized gray level image was divided into exemplars with size of 32 x 32.

Step 3: Divide 32 x 32 sized exemplars to image.

$$Exm^k = gray(i: i + 31, j: j + 31) \quad (3)$$
$$i = \{1, 33, \dots, row\} \quad (4)$$
$$j = \{1, 33, \dots, col\} \quad (5)$$
$$k = \{1, 2, \dots, \frac{row*col}{32^2}\} \quad (6)$$

Where Exm^k k^{th} exemplar, i and j describe indices of the image, row and col are number of rows and coloumns of the preprocessed Graylevel image.

3.3. Feature Extraction with Histogram of Oriented Gradients

HOG is one of the mostly known image descriptors and has been widely used for computer vision studies. The use of the HOG algorithm, which can also be called characteristic of the orientation (θ) and magnitude values of the pixels in the image in recent years, has received a lot of attention in many areas. It is an effective local image descriptor and calculates oriented gradients. The mathematical representation of the HOG method 0 is as Eqs. 7-10.

$$G_x = M_x \odot Exm \quad (7)$$

$$G_y = M_y \odot Exm \quad (8)$$

$$G = \sqrt{G_x^2 + G_y^2} \quad (9)$$

$$\alpha = arctan(\frac{G_x}{G_y}) \quad (10)$$

where M_x, M_y represent convuloution matrixes. To calculate graiends of and exemplar, convulotion process is used and $\odot$ represents convolution operator. G_x and G_y gradients of the used exemplar by using M_x and M_y. G is oriented graidents and α angel of the gradients. By using histograms G and α, HOG features are obtained.

In this study, we used HOG for automated classification of the microscopic images of the *Astragalus* chromosomes.
Step 4: Generate features using HOG and each examplers.
$$feat^k = HOG(Exm^k) \quad (11)$$

where $feat^k$ is k^{th} feature of kth exemplar with size of 324.
Step 5: Concatenate features of each exemplar to obtain level features.

$$feat^{level} = feat^1 | feat^2 \dots | feat^k \quad (12)$$

where $feat^{level}$ features of level, | is concatenation operator. We created five levelled pyramid by using average pooling.

3.4. Average Pooling

Average pooling has been widely used in the deep networks to create layers and reduce dimensionality of the features or images. It is basic and effective method. To create pyramid, average pooling is used.
Step 6: Reduce dimension of the image using average pooling with 2 x 2 sized non-overlapping blocks.

$$gray_{r,c}^{level+1} = \frac{gray_{i,j}^{level} + gray_{i,j+1}^{level} + gray_{i+1,j}^{level} + gray_{i+1,j+1}^{level}}{4} \quad (13)$$

where r and c indices of the pooled image.
Then, pooled image is divided into exemplars and features extraction process (HOG) is implemented. Steps 3-6 are repeated five times and each level features are concatenated to obtain final feature with size of 110,484.

3.5. Feature selection with NCA

The obtained final feature vector was very large. Therefore, a feature selector should be used. NCA was selected as feature selector in this work. NCA computes weights of each features and these weights represents strength of the used features. It was used for classification and regression because uses gradient optimization methods (stochastic gradients descent, adam) to calculate weights. Disctances from features were calculated by Manhattan distance. In this view, NCA similar to 1NN. Steps of the NCA based feature selection phase were given as below.

Step 7: Normalize final features in range of from 0 to 1. To calculate correct weights, normalization should be used

because NCA is a distance based feature selector.

$$feature = \frac{feature - feature_{min}}{feature_{max} - feature_{min}} \quad (14)$$

Step 8: Generate weights using NCA.

$$weights = NCA(feature) \quad (15)$$

Step 9: Sort weights from large to small and calculate indexes of the ordered weights.

$$[sorted, indexes] = sort(weights) \quad (16)$$

Step 10: Select 128 most disciriminative feature using $indexes$.

$$feat_i^S = feature_{indexes_i}, i = \{1,2,\ldots,128\} \quad (17)$$

where $feat^S$ defines selected 128 most discriminative features according to NCA.

3.6. Classification

Classification was final phase of the proposed HOG based exemplar pyramid method. 128 most discriminative features were forwarded to conventional classifiers. To use classifiers, MATLAB Classification Learner (MCL) Toolbox was chosen. There were 23 classifiers in the MCL and these classifiers can be grouped as Trees, Discriminants, Support Vector Machines, KNNs and Ensemble Classifiers. We selected Fine Tree (FT), Linear Discriminant (LD), Quadratic SVM (QSVM), Cubic SVM (CSVM), Medium Gaussian SVM (MGSVM), Fine KNN (FKNN), Weighted KNN (WKNN), Ensemble Subspace Discriminant (ESD), Ensemble Subspace KNN (ESKNN) and Ensemble Bagged Tree (EBT). 10 of the 23 classifiers of the MCL were used in the classification phase to obtain comprehensively results and default settings of these classifiers were used. Tests were validated using 10-fold cross-validation.

4. Results and Discussions

Various studies have shown that deep learning techniques achieve good results in plant classification. (Grinblat et al., 2016; Sun et al., 2017). For example, Grinblat et al. (2016) demonstrated the succesful results of deep learning in the plant identification by using leaf vein images. Also, Sun et al. (2017) used deep learning models to design plant classification in the natural habitats and they found that the achievement ratio of model was determined 91.78% (Sun et al., 2017). Similarly, Kayhan and Ergün (2020) classified the medicinal and aromatic plants by using several machine learning techniques and their results showed that plants were diagnosed in the correct classifications.

In this study, numerical results of the used 10 classifiers were presented. These 10 classifiers were used for validation to illustrate success of the proposed HOG based exemplar pyramid feature extraction method. The mostly used performance parameter for the classification was accuracy

(ACC). ACC has very simple mathematical. It is calculated by dividing the number of correctly predicted observations by the number of observation. Formula of ACC was given as Eq. 18.

$$ACC = \frac{NTPO}{NO} \quad (18)$$

ACC: Classification accuracy
$NTPO$: Number of true predicted observations.
NO: Number of observation.

We used 10 classifiers in five groups and these groups are tree, discriminant, SVM, KNN and ensemble. Tests were performed by using 10-fold CV. It means of the proposed method was a robust method. Calculated ACC are listed in Table 1.
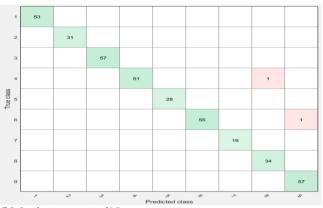
Table 1. Calculated ACC rates using tree, discriminant, SVM, KNN and ensemble classifiers.

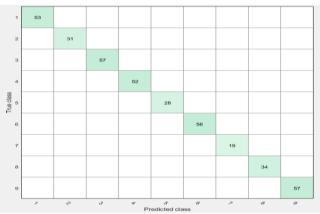| Group	Name	ACC	Training Time (sec)
Tree	Fine Tree	74.16%	1.43
Discriminant	Linear Discriminant	99.48%	0.64
KNN	Fine KNN	100.0%	4.88
	Weighted KNN	99.48%	5.61
SVM	Quadratic SVM	100.0%	4.55
	Cubic SVM	100.0%	4.43
	Medium Gaussian SVM	100.0%	7.43
Ensemble	Ensemble Subspace Discriminant	98.70%	9.88
	Ensemble Subspace KNN	100.0%	9.79
	Ensemble Bagged Tree	97.42%	11.05

As seen from Table 1, FKNN, QSVM, CSVM, MGSVM and ESKNN classifiers achieved 100.0% ACC. Table 1 clearly demonstrates that the used 5 of 10 classifiers achieve the best ACC. The worst classifiers was FT according to ACC. Ensemble classifiers has high training time because their training times are higher than 9 second. The fastest classifier was LD and it trained the selected feature set in 0.64 second and it achieved 99.48% success rate. Confusion matrixes of the used classifiers were shown as below (Fig 2a-j).
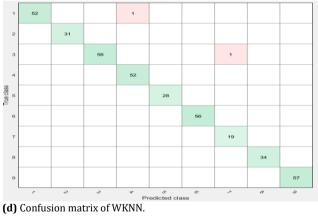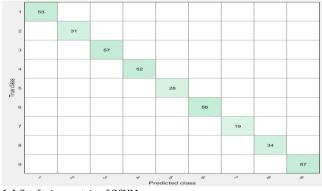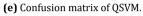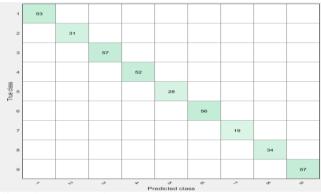

(a) Confusion matrix of FT.

(b) Confusion matrix of LD.

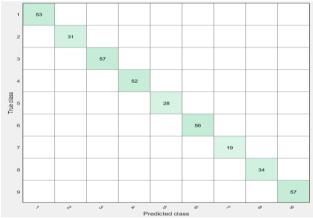**(c)** Confusion matrix of FKNN.

(d) Confusion matrix of WKNN.

(e) Confusion matrix of QSVM.
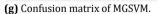

(f) Confusion matrix of CSVM.
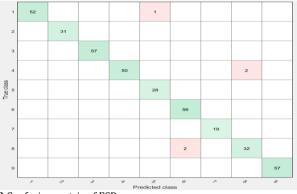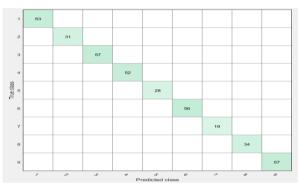

(g) Confusion matrix of MGSVM.
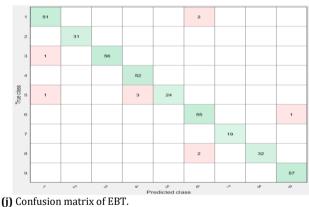
(h) Confusion matrix of ESD.

(i) Confusion matrix of ESKNN.

(j) Confusion matrix of EBT.
Figure 2a-j. Confusion matrixes of the used classifiers

It was indicated that the chromosome numbers of *Astragalus* were varied from x=7,8,11-15 and it was suggested that the basic chromosome numbers can be used as taxonomical tool to distinguish the *Astragalus* taxa (Sheidai et al., 2009; Gedik et al., 2019). A karyological study done by Gedik et al. (2019) demonstrated that *A. campylorhynchus, A. suberosus, A. densifolius* subsp. *densifolius* (has sat-chromosomes)*, A. cretaceus, A. leporinus* var*. hirsutus, A. anthylloides, A. odoratus* have 2n=16 while *A. lamarckii* has 2n=32 and *A. camptoceras*, 2n = 2x = 16 and 2n = 4x = 32 tetraploid chromosomes. Additionally, it was shown that *A. cretaceus* and *A. densifolius* subsp. *densifolius* were located same section according to the

Flora of Turkey (Davis, 1988)*,* And also, Emre et al. (2011) demonstrated that *A. camptoceras, A. suberosus, A. densifolius* and *A. cretaceus* have similar total and some individual saturated and unsaturated fatty acid compositions. However, it was showed that the used all of classifiers except for FT classified successfully separated and classified *Astragalus* taxa based on chromosome images (Table 1; Figure 2a-j).

4. Conclusions

In this study, a novel HOG based exemplar pyramid feature extraction method is presented. The main goal of the proposed HOG based exemplar pyramid iterative feature extraction is to generate distinctive features of the *Astragalus* species by using chromosome images. HOG is chosen as feature extractor and it generates features from each exemplar. By using pyramid method, features are extracted from exemplar of image and whole image. NCA reduced dimension of the extracted and concatanted features from 110,484 to 128. To illustrate success of the HOG based exemplar pyramid feature extraction method, 10 classifiers were chosen and five of them achieved 100.0% classification accuracy. Training times of the used classifiers were also listed in Table 1 to comprehensively comparison. Merits of the automated *Astragalus* chromosome image classification method;

- An *Astragalus* chromosome Image dataset is collected. This dataset acquires images of the 9 specifies.
- 100.0% classification accuracy was achieved. Therefore, it was suggested that the HOG can be contribute to taxonomical studies in addition to traditional taxonomical approaches including morphological and biochemical parameters.
- A cognitive biological image classification method is presented because there is no need set millions parameters.
- The proposed method has a high performance because 100.0% CA was obtained.
- Novel automated specifies classification application can be developed by using the proposed method.

References

Albayrak, S. and Kaya, O., 2019. Antioxidant, Antimicrobial and Cytotoxic Activities of Endemic *Astragalus argaeus* Boiss. From Turkey. Hacettepe J. Biol. & Chem., 47 (1), 87–97.

Çeçen, Ö., Aytaç, Z., and Mısırdalı, H., 2016. *Astragalus unalii* (Fabaceae), A New Species from Turkey. Turkish Journal of Botany, 40, 81-86.

Ebrahimzadeh, R., and Jampour, M., 2014. Efficient Handwritten Digit Recognition Based On Histogram of Oriented Gradients and SVM. International Journal of Computer Applications (0975 – 8887), 104(9), 10-13.

Elci, S., 1982. Observations and Reserarch Methods in Cytogenetics. Fırat University Press, Elazig.

Emre, İ., Şahin, A., Türkoğlu, İ., Yılmaz, Ö., Bahşi, M., and Kurşat, M., 2011. Compositions of Seed Fatty Acids in Some *Astragalus* L. Taxa from Turkey, Acta Botanica Gallica, 158(4), 487-491.

Fernandes, F., Weigel, L., Jung, C., Navaux, P., Carro, L., and Rech, P., 2016. Evaluation of Histogram of Oriented Gradients Soft Errors Criticality for Automotive Applications. ACM Transactions on Architecture and Code Optimization, 13 (4), 38:2-38:25.

Gedik, O., Kurşat, M., and Kiran, Y., 2019. Karyological Studies On Nine *Astragalus* L. taxa in Turkey. KSÜ Tarım ve Doğa Dergisi, 22(1), 35-44.

Grinblat, G., Uzal, L.C., Larese, M.G., and Granitto, P.M. 2016. Deep Learning for Plant Identification Using Vein Morphological Patterns. Computers and Electronics in Agriculture Journal, 127,418-424.

Kayhan, G. and Ergün, E. 2020. Medicinal and Aromatic Plants Identification Using Machine Learning Methods. Balkan Journal of Electrical & Computer Engineering, 8 (1), 81-87.

Liu, N. and Kan, J.M., 2016. Improved Deep Belief Networks and Multifeature Fusion for Leaf Identification. Neurocomputing, 216, 460–467.

Naresh, Y.G. and Nagendraswamy, H.S., 2016. Classification of Medicinal Plants: An Approach Using Modified Lbp with Symbolic Representation. Neurocomputing, 173,1789–1797.

Sheidai, M., Shahin, Z., and Jalal, I., 2009. New Chromosome Number Reports in Tragacanthic *Astragalus* Species. Caryologia, 62(1), 30-36.

Sing, M.M. and Singh, T.R. 2021. A Survey On Different Methods for Medicinal Plants Identification and Classification System. Revista Geintec, 11(4), 3191-3202.

Siraj, F., Salahuddin, M.A., and Yusof, S.A.M., 2010. Digital Image Classification for Malaysian Blooming Flower. Second International Conference on Computational Intelligence, Modelling and Simulation, 33-38.

Sun, Y., Liu, Y., Wang, G., Zhang, H., 2017. Deep Learning for Plant Identification in Natural Environment. Computational Intelligence and Neuroscience, 1-6.

Uchida, S., 2013. Image Processing and Recognition for Biological Images. Develop. Growth Differ., 55, 523–549.

Wäldchen, J. and Mäder, P. 2018. Machine Learning for Image Based Species Identification. Methods in Ecology and Evolution, 9, 2216-2225.

Wang, Z., Li, H., Zhu, Y., and Xu, T.F., 2017. Review of Plant Identification Based On Image Processing. Arch Computat Methods Eng., 24, 637–654.

Yigit, E., Sabanci, K., Toktas, A., and Kayabası, A., 2019. A Study On Visual Features of Leaves in Plant Identification Using Artificial Intelligence Techniques. Computers and Electronics in Agriculture, 156, 369–377.