



# Fully-automatic natural plant recognition system using deep neural network for dynamic outdoor environments

Masoud Fathi Kazerouni<sup>1</sup> · Nazeer T. Mohammed Saeed<sup>1</sup> · Klaus-Dieter Kuhnert<sup>1</sup>

© Springer Nature Switzerland AG 2019

## Abstract

The leaf, as the vital part of plants, can be affected by physical and physiological factors which might lead to changes in its shape, color and size. These unique parts play an essential role in the design and implementation of plant recognition systems, as the shapes of leaves vary among different plants. Weather type and related factors, such as light intensity, humidity, temperature and wind-speed, may have effects on the number of leaves that grow on a plant, the amount of fresh and dried leaves, the deformation of leaves, color of leaves, positions of leaves on branches, etc. In addition, photographing in outdoor environments with different weather conditions has undesired impacts on images. For instance, light scattering changes severely when there are water droplets during photographing which influences the images captured in rainy weather. Despite the importance of the proposed factors and the relevant effects on the plants, leaves and images taken from plants, a plant recognition system should be independent from all environmental and non-environmental factors to be practically useful and applicable for identifying plant species in uncontrolled outdoor environments. Moreover, changes in the time of day (morning, noon and evening), the distance between camera and plant species as well as the angle and illumination of the object and environment when the images are taken should be considered in the process of plant recognition. For instance, in a windy weather condition, for an observer even from a short distance, it is challenging to distinguish all single leaves and identify the plant type from the shape of its leaf. Furthermore, the unstructured background of the images is another challenge which affects the images captured in fields and outdoor environments. These are only some difficulties for identification of plants in natural environments such as farms, forests, etc. A consideration of mentioned factors contributes to developing a novel, efficient and accurate system for plant recognition that can be used in various uncontrolled outdoor environments. In the real-life, the images of the plants captured with the presence of these factors are very challenging for recognizing plant species and it is a desire to develop an efficient and accurate system that can be used in various natural conditions and uncontrolled situations. This paper presents the development and implementation of a convolutional neural network for automatic plant recognition in uncontrolled outdoor environments. The deep network has been used to recognize four different natural plant species. The proposed system brings the opportunity and transformative potential of deep neural networks to the plant recognition field. This fully-automatic system is efficient and generalized for recognition of plants in outdoor environments like forests and farms, and its the accuracy is 99.5%. Meanwhile, the final system has the functionality of being applied as a real-time one.

**Keywords** Deep learning · Convolutional neural networks · Natural plant recognition · Fully-automatic system · Dynamic environment

---

✉ Masoud Fathi Kazerouni, masoud.fathi@uni-siegen.de | <sup>1</sup>Institute of Real-Time Learning Systems, University of Siegen, Hölderlinstr. 3, 57076 Siegen, Germany.



SN Applied Sciences (2019) 1:756 | <https://doi.org/10.1007/s42452-019-0785-9>

Received: 28 March 2019 / Accepted: 14 June 2019 / Published online: 20 June 2019

## 1 Introduction

After several decades of using various traditional machine learning algorithms, like support vector machine (SVM) [1], naive Bayes [2] and traditional neural networks, deep learning is a tremendous and competitive area for researchers and experts. Strenuous effort has been consumed with simulating the human brain, visual, and auditory systems in reality. Astonishments are inevitable in the deep learning area as it is a completely new generation in the machine learning field. This area has obtained a significant and undeniable role in machine learning and pattern recognition fields, although it is still at the beginning of its existence [3]. Deep learning is spending its evolution path, and it is in continuous progress as an active area. Many applications and areas have been connected to deep learning such as speech recognition, object detection and recognition, natural language processing (NLP) [4], customer relationship management (CRM) [5], etc. Benefits of deep learning algorithms can be used to improve and enhance the current developed systems in areas such robotics, agriculture and health sector.

Plants from the Devonian period did not grow very high as evidenced by fossils discovered in different places of the world, and the simplicity in structure was another common feature. These plants had stems, bifurcations and possibly spines, without leaves. The levels of their complexity have changed during the course of their life history. Furthermore, plants have been spread all around the world and even some of them are living in the oceans. Due to the location and circumstances of plants, they have unique roles which affect their own life and the lives of other organisms. Considering their impacts on human life leads to a better understanding of their roles. Plants have been traditionally considered as huge sources of energy, which is a permanent fact, and history shows the development of agricultural science before Roman times.

In modern life, the role of plants has become more colorful, whereby many researchers from various fields are working with plants either directly or indirectly. Plants affect climate and ecosystems directly and indirectly. They have many applications in agriculture, energy, environment, health, medicine, etc. Since the twenty-first century, providing necessary food for the world has become more important due to population growth and the climate change, particularly in some regions. Therefore, more emphasis has been given to developing new agricultural methods for farmers. In addition, farm management has different aspects which cannot be ignored; for instance, optimization is one of

the significant aspects of this issue. In order to use the whole capacity of a farm and optimize the products, one possible solution is to identify weeds and pests in field crops. It is very challenging for farmers to know all types of weeds due to their large variety. Hiring plants specialists to recognize types of plants does not have any economic justification, and it is also impossible to host this expertise on farms all around the world. In addition to a lack of knowledge, human observations are often inaccurate when two different plants have the same shape of leaves. Therefore, it is difficult to separate observed human data in some cases.

Traditional methods are not always useful for plant recognition as they are mostly expensive, time-consuming and human interaction is also needed. Thus, a system should be developed to recognize and identify plants automatically. The generalization and characterization of such a system are important factors, limits and parameters to judge its applicability. Additionally, such a system can be applied and used as the backbone of modern farming.

In Germany, the Federal Ministry of Food and Agriculture decides on general regulations. It tries to increase the motivation of the local population to realize their own projects to secure and guarantee the future of farms and villages by organizing federal competitions, such as "Our Village has a Future" [6]. In this way, the importance of using new agricultural methods is spread throughout the country.

Plant recognition is always a challenge even for botanists who are specialists in botany and plant sciences. It is very glamorous research due to its importance in other fields such as medicine, the pharmaceutical industry, modern farming, etc. In order to automate plant recognition, tremendous efforts have been put into this task to find a good solution and create an exact and precise system [7–12].

In previous work [7], Fathi Kazerouni described and used modern description methods to recognize different plant species of the Flavia dataset. In order to do feature detection and extraction, the scale-invariant feature transform (SIFT) algorithm [13] has been used, and two other methods, including the features from accelerated segment test (FAST) [14] and HARRIS [15] methods, have been combined with the SIFT algorithm. The obtained result is exploited by SVM, which is a machine learning task. The researchers used the speeded up robust features (SURF) algorithm [16] and combined the SURF methods for recognition of plant species in [8].

In [17], color histograms and color co-occurrence matrices have been used to distinguish the similarity between two images. If the overall color or color pair distributions of two images were close, they were matched as similar in terms of their colors [17]. However,

plant color is not a reliable factor as it changes across different seasons and during times of the day. Moreover, light intensity affects the observed color of leaves. Geometrical features have been analyzed in [18] to do the plant recognition task. A series of measurements, like area, perimeter, maximum length, maximum width, compactness and elongation [18], have been carried out, but this proposed method cannot be utilized in all conditions and situations. In [19], leaf contour is performed by using a feature extraction method and plant identification has been carried out through the images. The method used can be applied and utilized for artificial images. The probabilistic neural network has been proposed to recognize plants in [20], whereby 12 features were extracted in order to fulfill the purpose of achieving an accuracy greater than 90%. In [21], two methods of feature extractions, linear discriminant analysis (LDA) [22] and principal component analysis (PCA) [23], were applied to plants. In [24], a novel approach called semi-supervised locally linear embedding (Semi-SLLE) was proposed and it could be coupled with simple classifiers to recognize plant species. In addition to efficient dimension reduction, the final results proved the great performance of the proposed approach on leaf images and the recognition rate was equal to 90.26%. One important point is that most proposed methods deal with a single leaf in each plant image. In some situations, it is impossible to deal with just one complete leaf in an image, therefore, it is very challenging to recognize plant species correctly.

Due to the necessities of the modern life and the development of technology, modern data is needed to solve new problems and challenges of human life and its environment. To overcome the real-life problems in different fields, new novel techniques have been proposed which are directly or indirectly connected to soft computing area. In 2002, the problem of the multi-person multi-objective conflict decision was addressed, and a model was proposed which was based on the reservoir flood control [25]. In 2011, Wu and Chau [26] proposed three models, linear regression (LR), artificial neural network (ANN), and modular artificial neural network (MANN), with singular spectrum analysis (SSA) for predicting daily Rainfall–Runoff (R–R) transformation which has been a hard task in hydrology. The models were examined by two different watersheds, Wuxi and Chongyang, and the final experiments proved that the ANN R–R model led to promising results. In 2018, another interesting work was proposed in [27] and the lack of the measured evaporation data was investigated in meteorological stations of two cities in Iran, Rasht and Lahijan, over a period, from 2006 to 2016. A support vector regression (SVR) model [27] and a hybrid SVR-based

firefly algorithm (SVR-FA) model were used to simulate the evaporation process.

There are still unsolved and complex problems in real-life applications. Another challenging task, direct measurement of solar radiation, has been considered in [28] and an excellent examination of the performances of different data-driven techniques, SVR, model trees (MT), gene expression programming (GEP) and adaptive neuro-fuzzy inference system (ANFIS) was done. In addition, six equations in predicting global solar radiation (GSR) at the same synoptic station in Tabriz, Iran, were proposed and they were actually empiric. It is worth mentioning that the GSR was daily measured at the mentioned station from the beginning of 2011 to the end of 2013. Another interesting real-life problem related to a set of different fields, agriculture, hydrology, water resources engineering and early flood warning has been proposed in [29] and enhanced version of extreme learning machine (EELM) model [29] was proposed to do river flow forecasting. In order to do the experiment and the evaluation, the proposed model was applied in a tropical environment and several metrics such as the coefficient of determination ( $r$ ), Nash-Sutcliffe efficiency (Ens), Willmott's Index (WI), root-mean-square error (RMSE) and mean absolute error (MAE) were computed [29]. Recently in 2019, Baghban et al. [30] proposed a new approach to develop a general model for predicting nanofluid relative viscosity (NF-RV) [30]. To achieve the goal, expansion of an adaptive network-based fuzzy inference system (ANFIS) was performed [30]. The proposed model can be used as a tool for helping chemists and engineers who are involved with nanofluids in their works [30].

Another complex problem in agriculture and botany is natural plant recognition in real-life environments, and it has been remained an unsolved problem by considering different environmental limits and factors. Hence, one modern dataset can evolve in terms of size, complexity, generality, etc. The modern dataset [31] is completely different from other available datasets, and more details are explained in Sect. 3. In the controlled conditions, like a laboratory, artificial environment and indoor environment, it is feasible to capture leaf images by considering constant and fixed limits and factors during photographing. Available plant datasets like [20, 32] contain mostly single leaf images with homogeneous backgrounds, usually a simple white background, and images taken of similar leaves without petiole. Although the size or color of leaves might be changed within images captured from one specific plant species, similarity among the leaves of the plant species still remains. However, it is even hard for the human eye to recognize different plant species in such plant datasets due to the complexity of single leaf images. Furthermore, a certain lighting condition (brightness and light intensity) is usually considered and different related

factors, such as the point of view and angle, are kept fixed when capturing images from leaves.

The points mentioned create a gap between existing plant recognition systems and a system that is a demand in real-life. Therefore, a major task is to first find the real challenges and difficulties of plant recognition in uncontrolled conditions and then prepare a new dataset by considering the main challenges in natural environments. In nature and outdoor environments, the first factors that affect photographing are lighting condition and light intensity. If two pictures are taken of a plant, on two different sunny days, but in an outdoor environment with the same camera setting, angle, point of view, etc., as well as the same distance from the camera to the plant, there is no doubt that the images will not be completely the same because the light intensity, brightness and lighting condition might be different and the position of the main source of light, the sun, will affect the images and photographing process. The change of illumination and its effects cannot be neglected in natural color images captured in natural outdoor environments.

Another effective factor in the plant recognition process is actually the weather type, but its importance has not been considered in existing common datasets. For instance, the windy weather might result into the movement of the leaves while taking pictures. It results in a reduction in the clarity of the objects, the leaves, and an increase in the number of the deformed leaves in the images. Consequently, the final captured images will most probably be blurry. Whereas the fog leads to a decrease in the contrast of the images, and light scattering and blocking might be caused by small water droplets in the outdoor environment. As a result, other parameters, such as the amount of light, contrast, visibility, etc., will be changed. Furthermore, clouds change the amount of absorbed and diffused light, and visual effects might result as there is no direct light in the environment. All mentioned points add new challenges and difficulties to the plant recognition process.

Another important factor is the time that pictures are taken in uncontrolled conditions. If the setting of the camera, distance, etc., have been kept fixed and an image has been taken from one plant in an urban environment in the morning, the image captured from the same plant with the same conditions in the afternoon will be completely different as the light source and the amount of shadow have not remained the same. If the images of plants have been taken on different days, more challenges are added to the plant recognition process. For instance, the number of dried and fresh leaves might be changed, and the color of leaves does not stay the same even if the specific parts of plants have been considered as the region of interest for capturing photos.

Before explaining the other effects of weather types and time, it is essential to consider another factor that has effects on plant recognition which has been neglected in available datasets. A plant recognition system can be called a general one if it can be utilized at different distances. The definition of distance in this case is actually the distance between camera and plant. To meet the goal of developing a plant recognition system without depending on the distance, it is necessary to have a dataset which includes images taken from different distances. Being independent from the distance increases the system's efficiency in real-life application. It should also be pointed out that human visual system is not able to identify the shapes of leaves and plant species if the distance between the individual and the plant is great. For instance, if the observer is looking at a plant from 2 m, it is certainly impossible to identify the shape of a single leaf with enough detail. Furthermore, the observer would be looking at a scene with a bunch of leaves that might not be countable at all. An image taken of a plant from 150 cm also contains an uncountable number of leaves with the additional complexity of unwanted objects in urban and outdoor environments which add new challenges to the recognition task. Although the distance is an extra challenge, adding this factor to a plant dataset is a big jump for the next generation of plant recognition systems.

The effects of the weather condition and the distance can be considered together. It is possible to investigate weather types due to the physical properties and visual effects and make classifications referring to steady types, like fog, mist and haze, and dynamic types, like rain, snow and wind. By considering the steady class, it is found that droplets are too small, and they lie in a range from 1 to 10  $\mu\text{m}$ . Due to the mentioned range, droplets cannot be detected in an image captured at a long distance, but they obviously have effects on the plant recognition task. On the other hand, an investigation of the dynamic class proves that weather dynamics make images much more complex. Considering two images taken of plants in outdoor environments, like farms, shows that wind makes leaves invisible and rain produces a sharp intensity in images. In addition, a raindrop consists of small particles which are 1000 times larger in size, from 0.1 to 10 mm, if it is compared to a droplet in the steady class.

During photographing in the natural environment, two other important factors include the point of view and the angle. In many existing plant datasets, these two factors do not change, and they stay the same for all images. However, if the goal is to use a plant recognition system as a real-time system on farms, the story will be changed, and there is no guarantee of taking pictures from the same point of view or angle. For this purpose, it is necessary to prepare a dataset consisting of different images taken

from various points of view and angles. An effective solution is to capture the images from the plants in the natural environments randomly which can be interpreted as taking the pictures without a dependency on the point of view and the angle and trying to increase the variety of the images taken at various angles and points of view. Hence, the final system should not have an angle-dependent mechanism to be able to identify the plant species correctly without any prior knowledge about the complete shape of leaves.

By taking random images of plants, leaves might overlap which result in the deformation of the shape of the leaves in images. It is a hard challenge to develop a system that is able to perform recognition task when it is almost impossible to extract complete and enough information from the image about the shape of the object which should be recognized. In images captured from plants in outdoor environments, all leaves within images may not be complete and visible, thus the recognition process cannot be done correctly if the plant recognition system relies on information extracted from the whole contour of the complete leaf. In order to overcome this challenge, it is a demand to design a system that is able to identify the types of the plants in tough uncontrolled environments. Due to the mentioned points, the modern natural dataset has been made in an appropriate way to meet the requirements.

Another important factor is time of taking pictures in uncontrolled conditions. If the setting of the camera, distance, etc. have been kept fixed and an image has been taken from one plant in urban environment on the morning, the image captured from the same plant with the same conditions on afternoon would be completely different as the light source and amount of shadow haven't remained the same after several hours. If images from plants have been taken in different days, additional challenges would be added to plant recognition process. Number of dried and fresh leaves might be changed, and color of leaves doesn't remain the same even if specific parts of plants have been considered as region of interest for capturing photos.

The mentioned factors add new challenges to the plant recognition task and they have been introduced and identified to make a real natural plant dataset and subsequently develop a real-life application to identify plant species. The prepared dataset [31] in this work has unique characteristics from different aspects and the golden key to solving the problem of natural plant recognition is large variability seen among its images. It is also worthy to mention that a similar dataset is not available and the details of the modern dataset will be provided in Sect. 3.

This work is the continuous works of [9, 10] where combined methods has been used for plant recognition

systems. Alongside with the high accuracy of 94.9404% of the previous works, their limitation were their dependency on different conditions and insufficiency for real-time applications. In this work, however, the state of the art, deep learning method, has been employed for the natural plant recognition system to overcome the limitation of the previous works and also to increase the overall accuracy of the system.

In [9, 10], modern approaches like FAST-SIFT, FAST-SURF, HARRIS-SURF, etc., have been applied to build automatic natural plant recognition systems where the modern dataset which will be explained in Sect. 3 has been utilized and six different systems have been implemented finally. Each system was based on building a unique vocabulary for each distance. As an example, the system based on the HARRIS-SURF method has one vocabulary for the distance 25 cm and another vocabulary for the distance 50 cm. If the testing step is performed on the images taken from 25 cm, it is essential to use the vocabulary built for this specific distance and it is not possible to utilize the vocabulary which has been created for distance 50 cm. The experiments and the tests showed promising results. For instance, the accuracy of the system which has used the SIFT approach as one of the components is 94.9404%. One lack of the developed systems for the natural plant recognition was the dependency of the systems on the pre-information about the distance between the camera and the plant. Hence, it was essential to use the correct vocabulary for each distance to recognize the type of the plant for the new test.

Despite the efficiency and good performance of the proposed systems based on the combined methods, new goals have been defined and the current work will fulfill the remaining goals. In fact, the proposed systems are the first members of the new generation of plant recognition systems which are able to identify plant species in natural environments by considering new factors like distance, time of day, etc. One goal is to build an improved system with higher accuracy; therefore, it has been decided to use deep neural network by the human brain and vision systems and extract more useful information from natural plant images. In addition, one goal of the current work is to make natural plant recognition closer to the human vision system.

In addition, it is a desire to have a general algorithm which is able to detect and extract features efficiently in any challenging condition in a natural environment. This generality also contributes to the other goals of the new system as it should finally be used as mobile and real-time plant recognition system. Instead of using just 2-dimensional information, it is feasible to add depth to the input images which results in obtaining 3-dimensional volumes in the training process and getting high level features and

spatial information which are actually the main need for the classification problems. Furthermore, it leads to predict the outputs through an advanced and effective process.

Another goal is to build a distance-independent system which can be applied as a real-time system for the plant recognition in natural outdoor environments. In order to use the proposed systems in [9, 10] as real-time systems, it is necessary to have pre-information about the distance between the camera and the plant to use the correct vocabulary for testing the new sample image in the real-time application. The distance can be measured by human or adding a new pre-processing part to the system for performing the measurement. In both cases, it is necessary to add a new method to select automatically the correct vocabulary for testing the input natural image for the real-time application. Hence, the complexity of the proposed algorithms increases and the cost of making real-time systems is high due to the increase of the total computations. In the current work, one goal is to develop a real-time system which doesn't need any pre-information. It is worth mentioning that there is no overlap between the new proposed system and the systems proposed in [9, 10] as the current work doesn't apply traditional machine learning algorithms. In this work, the feature engineering process follows a new approach that differs widely from previous proposed systems. The mentioned points and goals will be considered in this work and such goals will be pursued as well.

In order to use the potential of the deep learning methods, the intention is to classify plant species by convolutional deep neural networks [33]. Convolutional deep neural networks are actually fully connected at the top and use some ideas such as local receptive fields, shared weights and pooling. Convolutional deep neural networks have been used in various deep learning frameworks. Deep learning frameworks are under the pressure of increasing their characteristics, features, capacity, functionality, adaptability, and availability for new ideas to be useful and applicable in different aspects and tasks. Modern convolutional neural networks (CNNs) [34] have been applied in [35, 36] and have demonstrated their huge power and capacity in order to recognize images.

With an increasing number of fully-automatic systems, the demands of these systems have greatly expanded in different scientific and industrial fields and ordinary people are also interested in utilizing user-friendly technologies and applications. As a result, there is an intense competition between companies to industrialize deep learning concepts and methods and use them in their products practically. On farms, robot technology can play a fundamental role in different applications to improve farming tasks. In order to use the potential of deep learning methods, it is intended to classify plant species and rely on this generation of machine

learning. In the scope of the future work of the system, it will be used as a part of a mobile robot system for identifying plant species in uncontrolled outdoor environments.

The dataset proposed in this work consists of 1000 images which are taken at different distances, weather types, times, illuminations and light intensities for four different plant species. These plant species, Hydrangea, Amelanchier Canadensis, Acer Pseudoplatanus and Cornus, belong to a region named Siegerland in Germany. They are common plants in this region, and the variation in the appearance of leaves is usually high over time. This dataset is a novelty since there are no other datasets using real-life images of plant species in different possible conditions. Therefore, this paper is the first to use deep neural networks for such complex data and it cannot be directly compared to other studies using artificial datasets.

From the viewpoint of implementing the system based on the deep learning, it is certainly a good solution to the need of recognizing plant species in outdoor environments even in the challenging weather condition and building a mobile and real-time application with the feasibility of being used in mobile robot and semi-robot systems. The novelty of the implemented system lies in the fully-automatic recognition of plant species in very complex and challenging natural conditions and dynamic outdoor environments. For instance, this novel deep neural network system recognizes and identifies plant species at long distances, like 200 cm, and in difficult weather conditions with a high accuracy of 99.5%. In this paper, the developed system employs a convolutional deep neural network architecture to carry out plant recognition and classification tasks. The evaluation of the system is then investigated and some experiments have also been performed to compare the results to the systems with traditional machine learning algorithms in [9, 10] which have been implemented through other pipelines as discussed before. In addition, the proposed system provides a precious feature, which is flexibility, and it can be used in lack of expensive hardware equipment (more details are given in the next section).

The rest of the paper is organized as follows: Sect. 2 presents the plant recognition task; Sect. 3 describes the details and important points of the dataset; Sect. 4 represents the details of the approach and reports how the model is working in order to recognize plant species; Sect. 5 explains the conducted experiments; and Sect. 6, concludes the paper.

## 2 Plant recognition task

The orientation of the plant recognition task is to identify the plant species of a given observation which is an image taken with the Canon EOS 600D camera. The images of the

dataset have been taken at different distances, weather conditions, lighting and background. There are some other points with regard to the dataset. Moreover, the images are taken at different times, mornings and evenings, and days. Additionally, points of view and perspectives are also different from one image to another. Furthermore, light intensity and illumination are not constant in all images. By using the mentioned camera, all images of the dataset are in red-green-blue (RGB) format. Due to the briefly mentioned characteristics, the dataset is complex; therefore, the plant recognition task is challenging. Each observation is actually an image captured from one of four plant species in an uncontrolled natural environment.

Furthermore, one significant challenge is to train large deep neural structures. This point is highlighted and presented in Sect. 4.

### 3 Modern natural plant dataset (MNPd)

The created dataset, called the modern natural plant dataset (MNPd) [31], contains color images which include considerably different characteristics, percentage of homogeneous regions, details, etc. Some points have been considered as general rules of preparing the dataset. In order to take pictures, similar protocols have not been used to acquire the images. The attempt has been to take pictures of the leaves of the same species from distinct plants in various conditions and situations at different times. No special consideration has been assumed in the camera selection process, therefore, there is no dependency on the used camera. Furthermore, the size of images is not kept constant at all and it varies from an image to another image.

To have a useful natural dataset, different aspects or components of the natural environment should be taken into account. Adding these continuum aspects leads mainly to create and provide the logical and efficient collection of data to solve the problem of plant recognition in a natural outdoor environment and compensate for the lack of a modern natural dataset. A careful exploitation of information can be of help to enrich numerous applications with fresh insights into plant identification. In other words, the dataset provides a wealth of information that reveals insights into current challenges.

The distance is a significant factor which has been considered during the preparation of the dataset. This factor contributes to garnering useful images for a modern dataset and to shortening the gap between natural real-world and virtual-world. Images have been captured from different distances as below:

- 25 cm

- 50 cm
- 75 cm
- 100 cm
- 150 cm
- 200 cm

Furthermore, the change of viewpoint, variation of illumination, change of light intensity, complex scenes with various backgrounds in varying weather types, etc., are only some special characteristics of the images in this natural and challenging dataset. In Fig. 1, four different images of the dataset are represented.

It is also worth mentioning that one of the added factors of this dataset is unique in comparison to available datasets which is the changes of weather types. For instance, a number of images has been captured in cloudy and windy weather conditions, and as described earlier clouds absorb a part of light and diffuse the rest in a cloudy day, so there is usually no direct light on objects in natural environments and this causes visual effects. On a windy day, the movements of leaves are undeniable, even the human eye cannot distinguish the shapes of the leaves correctly. In this particular weather type, the images can be blurry, and movements of leaves may lead to leaf deformation. In addition, there was no force to exclude any unwanted effect because naturalness is important for the defined goals to build an automatic natural plant recognition system, and the key is to have natural images without any additional consideration by a human.

An increase of change in the appearance of the leaves is fruitful to solve the complicated task of plants identification and to correctly develop natural plant recognition systems.



Fig. 1 Four different samples of the dataset

### 4 Proposed approach

The core of the proposed approach for plant recognition is deep learning, which is investigated to perform the desired task. In fact, a deep CNN model, a hierarchy with many layers, is applied, and the proposed approach is based on nonlinear transformation functions and input data. In this section, an explanation of the proposed approach and architecture or topology of the CNN model is provided. It is indisputable that the selection of the approach depends on various factors including dataset, computing power, intended application and ideas. Moreover, hardware availability and abundance of datasets are exploited by deep learning concepts. Therefore, millions of parameters and hidden layers can be employed to implement deep CNN models.

A deep CNN model is developed and used to do automatic plant species classification without any user interaction. A CNN model can be composed of many different types of layers, such as convolutional layers, pooling layers, fully connected layers, etc. These layers contribute to making biological concepts a reality. Convolutional layers are the cores of a CNN model and consist of a set of learnable filters. Although each filter is not spatially big, they play the role of extending through the full depth of the input data. Therefore, it is feasible to achieve the goal of having a volume of neurons.

The proposed network consists of two different types of layers which are convolutional and fully connected layers. Figure 2 represents these two types of layers for CNNs. The type of the first five layers is a convolutional layer, while the layer type of last three layers is a fully connected layer. A fully connected layer has been utilized to connect current neurons to all the neurons of the previous layer. Overall, the architecture of the proposed deep learning network and the layers of the

implemented network are explained in the details that follow. Considering the output of the convolutional layer indicates that high level features are actually the output of this type of layers and somehow it can be interpreted as the feature extraction process. The convolutional layer should be trained for extracting meaningful patterns from the input natural images where the outputs of the lower layers, first and second convolutional layers, are similar to the extracted edges of the natural image and such similarity can be observed in the visualization part of the system which is provided in next section. By comparing the role of the fully connected layer to the convolutional layer, the fully connected layer acts like a classifier in the traditional machine learning algorithm. Despite the important role of the fully connected layer, this layer increases the complexity of the model and it is computationally expensive without any doubt.

The input data (input image) is  $227 \times 227 \times 3$  and it is filtered by a convolutional layer; it is the first convolutional layer of the deep network. This layer filters the input data by using 96 kernels. The size of the kernels is  $11 \times 11 \times 3$  where a stride of 4 pixels has been selected. The stride is the distance between the receptive field centers of neighboring neurons in a kernel map. If the value of the stride was larger than the defined value, the probability of losing information would be increased. In such case, the overlap to the receptive fields would be reduced and spatial dimensions would be consequently decreased. The type of the second layer is also a convolutional layer. Furthermore, normalization and pooling are also performed in this part. The first convolutional layer is followed by a response-normalization layer as rectified linear unit (ReLU) [37] neurons have been utilized. One characteristic of these neurons is their unbounded activations; thus, normalization is an essential step after using ReLU neurons. Then, the local response normalization layer is followed by another layer, called pooling

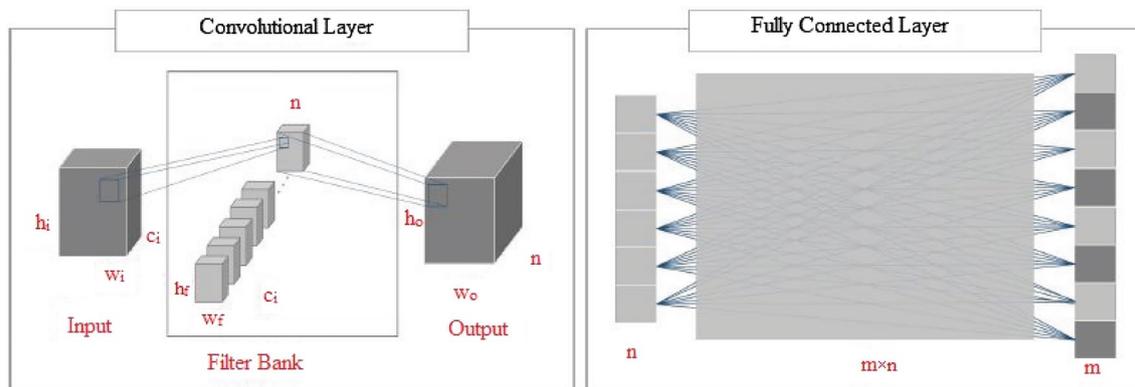


Fig. 2 Two types of CNN layers

layer, and the type of this layer is a max-pooling layer. Its main responsibilities are to carry out down-sampling operation and reduce the number of parameters leading to the decrease of the computational cost. Consequently, it contributes to preventing overfitting as it provides an abstracted form of the image representation. When the size of the kernel is equal to  $3 \times 3$ , it means that a region with this size will be pooled over. By using this type of layer, an overlap will be available and important information about the location of object within the image won't be lost. The weights in the first layer are initialized from a Gaussian distribution, where its mean value is zero and its standard deviation value is 0.01. By investigating the value of the standard deviation, setting smaller value results in chocking activations and applying larger value leads to the explosion of the activation. In this layer, neuron bias has been initialized by using the constant 0. The output of the first layer is the input of the second layer.

The ReLU activation function is:

$$y_i = \max(0; x_i) \quad (1)$$

where  $x_i$  is the input of the  $i$ th channel. It is a simple function and accelerates the convergence of stochastic gradient descent. Consequently, faster learning becomes possible as it removes negative values by using a very simple process with less computation cost and maintains memory consumption efficiently. This function takes on its responsibility very well and contributes to having the bottom and the top alike at the same size. Subsequently, its operation is not expensive in comparison to other functions. Other functions, such as sigmoid and hyperbolic tangent (tanh), face the problem of the gradient vanishing where values move away from zero. As an example, the gradient of the sigmoid becomes increasingly smaller as the absolute value increases. On the other hand, the ReLU function solves the gradient vanishing problem, and this function has another important priority over other possible functions by considering the other aspect when considering computational cost.

Before explaining the next layer, it should be pointed out that there are two batches; one belongs to the training phase and the other one belongs to the testing phase. The size of the batch means the number of inputs in one pass for processing. If the size of the batch is set to high value like 250, error might happen with regard to the memory of the GPU; thus the size of the batch should be set to a lower value. In this model, the size of the batch in the training phase is set to 50 and its value in the testing phase equals 10. Dividing the number of the training images by 50, the result is equal to 16, thus it is possible to use a coefficient of 16 as the test interval value and it has been set to 32. Furthermore, the number of the testing images are divided the size of the batch in

the testing phase and the result is equal to 20 which is used as the value of the test iteration.

The second convolutional layer performs filtering too. 256 kernels with  $5 \times 5 \times 48$  have been applied for filtering. This part also has an intervening normalization and pooling layer. In this layer, the local response layer and max pooling layer have been utilized like the previous convolutional layer. It is worth mentioning that a neurobiological concept which is called the lateral inhibition is implemented by the local response layer. It plays an important role because of the need of normalizing the ReLU function as an unbounded activation. In addition, its role is to detect high frequency features with a large neuron response and simultaneously damp large responses that are uniformly large in any local neighborhood. Furthermore, initialization of the weights and neuron bias has been performed by means of a zero-mean Gaussian distribution with the standard deviation 0.01 and the constant 0.1, respectively. The next layer, the third one, is a convolutional layer without any local response normalization or pooling layer. Normalized and pooled outputs of the second layer have been connected to the third layer, and this third layer has 348 kernels with the size of  $3 \times 3 \times 256$ . The type of the weight filter is Gaussian with the standard deviation of 0.01 and the mean of zero, and the bias layer is the same as the first convolutional layer. A ReLU layer is also used in this part in order to introduce non-linearity to the proposed deep network. The fourth and fifth layers have 384 and 256 kernels, respectively. The size of the fourth convolutional layer's kernels is  $3 \times 3 \times 192$  and this size equals to  $3 \times 3 \times 192$  for the fifth convolutional layer, too. The fifth convolutional layer is followed by max-pooling layer. The number of neurons of the fully connected layers, sixth layer and seventh layer, are 4096. Moreover, weight initialization in these mentioned layers has been done by using a Gaussian distribution with the standard deviation 0.01. Additionally, neuron biases have been performed by the constant 0.1.

In sixth and seventh layers, there are also ReLU and dropout layers. Dropout is added into these layers to solve one of the most important problems of neural networks, overfitting. The dropout method [38] contributes to the implemented network, conquering the mentioned problem. The dropout layer is also a biological inspiration like neural networks. The advantage of the dropout is to randomly deactivate neurons along with all of their incoming and outgoing connections from the network during the training process; as a result, it is possible to create multiple independent internal representations of the relationship in the data. A dropped neuron has no effect on the loss function and the gradient that flows through it during back-propagation will be effectively zero and its weights will not be updated. It can be applied instead of the common

method of combining the predictions of different models which has been used for test error reduction in [39, 40]. In order to converge, dropout is helpful as it changes the required iteration value which is approximately a double one and it also leads to better generalization.

Finally, the output of the last fully connected layer, 8th layer, is fed to one 4-way softmax with loss as the number of classes (labels) are equal to 4 different plant species. Neuron biases of the fully connected layers have been initialized by the constant 0.1. For initializing the weights, a zero-mean Gaussian distribution with the standard deviation 0.005 has been utilized for both sixth and seventh layers. In last layer, weight initialization has been done by one zero-mean Gaussian distribution with the standard deviation 0.01. In general, it does computation of a probabilistic likelihood per class and uses that to calculate the error that the network has created. In order to find the score of the network of the present batch, an accuracy layer is added.

To train the model, stochastic gradient descent has been applied and a batch size of 50 examples is set. Other important parameters, momentum and weight decay, are equal to 0.9 and 0.0005, respectively. Weight decay is a parameter which governs the regularization term of the deep network. Additionally, the low value of this parameter is helpful for error reduction of the model's training. Setting the value of the weight decay depends practically on the network and the goals. Setting this parameter to such low value fulfills the desired goal for caring about the predictions and getting high accuracy.

The used learning rates for the implemented layers have the same values. Furthermore, the base learning rate of the network has been set to 0.00001 and the advantage of setting a low value for this parameter is to benefit from a more reliable training process.

## 5 Results and discussion

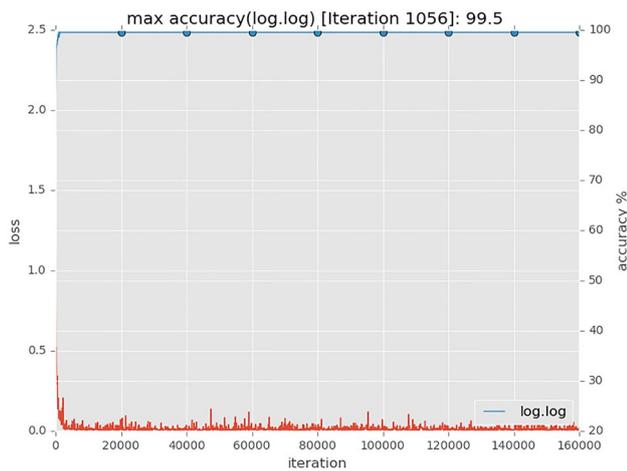
The experiments have been conducted on the MNPD. For the experiment section, the MNPD is divided into two groups which are train and test datasets and the selection of images for each dataset is randomly done. Train dataset contains 800 images and the test dataset has 200 images. Some factors, such as non-uniform illuminations, including shadows, underexposure and overexposure, background clutter and pose, vary significantly among the images of this dataset and such a large range of variation in both train and test datasets is suitable to explore

various aspects of the problem and find a suitable solution for overcoming the challenges of recognizing plant species in natural environments. In addition, each image might be affected by several factors and there is mainly no focus on the effects of only one factor as there isn't any control on environmental factors during the recognition task. Table 1 shows the number of images in different distances. For the experiment, the specifications of the used machine are Intel Core i7 (4820K), central processing unit (CPU) [41] 3.70 GHz, installed memory (random-access memory called RAM) 16.0 GB, and graphics GeForce GTX 760. In the test phase, natural image inputs have been fed into the implemented network without doing any additional preprocessing operation, such as cropping, scaling, etc., on them. It should be pointed out that Caffe [42] is the platform used for the implementation of the network, and it is well-known as a powerful platform for building deep neural networks. As a popular and active platform for classification tasks, it provides flexibility for CNN implementation, and it can be extended by linking the deep model to other relevant toolboxes. Another important reason for selecting the Caffe is the possibility of switching between CPU and graphics processing unit (GPU). The results is that the implemented model by the Caffe can be used in the absence of sufficient hardware equipment and related facilities. Hence, a lack of GPUs does not have any effect on the applicability of the final deep natural plant recognition system and it is a great advantage when the system should be employed by small field robots with limited hardware equipment.

The primary concern in image classification problems is to measure accuracy and evaluate the quality of the natural plant recognition system. In this experiment, accuracy is the ratio of the number of correctly predicted natural plant images according to the total number of predicted natural plant images in the test dataset which equal 200 images taken from four plant species in an uncontrolled outdoor environment. As test images have been randomly selected from the modern dataset and large variations can be observed among natural plant images of each plant species, there is no dependency on the distribution of images of each plant species in the test dataset. In the test dataset, there is no consideration about the number of images taken in each weather type, and the whole process of choosing images for the test dataset is performed randomly, and one class of plant species might contain 2 images taken on sunny days and another class of plant species contains probably 8 images taken from this

**Table 1** Number of the images at different distances

Dataset (cm)	25	50	75	100	150	200
Number of the images	240	240	240	240	20	20



**Fig. 3** Maximum accuracy, iterations and accuracy

specific plant on sunny days. As a result, this is an advantage that makes the system more reliable than when a user decides on the selection of test images for each class of plant species.

Figure 3 shows the accuracy of the deep network in different iterations. It represents the maximum accuracy and changes of iteration and accuracy from the first iteration to the last. Additionally, the total error of the deep network is observable in the color red and the accuracy can be observed in blue. As illustrated, the maximum accuracy occurred in iteration 1056. This means that the highest accuracy has been obtained in 1056th iteration for the first time, and it remains constant in the next iterations. The training phase has been completed in this iteration and the final model has been obtained. Furthermore, the first y-axis belongs to the loss parameter and the second y-axis belongs to accuracy values.

The proposed deep neural network has been spread across the GPU to speed up the training phase. The time needed for the training phase has been computed and it is equal to 1248.5088 (s) for the 1056th iteration. Perhaps the needed time of the training phase seems to be high, but it is important to consider how long it takes to train the proposed deep neural network if the CPU has been utilized in the training phase.

To address this training question, an extra experiment has been performed and the needed time for training the network is computed when the used unit is the CPU. The specifications of the used machine are Intel Core i7 (4790K), CPU 4.00 GHz, and 16.0 GB is the installed memory RAM. It took more than 5 weeks to train the recognition system. The amount of time needed is not comparable to what was necessary with the GPU. In any case, it is possible to train the model with a CPU in the case of a lack of a GPU.

**Table 2** Confusion matrix of the proposed deep network

Deep learning	Hy.	Am.	Ac.	Co.
Hydrangea (Hy.)	50	0	0	0
Amelanchier (Am.)	0	50	0	0
Canadensis				
Acer (Ac.)	0	0	50	0
Pseudoplatanus				
Cornus (Co.)	0	1	0	49

As has been discussed in Sect. 4, overfitting is one of the most important issues. Therefore, investigation of this issue has been done by doing a test. Firstly, the dataset has been divided into 5 groups of images, namely group A, group B, group C, group D and group E. Each group contains 200 natural plant images of the original modern dataset. Then group A has been considered as the testing dataset and the other 4 groups constitute the training dataset containing 800 images of the dataset. After that, the entire training and testing phases have been conducted to get the final accuracy. In order to do the next step of the experiment, group B has been considered as the testing dataset and other groups constitute the training dataset. Again, the proposed deep model has been trained and tested by these new training and testing datasets and the final accuracy is calculated. This procedure should be continued for the other groups and each group is then considered once as the testing dataset and the rest as the training dataset. Consequently, comparing the obtained accuracy values of the experiment proves that no overfitting happened. Moreover, the issue of overfitting has been taken into account during the design of the architecture and implementation of the model by using a dropout layer which both prevents overfitting and limits the size of the weights by means of a parameter called weight decay. Both the dropout layer and weight decay have already been explained in a previous section, Sect. 4.

To evaluate and visualize the performance of the implemented system for natural plant recognition, an interesting matrix, named the confusion matrix, is constructed [43]. In the case of the classification task for four different natural plant species, the confusion matrix is one  $4 \times 4$  matrix containing information about the actual classification results and different category labels through the classification in its rows and columns. Table 2 is the obtained confusion matrix of the classification experiment.

During the classification experiment, one misclassification has been observed for one of the classes, and the plant species' name is Cornus, and a Cornus sample has been wrongly predicted as Amelanchier Canadensis. The accuracy percentage of the implemented deep network is

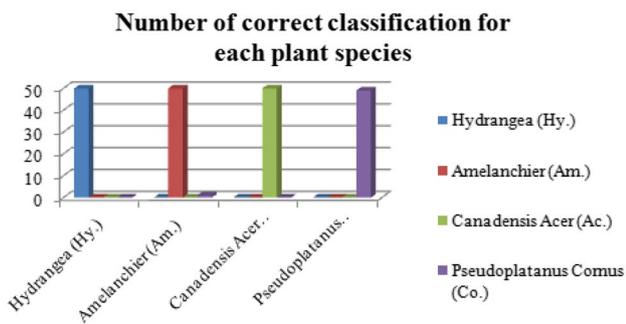


Fig. 4 Number of correct classifications for each plant species

99.5%. This accuracy is higher than all six systems in [9, 10] which have utilized modern combined methods. Figure 4 shows the number of correct classifications for each plant species.

In [9], different modern detection and description techniques, SIFT, HARRIS-SIFT, and FAST-SIFT have been utilized to create plant recognition systems. The system that has used the SIFT algorithm owns the highest accuracy, 94.9404%, among the three implemented systems. Distance is a factor which impacts these systems. If we consider the implemented system with FAST-SIFT, one vocabulary has been constructed for each distance; therefore, there is still a dependence on the distance between the image and the camera. For instance, there is a vocabulary for the distance 50 cm and another vocabulary is built for the distance 75 cm. If the intention is to identify the type of plant for a new sample image, it is essential to know the distance between the camera and sample plant as a pre-information. In [10], three new systems have been implemented where the detection and description algorithms are modern combined methods, SURF, HARRIS-SURF and FAST-SURF. There is also a dependency on the distance for these automatic plant recognition systems and each system has its own vocabularies at defined distances. The implemented system with SURF has the highest accuracy that is equal to 93.9575%. The second highest accuracy has been obtained by the system with the FAST-SURF method and the value of the accuracy is 90.9375%. The system with the HARRIS-SURF method has the lowest accuracy, 90%, although its accuracy is still acceptable due to the characteristics and nature of the system.

There is an important difference between the implemented systems by SVM classifiers and modern detection and description methods [9, 10] and the current proposed system based on a deep neural network. The number of training images in [9, 10] are equal to 664 images where the number of test images are 336. As explained before, the prepared dataset is actually divided into two sub-datasets in the proposed approach, 800 images in the training

### Number of training and test images in systems based on combined modern detection and description methods

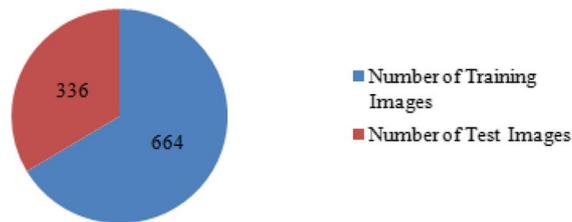


Fig. 5 Number of training and test images in [9, 10]

dataset, and 200 images for the test dataset. Figures 5 and 6 represent how the dataset is divided in the proposed system in this work and the other previously implemented systems in [9, 10].

In addition, the accuracy for each system is shown in Table 3. In the two proposed systems in [9, 10], more images have been considered as test images if compared to the proposed system based on the deep neural network. 20.0000% of the dataset are test images as shown in Table 3, and the accuracy of the recognition system is 99.5%.

For a thorough assessment of the confusion matrix, three evaluation metrics are extracted which are recall, precision, and F-score [44]. These metrics can provide useful information of the developed natural plant recognition system. The precision and recall measurements are shown in Figs. 7 and 8 where 1, 2, 3, and 4 in x-axis are Hydrangea, Amelanchier Canadensis, Acer Pseudoplatanus and Cornus, respectively. Higher values of precision and recall for a plant species mean better performance of the model for this plant species. For instance, the recall value for the Cornus is less than 1 and it equals to 0.98, whereas the recall value of Hydrangea is exactly 1 and it is the best recall value. The area under the

### Number of training and test images in proposed approach

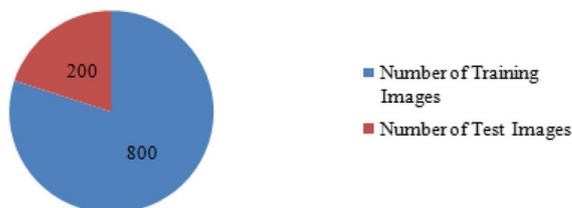
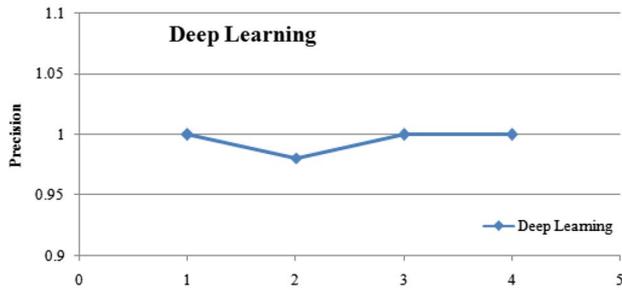


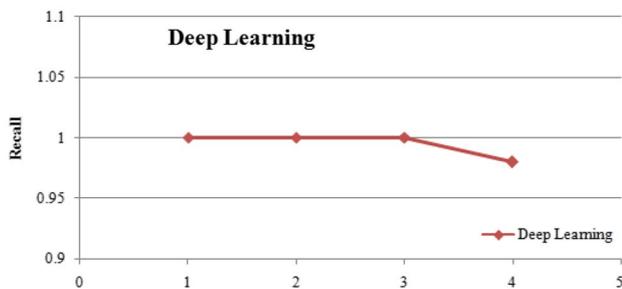
Fig. 6 Number of training and test images in the proposed approach

**Table 3** Accuracy of different systems and the number of the training and test images

Proposed system	Percentage of training images (%)	Percentage of test images (%)	Correct predictions (%)	Wrong predictions (%)
System based on the SIFT in [9]	66	34	94.9404	5.0596
System based on the SURF in [10]	66	34	93.9575	6.0403
Proposed system based on the deep neural network	80	20	99.5000	0.5000



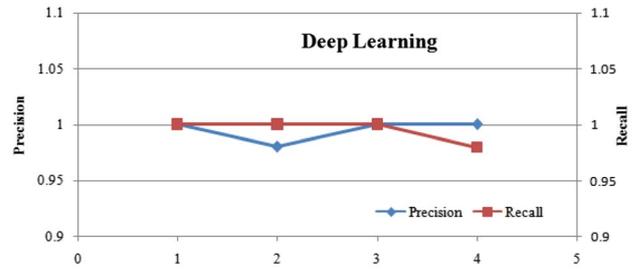
**Fig. 7** Precision measurement for the proposed system



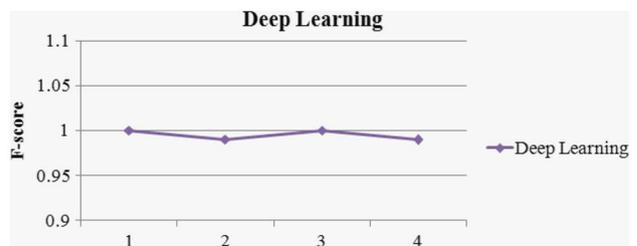
**Fig. 8** Recall measurement for the proposed system

plotted precision measurements in Fig. 7 would have maximum value if all test images were predicted and classified correctly, then this area would be equal to 3.00. As there is only one wrong prediction in the test images, this area is less than its maximum value, but it is the highest possible value by consideration of one wrong classification and the value of this area is equal to 2.98039. The area under the plotted recall measurements can also be investigated in Fig. 8. If no wrong prediction was made in the test images, the area under the plotted recall measurements would be equal to 3.00. In the performed experiment, the area under the recall measurements is equal to 2.99 and it is the highest possible value when one wrong classification has occurred.

$$precision = \frac{\text{Number of correct positive predictions}}{\text{Total number of positive predictions}} \quad (2)$$



**Fig. 9** Precision and recall measurements for the proposed system



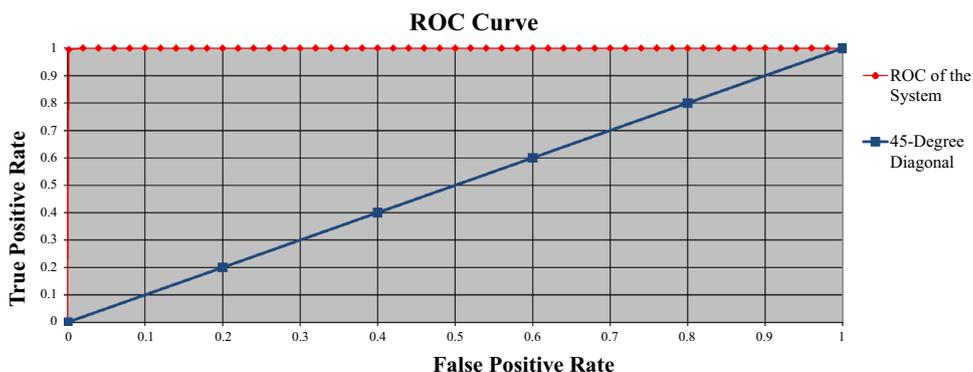
**Fig. 10** F-score measurement for the proposed system

$$recall = \frac{\text{Number of correct positive predictions}}{\text{All predictions in actual class}} \quad (3)$$

Both recall and precision measurements are shown in the same figure, Fig. 9, which helps to compare them simultaneously in one figure and investigate the behavior of the metric at the same time.

The last metric is the F-score that can be measured by two other metrics, precision and recall, and it is a harmonic mean of precision and recall measurements. The largest and smallest possible values of F-score are 1 and 0 respectively, and this measurement considers both precision and recall values; therefore, it is a function of both mentioned measurements. Figure 10 shows the F-score measurements for each plant species where 1, 2, 3, and 4 in x-axis are Hydrangea, Amelanchier Canadensis, Acer Pseudoplatanus and Cornus, respectively.

**Fig. 11** The ROC curve represented in red and the 45-degree diagonal of the ROC space in dark blue

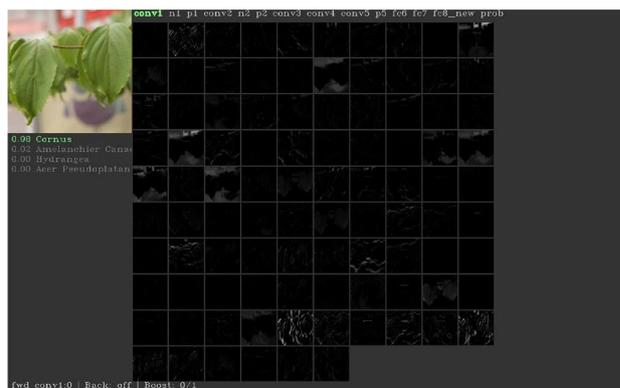


In Fig. 10, two classes, Hydrangea and Acer Pseudoplatanus, have obtained the highest possible value and the minimum value belongs to Cornus where this value is equal to 0.9899. The F-score of the Amelanchier Canadensis equals 0.9901 and its value is less than 1 because one Cornus sample has been recognized as Amelanchier Canadensis by the plant recognition system. The plant species of the Hydrangea and Acer Pseudoplatanus are more robust if they are compared to other plant species of the dataset, Cornus and Amelanchier Canadensis. The advantage of this measurement is to consider both false positive predictions and false negative predictions at the same time which gives a new sense to the extracted metrics from the confusion matrix. The definition of this metric is mathematically provided by Eq. 4.

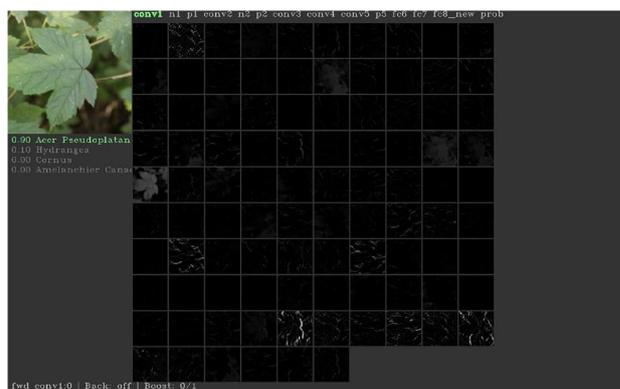
$$F - score = \frac{2 \textit{precision recall}}{\textit{Precision} + \textit{recall}} \tag{4}$$

In order to evaluate the performance of the natural plant recognition system, a new graph called the Receiver Operator Characteristic (ROC) curve [45] is plotted. This graph is based on the relationship between the true positive rate and the false positive rate [45]. In other words, it shows the trade-off between the true positive rate and the false positive rate. When the curve comes closer to the 45-degree diagonal of the ROC space, it means less accuracy. Figure 11 represents the ROC curve of the test and it is so close to an ideal classifier which is able to distinguish between the classes. The 45-degree diagonal of the ROC space is also shown in Fig. 11.

Another experiment performed is layer and output visualization. In [46], a visualization tool has been introduced for helping in the interpretation of the trained neural networks. This tool has been applied to visualize the layers of the implemented deep network and show the result visually. Additionally, it is possible to represent and compare several test images. In order to utilize the mentioned tool, the intention is to make the whole process fully-automatic, which has been performed. In Figs. 12, 13, 14 and 15, the results of four different samples are shown.



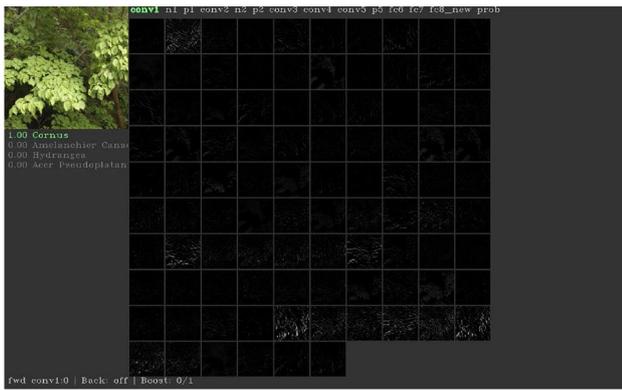
**Fig. 12** Visualization and classification example of the Cornus sample at a short distance



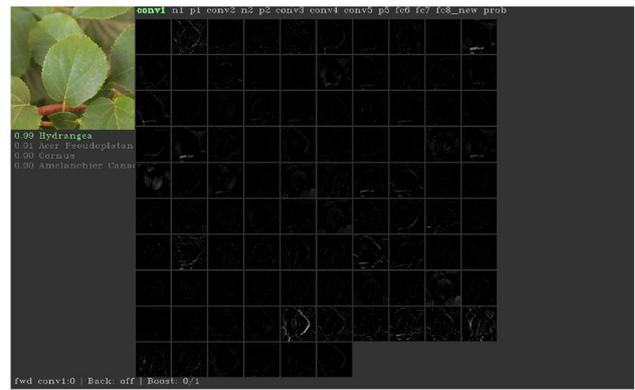
**Fig. 13** Visualization and classification example of the Acer Pseudoplatanus sample at a short distance

One of the images that has been taken in windy weather is represented as one of the examples, and Fig. 15 shows the result of this sample in windy weather.

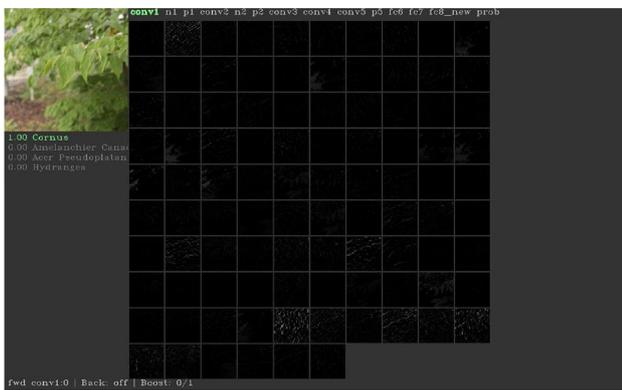
For instance, the input test image is shown on the top left corner of the Fig. 12, where the names of four plant species are under this image. The first name is Cornus and one value is written on the left side of it which is equal to 0.98. This value means the probability of being Cornus is



**Fig. 14** Visualization and classification example of the Cornus sample at a long distance



**Fig. 16** Visualization of the first convolutional layer of the proposed network



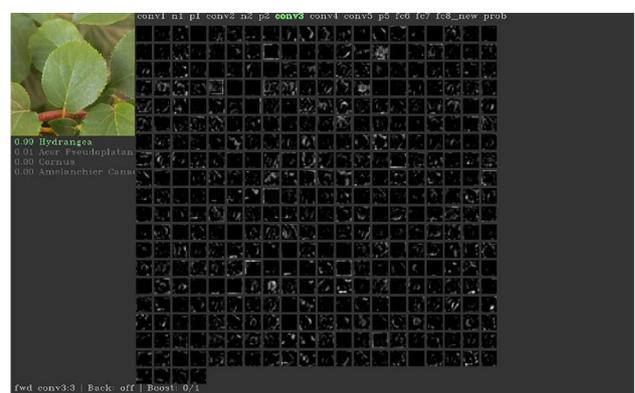
**Fig. 15** Visualization and classification example of the Cornus sample in the windy weather



**Fig. 17** Visualization of the fourth convolutional layer of the proposed network

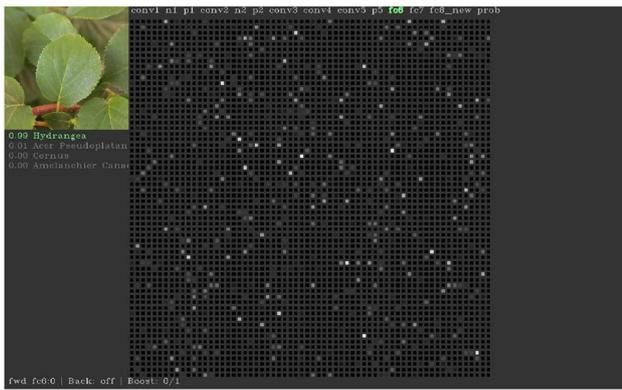
98%. The second choice is *Amelanchier Canadensis*, and the value on the left side of it is 0.02. It means that the probability of being *Amelanchier Canadensis* for the input test image is 2%. The next choices are the names of the two other plant species and their values are zero. The prediction of the input test image is visualized as discussed.

Another advantage is the possibility of visualizing the layers of the deep network and checking out the important parts of the natural plant images in each layer individually. Looking into different layers proves that the higher layers are more complex and an increase of complexity is undeniable by going through the deep model from the first convolutional layer to higher layers, like the third or fourth convolutional layer. In addition, invariant representations have increased if a lower layer is compared to a higher one. Figures 16 and 17 show the first and the fourth convolutional layers of the deep model that have been visualized by using the toolbox as a part of the system. Investigation of the fully connected layers, like the sixth layer, represents a greater increase in the variations of patterns among the lower layers, like the third convolutional

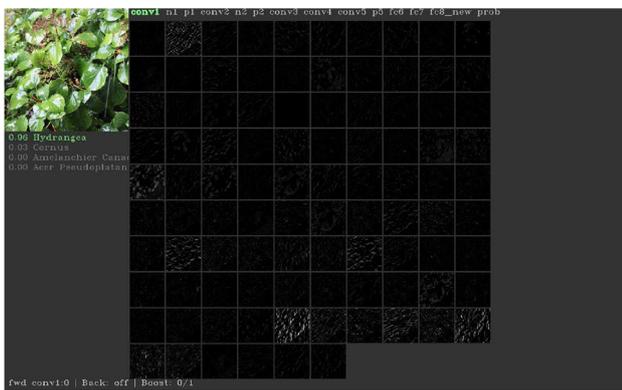


**Fig. 18** Visualization of the third convolutional layer of the proposed network

layer. Furthermore, the visualization part helps to inspect the representation of the input natural image in various layers and view the abstracted natural image in each layer. Figures 18 and 19 show different representations of layers when one layer is a convolutional layer and the



**Fig. 19** Visualization of the sixth layer of the proposed network



**Fig. 20** Visualization and classification example of the Hydrangea sample in rainy weather

other one is a fully connected layer with large variation; consequently, the recognition of the pattern of the input natural image is not possible in a fully connected layer if it is compared to lower layers, like the first and second convolutional layers.

Figure 20 shows the result of an input captured from Hydrangea in rainy weather, and the input natural image has been recognized as Hydrangea with the probability of 96%.

In order to monitor the products, it is essential to know all species which have been grown in the fields. During harvesting, it is so important to recognize plant species too. The automatic recognition of plants is a demand on modern farms and fields where time is a vital factor. To attain peak efficiency, farmers need to use advanced equipment in mobile platforms. This work contributes to natural plant recognition for real-time application and one camera can be mounted on a mobile platform. It is worth mentioning that the schematic of the mobile platforms has been drawn and the goals have been defined that would be used in the future for the real-time applications of the

proposed system for automatic recognition of plants in dynamic outdoor environments.

## 6 Conclusion

In this work, a convolutional neural network has been designed and implemented for classification of very challenging natural plant species, as different factors and natural effects have been considered according to the characteristics and properties of the images of the modern natural dataset. One of the important properties of the implemented system is its generality which makes it unique and applicable to different natural conditions in uncontrolled outdoor environments. The deep network classifies four different plant species with a high accuracy, which is equal to 99.5%. In comparison to modern combined methods and traditional machine learning approaches for natural plant recognition, the deep-learning-based system has higher accuracy with satisfactory performance. It is more efficient in different aspects, like compatibility, flexibility and generality. It is practical to use the system in natural outdoor environments and challenging conditions as a real-time system. It can use both GPU and CPU, with the possibility of switching between the CPU and the GPU. In addition, the proposed system is distance-independent and robust in uncontrolled natural environments (influenced by various natural factors, like various weather types, complex backgrounds, time of photographing, large viewpoint changes, change of light intensity, etc.), although the aforementioned systems are really useful in many cases too. Contrary to the existing literature, the proposed system in this work compensates for the gap between the previously proposed systems and the real-life applications by considering new factors such as distance, time, weather condition, etc. The proposed fully-automatic natural plant recognition system can be used in different fields such as agriculture, medicine, drugs, etc. and applied as real-time and mobile systems like robots and semi-robots. Due to the application of the proposed system in outdoor environments, the future line of work is to use this fully-automatic plant recognition system with a mobile platform on a robot or semi-robot for recognizing of plants in challenging outdoor environments.

Recently, a new work has been proposed in [47] and the idea is to build neural network with new component called capsule where new advantages like fewer parameters and viewpoint invariance have been added to the model. By considering this concept, it is feasible to design and develop a new neural network model and check whether a model with capsules and the concepts behind it can outperform the current work and improves the performance of the system. Another future line of the work could be

extending the system to be able to recognize a greater number of plant species in uncontrolled outdoor environments. Future research in this area can focus on potential inclusion of neural networks and histogram equalization techniques and make effective data augmentation approaches and algorithms to increase both number and diversity of natural images.

**Acknowledgements** The prepared modern dataset belongs to the Institute of Real-time Learning Systems, University of Siegen, Hölderlinstr. 3, 57076 Siegen, Germany.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Cortes C, Vapnik V (1995) Support vector machine. *Mach Learn* 20(3):273–297
- Quinlan J (1993) C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc, San Francisco
- Arel I, Rose DC, Karnowski TP (2010) Deep machine learning—a new frontier in artificial intelligence research [research frontier]. *IEEE Comput. Intell. Mag.* 5(4):13–18
- Manning CD, Manning CD, Schütze H (1999) Foundations of statistical natural language processing. MIT Press, Cambridge
- Chen IJ, Popovich K (2003) Understanding customer relationship management (CRM) people, process and technology. *Bus. Process Manag. J.* 9(5):672–688
- Bundesministerium für Ernährung und Landwirtschaft. <https://www.bmel.de>. Accessed 28 Nov 2018
- Fathi Kazerouni M, Schlemper J, Kuhnert KD (2015) Comparison of modern description methods for the recognition of 32 plant species. *Signal Image Process* 6(2):1–13
- Fathi Kazerouni M, Schlemper J, Kuhnert KD (2015) Efficient modern description methods by using SURF algorithm for recognition of plant species. *Adv Image Video Process* 3(2):10–24
- Fathi Kazerouni M, Schlemper J, Kuhnert KD (2017) Modern detection and description methods for natural plants recognition. *Int J Comput Electr Autom Control Inf Eng* 10(8):1497–1512
- Fathi Kazerouni M, Schlemper J, Kuhnert KD (2017) Automatic plant recognition system for challenging natural plant species. In: Short papers proceedings of 25. International conference in central europe on computer graphics, visualization and computer vision WSCG 2017, pp 81–89
- Abbasi S, Mokhtarian F, Kittler J (1997) Reliable classification of chrysanthemum leaves through curvature scale space. In: *Scale-Space Theory in Computer Vision*, pp 284–295
- Lee CL, Chen SY (2006) Classification of leaf images. *Int J Imaging Syst Technol* 16(1):15–23
- Lowe DG (1999) Object recognition from local scale-invariant features. In: *The proceedings of the seventh IEEE international conference on computer vision*, vol 2. IEEE, pp 1150–1157
- Rosten E, Drummond T (2006) Machine learning for high-speed corner detection. In: *European conference on computer vision*. Springer, Berlin, pp 430–443
- Harris C, Stephens M (1988) A combined corner and edge detector. In: *Alvey vision conference*, vol 15, no 50, pp 10–5244
- Bay H, Tuytelaars T, Van Gool L (2006) Surf: speeded up robust features. In *European conference on computer vision*. Springer, Berlin, pp 404–417
- Kebapci H, Yanikoglu B, Unal G (2010) Plant image retrieval using color, shape and texture features. *Comput J* 54(9):1475–1490
- Sakai N, Yonekawa S, Matsuzaki A, Morishima H (1996) Two-dimensional image analysis of the shape of rice and its application to separating varieties. *J Food Eng* 27(4):397–407
- Gwo CY, Wei CH (2013) Plant identification through images: using feature extraction of key points on leaf contours. *Appl Plant Sci* 1(11):1–9
- Wu SG, Bao FS, Xu EY, Wang YX, Chang YF, Xiang QL (2007) A leaf recognition algorithm for plant classification using probabilistic neural network. In: *2007 IEEE international symposium on signal processing and information technology*. IEEE, pp 11–16
- Gaber T, Tharwat A, Snasel V, Hassanien AE (2015) Plant identification: two dimensional-based vs. one dimensional-based feature extraction methods. In: *10th international conference on soft computing models in industrial and environmental applications*. Springer, Switzerland, pp 375–385
- Fisher RA (1938) The statistical utilization of multiple measurements. *Ann Eugen* 8(4):376–386
- Fukunaga K (2013) Introduction to statistical pattern recognition. Elsevier, Amsterdam
- Zhang S, Chau KW (2009) Dimension reduction using semi-supervised locally linear embedding for plant leaf classification. In: *International conference on intelligent computing*. Springer, Berlin, pp 948–955
- Chuntian C, Chau KW (2002) Three-person multi-objective conflict decision in reservoir flood control. *Eur J Oper Res* 142(3):625–631
- Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *J Hydrol* 399(3–4):394–409
- Moazenzadeh R, Mohammadi B, Shamshirband S, Chau KW (2018) Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Eng Appl Comput Fluid Mech* 12(1):584–597
- Samadianfard S, Majnooni-Heris A, Qasem SN, Kisi O, Shamshirband S, Chau KW (2019) Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate. *Eng Appl Comput Fluid Mech* 13(1):142–157
- Yaseen ZM, Sulaiman SO, Deo RC, Chau KW (2019) An enhanced extreme learning machine model for river flow forecasting: state-of-the-art, practical applications in water resource engineering area and future research direction. *J Hydrol* 569:387–408
- Baghban A, Jalali A, Shafiee M, Ahmadi MH, Chau KW (2019) Developing an ANFIS-based swarm concept model for estimating the relative viscosity of nanofluids. *Eng Appl Comput Fluid Mech* 13(1):26–39
- Modern natural plants dataset of the institute of real-time learning systems. University of Siegen, Siegen. <http://www.eti.uni-siegen.de/ezls>. Accessed 1 Dec 2018
- Söderkvist O (2001) Computer vision classification of leaves from swedish trees
- Lab L (2017) Convolutional neural networks (lenet) deeplearning 0.1 documentation. <http://deeplearning.net/tutorial/lenet.html>. Accessed 3 Dec 2018
- LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 1097–1105

36. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
37. Nair V, Hinton GE (2010) Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 807–814
38. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR (2012) Improving neural networks by preventing co-adaptation of feature detectors. arXiv Preprint [arXiv:1207.0580](https://arxiv.org/abs/1207.0580), pp 1–18
39. Bell RM, Koren Y (2007) Lessons from the Netflix prize challenge. *Acm Sigkdd Explor Newsl* 9(2):75–79
40. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
41. Bernaschi M, Di Lallo A, Fulcoli R, Gallo E, Timmoneri L (2011) Combined use of graphics processing unit (GPU) and central processing unit (CPU) for passive radar signal and data elaboration. In: 2011 proceedings international radar symposium (IRS), pp 315–320
42. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on multimedia, pp 675–678
43. Provost F, Kohavi R (1998) Glossary of terms. *J Mach Learn* 30(2–3):271–274
44. Sasaki Y (2007) The truth of the F measure. *Teach Tutor Mater* 1(5):1–5
45. Hanley JA (2014) Receiver operating characteristic (ROC) curves. Wiley StatsRef, Statistics Reference Online
46. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H (2015) Understanding neural networks through deep visualization. arXiv Preprint [arXiv:1506.06579](https://arxiv.org/abs/1506.06579), pp 1–12
47. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. In: Advances in neural information processing systems, pp 3856–3866

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.