A modern deep learning framework in robot vision for automated bean leaves diseases detection

Sudad H. Abed¹ · Alaa S. Al-Waisy² · Hussam J. Mohammed¹ · Shumoos Al-Fahdawi³

Received: 24 September 2020 / Accepted: 17 April 2021 / Published online: 30 April 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2021

Abstract

The bean leaves can be affected by several diseases, such as angular leaf spots and bean rust, which can cause big damage to bean crops and decrease their productivity. Thus, treating these diseases in their early stages can improve the quality and quantity of the product. Recently, several robotic frameworks based on image processing and artificial intelligence have been used to treat these diseases in an automated way. However, incorrect diagnosis of the infected leaf can lead to the use of chemical treatments for normal leaf thereby the issue will not be solved, and the process may be costly and harmful. To overcome these issues, a modern deep learning framework in robot vision for the early detection of bean leaves diseases is proposed. The proposed framework is composed of two primary stages, which detect the bean leaves in the input images and diagnosing the diseases within the detected leaves. The U-Net architecture based on a pre-trained ResNet34 encoder is employed for detecting the bean leaves in the input images captured in uncontrolled environmental conditions. In the classification stage, the performance of five diverse deep learning models (e.g., Densenet121, ResNet34, ResNet50, VGG-16, and VGG-19) is assessed accurately to identify the healthiness of bean leaves. The performance of the proposed framework is evaluated using a challenging and extensive dataset composed of 1295 images of three different classes (e.g., Healthy, Angular Leaf Spot, and Bean Rust). In the binary classification task, the best performance is achieved using the Densenet121 model with a CAR of 98.31%, Sensitivity of 99.03%, Specificity of 96.82%, Precision of 98.45%, F1-Score of 98.74%, and AUC of 100%. The higher CAR of 91.01% is obtained using the same model in the multi-classification task, with less than 2 s per image to produce the final decision.

Keywords Deep learning · U-Net architecture · Bean leaves diseases · Transfer learning · Robot vision · ResNet34 model

1 Introduction

Agriculture is considered as one of the most important economic resources for farmers and countries. The beans are one of the essential agricultural crops which provide dietary fibre as the second significant source and the third important source of calories for human beings (Pamela et al. 2014). However, these crops suffer from different kinds of diseases

such as angular leaf spots and bean rust, which can cause big damage to bean crops and decrease their productivity. To date, the crops diseases detection technique is achieved using the human's eyes by the experts to observe and identify them (Singh and Misra 2017). This is a very tedious, time-consuming, and costly technique that requires many experts as well as additional efforts. Furthermore, it is not easy to monitor and treat crops in large fields. In many countries, a large number of farmers lack the facilities and knowledge to deal with crop diseases as well as having difficulties in finding and consulting experts (More et al. 2016). This has resulted in the introduction to further innovative technological methods and materials (e.g., robotics) for the early diagnosis of crop diseases and to overcome some of the inherent limitations of traditional techniques. This can play a significant role in the early diagnosis and treatment of the crops symptoms in an automated way with less effort and cost.

✉ Sudad H. Abed
shabed@uoanbar.edu.iq

¹ Computer Center, University Of Anbar, 11, Ramadi, Anbar, Iraq

² Computer Technologies Engineering Department, Information Technology Collage, Imam Ja'afar Al-Sadiq University, Baghdad, Iraq

³ Computer Science Department, Al-Ma'aref University College, Ramadi, Anbar, Iraq

Robotics have been deployed in agriculture to increase productivity and decrease labour costs. Besides, they can aid to increase the efficiency in treating many crop diseases including the bean crops diseases by targeting the infected crops only (Larese et al. 2013). This can be achieved using image processing and computer vision approaches to classify and identify the normal from the infected leaf. The key step of computer vision to diagnose leaf diseases is to take clear and high-resolution images. There are several factors associated with the image acquisition process that can significantly affect the quality of the captured images including complex weather conditions, noisy background, shooting angle, etc. (Zhang et al. 2018). This can also affect the accuracy and outcome of the other subsequent stages in the automatic detection system, such as image segmentation and classification stages.

In this regard, various methods of unsupervised and supervised deep learning approaches have effectively been utilized to address the image segmentation and classification tasks. Convolutional Neural Network (CNN) is one of the most common supervised deep learning approaches, which has been applied successfully to solve many challenging computer vision tasks [e.g., face recognition (Al-Waisy et al. 2018a) iris recognition (Al-Waisy et al. 2018b), COVID-19 detection (Hemdan et al. 2020), etc.] due to its ability to learn high-level features representations from the input image. Besides, CNNs have recently outperformed the performance of humans and other methods in both segmentation and classification tasks (Al-Waisy et al. 2018b). In the segmentation process, various studies have utilized U-Net architecture due to its ability to obtain a high image segmentation accuracy by correctly predicting the class of each pixel in the input image. U-Net is a fully convolutional network, which consists of two parts including contracting and expanding to produce an accurate image segmentation, especially in the medical imaging field (Liu et al. 2020). On the other hand, many CNNs models with different architectures have been developed to address the image classification task. For instance, the residual neural networks (ResNets) produce a great performance using less number of hyperparameters along with an elegant design of stacking the CNN's layers (He et al. 2020).

In this paper, a modern fully-automated and fast robotic perception framework is proposed to diagnosis the bean leaves diseases, and classify them into either healthy or unhealthy leaves. Firstly, the U-Net based ResNet model is applied to accurately and automatically detect the bean leaves from an input image captured in uncontrolled environmental conditions. Once the bean leaves are detected the features extraction and classification steps are applied, respectively, to classify the leaves of the beans into either healthy or unhealthy leaves. The primary contributions of this research can be outlined as follows:

1. A modern fully-automated and fast robot perception framework is proposed to diagnose the healthiness status of the bean crops by classifying them into different classes (e.g., healthy, angular leaf spot, and bean rust). Herein, an efficient deep learning model for fast and accurate segmentation of bean leaves is proposed. This work proves that the accuracy of the U-Net architecture can be enhanced using the pre-trained ResNet34 encoder. Unlike, other manual and semi-automatic segmentation approaches, there is no need for user intervention by using the proposed image segmentation model. To the best of the authors' knowledge, this is the first work that investigates the potential of using the U-Net architecture to detect the leaf of the bean images that were captured in uncontrolled environmental conditions.
2. The performance of most popular deep convolutional feature extraction and classification models is developed to extract highly discriminative feature representations and classify the bean leaves into different classes (e.g., healthy, angular leaf spot, and bean rust). Using deep learning features descriptors can significantly reduce the demand for manual extraction of features due to their ability to automatically learn the essential feature representations in a given image. To the best of the authors' knowledge, this is the first attempt to introduce such a solution for detecting the infected bean leaves using various deep learning models.
3. The efficiency and reliability of the proposed approaches are demonstrated using a very challenging dataset, in which all the images of the bean leaves are captured in uncontrolled environmental conditions. Furthermore, the possibility of using the proposed system in a real-world robot application for early diagnosis of the infected leaves of the bean has been demonstrated where less than 2 s per image are required to produce the final decision.

The rest of the paper is organized as follows: Sect. 2 highlights related work in the area of leaf diseases segmentation and classification. Section 3 explains the proposed methodology. Section 4 presents the experimental results of the different tests undertaken to evaluate the proposed approach. Section 5 discusses the findings and possible future work. Finally, Sect. 6 contains the conclusions and the directions of the future research.

2 Related work

Recently, several efforts by researchers have been presented to cope with the crop's disease identification issues aiming to propose an ideal solution that can help to identify the infected leaf automatically. Therefore, several approaches

have been proposed using machine learning methods. For instance, Muthukannan et al. (Muthukannan et al. 2015) proposed a framework to classify the crops depending on their disease using neural network methods, such as Learning Vector Quantization (LVQ), Feed Forward Neural Network (FFNN), and Radial Basis Function Networks (RBF). The performance of this framework was evaluated by collecting images and extracting features of two types of leaves, such as Bean and Bitter gourd. Afterward, the neural network algorithms were utilized independently to classify the type of disease. Their experiments showed that the FFNN method recorded the best accuracy that reached about 90%. However, the dataset used is too small to evaluate the proposed method properly where the number of samples was 118 only. Ramesh et al. (Ramesh and Vydeki 2019) suggested an approach by optimizing Deep Neural Network (DNN) with Jaya algorithm (DNN-JOA) to classify paddy leaf diseases. This approach consists of several phases including image pre-processing, image segmentation, and classification. In the image pre-processing step phase, the size of the input image was reduced and a background subtraction procedure was applied. The K-means clustering method was utilized to separate and segment the health part from the unhealthy part. Finally, the optimized DNN-JOA approach was used to classify the unhealthy parts. Their experiment was conducted using a dataset of rice plant leaf images, which contains 650 images for four different diseases. The obtained results showed that the best accuracy rate of 95% across four classes of diseases was achieved. Although the classification task depends completely on the segmentation part, there is not any experiment to assess the accuracy of the proposed image segmentation procedure. Besides, the number of images within the dataset for each class was relatively small to reveal the real performance of the proposed approaches. Mohanty et al. (Mohanty et al. 2016) developed a smartphone-assisted disease framework using deep learning approaches to classify and diagnose crop diseases. Their approach utilized AlexNet and GoogLeNet as CNN models. They demonstrated that these two models can handle and solve large scale image recognition issues efficiently. The PlantVillage dataset, which comprises 54,306 images of 26 diseases in 14 crop species, was employed to validate the performance of the proposed framework. The experimental results showed that the proposed framework can classify the crops diseases with a very promising accuracy rate of 99%. However, the performance of the proposed framework was only tested using the PlantVillage dataset, in which all the images were captured in controlled environmental conditions.

Ruchita and Bhosale (More et al. 2016) developed a new approach using an Agrobot robot to detect leaf diseases in an automated way. This approach was divided into two stages including the detection of leaf plant and the diagnosis of

the diseases within the detected leaf. The first stage was conducted by detecting the leaf within the input images, which were collected from the real farms. The quality of the collected images was enhanced and a set of discriminative features were extracted and used as input vectors to artificial neural network (ANN) to detect the leaf in the input image. This is followed by applying the K-means algorithm to detect the infected area within each detected leaf. Afterward, the ANN algorithm was used to identify the leaf diseases in their early stage. However, there were no experimental results were conducted to evaluate and validate the proposed approaches. In a similar context, Olsen et al. (2019) contributed a large dataset of weed species images from the Australian rangelands, named the DeepWeeds dataset, which contains 17,509 labelled images of eight different classes (e.g., Chinese Apple, Lantana, Parkinsonia, Parthenium, Prickly Acacia, Rubber vine, Siam Weed, and Snake Weed). They highlighted that the dataset can be used to develop robust robotic frameworks to control the weed. The performance of two different deep learning models (e.g., Inception-v3 and ResNet-50) was tested using the DeepWeeds dataset to classify the input images into eight different classes. Their experiments showed an accuracy rate of 95.1% and 95.7% which are achieved by using the Inception-v3 and ResNet-50 model, respectively.

Throughout this review, one can see that several comprehensive frameworks were developed to tackle and address the leaf disease identification task. In most cases, image processing and computer vision techniques play an essential role in increasing the productivity of the crops by identifying and treating the disease in the early stages. Furthermore, the DNNs prove their ability to detect the region of the leaf diseases and classify them into the normal or abnormal class. These techniques can be compatible with robotics to achieve their purpose in the agriculture field. However, most of the presented studies apply their approaches using datasets composed of images were collected under controlled environmental conditions and may not reveal the actual performance of the proposed approaches.

2.1 The proposed methodology

The proposed framework aims to classify the leaves of the bean into healthy and unhealthy based on the images provided by the camera of a robot. Then actions are taken depending on the result obtained. The block diagram of the proposed system is depicted in Fig. 1. To identify the unhealthy leaves and treat them, the robot must be supported by a camera that captures images of the plant to be used as input to the proposed system and make a decision. In other words, the first step is collecting labelled images of both healthy and unhealthy bean leaves plants. This is followed by detecting and isolating the leaves of the bean

Fig. 1 The proposed method of identifying plants' healthiness and taking action

from the background and other surrounding objects, such as the leaf for other plants, stones, ground, etc. This is due to the symptoms and color of unhealthy leaves which can be similar to the background and other surrounded objects as the images are taken in a real field, and thereby it might negatively affect the performance of the classification model. Thus, the classification model should be targeting the leaf in the images only. Once, the leaf is detected, it is used as input to the classification model to identify the healthiness status of the plant.

2.2 Image segmentation

Image segmentation is a pixel-based classification task where the model tries to assign each pixel in the image to its corresponding class label. Recently, DNNs have been used to address many challenging image segmentation problems and have given outstanding results (Iglovikov and Shvets 2020; Falk et al. 2019). For instance, the U-Net architecture was introduced in 2015 for segmenting biomedical images and achieved a good performance winning on the ISBI challenge (Ronneberger et al. 2020). Compared with other CNNs, the main architecture of the U-Net was created to work on a few training samples with a more accurate segmentation outcome. The U-Net architecture is split into two main parts forming the U shape. The left part of the model is called the contraction path, which represents the encoder, and the asymmetric right part is called the expansion path, which represents the decoder (Ronneberger et al. 2020). The encoder follows the typical structure of a CNN architecture and composes five convolutional blocks. Each convolutional block includes two convolutional layers with a kernel of size (3×3) pixels. This is followed by a rectified linear unit (ReLU) and max-pooling layer with a stride of 2 that are implemented at the end of each block except the last block, which is intended for down-sampling the size of the special data. Through the contraction path, the number of feature channels is increased by double at each convolutional block. In the expansion path (decoder), an up-sampling process is taking place in every block that is followed by (2×2)

up-convolution layer to reduce the number of features maps, the concatenation process with the symmetrically high-resolution and discriminative cropped feature map from the contraction path (via skip-connections), and two (3×3) up-convolutions layers, each followed by applying a ReLU function. At the last layer, a (1×1) pixels, convolution is performed to map every vector of feature to the anticipated number of classes. The final output of the U-Net architecture is a pixel-by-pixel mask, which assigns a class label for each pixel in the input image.

Typically, to train the U-Net from scratch using randomly initialized weights requires an extremely large-scale dataset that contains millions of images to avert the overfitting issue and achieve satisfactory performance. Therefore, the weight configurations of different DNNs, that are trained on the ImageNet dataset (Russakovsky 2015) are extensively employed recently as a beginning step to initialize the weights of the U-Net and address many image segmentation tasks.

In this work, a U-Net based on ResNet34 model, where pre-trained by employing the ImageNet dataset, is used as an encoder to address the bean leaves segmentation task. This is to accelerate the network convergence and enhance the performance of the U-Net performance (see Fig. 2). The ResNet is one of the most dominant DNNs based on the CNNs structure and the champion of the ILSVRC ImageNet classification competition in 2015 (Russakovsky 2015). Similar to other CNN models, the ResNet34 is composed of convolution layers, pooling layers, and fully connected layers, which are stacked sequentially on top of each other. However, it differs from typical CNN models where an identity connection is originated from the input layer to the end of each residual block, as shown in Fig. 3b.

The encoder of the proposed U-Net receives a colored image of size (384×384) pixels and generates a 1-channel binary image of the same size as the input image. The initial block of the ResNet34 encoder starts by applying a convolutional layer with a kernel of size (7×7) pixels and stride 2, a Batch Normalization (BN) layer, ReLU layer, and a max-pooling layer with a stride of 2 to reduce the

Fig. 2 The proposed bean leaves segmentation model using the U-Net architecture using pre-trained ResNet34 encoder

Fig. 3 The difference in building neural blocks: **a** a typical block, and **b** a residual block

size of the input image. Then, repetitive residual blocks of the ResNet34 model are applied, as presented by the green blocks in Fig. 2. Within each residual block, the first convolutional layer is performed with stride 2 to reduce the size of the input data, while the rest convolutional layers use stride 1. Herein, the average pooling and the fully-connected layers are removed from the end of the ResNet34 model and replaced with the decoder of the U-Net. To avoid a rapid decline in the size of the input data while moving toward the lowest layers, a zero-padding of 1-pixel value is applied at each convolutional layer.

The decoder is composed of several decoder blocks to perform up-sampling, which are represented in purple color as illustrated in Fig. 2. The input to each decoder block is a channel-wise fused tensor of both the output of the previous decoder block and the activation map that is produced from the corresponding encoder block. These operations are represented by the horizontal arrows connecting the yellow blocks on both sides. In the decoder path, the size of the produced activation maps is increased to double while the

number of feature channels is reduced to half. Each decoder block comprises the BN layer, the ReLU layer, and transpose convolutional layer with a kernel size of (2×2) pixels and a stride of 2 to perform the up-sampling process. During the training process, 80% of bean leaves dataset images are utilized along with their annotated binary masks as a training set, the rest 20% images are used for the testing set. Then, 20% of the training set is randomly chosen and utilized as a validation set to evaluate the generalization knowledge of the last trained model and save the weight configurations that give a minimum error rate. The parameters of the network are trained using Adam optimizer (Kingma and Lei Ba 2020) for 20 epochs with a mini-batch size of 10, a weight decay value of $1e-2$, a learning rate set to be $1e-5$, and a momentum value of 0.9. In this work, a data augmentation procedure is also applied to increase the number of images in the dataset artificially. This can prevent the overfitting

problem and achieve better generalization during the learning process efficiently. Firstly, the size of the original bean images is fixed to (384×384) pixels, and then two images are produced from each resized image by applying horizontal and vertical flipping. This is followed by rotating both the original images and their horizontal and vertical flipping versions 45 degrees clockwise and counter-clockwise. Therefore, each original image in the dataset is expanded into eight different images. The proposed procedure is implemented after splitting the dataset into two commonly independent sets (e.g., training and testing to prevent tendentious outcomes. The main reason for applying the data augmentation to the testing set is to examine and validate the model generalization ability to new unseen data using more image diversity. Figure 4 shows the output of the proposed data augmentation procedure for both the raw images and the mask images (ground truth).

Fig. 4 The proposed data augmentation procedure: **a** the original raw and ground truth image, **b** the eight generated images from both the raw and ground truth image

2.3 Image classification

Transfer learning is a widely applied strategy to leverage knowledge previously learned to address a certain problem in a particular field and to solve a new problem in another related field (Tan et al. 2020). Several studies demonstrate the impact of applying transfer learning in enhancing the capability of DNNs to gain additional discriminative representation learning and tackle various arduous computer vision tasks (Mehra 2018). It empowers the DNNs to detect common feature representations (e.g., edges, lines, curves, etc.) from the large-scale datasets of the previous task that might not be detected because of the insufficient number of the training samples in the present task. In this study, the performance of the current state-of-the-art CNN architectures is assessed for identifying the abnormalities in the bean leaves. Herein, the weight configurations of five different pre-trained CNN models (e.g., Densenet121, ResNet34, ResNet50, VGG-16, and VGG-19), which are learned on a large-scale ImageNet dataset and transferred into the current task instead of training each model from scratch due to the limited size of the employed bean leaves dataset. In the learning process, all the weights are initialized with the weight configurations of the pre-trained models, and then they are optimized over the current task using the Adam optimizer. The main idea is that the weight configurations of the pre-trained models are already learned on how to detect powerful and discriminative features (e.g., curves, edges, lines, etc.). Significantly, it can reduce the training time, enhance the model generalization ability, and prevent overfitting. In this work, the proposed deep CNN models are tested in two different scenarios. The first scenario classifies the input bean leaves image into either healthy or unhealthy, while the second scenario classifies the input bean leaves image into one of three different classes, namely, healthy class, angular leaf spot class, and bean rust class. Through the learning process, the adopted deep CNN models were trained using 60% randomly selected samples from the dataset while the remaining 40% is equally divided into validation and testing set, which are used to monitor the performance of the CNN model during the training process and report the final performance, respectively. The same data augmentation procedure described in the previous section, was applied for each individual set (e.g., training, validation, testing set). The main steps of the proposed training procedure can be summarized as follows:

1. Dividing the dataset into three different sets (e.g., Training set, Validation set, and Testing set).
2. Determining the hyper-parameters starting values (e.g., learning rate, number of epochs, weight decay, etc.).
3. Using the starting values from step 2 to train the network.

4. Evaluating the network performance by utilizing the validation set through the learning process.
5. Iterating on steps 3 and 4 for 20 epochs.
6. Choosing the best-trained model with a less error rate on the validation set.
7. Finding the actual performance of the trained model by employing the testing set.

3 Experimental results

In this section, several comprehensive experiments are conducted to demonstrate the effectiveness of the proposed image segmentation and classification approaches. Firstly, a brief description of the employed beans dataset in these experiments is given. Secondly, a detailed evaluation of the proposed image segmentation model is presented and compared with the ground truth images. Finally, an extensive comparison study among five different DNNs is performed to address both the binary and multi-classification bean leaves diseases tasks. The code of the proposed bean leaves diseases classification system is executed with the Python programming language and Google Colab server is used as a running platform with 69K GPU graphics card, and 16 GB of RAM on the Windows 10 operating system, Intel(R) Core(TM) i7-4510U GHz CPU.

3.1 Dataset description

Although the need for accessing a large number of images for bean leaves diseases is essential to make the entire experiment more reliable, the availability of real datasets is very limited. Thus, only one real and challenging bean leaves dataset is used in this study to examine the ability of the proposed framework in accurately identifying the healthiness of the bean leaves. This dataset is created by the Makerere AI research lab and released on Jan 20, 2020 (Lab 2020). The main aim of this dataset is to develop an accurate machine learning model, which can differentiate between diseases in the bean plants. The dataset consists of bean leaves images, which are taken in the field using a smartphone camera with an image size of (500×500) pixels and stored in JPG format. The images in this dataset are annotated by experts from the National Crops Resources Research Institute (NaCRRI) in Uganda and classified into three different classes namely, Healthy class, Angular Leaf Spot class, and Bean Rust class, as shown in Table 1 (Lab 2020). Some samples of the healthy and unhealthy classes are depicted in Fig. 5.

3.2 Image segmentation evaluation

As mentioned before, all the images provided by the bean dataset are captured in the real fields under uncontrolled

Table 1 Number of images per class in the Bean leaves dataset

Healthy class	Unhealthy classes		Total
	Angular leaf spot	Bean rust	
428	432	436	1295

environmental conditions. Therefore, in most cases, the images hold undesired details that can negatively and directly affect the accuracy of the classification model. As presented in Fig. 5, it is clear that the soil portion in the image is not related to the classification problem. Besides, several images are taken at a particular angle where various leaves belonging to multiple plants are presented within a single image. Hence, the bean leaves need to be detected and isolated from the background and other surrounding extraneous features before applying the image classification process. Firstly, a corresponding version of the dataset that contains the ground truth images is constructed by extracting the targeted bean leaves in the images and discarding the other details (e.g., background, other plant leaves, etc.). The main purpose of this contracted dataset is to be used in developing an automatic

and fast image segmentation model and assessing its accuracy in detecting the bean leaves.

In this work, an efficient image segmentation method by Rother et al. (2004), named the GrabCut algorithm is used to build the corresponding ground truth dataset by effectively isolating the bean leaves in the given image from the background and other undesired details. The GrabCut algorithm is essentially based on the graph cuts technique where the user needs to draw a bounding box around the foreground objects to be segmented within a given image. Then, the color distribution of the foreground object is estimated using the Gaussian Mixture Model (GMM). According to the data of a given image, the GMM learns and produces labels for all the unknown pixels in the image where each pixel is labelled either foreground or background pixel in terms of color statistics (Basavaprasad and Hegadi 2014).

As shown in Fig. 6, the GrabCut algorithm considers the image as a graph where the vertices represent the pixels of the image; the edges between the vertices represent the feature connection between these pixels (Basavaprasad and Hegadi 2014). The GrabCut algorithm iterates on the image pixels and cuts the weak connection between the pixels and assign them to either the foreground or background. The

Fig. 5 Some Samples of the bean leaves images dataset

Fig. 6 A segmented graph into foreground and background using GrabCut algorithm (Basavaprasad and Hegadi 2014)

determined bounding box can significantly affect the algorithm accuracy since the area outside the box holds the most features of the background. If the features of a selected pixel inside the bounding box are close to the features of the pixels out of the bounding box, the selected pixel is considered belonging to the background. Otherwise, the selected pixel is considered a part of the targeted foreground object to be extracted. Herein, the bounding box is drawn and fed into the GrabCut algorithm using fixed coordinates (X, Y, H, W),

which are set to be (50, 50, 400, 400). These coordinates are applied to all the images in the beans dataset to form a rectangle that contains the targeted foreground object (bean leaves). All the segmented images obtained from the GrabCut algorithm are checked and corrected manually in a few cases to ensure that the bean leaves are extracted properly. Figure 7b shows the generated ground truth images generated using the GrabCut algorithm.

In this section, several experiments are conducted to test the accuracy of the bean leaves segmentation model. Initially, the performance of the bean leaves segmentation model is tested against the ground truth images that are generated from the GrabCut algorithm (see Fig. 7b, c). This experiment follows the same evaluation protocol that is described in (Al-Fahdawi 2018). The performance of the proposed automatic segmentation model is evaluated by calculating seven quantitative performance measures which are: Probabilistic Rand Index (PRI) (Kaur et al. 2012), Structural SIMilarity (SSIM) Index (Wang et al. 2004), gradient magnitude similarity deviation (GMSD) (Xue et al. 2013), variation of information (VoI) (Meilă 2007), mean square error (MSE), normalized absolute error (NAE) (Mallikarjuna et al. 2016) and global consistency error (GCE) (Martin et al. 2001). These are the most common seven metrics used to evaluate the performance of most image segmentation approaches in the literature. As illustrated in Fig. 8, a compatible performance is achieved using the automatic segmentation model with corresponding ground truth images. For instance, a very good similarity with ground truth images has noticeably been achieved with 0.9937, and 1.0 using

Fig. 7 Some examples of the manual and automatic bean leaves segmentation output: **a** raw images, **b** generated ground truth image using the GrabCut algorithm, **c** the binary version of the ground truth image, and **d** the output of the proposed automatic segmentation model

Fig. 8 Descriptive statistics of the segmentation model performance on Beans dataset, where a higher value of PRI and SSIM is better and a lower value of GMSD, Vol MSE, NAE and GCE is better

the PRI and SSIM metrics, respectively. On the other hand, the proposed automatic segmentation model manages to achieve a distance score of 0.0809, 0.0451, 0.0565, 0.0225, and 0.0215 using the GMSD, Vol MSE, NAE, and GCE, respectively.

In the second experiment, the automatic estimations of two geometrical measures, namely the mean leaf area (MLA) and mean leaf perimeter (MLP) are computed from automatically detected bean leaves and compared with reference values of the same geometrical measures. These values are estimated from the binary image generated from the GrabCut algorithm (see Fig. 7c). Table 2 shows the overall mean, standard deviation (SD), max, and a min of each geometrical measure for both manual and automated bean leaves images, along with the difference (Diff) and the percentage difference (Diff %) between them. As shown in Table 1, an excellent agreement is obtained using the proposed

automatic segmentation model with reference values of both geometrical measures. The Diff values between manual and automatic calculations are less than 5.5 and 8.5 for MLA and MLP, respectively. At the same time, the percentage difference values between manual and automatic estimations are less than 0.006% and 0.5 for MLA and MLP, respectively, with no geometrical measure with a proportional difference ($> 0.5\%$) between the manual and automatic estimations. Finally, the Bland–Altman analysis is performed between differences against means for these two geometrical measures to further demonstrate the agreement between manually and automatically estimated MLA and MLP measure. As shown in Fig. 9, the proposed segmentation model is managed to achieve a very good agreement where more than 95% of the data fell between 2-SD agreements borders. The present findings are very promising and representing an excellent initial step toward applying the proposed automatic segmentation model in real robotic application to solve the automatic detection for the bean leaves under uncontrolled environmental conditions of the fields.

In this work, an experiment is conducted to demonstrate the usefulness and the significant effects of detecting and isolating the leaf of the beans from the background and other surrounding extraneous features before applying the image classification step. As shown in Table 3, a comparison study is conducted to assess the performance of the ResNet34 model in identifying the healthiness status of the beans leaf with and without applying the image segmentation step. In this experiment, the same training procedure described in (Sect. 3.2) is employed to train the ResNet34 model, and its performance in addressing the binary image classification task was evaluated by computing the average values of six quantitative measures described in (Sect. 4.3). From Table 3, one can see that by training the ResNet34 model on the top of the whole image without detecting and isolating the leaf of the beans

Table 2 Performance comparison conducted between the manual and automatic estimations of two different parameters

	Manual measurement	Automatic measurement		
	MLA (μm^2)	MLA (μm^2)	Diff	Diff %
Average	100,509.5	100,514.8	5.33	0.005303
STD	19,417.82	19,417.48	−0.34	−0.00175
Max	136,898	136,900	2	0.001461
Min	11,269	11,280	11	0.097565
	MLP (μm)	MLP (μm)	Diff	Diff %
	MLP (μm)	MLP (μm)	Diff	Diff %
Average	1628.17	1636.17	8	0.490145
STD	441.13	441.11	−0.02	−0.00453
Max	3108	3118	10	0.321234
Min	420	423	3	0.711744

The (Diff) refers to the differences between the manual and automatic estimations, while the (Diff %) refers to the percentage difference

Fig. 9 Bland–Altman plots showing difference versus mean for each pair of manual and automatic estimations: **a** leaf area, and **b** leaf perimeter

Table 3 Performance comparison to assess the performance of the ResNet34 model in identifying the healthiness status of the bean leaves with and without applying the image segmentation step

Quantitative measures	With image segmentation	Without image segmentation
CAR	97.27	58.78
Sensitivity (recall)	98.64	60.14
Specificity	94.44	55.35
Precision	97.32	58.92
F1-Score	97.97	61.19
AUC	99.49	59.99

from the background and other surrounding objects very poor results are obtained with a CAR of 58.78%, Sensitivity of 60.14%, Specificity of 55.35%, Precision of 58.92%, F1-Score of 61.19%, and AUC of 59.99%. The main reason is due to the undesired details presented in the original images that can negatively and directly affect the accuracy of the classification model. Furthermore, after a deep investigation, it is found that most of the healthy bean leaves are classified incorrectly as unhealthy leaves due to the presence of the background and other surrounded objects that have similar symptoms and color to the unhealthy bean leaves. Figure 10 shows some examples of the bean leaves, which are correctly classified using the proposed system after accurately detecting the bean leaves in the input image and misclassified them without applying the image segmentation step.

3.3 Image classification evaluation

In this section, a comparison study is conducted to assess the performance of five various deep learning models (e.g., Densenet121, ResNet34, ResNet50, VGG-16, and VGG-19) in solving the binary and multi-image classification problems. In the binary image classification task, the performance of the chosen five deep learning models is evaluated via calculating the mean values of six quantitative measures using the testing set, including classification accuracy rate (CAR), sensitivity (recall), specificity, precision, F1-score, and area under the curve (AUC). The AUC measure can be interpreted as the probability of ranking a random positive sample by the model more highly than a random negative sample. The first five quantitative measures can be computed as follows:

$$\text{CAR} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}, \quad (1)$$

$$\text{Sensitivity (Recall)} = \frac{(\text{TP})}{(\text{TP} + \text{FN})}, \quad (2)$$

$$\text{Specificity} = \frac{(\text{TN})}{(\text{TN} + \text{FP})}, \quad (3)$$

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP} + \text{FP})}, \quad (4)$$

Fig. 10 some examples of the bean leaves, which are classified correctly using the proposed system after accurately detecting the bean leaves in the input image and misclassified them without applying the image segmentation step

$$\text{F1 Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}. \quad (5)$$

where TP, TN, FP, and FN are referring to the True Positives, True Negatives, False Positives, and False Negatives, respectively. In these experiments, the same hyper-parameters are used for all the adopted CNNs models. Particularly, the ReLU activation function is used to activate all the convolutional layers. A dropout ratio of 0.5 is applied in the fully-connected layers to avoid the overfitting problem by preventing the hidden units from complex co-adaptations. Adam optimization is employed with weight-decay factor $1e-2$ and learning rate of $1e-5$. Each deep learning model was trained for 20 epochs and a batch size of 2. Table 4 shows the performance comparison of the adopted five deep learning models. From this table, one can see that the best performance is achieved using the Densenet121 model with a CAR of 98.31%, sensitivity of 99.03%, specificity of 96.82%, Precision of 98.45%, F1-Score of 98.74%, and AUC of 100%. Although the ResNet50 model manages to achieve slightly higher results in terms of specificity and precision compared to the Densenet121 model, it gets inferior results on the other four quantitative measures.

Figure 11 shows the ROC curves of the five deep learning models by calculating the true positive rate (TPR) and false positive rate (FPR) using different accuracy threshold values. From this figure, it can be seen that the Densenet121 model manages to achieve an AUC of 100% compared to 99% is achieved by other models such as ResNet34, ResNet50, and VGG-19 model. It is worthy to notice that the high value of AUC of 100% obtained by the Densenet121 model, it is tremendously significant to decrease the number of misclassified healthy bean leaves as an unhealthy leaf. Finally, the confusion matrices of the healthy and unhealthy bean leaves of the five deep learning CNNs models are presented in Fig. 12. It was recognized that utilizing the Densenet121 model are only 8 out of 252 images of healthy bean leaves which is misclassified while there are only 5 out of 516 images of unhealthy bean leaves are misclassified as a healthy leaf. A slightly less performance is achieved using the ResNet34 model where 14 out of 252 images of healthy bean leaves are misclassified as an unhealthy leaf, and 7 out of 516 images of unhealthy bean leaves are misclassified as a healthy leaf. These results obtained confirm the reliability of the proposed training methodology and the adopted strategy of transfer learning where the knowledge (model

Table 4 Performance comparison of five different deep learning models in the binary classification task

Quantitative measures	Densenet121	ResNet34	ResNet50	VGG-16	VGG-19
CAR	98.31	97.27	90.10	89.71	92.97
Sensitivity (recall)	99.03	98.64	86.43	86.62	91.27
Specificity	96.82	94.44	97.61	96.03	96.42
Precision	98.45	97.32	98.67	97.81	98.12
F1-Score	98.74	97.97	92.14	91.88	94.57
AUC	100	99.49	98.71	98.22	99.01

Fig. 11 Comparisons of the ROC curves for five different deep learning models in the binary classification task: **a** Densenet121, **b** ResNet34, **c** ResNet50, **d** VGG-16, and **e** VGG-19

configurations) from the source dataset (e.g., ImageNet dataset) is utilized in the current problem to increase the prediction precision of the proposed deep learning approaches

even with the lack in the size of the employed bean leaves dataset. In addition to the probability of using the proposed bean leaves disease system (e.g., using the Densenet121

Fig. 12 Comparisons of confusion matrices for five different deep learning models in the binary classification task: **a** Densenet121, **b** ResNet34, **c** ResNet50, **d** VGG-16, and **e** VGG-19

model) in real-world applications including robotics to efficiently detect the unhealthy bean leaves in uncontrolled environments, with less than 2 s per image needed to get the prediction outcome.

Furthermore, the same five deep learning models are also employed to address the multi-image classification task, in which the input bean leaves image is classified into one of three various classes (e.g., healthy class, angular leaf spot class, and bean rust class). In these experiments, the five deep learning models were trained using the same hyper-parameters mentioned before, except the number of the units in the last layer is changed to three units instead of one unit to represent the number of the desired classes. As shown in Table 5, although a comparable performance is achieved using the ResNet34, ResNet50, VGG-16, and VGG-19 model, a higher CAR of 91.01% is obtained using the Densenet121 model. A closer examination of the results reveals some interesting results and hence the confusion matrix of each deep learning model is shown in Fig. 13. From this figure, one can see that the best results

Table 5 Performance comparison of five different deep learning models in the multi-image classification task

Quantitative measures	Densenet121	ResNet34	ResNet50	VGG-16	VGG-19
CAR	91.02	86.52	82.23	79.69	81.25

are achieved using the Densenet121 model and it can identify the healthy from unhealthy bean leaves efficiently. The highest CAR of 100% is obtained for a healthy class and the lowest CAR of 82.56% for beans rust class, which is usually confused with angular leaf spot class.

4 Conclusions and future work

This paper proposed a real-time and fully automated robotic perception framework to identify the healthiness of the bean leaves using DNNs. Initially, the bean leaves are

Fig. 13 Comparisons of confusion matrices for five different deep learning models in the multi-image classification task: **a** Densenet121, **b** ResNet34, **c** ResNet50, **d** VGG-16, and **e** VGG-19

automatically detected in the given image using the U-Net with a pre-trained ResNet34 model. Once, the bean leaves are detected, their health condition is evaluated using the Densenet121 model by classifying it into different classes (e.g., healthy class, angular leaf spot class, and bean rust class). The results obtained have demonstrated the reliability of the proposed automatic segmentation model by achieving an excellent agreement with ground truth images. Furthermore, the percentage difference values between manual and automatic estimations of two geometrical measures are less than 0.006% and 0.5 for MLA and MLP, respectively, with no geometrical measure with a proportional difference ($> 0.5\%$) between the manual and automatic estimations. On the other hand, a CAR of 98.31% and 91.01% is obtained using the Densenet121 model in both binary and multi-classification tasks. The present findings have important implications for solving the problem of identifying the healthiness status of the bean plant in the very large fields with fewer efforts required by the farmers using a robot-supported with a camera to capture images of the bean leaves. Further

experiments will be required to prove the efficiency and reliability of the proposed robotic perception framework using a larger and more challenging bean leaves dataset. This is an important and vital issue for future research.

References

- Al-Fahdawi, S., et al.: A fully automated cell segmentation and morphometric parameter system for quantifying corneal endothelial cell morphology (2018). <http://hdl.handle.net/10454/15684>. Accessed 9 Aug 2020
- Al-Waisy, A.S., Qahwaji, R., Ipson, S., Al-Fahdawi, S.: A multimodal deep learning framework using local feature representations for face recognition. *Mach. Vis. Appl.* **29**(1), 35–54 (2018a). <https://doi.org/10.1007/s00138-017-0870-2>
- Al-Waisy, A.S., Qahwaji, R., Ipson, S., Al-Fahdawi, S., Nagem, T.A.M.: A multi-biometric iris recognition system based on a deep learning approach. *Pattern Anal. Appl.* **21**(3), 783–802 (2018b). <https://doi.org/10.1007/s10044-017-0656-1>
- Basavaprasad, B., Hegadi, R.S.: Improved grabcut technique for segmentation of color image. *Int. J. Comput. Appl.* (2014). <http://cites>

- eerx.ist.psu.edu/viewdoc/download?doi=10.1.1.672.6067&rep=rep1&type=pdf. Accessed 16 May 2020
- Falk, T., Mai, D., Bensch, R., Çiçek, Ö., et al.: U-Net: deep learning for cell counting, detection, and morphometry (2019). <https://sci-hub.im/https://www.nature.com/articles/s41592-018-0261-2>. Accessed 17 May 2020
- He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2020). <http://image-net.org/challenges/LSVRC/2015/>. Accessed 15 Jul 2020
- Hemdan, E.E.D., Shouman, M.A., Karar, M.E.: COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in X-ray images (2020). <http://arxiv.org/abs/2003.11055>. Accessed 6 Aug 2020
- Iglovikov, V., Shvets, A.: TeraNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation (2020). <http://arxiv.org/abs/1801.05746>. Accessed 17 May 2020
- Kaur, J., Agrawal, S., Vig, R.: Integration of clustering, optimization and partial differential equation method for improved image segmentation. *Image Graph. Signal Process* **11**, 26–33 (2012). <https://doi.org/10.5815/ijjgsp.2012.11.04>
- Kingma, D.P., Lei Ba, J.: Adam: a method for stochastic optimization (2020). <https://arxiv.org/abs/1412.6980>. Accessed 9 Aug 2020
- Lab, M.A.I.: Bean disease dataset (2020). <https://github.com/AI-Lab-Makerere/ibean/>
- Larese, M.G., Namías, R., Craviotto, R.M., Arango, M.R., Gallo, C., Granitto, P.M.: Author's accepted manuscript automatic classification of legumes using leaf vein image features automatic classification of legumes using leaf vein 1 image features. *Pattern Recognit.* (2013). <https://doi.org/10.1016/j.patcog.2013.06.012>
- Liu, X., Zhang, Y., Jing, H., Wang, L., Zhao, S.: Ore image segmentation method using U-Net and Res_Unet convolutional networks. *RSC Adv.* (2020). <https://pubs.rsc.org/en/content/articlehtml/2020/ra/c9ra05877j>. Accessed 23 Jul 2020
- Mallikarjuna, K., Prasad, K.S., Subramanyam, M.V.: Image compression and reconstruction using discrete rajan transform based spectral sparsing. *Image Graph. Signal Process* **1**, 59–67 (2016). <https://doi.org/10.5815/ijjgsp.2016.01.07>
- Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics (2001). <https://ieeexplore.ieee.org/abstract/document/937655/>. Accessed 11 Jul 2020
- Mehra, R.: Breast cancer histology images classification: training from scratch or transfer learning? *ICT Express* (2018). <https://sci-hub.st/https://www.sciencedirect.com/science/article/pii/S2405959518304934>. Accessed 25 Jul 2020
- Meilă, M.: Comparing clusterings—an information based distance. *J. Multivariate Anal.* (2007). <https://www.sciencedirect.com/science/article/pii/S0047259X06002016>. Accessed 11 Jul 2020
- Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. *Front. Plant Sci.* (2016). <https://doi.org/10.3389/fpls.2016.01419>
- More RB, Bhosale PDS (2016) Agrobot- A Robot for Leaf Diseases Detection. *Int J Eng Sci Comput* **6**:7352–7355. <https://doi.org/10.4010/2016.1743>
- Muthukannan, K., Latha, P., Selvi, R.P., Nisha, P.: Classification of diseased plant leaves using neural network algorithms. *J. Eng. Appl. Sci.* (2015). <https://pdfs.semanticscholar.org/66b9/802a3d11706bbab12e707427b95da8c39726.pdf>. Accessed 16 May 2020
- Olsen, A., Kononov, D., Philippa, B., Ridd, P., Wood, J.C.: Deep-Weeds: a multiclass weed species image dataset for deep learning. *Sci. Rep.* (2019). <https://www.nature.com/articles/s41598-018-38343-3>. Accessed 16 May 2020
- Pamela, P., Mawejje, D., Ugen, M.: Severity of angular leaf spot and rust diseases on common beans in Central Uganda. *Uganda J. Agric. Sci.* **15**(1), 63–72 (2014). <https://www.ajol.info/index.php/ujas/article/view/126189>. Accessed 16 May 2020
- Ramesh, S., Vydeki, D.: Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Inf. Process. Agric.* (2019). <https://www.sciencedirect.com/science/article/pii/S2214317319300769>. Accessed 16 May 2020
- Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation (2020). <http://lmb.informatik.uni-freiburg.de/>. Accessed 18 May 2020
- Rother, C., Kolmogorov, V., Blake, A.: GrabCut—interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **23**(3), 309–314 (2004). <https://doi.org/10.1145/1015706.1015720>
- Russakovsky, O., et al.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
- Singh, V., Misra, A.K.: Detection of plant leaf diseases using image segmentation and soft computing techniques. *Inf. Process. Agric.* (2017). <https://www.sciencedirect.com/science/article/pii/S2214317316300154>. Accessed 16 May 2020
- Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning (2020). https://link.springer.com/chapter/10.1007/978-3-030-01424-7_27. Accessed 25 Jul 2020
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity (2004). <http://www.cns.nyu.edu/~lcv/ssim/>. Accessed 11 Jul 2020
- Xue, W., Zhang, L., et al.: Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Trans. Image Process.* (2013). <http://www4.comp.polyu.edu.hk/~cslzhang/IQA/>. Accessed 11 Jul 2020
- Zhang, J., Kong, F., Wu, J., Han, S., Zhai, Z.: Automatic image segmentation method for cotton leaves with disease under natural environment. *J. Integr. Agric.* (2018). <https://www.sciencedirect.com/science/article/pii/S209531191861915X>. Accessed 16 May 2020

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

Sudad H. Abed is a full-time instructor in the University of Anbar. He received his master degree in computer science from Ball State University in 2016. His research area of interest is machine learning, neural networks, deep learning, and data analytics. He has a solid experience in web application development and currently serves as the director of the software development department in the University of Anbar.

Alaa S. Al-Waisy is a doctor of philosophy in computer science. He received his BSc in 2009 and MSc in 2011 in computer science from Al-Anbar University. In 2018, he received PhD in digital imaging and visualization from the University of Bradford. His research interests include pattern recognition, image processing, computer vision, medical imaging, designing and implementing unimodal and multimodal biometric systems.

Hussam J. Mohammed received his BSc and MSc in computer science from University of Anbar, Iraq, and his PhD in digital forensic and cybersecurity from University of Plymouth, UK. He currently works at the Computer Center, University of Anbar. Hussam does research in the digital forensics, machine learning and AI. His current research interest in cybersecurity and digital evidence.

Shumoos Al-Fahdawi is a doctor of philosophy in computer science. She earned her BSc in 2009 and MSc in 2011 in computer science from Al-Anbar University. In 2018, She received PhD in digital imaging and visualization from the University of Bradford. She has worked as a lecturer at Al-Ma'aref University (Iraq) for one year. Through this period she has taught three subjects, such as Software engineering, Logic Design, and Computational Theory and she has supervised on two of the BSc's projects in the image processing field. Her current research is image processing, medical image, biometric system, pattern recognition, segmentation, and classification.