Developing an automatic classifier of different plant genera of the subspecies Acer Palmatum

Rúbia E. O. Schultz Ascari
rubia@utfpr.edu.br
Federal University of Technology - Paraná (UTFPR)
Pato Branco, Paraná, Brazil

Dalcimar Casanova
dalcimar@utfpr.edu.br
Federal University of Technology - Paraná (UTFPR)
Pato Branco, Paraná, Brazil

Eliane Maria De Bortoli Fávero
elianedb@utfpr.edu.br
Federal University of Technology - Paraná (UTFPR)
Pato Branco, Paraná, Brazil

Patrick de Souza Sagioratto
patricksagioratto@alunos.utfpr.edu.br
Federal University of Technology - Paraná (UTFPR)
Pato Branco, Paraná, Brazil

Vinicius Caron Braun
viniciusbraun@alunos.utfpr.edu.br
Federal University of Technology - Paraná (UTFPR)
Pato Branco, Paraná, Brazil

ABSTRACT

Plant species identification is a task of interdisciplinary interest, desirable in many contexts, such as gardening, botanical research, and agriculture. For some plant species such as the Acer Palmatum, the characteristics of leaves, petioles, and trunks can drastically vary among the different genera of the same subspecies. Computer Vision and Machine Learning research areas made possible the creation of different classifiers trained to assist in species plant recognition based on digital images. However, the success of the training of a classification model is directly linked to the quality and adequacy of the dataset used. For the classification of Acer Palmatum plants, datasets composed of samples regarding the different varieties within this genus were not identified. Thus, in this paper, we proposed the creation of a new dataset and of a classifier to support the identification of distinct plant genera of the subspecies Acer Palmatum. We believe that our proposal aggregates relevant information not currently available, and will encourage further work aimed at automatically classifying between genera of some plant species which task is considered non-trivial even for experienced growers.

KEYWORDS

Dataset, Computer Vision, Machine Learning, Acer Palmatum plant

1 INTRODUCTION

To study a plant effectively, its classification and identification are of great importance.The classification of plants is the foundation of the science of botany; it is also the basis of plant genetics, plant ecology, plant medicine, and archival science [12]. Each species of plant has specific characteristics, which may be similar to characteristics of plants of other species, and that also can vary drastically among different genera of the same species. These characteristics may be linked to specifics of leaves, petioles, trunks, flowers, and fruits, which vary in shape, texture, margin, and color. Thus, the identification and classification of species is not always a trivial task, especially for non-specialized people.

The rapid development of computer image processing and pattern recognition technology has made possible automatic computer recognition of plant species based on image processing [48]. So, making use of the potential offered by the union of the areas of knowledge of Computer Vision and Machine Learning, an automatic plant identification tool can be developed and be very useful both for lay users and for experienced botanists. However, there is still no such developed system that can identify all plant species [16], so there are still many challenges that can be further explored.

The plants of the Acer genus, commonly known as maple, contain approximately 200 species with leaves varying in size, shape, and color. These Acer species are widely distributed in the northern hemisphere and are planted as ornamental plants all over the world [15]. The subspecies Acer Palmatum is original and popular on the Asian continent (China, South Korea, and Japan), but actually is widespread in various parts of the world and is subdivided into several varieties, which differ mainly in terms of leaf size, appearance, and color.

The subspecies Acer Palmatum is highly appreciated for the practice of bonsai art and the correct identification of the variety can be useful in the search for information about the correct management to be given to the plant. Many datasets composed of images of leaves, or even images of plants as a whole, are currently available for public use. However, based on the results of a systematic mapping of the literature carried out with the objective of identifying samples of different genera of the plant Acer Palmatum, we verified that no dataset contains samples subdivided between the different existing genera. The classification "Acer Palmatum" is available in many datasets, however, they do not encompass the different variants of Acer Palmatum, which correspond to a few hundred, such as Athropurpureum, Butterfly, Katsura, and Deshojo, among others. In this context, this paper presents a proposal for the creation of a new dataset and of a classifier to support the identification of different plant genera of the subspecies Acer Palmatum.

2 SYSTEMATIC MAPPING: MACHINE LEARNING ALGORITHMS FOR CLASSIFICATION OF PLANTS OF THE GENUS ACER PALMATUM

The possibilities of using Machine Learning to support the classification of plant species based on images are diverse. Digital images of plants have been widely used currently to assist Computer Vision systems that can be useful in different contexts, such as for flower classification [34] [32] [45], quality assessment of grains and seeds [7] [6] [24], classification of bonsai styles [28], and classification or identification of plants in general [39] [49].

Considering a large number of existing studies related to this research topic, we carried out a systematic mapping of the literature aiming to investigate recent research initiatives in the field of Machine Learning applied to support the classification of plant species, in which plant varieties of the Acer Palmatum were considered. The systematic mapping was developed based on the guidelines of Petersen et al. [33]. Thus, to select publications were established the following steps:

- Study protocol: planning and formalization of a study protocol, specifying the research problem, objective, general question, and research questions.
- Search expression: modeling and calibration of a search expression.
- Selection criteria: definition of selection criteria for publications returned by the search.
- Automatic search execution: the search expression was applied in the selected database and the search results were stored, containing the publications metadata, such as title, authors, year, abstract, keywords.
- First filter: reading of the title and abstract of each publication returned from the search and analysis of the data according to the selection criteria.
- Second filter: full reading of each publication and new analysis of the data according to the selection criteria.
- Data Extraction: data were extracted from each selected publication to answer the research questions established.

Table 1 presents details about the study protocol employed.

Table 1: Study Protocol used to conduct systematic mapping, specifying the research problem, objective, general question and research questions.

Research problem: Investigate research initiatives that used machine learning to classify images of plant leaves, specially plant varieties of the genus Acer Palmatum.
Objective: To map recently developed research applying machine learning for plant leaf classification, in which plant varieties of the genus Acer Palmatum were considered.
General question: What are the practical approaches developed applying machine learning for plant leaf classification, especially for plant varieties of the genus Acer Palmatum?
Research questions: 1) How has the machine learning been exploited to plant leaf classification? 2) What types of machine learning algorithms have been used? 3) Which datasets have been used for perform experiments? 4) In what context is plant varieties of the genus Acer Palmatum mentioned in the work? 5) There are studies specifically aimed at classification of plant varieties of the genus Acer Palmatum? 6) There are datasets composed of different plant varieties of the genus Acer Palmatum?

For the mapping, were considered scientific articles/papers indexed by the Google Scholar [1], published until June 2021, when an automatic search was performed. The search returned 50 studies, being the first filter excluded 20 papers, and the second filter excluded 5 papers, resulting in a set of 25 studies, categorized into themes of research interest: classifiers evaluation, classification of plant species, color variation in plant leaves, identification of Ayurveda's herbaria, linguistic descriptions of numerical data, new dataset of plant species, range expansions of Native and Exotic Plants. The search expression used was composed of: "leaf" AND "image" AND "acer palmatum" AND "machine learning".

The selection criteria were defined and applied as a first filter for the retrieved studies. The selection criteria were divided into inclusion and exclusion criteria and used to classify the studies according to their metadata (title, abstract, and keywords). Studies that met at least one of the inclusion criteria were included, and studies that met at least one of the exclusion criteria were excluded.

Table 2 presents the inclusion and exclusion criteria used to filter studies in the first filter.

Table 2: Inclusion and exclusion criteria used in the first filter of systematic mapping for study selection.

1st Filter	
Inclusion criteria	**Exclusion Criteria**
IC1: The study defines or presents practical approaches aimed to plant leaf classification, using machine learning and considering plant varieties of the genus Acer Palmatum.	**EC1**: The study does not present instruments, investigations, comparisons, or evaluations of practical approaches aimed to plant leaf classification, using machine learning and considering plant varieties of the genus Acer Palmatum.
IC2: The study investigates, compares or evaluates practical approaches aimed to plant leaf classification, using machine learning and considering plant varieties of the genus Acer Palmatum.	**EC2**: The study does not refer to a full paper.
IC3: The study presents instruments aimed at plant classification using machine learning, considering features from leaf image, citing plant varieties of the genus Acer Palmatum.	**EC3**: The study was not published in English.
IC4: The study presents instruments aimed to evaluate machine learning algorithms applied to the classification of plant species from the image of their leaves, citing plant varieties of the genus Acer Palmatum.	**EC4**: The study is not available for full access.

The selection criteria were applied again after the complete reading of the studies resulting from the first filter. Thus, after a second filter, only the studies dealing with topics related to at least one of the selection criteria were maintained and were submitted to data extraction.

Table 3 shows the data extraction form used to standardize the data extracted from the publications read, to reduce the bias of the results and the informality of the process. This data extraction aimed maintaining only the most relevant information for the research domain of this work.

The extraction was carried out by tracking the information on the extraction form according to declarations of each article. Based on the data extracted from the selected studies, we identified some methods and approaches used, as well as some characteristics of the researches that applied machine learning to plant leaf classification, considering plant varieties of the genus Acer Palmatum. From the results obtained with the mapped studies, the research questions

[1] https://scholar.google.com/

Table 3: Data extraction form used to standardize the recording of information obtained from the studies selected in the systematic mapping.

DATA EXTRACTION FORM
Year of publication.
Country of of the principal researcher.
Keywords.
Study objective.
Method and technologies used.
Dataset used.
Results obtained.
Context in that Acer Palmatum is used.

defined for the systematic mapping study were answered, and the main results will be presented, regarding the datasets and Machine Learning algorithms used in the selected studies.

2.1 Datasets containing Acer Palmatum samples

Based on the results of the mapping, we observed that some authors made use of preexisting datasets, available for free access, while others generated their own datasets, capturing the samples taking into account the requirements of the study in question. Regarding the datasets used, Table 4 presents the dataset and the studies that used each one, considering that some studies used two datasets.

Table 4: Datasets used in the selected studies of the systematic mapping.

Dataset	**Studies**
Flavia dataset [49]	[8], [17], [23], [25], [29], [36], [43]
ICL leaf dataset [22]	[46]
LeafSnap dataset [21]	[19]
Own-authored dataset	[1], [3], [10], [13], [26], [29], [35], [42]
UCI Iris dataset [9]	[26]
UCI Leaf dataset [38]	[2], [27], [30], [41], [51], [44]
UCI One-hundred plant species leaves dataset [26]	[20], [51]
Uninformed	[4], [31]

We realized that a wide variety of datasets was used in the selected studies, and many authors chose to create their own datasets (8 studies of 25), possibly by not finding available datasets with the desired characteristics or information. Even so, public datasets were also widely used, with Flavia dataset [49], and UCI Leaf Dataset [38] being the most used (7 and 6 studies of 25, respectively). This is a very common choice, since the use of a public dataset allows the experiments repeatability for other researchers, in addition to enabling comparisons of performance with existing approaches in the literature. It's important to highlight that the choice of datasets that contained samples of plants of the Acer Palmatum species, and the citation of these samples in the text were responsible for some articles being selected (they fit the inclusion criteria), even if the focus of the studies were not directed towards identification of specific species of that genus.

Some studies [1, 10, 19] used leaves from Acer species plants to illustrate samples of the datasets used or to illustrate some steps taken in their approaches. The species Acer Palmatum is mentioned in the following works, as an example of the species available on the dataset used: [2], [3], [13], [20], [23], [25], [26], [27], [29], [30], [31], [35], [36], [44], [41], [42], [43], [46], [51].

In addition to Acer Palmatum, other Acer species are also mentioned, such as Acer Buergerainum [13, 23, 25, 29, 43, 46], Acer Negundo [3, 42, 44], Acer Rubrum [19, 51], Acer Saccharinum [19, 51], Acer Platanoides [19, 51], Acer Pseudoplatanos [19], Acer Pensylvanicum [19], Acer Campestre [21, 26, 51], Acer Capillipes [21], Acer Circinatum [26, 51], Acer Mono [26, 51], Acer Opalus [31], Acer ginnala Maxim. Subsp. Theiferum (Fang) [46], Acer mono Maxim. [46], and Acer Rufinerve [51].

Even though the selection of studies was not focused on different species of Acer plants, we observed a great variety of these species in the datasets used. Likewise, we observed that the different plant varieties of the genus Acer Palmatum were not mentioned in the selected studies. Based on the answers obtained for the research questions established in the systematic mapping, no studies were identified that used Machine Learning algorithms to perform the automatic identification and classification of plants of the genus Acer Palmatum, which indicates that this is a gap that can be further investigated and exploited.

We observed also there are various datasets composed of leaves of plants of the Acer Palmatum, whether datasets of own authorship or authored by third parties. However, the non-existence of a dataset composed of leaves of the different genera of the species Acer Palmatum and even the non-existence of studies specifically focused on classification within this subspecies suggests the need to create such a structure, which motivated the proposal presented in this paper.

2.2 Machine Learning algorithms used for species classification

The identification of plants based on images generally involves four steps, that is, image acquisition, pre-processing, feature extraction and classification [16], as shown in Figure 1. A variety of Machine Learning algorithms have proven successful for the classification step of leaf species using image features [29].

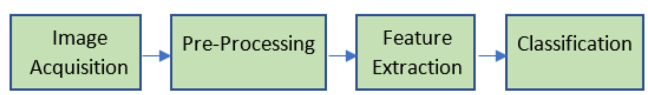

Figure 1: Generic steps in an image-based plant classification process using Machine Learning.

All the articles selected in the systematic mapping employed Machine Learning algorithms in some of the steps of the proposed approach. Table 5 presents the identified algorithms, associated with the studies that used them. Some studies used more than one algorithm for the classification step.

We observed that in the classification step, the classifier used most frequently was the Support Vector Machine, followed by the K-NN and Naive Bayes, CNN, LSVM, MLP, and Transfer Learning (with AlexNet). It can be highlighted that the choice for a particular classifier is closely linked to the resources available or that are considered important for the context in question. In situations where there is no prior knowledge about the important characteristics of

Table 5: Machine Learning algorithms used in the selected studies of the systematic mapping.

Machine Learning algorithm	Studies
Adaptive Boosting (AdaBoost)	[25]
Capsule Network (CapsNet)	[31]
Data-Driven Conceptual Spaces*	[2]
Convolutional Neural Network (CNN)	[1], [41]
Decision trees	[25]
Deep Neural Network (DNN)	[41]
Direct Acyclic Graph-based Multi-class Least Squares Twin Support Vector Machine	[42]
Extreme Learning Machines (ELM)	[43]
Fuzzy Relevance Vector Machine (FRVM)	[46]
Gaussian Support Vector Machines (GSVM)	[29]
j48	[8]
k-Nearest Neighbours (K-NN)	[19], [26], [25]
Linear Support Vector Machines (LSVM)	[29], [3]
Logistic Regression	[29]
Multi-Class Kernel Support Vector Machine (SVM)	[36]
Multilayer Perceptron (MLP)	[25], [27]
Neural Network using radial basis function	[17]
Neural Network (NN)	[29]
Support Vector Machine (SVM)	[19] (Leaf/Non-Leaf Classification), [13], [8], [10], [25], [27], [35]
Transfer Learning approach with AlexNet	[21], [23]
Transfer Learning approach with ResNet, GoogLeNet and VGGNet	[23]
Naive Bayes (NB)	[8], [27], [30]

* This study employed Machine Learning algorithms for the task of identifying relevant features and concepts in a numerical dataset, in order to shape the domains and quality dimensions of conceptual space.

the leaf, the use of deep learning, and automatic feature extraction can be a good option.

2.3 Challengins on identification and classification of Acer Palmatum subspecies

The Acer Palmatum species leaves available in datasets used by the mapped studies, correspond to its most common form, that is, an image of a green leaf, isolated, on a white background, such as presented in Figure 2. However, this plant, like many deciduous plants, undergoes changes in the color of its leaves, and not only related to autumn. Thus, the color of the leaves can vary greatly, taking on shades of red, and orange, among others. These variations in leaf color make the task of classifying species based on leaf images even more challenging, and in this context, color is a characteristic that cannot be considered in isolation. Thus, there are other important characteristics that need to be taken into account, such as the shape and texture of the leaves, or even the characteristics of the trunks and petioles.

Figure 2: Acer Palmatum leaf sample of Flavia Dataset [49].

Among the genus of Acer Palmatum, there are some of them with specific characteristics, such as the Acer Palmatum Atrorpurpureum, which has leaves in reddish tones throughout the year, suffering variations in the autumn period. Figure 3 shows two leaf samples captured from the same tree during different periods of the year.

Figure 3: Leaf samples of Acer Palmatum Athorpurpureum captured from the same tree in different periods of the year.

As another example, we can compare the common Acer Palmatum (cited in most of the datasets used in the studies returned via systematic mapping) and the Acer Palmatum Dissectum. Both plants have leaves with shades of green during the budding period. However, the shape of the leaves is quite different, as shown in Figure 4.

Figure 4: Leaf samples of Acer Palmatum Common and Acer Palmatum Dissectum.

Another genus of Acer Palmatum with interesting characteristics is the Acer Palmatum Deshojo, which has red shoots, and after the budding period, the leaves become green, as shown in Figure 5.

Figure 5: Leaf samples of Acer Palmatum Deshojo. On the left, is the leaf with red coloration after budding, and on the right, the already mature leaf, with green coloring.

Figure 6 shows samples of some genera of Acer Palmatum plants, which present visually perceptible differences. However, there are some genera of Acer Palmatum that have very similar characteristics, as shown in Figure 7 (Leaf samples of Acer Palmatum Atropurpureum and BloodGood.), and in these cases, the usefulness of

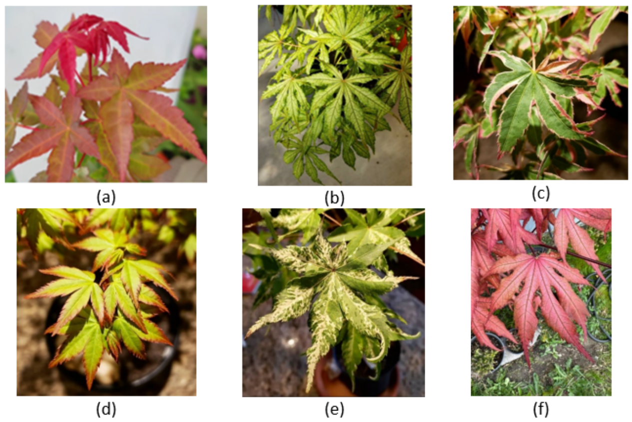

Figure 6: Leaf samples of Acer Palmatum regarding the genera (a) Deshojo, (b) Elmwood, (c) Beni Schichihenge, (d) Komachi Hime, (e) Orido Nishiki, (f) Amber Ghost.

Figure 7: Leaf samples of Acer Palmatum regarding the genera (a) Atropurpureum and (b) BloodGood.

using an intelligent algorithm is more evident due to the possibility of identifying patterns that are not so perceptible.

Thus, these inherent peculiarities of plants of the Acer Palmatum genus motivated the proposal presented in this paper. The research is in progress, and the methodology used to create the proposed dataset, and subsequent development of a classifier based on Machine Learning algorithms are presented below.

3 METHODOLOGY

Although the process as a whole can vary according to the techniques or methods used, in general, in an automatic classification process, there are image files, which are used as input to a Machine Learning-based system. A feature set is extracted from each input image and represented as a feature vector. In the classification step, a correspondence is performed between the feature vector of the query image and the feature vector extracted from the base images that make up a bank of known images. Based on the similarity measures, the system can identify which class the query file belongs to, presenting this information as a result.

To create a classifier capable of identifying different plant genera of the Acer Palmatum subspecies, a model needs to be trained from a dataset composed of different images of each genus, adequately labeled, mainly when supervised Machine Learning is employed. Thus, the first step in creating the classifier refers to the creation of a specific dataset for the context in question.

3.1 Image acquisition: Own dataset generation

Currently, many images are available for public use on the Internet, including pictures of plants and diverse datasets composed of a very expressive number of species. However, as mentioned above, the authors are unaware of a dataset with specific samples from the different genera of Acer Palmatum. There are many images of different genera available, however, there is no correct identification of the corresponding genus or variety, which makes the classification process very difficult. Thus, with the objective of creating such a structure, the authors of this research searched for people who grow plants of the subspecies Acer Palmatum and invited them to collaborate in the assembly of a dataset, through the sharing of pictures of their plants, including images of leaves, trunks and of the plant as a whole.

After sending several invitations through email messages and contacts on social networks, we received some positive responses, and at the time of writing this document, images were sent by twelve growers residing mainly in the south of Brazil and some other countries, generally in regions where there is a very severe winter, a desirable characteristic for the good cultivation of subspecies Acer Palmatum. Thus, the dataset is under generation, but already there are images of 59 different genera of the subspecies Acer Palmatum that have been received so far, with the number of samples ranging from only one sample to about forty samples. Table 6 presents the description of each genus as shared by the growers.

Table 6: Genera of subspecies Acer Palmatum contained in the dataset at the moment.

Genera		
Aconitifolium	Elena's Coral	Osakazuki
Amagi Shigure	Elizabeth	Red Emperor
Amber Ghost	Elmwood	Ruby Stars
Arakawa	Emperor I	Ryuzu
Atropurpureum	Fascination	Sango Kaku
Beni Kawa	Filigree	Seigen
Beni Maiko	GreenMist	Seriyu
Beni Schichihenge	Hogyoku	Shigi No Hoshi
Benichidori	Jordan	Shigitatsu sawa
Bihou	Katsura	Shirasawanum Aureum
Bloodgood	Kiyohime	Shirasawanum Moonrise
Butterfly	Komachi Hime	Shishi Yatsubusa
Common	KotoNoIto	Shishigashira
Crimson Carole	Kumoi Nishiki	Suminagashi
Deshojo	Mikawa yatsabusa	Thunbergii
Dissectum	Mikazuki	Tsumagaki
Dissectum Atropurpureum	Nomura	Villa Taranto
Dissectum Garnet	Orange Dream	Winterflame
Dissectum Rubrum	Orido Nishiki	Yama Nishiki
Dissectum Viridis	Orion	

The images received vary a lot, as there are samples corresponding to the leaves of the plants, others corresponding to the trunks, and others to the plant as a whole. Some samples were captured in a controlled environment, with a white background, while others were captured with a complex background, which may require some pre-processing and segmentation work before being used as input to a classifier based on Machine Learning.

3.2 Pre-processing

To build a dataset for use in Machine Learning algorithms, many samples of each genre to be studied are needed. Often the acquisition of these samples does not follow a standardized way, and may require the application of digital image processing techniques, such as segmentation, for example. From an image with a complex background, as shown in Figure 8 (a) it is possible to extract via segmentation only interest regions, such as the image of an isolated leaf, with a standardized background, as shown in Figure 8 (b), regarding the Acer Palmatum Katsura genre.

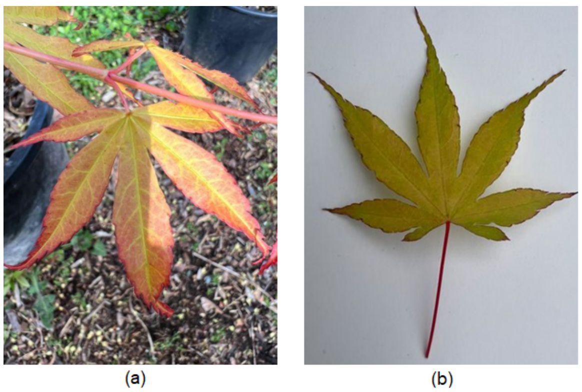

Figure 8: Acer Palmatum Katsura leaf sample (a) with a complex background and (b) with a white background.

In this way, it is intended to generate better-quality samples, favoring the automatic identification of the classifier to be used. Employing segmentation techniques, from images such as those available in Figure 6, would allow the extraction of other information, related to petioles, trunks, flowers, and seeds that allowed the creation of an intelligent classifier based on diverse resources from the same plant. In the study by Yang, Zhong, and Li [50] deep learning techniques are applied to perform the segmentation step, which is a viable strategy to increase the number of samples available for training the model to be created.

Another option to be used, as it is common in approaches that are based on Machine Learning algorithms, is the use of Data Augmentation, proposed by Tanner and Wong [40] where additional samples are created from existing data. For this research, each sample of training data will be augmented by creating variations by rotating and scaling the original image, such as presented in Figure 9. This process is useful in a scenario of insufficient training data to improve the classifier's performance.

3.3 Feature extraction

Regarding the feature extraction step, based on the systematic mapping carried out, it was observed that the shape and texture of the leaves are dominant characteristics for identifying plants. Other features such as color and venation pattern were also used. Itakura et al. [13] used multi-input for the tree species classification extracting features of leaf and tree trunk images. Turkoglu and Hanbay [43], used features extracted from leaves divided into two or four parts, instead of extracting from the whole leaf. We believe that

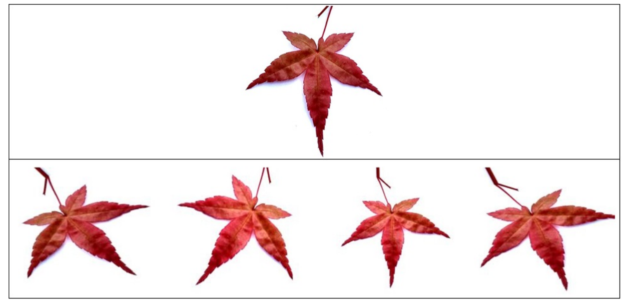

Figure 9: Example of Data Augmentation process. From a representative plant leaf image (original image, presented on the top), another four variations are generated through rotation and scaling operations.

merging several features could improve classification precision, and experiments aimed at evaluating this will be conducted. Furthermore, in many studies where deep learning is used for the automatic classification process, the feature extraction step can be omitted, since the model extracts the useful leaf characteristics directly from the input image, taking into account different features. The use of deep learning is also a promising alternative for working with in situ images, captured in uncontrolled environments. Our research aims to evaluate the performance of different Machine Learning algorithms, including the use of deep learning.

3.4 Classification

After selecting the desired features from each sample, the detection and recognition of these features through Machine Learning algorithms, called classifiers, are the next step to be performed. Classifiers identify the set of categories that the data subsets support and use this information to assign a class to an unidentified or unknown sample [5]. Different classification algorithms often require different sets of feature representations.

A Machine Learning process consists of two phases: the learning phase, in which the system analyzes the data and generates rules, finding some similarities between the data, and the validation phase, in which the generated rules must be verified, computing some performance evaluation function on a new dataset [14]. Machine Learning Algorithms can employ different learning approaches [18], being important to highlight supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning [37].

Supervised learning algorithms use pattern detection to establish predictions and can be subdivided into classification algorithms and regression algorithms [37]. Classification algorithms (commonly used in CV solutions) aim to classify something (items or samples) into a distinct set of classes or categories, according to the characteristics observed by the supervisor. Regression algorithms work by understanding the machine's relationship to variables to predict the values of a continuous variable. Our proposal aims to develop a classification algorithm based on supervised Machine Learning.

Every classifier needs to be evaluated for its predictive performance, taking into account mainly the proportion of correct classifications to the total samples available for classification. The estimation of the classifier's performance is performed using a set of images whose genres (labels) are known, that is, the test dataset. In an ideal classifier, the number of hits is equal to the number of samples used for testing, that is, the classifier does not make errors. Thus, through comparative analysis, the percentage of samples correctly classified for the set of samples in question can be calculated. If the model's accuracy is estimated from the same examples used for its construction, an optimistic estimate will probably be obtained. Thus, the methodology that will be employed in our approach will use examples that were not part of the training set, that is, data from the test set, to estimate the performance of the classifier.

There are different quality and accuracy metrics for evaluating the performance of a classifier, and many of them are based on the confusion matrix, a table that contains information about the actual values and predicted values for a classifier. From these data, precision, sensitivity or false positive rate, misclassification rate, specificity, and ROC curve (Receiver Operating Characteristic) [5] can be calculated. If the performance test used produces results considered acceptable for the context in question, according to the desired metric, the classifier can be used in the recognition of new samples, that is, images of plants whose classes (labels) are unknown.

4 DISCUSSION

When comparing images of plants of different species, it is common to observe relevant differences in relation to the shape of the leaves of the plants that make up the datasets used, which can facilitate the identification of similarities in some cases.

Comparison based on images between Acer Palmatum plants will in general require a comparison between leaves with the same shape, and very often similar color shades. Furthermore, the wide variety of existing Acer Palmatum plant subspecies, and the existence also of feature variations between them (Figure 6), show that a more in-depth assessment of these specificities is a relevant topic for research.

A classifier able to consider specific characteristics, such as vein pattern, petiole color, texture, and possibly attributes of the plant's trunk, demands a rich and varied dataset of plant images, composed of a very representative number of samples, especially to use Deep Learning. Convolutional Neural Networks (CNN) and their success in image pattern recognition problems have brought significant advances in several areas. However, there is a challenge related to how to adjust CNN for a small dataset while maintaining similar performance to the large-scale dataset. A common method for using CNN in small datasets is Transfer Learning, which eliminates the classifier layer of a pre-trained CNN and fits it into the target dataset [11]. The goal is to improve learning in the target task, taking advantage of the knowledge of the source task, which expands the application possibilities of CNN. Thus, Transfer Learning is a technique that we intend to use in our research, to enable the use of Deep Learning, while the dataset does not have such a significant number of available samples. Another reason to consider the use of Deep Learning important is that it is able to automatically extract the relevant feature sets from the images, which can enable the use of non-segmented images. Wäldchen and Mader [47] suggest that for real-life application, studies should use more realistic images containing multiple leaves, having a complex background, and being taken in different lighting conditions.

5 CONCLUSION

This paper presented a proposal in construction, toward the creation of a new dataset and an automatic classifier to support the identification of distinct plant genera of the subspecies Acer Palmatum, widely used for its ornamental character. Plants identification is an important task as it provides valuable information about their characteristics. Due to the existence of a wide variety of plants of the genus Acer Palmatum, the correct identification of these varieties through an automated classification system can be useful for both lay users and professionals, mainly for the correct management and conservation of the plants, and also for helping in monitoring the biodiversity. Some results were also presented from a systematic mapping of the literature on scientific contributions where Machine Learning was applied in the automatic classification of plants based on leaf image, emphasizing the identification of plant species of the genus Acer Palmatum. The mapping resulted in the reading and categorization of 25 papers published between 2011 and 2021. Results showed the lack of initiatives with similar objectives to the proposal presented in this document, and the inexistence of datasets composed of leaves of the different genera of Acer Palmatum, suggesting that there is much space to be explored in this research topic. According to the nature of Acer Palmatum with many varieties, the color and texture feature becomes very important in the classification process. To construct a robust algorithm for the classification process, the pre-processing techniques must be efficient so that discriminant data features can be constructed.

6 ACKNOWLEDGMENTS

The authors thank CAPES and CNPq for supporting this research.

REFERENCES

[1] K Anuradha, DPM Laksha, and RPS Kathriarachchi. 2020. A Study on the Ayurveda Plant Recognition for Remedial Medications Using Image Processing Techniques. (2020).

[2] Hadi Banaee, Erik Schaffernicht, and Amy Loutfi. 2018. Data-driven conceptual spaces: creating semantic representations for linguistic descriptions of numerical data. *Journal of Artificial Intelligence Research* 63 (2018), 691–742.

[3] Malay S Bhatt and Tejas P Patalia. 2017. Species and Variety Classification of Leaves from Images. *International Journal of Computer Science and Network Security (IJCSNS)* 17, 1 (2017), 109.

[4] Zhu Chen, Xiaoyu Lu, Yun Xuan, Fei Tang, Jingjing Wang, Dan Shi, Songling Fu, and Jie Ren. 2019. Transcriptome analysis based on a combination of sequencing platforms provides insights into leaf pigmentation in Acer rubrum. *BMC plant biology* 19, 1 (2019), 1–16.

[5] Sibanjan Das. 2016. Classification methods. In *Data Science Using Oracle Data Miner and Oracle R Enterprise*. Springer, 189–237.

[6] André Dantas de Medeiros, Márcio Dias Pereira, TFSN Soares, Bruno Gomes Noronha, and Daniel Teixeira Pinheiro. 2018. Computer vision as a complementary method to vigour analysis in maize seeds. *Journal of Experimental Agriculture International* 25, 5 (2018), 1–8.

[7] Emanuelle Morais de Oliveira, Dimas Samid Leme, Bruno Henrique Groenner Barbosa, Mirian Pereira Rodarte, and Rosemary Gualberto Fonseca Alvarenga Pereira. 2016. A computer vision system for coffee beans classification based on computational intelligence techniques. *Journal of Food engineering* 171 (2016), 22–27.

[8] Heba F Eid. 2016. Performance improvement of plant identification model based on pso segmentation. *International Journal of Intelligent Systems and Applications* 8, 2 (2016), 53.

[9] Ronald A Fisher. 1936. The use of multiple measurements in taxonomic problems. *Annals of eugenics* 7, 2 (1936), 179–188.

[10] Bo Fu, Manman Mao, Xilin Zhao, Zhilei Shan, Zhang Yang, Li He, and Zipeng Wang. 2019. Recognition of Plants with Complicated Background by Leaf Features. In *Journal of Physics: Conference Series*, Vol. 1176. IOP Publishing, 032053.

[11] Dongmei Han, Qigang Liu, and Weiguo Fan. 2018. A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications* 95 (2018), 43–56.

[12] EM Imah, YS Rahayu, and A Wintarti. 2018. Plant leaf recognition using competitive based learning algorithm. In *IOP Conference Series: Materials Science and Engineering*, Vol. 288. IOP Publishing, 012058.

[13] Kenta Itakura, Teruhito Hata, and Fumiki Hosoi. 2020. Tree Species Classification Using Leaf and Tree Trunk Images. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 4339–4342.

[14] Lakhmi C Jain and Janusz Kacprzyk. 2013. *New Learning Paradigms in Soft Computing*. Vol. 84. Physica.

[15] Shi-Bao Ji, Masato Yokoi, Norio Saito, and Long-Sheng Mao. 1992. Distribution of anthocyanins in Aceraceae leaves. *Biochemical Systematics and Ecology* 20, 8 (1992), 771–781.

[16] Surleen Kaur and Prabhpreet Kaur. 2019. Plant species identification based on plant leaf using computer vision and machine learning techniques. *Journal of Multimedia Information System* 6, 2 (2019), 49–60.

[17] Nishant Deepak Keni and Rizwan Ahmed Ansari. 2017. Content based image retrieval for leaf identification using structural features and neural networks. In *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE, 298–303.

[18] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee, and Khairullah Khan. 2010. A review of machine learning algorithms for text-documents classification. *Journal of advances in information technology* 1, 1 (2010), 4–20.

[19] W John Kress, Carlos Garcia-Robledo, João VB Soares, David Jacobs, Katharine Wilson, Ida C Lopez, and Peter N Belhumeur. 2018. Citizen science and climate change: mapping the range expansions of native and exotic plants with the mobile app Leafsnap. *BioScience* 68, 5 (2018), 348–358.

[20] Deepak Kumar and Chaman Verma. 2021. Automatic Leaf Species Recognition Using Deep Neural Network. In *Evolving Technologies for Computing, Communication and Smart World*. Springer, 13–22.

[21] Neeraj Kumar, Peter N Belhumeur, Arijit Biswas, David W Jacobs, W John Kress, Ida C Lopez, and João VB Soares. 2012. Leafsnap: A computer vision system for automatic plant species identification. In *European conference on computer vision*. Springer, 502–516.

[22] Intelligent Computing Laboratory. 2015. Intelligent Computing Laboratory (ICL) Plant Leaf Dataset [Online]. http://www.intelengine.cn/English/dataset/index.htm

[23] Wei Liu, Wenlong Feng, Mengxing Huang, Guilai Han, and Jialun Lin. 2020. Plant taxonomy in hainan based on deep convolutional neural network and transfer learning. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 1461–1466.

[24] Zhengguang Luan, Chunlei Li, Shumin Ding, Miaomiao Wei, and Yan Yang. 2020. Sunflower seed sorting based on convolutional neural network. In *Eleventh International Conference on Graphics and Image Processing (ICGIP 2019)*, Vol. 11373. International Society for Optics and Photonics, 113731K.

[25] Shubham Mahajan, Akshay Raina, Xiao-Zhi Gao, and Amit Kant Pandit. 2021. Plant Recognition Using Morphological Feature Extraction and Transfer Learning over SVM and AdaBoost. *Symmetry* 13, 2 (2021), 356.

[26] Charles Mallah, James Cope, James Orwell, et al. 2013. Plant leaf classification using probabilistic integration of shape, texture and margin features. *Signal Processing, Pattern Recognition and Applications* 5, 1 (2013).

[27] Azra Music and Sadina Gagula-Palalic. 2016. Classification of leaf type using multilayer perceptron, naive bayes and support vector machine classifiers. *Southeast Europe Journal of Soft Computing* 5, 2 (2016).

[28] Guilherme HS Nakahata, Ademir A Constantino, and Yandre MG Costa. 2020. Bonsai Style Classification: a new database and baseline results. In *2020 IEEE International Symposium on Multimedia (ISM)*. IEEE, 104–110.

[29] Alex Olsen, Sunghyu Han, Brendan Calvert, Peter Ridd, and Owen Kenny. 2015. In situ leaf classification using histograms of oriented gradients. In *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 1–8.

[30] Francis Rey F Padao and Elmer A Maravillas. 2015. Using Naïve Bayesian method for plant leaf classification based on shape and texture features. In *2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*. IEEE, 1–5.

[31] Sandeep Dwarkanath Pande and Manna Sheela Rani Chetty. 2021. Fast Medicinal Leaf Retrieval Using CapsNet. In *International Conference on Intelligent and Smart Computing in Data Analytics: ISCDA 2020*. Springer Singapore, 149–155.

[32] Isha Patel and Sanskruti Patel. 2019. Flower identification and classification using computer vision and machine learning techniques. *International Journal of Engineering and Advanced Technology (IJEAT)* 8, 6 (2019).

[33] Kai Petersen, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and software technology* 64 (2015), 1–18.

[34] Alireza Soleimani Pour, Gholamreza Chegini, Payam Zarafshan, and Jafar Massah. 2018. Curvature-based pattern recognition for cultivar classification of Anthurium flowers. *Postharvest Biology and Technology* 139 (2018), 67–74.

[35] Shitala Prasad, Piyush Kumar, and RC Tripathi. 2011. Plant leaf species identification using curvelet transform. In *2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011)*. IEEE, 646–652.

[36] Abeer Toheed Quadri and Mehreen Sirshar. 2015. Leaf recognition system using multi-class kernel support vector machine. *International Journal of Computer and Communication System Engineering* 2, 2 (2015), 260–263.

[37] Gopinath Rebala, Ajay Ravi, and Sanjay Churiwala. 2019. *An introduction to machine learning*. Springer.

[38] Pedro FB Silva, Andre RS Marcal, and Rubim M Almeida da Silva. 2013. Evaluation of features for leaf discrimination. In *International Conference Image Analysis and Recognition*. Springer, 197–204.

[39] Gjorgji Strezoski, Dario Stojanovski, Ivica Dimitrovski, and Gjorgji Madjarov. 2015. Deep learning and support vector machine for effective plant identification. In *Proceedings of ICT Innovations 2015 Conference Web Proceedings*. 221–233.

[40] Martin A Tanner and Wing Hung Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association* 82, 398 (1987), 528–540.

[41] Shamik Tiwari. 2020. A comparative study of deep learning models with handcraft features and non-handcraft features for automatic plant species identification. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)* 11, 2 (2020), 44–57.

[42] Divya Tomar and Sonali Agarwal. 2016. Leaf recognition for plant classification using direct acyclic graph based multi-class least squares twin support vector machine. *International Journal of Image and Graphics* 16, 03 (2016), 1650012.

[43] Muammer Turkoglu and Davut Hanbay. 2019. Recognition of plant leaves: An approach with hybrid features produced by dividing leaf images into two and four parts. *Appl. Math. Comput.* 352 (2019), 1–14.

[44] Kemal TÜTÜNCÜ, Murat KÖKLÜ, and Özcan ÇATALTAŞ. 2016. Comparison of Classification Techniques on Leaf Dataset. In *4th International Conference on Advanced Technology Sciences (ICAT'Rome)*. 23–25.

[45] Jhon Edwin Vera Vera, Sergio Mora Martinez, Andrés Torres Pérez, and Jonathan Avendano. 2019. Classification of Gerbera Type Flowers Based in Decision Tree Rules. In *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*. IEEE, 1–4.

[46] Balasubramanian VijayaLakshmi and Vasudev Mohan. 2016. Kernel-based PSO and FRVM: An automatic plant leaf type detection using texture, shape, and color features. *Computers and Electronics in Agriculture* 125 (2016), 99–112.

[47] Jana Wäldchen and Patrick Mäder. 2018. Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering* 25, 2 (2018), 507–543.

[48] Zhaobin Wang, Huale Li, Ying Zhu, and TianFang Xu. 2017. Review of plant identification based on image processing. *Archives of Computational Methods in Engineering* 24, 3 (2017), 637–654.

[49] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, and Qiao-Liang Xiang. 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. In *2007 IEEE international symposium on signal processing and information technology*. IEEE, 11–16.

[50] Kunlong Yang, Weizhen Zhong, and Fengguo Li. 2020. Leaf segmentation and classification with a complicated background using deep learning. *Agronomy* 10, 11 (2020), 1721.

[51] Qi Zhang, Shaoning Zeng, and Bob Zhang. 2019. Initial investigation of different classifiers for plant leaf classification using multiple features. In *Eleventh International Conference on Digital Image Processing (ICDIP 2019)*, Vol. 11179. International Society for Optics and Photonics, 1117922.