NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., 2025; 14(1), 251-262

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences

Araştırma makalesi / Research article

www.dergipark.org.tr/tr/pub/ngumuh / www.dergipark.org.tr/en/pub/ngumuh

Harnessing deep learning for multi-class weed species identification in agriculture

Tarımda çok sınıflı yabancı ot türlerinin tanımlanması için derin öğrenmenin kullanımı

Ebru Ergün[1,*]

[1] *Recep Tayyip Erdogan University, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Rize, Türkiye*

Abstract

Effective identification of weed species is critical for efficient agricultural management, enabling targeted eradication and optimized farming practices. In this study, ResNet, VggNet and DenseNet were used to evaluate the performance of deep learning models in accurately classifying different weed species. The dataset consisted of high-resolution images of different weed species taken under different environmental conditions. The experimental results demonstrated the ability of these models to identify multiple weed species with high accuracy. Evaluation metrics, accuracy, precision, recall and confusion matrices, validated the effectiveness of the models in discriminating between species. Of the convolutional neural network architectures tested, VggNet showed the highest classification accuracy of 99.21%. The results underscored the potential of deep learning-based classification systems in advancing scalable and efficient weed species identification and management for agricultural applications.

Keywords: Agricultural management, Classification, Deep learning, Multi-class, Precision agriculture, Weed species identification

Öz

Yabancı ot türlerinin etkili bir şekilde tespiti, verimli tarımsal yönetim için kritik öneme sahiptir ve hedefe yönelik yok etme ve optimize edilmiş tarım uygulamalarını mümkün kılar. Bu çalışmada, ResNet, VggNet ve DenseNet derin öğrenme modellerinin farklı yabancı ot türlerini doğru bir şekilde sınıflandırmadaki performanslarını değerlendirmek için kullanılmıştır. Veri seti, farklı çevresel koşullarda çekilmiş çeşitli yabancı ot türlerine ait yüksek çözünürlüklü görüntülerden oluşmaktadır. Deneysel sonuçlar, bu modellerin birden fazla yabancı ot türünü yüksek doğrulukla tespit etme yeteneğini göstermiştir. Değerlendirme metrikleri, doğruluk, hassasiyet, hatırlama ve hata matrisi, modellerin türler arasındaki ayrımı yapmadaki etkinliğini doğrulamıştır. Test edilen evrişimli sinir ağı mimarileri arasında, VggNet %99.21'lik en yüksek sınıflandırma doğruluğunu sergilemiştir. Sonuçlar, derin öğrenmeye dayalı sınıflandırma sistemlerinin, tarımsal uygulamalar için ölçeklenebilir ve verimli yabancı ot türü tespiti ve yönetimi konusunda büyük bir potansiyele sahip olduğunu vurgulamıştır.

Anahtar kelimeler: Tarımsal yönetim, Sınıflandırma, Derin öğrenme, Çoklu sınıf, Hassas tarım, Yabancı ot türlerinin tanımlanması

1 Introduction

Agriculture is the practice of growing plants to fulfill human needs for food, fiber, and bioenergy. Various factors can limit the crop production process in agriculture [1], with weeds being one of the most significant. Weeds generally refer to plant species that grow spontaneously in natural environments without human cultivation. These plants are not intentionally grown for agriculture or horticulture but emerge naturally. Weeds may be native to a specific region and are often considered undesirable in fields, gardens, or lawns [2]. While some weeds can be harmful, others contribute to the natural balance. Certain weeds offer nutritional value and can be utilized in traditional medicine or cooking. Despite their occasional benefits, weeds are typically seen as undesirable due to the damage they cause to agricultural fields, gardens, and ecosystems. One of the primary negative effects of weeds is their competition with agricultural crops and native plants. Weeds compete for essential resources such as water, nutrients, and light, thereby inhibiting the growth of crops [3]. Furthermore, weeds can reduce agricultural yield, as their presence can negatively impact crop development, leading to a smaller harvest. In addition, some weeds serve as hosts for diseases and pests, facilitating the spread of plant diseases and pests. Weed infestations can also increase the labor and costs associated with weed control and land cleaning, resulting in additional time, financial resources, and effort required by agricultural enterprises [4]. Moreover, in natural ecosystems, weeds can outcompete native plant species, disrupting ecological balance and causing biodiversity loss. The negative consequences of weeds on both ecosystems and agricultural land emphasize the importance of effective weed control and management. Among the various control methods, herbicides play a vital role in managing weed

* Sorumlu yazar / Corresponding author, e-posta / e-mail: ebru.yavuz@erdogan.edu.tr (E. Ergün)
Geliş / Recieved: 03.06.2024 Kabul / Accepted: 16.12.2024 Yayımlanma / Published: 15.01.2025
doi: 10.28948/ngumuh.1495040

growth [5]. As technological advancements continue to progress, the agricultural sector, in particular, is increasingly adopting technology to replace labor. Weed control methods are evolving alongside these technological developments. One of the most significant innovations in weed management is the use of data analytics. Data analytics employs modern technologies to identify and assess weed densities in agricultural fields. Drones or sensors capture images that detect weeds, and this data can be analyzed using artificial intelligence or machine learning techniques to map weed distribution and assist in targeting herbicide application [6], [7].

The literature on weed species identification and categorization has mainly focused on data analysis-based methods. Numerous studies have investigated the application of artificial intelligence approaches, including deep learning and machine learning, to weed detection [8]. Using techniques such as pattern recognition and data analysis, these methods have identified and categorized different weed species. For example, Espejo-Garcia et al. had introduced a plant identification system that combined features extracted from traditional machine learning classifiers such as support vector machines (SVM), XGBoost and logistic regression. Their system had also used pre-trained convolutional neural networks (CNNs) such as Xception, Inception-ResNet, VggNet, MobileNet and DenseNet. They had evaluated the approach using a dataset containing two crop species - cotton and tomato - and two weed species - black nightshade and velvetleaf. Their results showed that the combination of an SVM with a fine-tuned DenseNet resulted in a micro-F1 score of 99.29% [9]. Hu et al. had proposed a novel graph-based deep learning architecture, Graph Weeds Net (GWN), to identify different weed species from RGB images of complex pastures. Their approach achieved a classification accuracy (CA) of 98.10% [10]. Similarly, Tang et al. had used convolutional neural networks combined with K-means feature learning for weed identification. By using K-means pre-training, they achieved a CA of 92.89% [11]. Trong et al. had developed a classification strategy based on late fusion of deep neural networks (DNNs) to improve the accuracy of weed identification. They had applied their method to two datasets, including the Chonnam National University (CNU) seedling and weed datasets, using five different DNN models such as NasNet, ResNet, Inception-ResNet, MobileNet and VggNet. Their results had shown a CA of 98.77% for the CNU weed dataset and 97.31% for the seedling dataset [12]. Olsen et al. had used an openly available multi-class image dataset of Australian rangeland weed species. This dataset contained 17509 labelled images of eight prominent weed species from northern Australia. They had used benchmark deep learning models, specifically Inceptionv3 and ResNet50, to assess classification performance. These models had achieved average classification accuracies of 95.10% and 95.70% respectively [13]. In addition, Jin et al. had investigated the use of grid cells to train deep learning models for accurate identification of weed locations in photographs. They evaluated several neural networks, including DenseNet, EfficientNetV2, ResNet, RegNet and VggNet, for the detection and differentiation of weed species in turfgrass. VggNet stood out by achieving an F1 score of 0.950 for identifying common dandelion and F1 scores of ≥0.983 for recognizing subspecies such as dallisgrass, purple nutsedge and white clover. In addition, multi-classification approaches using DenseNet, EfficientNetV2 and RegNet achieved F1 scores of ≥0.984 for the detection of dallisgrass and purple nutsedge. In their study, EfficientNetV2 also excelled in a binary classification framework, achieving F1 scores of ≥0.981 for distinguishing turfgrass from weeds [14].

These studies have highlighted the critical importance of effective weed control strategies in achieving high agricultural yields. To control weeds with uneven spatial distribution patterns, herbicides were applied by ground or aerial robots. Before herbicides are applied, computer vision algorithms detect weeds in the field, making these algorithms central to the herbicide application process. As a result, large datasets of agricultural weeds were essential for the development of advanced computer vision algorithms. This work used deep learning algorithms to detect weeds for precision herbicide application. Weed classification was performed using an image dataset previously reported in [15]. This study further advanced computer vision-based weed detection techniques for in-field spot spraying applications. In addition, by emphasizing the ability of computer vision models to extract weed-related information from complex backgrounds, it extended the applicability of weed identification to difficult areas within crop fields.

The structure of this manuscript is organized as follows: Section 2 presents the experimental setup, including a comprehensive description of the image dataset and the methodologies employed. Section 3 discusses the experimental findings and evaluates the performance of the proposed approach against several deep learning classification methods. Lastly, Section 4 concludes the study with a summary of the key insights and implications.

2 Materials and methods

2.1 Comprehensive data description

The dataset used in this study focused on categorizing weed species to facilitate precision pesticide application. The data was collected by Rai et al. [15] who collected image data in two categories: aerial weed images and individual weed images. However, only the individual weed category was used in this study. To replicate real world conditions, plants were placed close together to mimic a natural field environment. Aerial photographs were then used to identify and extract individual weeds for further analysis. Data were collected from three different sites, all following consistent field organization and image collection protocols. North Dakota State University (NDSU) collected the data from these sites in the United States [15].

The dataset contains 3975 photographs of five common weed species in North Dakota: waterhemp (amaranthus tuberculatus), kochia (bassia scoparia), ragweed (ambrosia artemisiifolia), horseweed (erigeron canadensis), and redroot pigweed (RRPW) (amaranthus retroflexus). The greenhouse data set was collected under varying light and background conditions over several days and times. Field data were

collected using a DJI Phantom 4 Pro (V2.0), a commercially available unmanned aerial system (UAS). To ensure that the data accurately represented real-world conditions, several field factors were taken into account, including different background textures, vegetation masking, shadows, image blur, and the presence of similar-looking objects and weeds. The dataset is available in two compressed files in the Mendeley repository: Aerial_Weeds.rar and Individual_Weed.rar.

The Aerial_Weeds.rar file contains the 'images' and 'labels' folders. The 'labels' folder is further subdivided into three subfolders for different label formats (JSON, TXT, XML), while the 'images' folder contains high resolution JPG aerial images. A similar hierarchical structure is followed in the Individual_Weed.rar file, where photos and labels for each weed species are organised within their respective species classes. Figure 1 illustrates the class design of a particular weed category as defined by Rai et al.

The aerial photographs were taken at a resolution of 5472×3648 pixels and saved in JPG format. Details of these images are given in Table 1, with the data divided into four weed species types. The individual weed dataset was developed by extracting specific classes of each weed species from aerial images, followed by manual annotation in multiple formats. To increase the diversity of the training set, a small number of photographs taken in greenhouses using a Canon 90D handheld camera were also included. After manually annotating 3424 photographs, a total of 7700 trials were generated across the four weed groups. Figure 2 shows the weed species trials together with the annotated photographs exported in various formats.

To increase the size and diversity of the dataset, data augmentation techniques were used by Rai et al. [15], [16]. The original dataset consisted of 785 photos of kochia, 448 of horseweed, 355 of ragweed and 115 of RRPW. Various augmentation methods were applied to these images and the final output is summarized in Table 1. Furthermore, Rai et al. conducted image collection in Casselton during late May and early June of 2021, while data from Carrington was gathered between mid-July and late August of the same year. At Grand Farm, image acquisition took place from mid-August to the end of September in 2022. Altogether, this effort resulted in the collection of two years' worth of image data.

Table 1. Distribution of images for each weed species [15]

Individual Weed	Weed Species				
	Kochia	Horseweed	Ragweed	RRPW	**Total**
Total images	1150	1032	878	364	3424
Total instances	2600	1700	2000	1400	7700

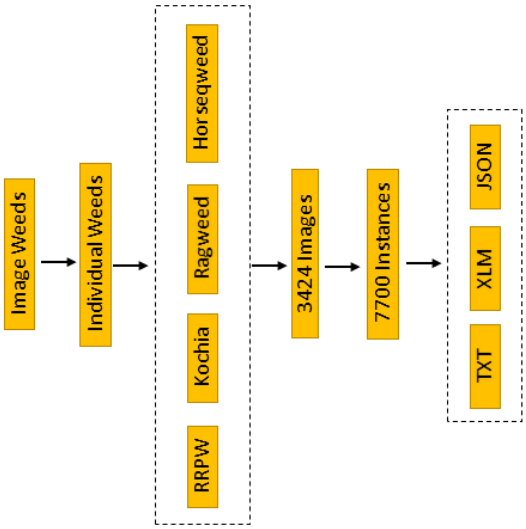


Figure 1. Flowchart folder representation of weed species [15]

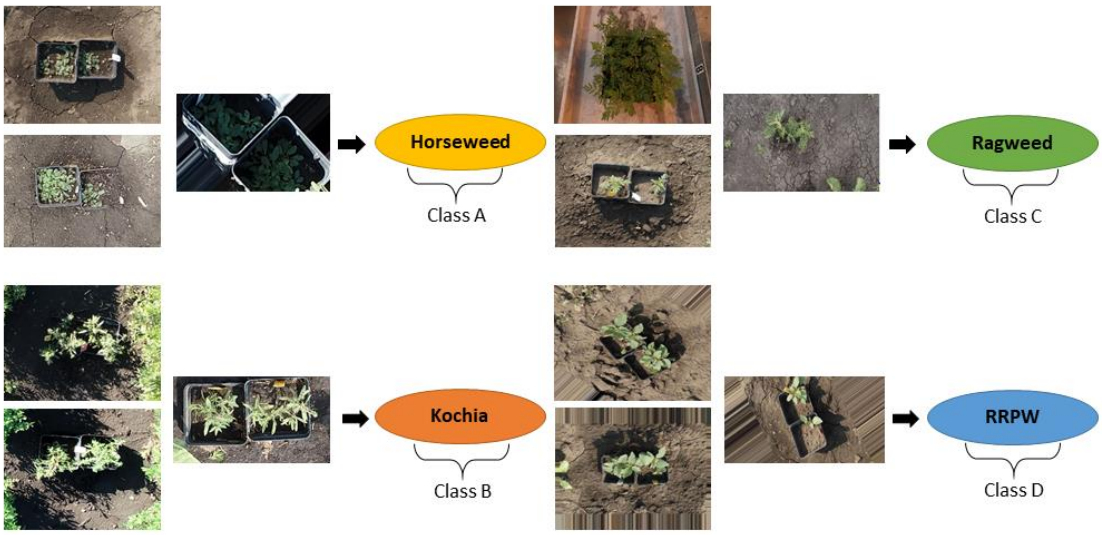


Figure 2. Individual weed species from cropped original images [15]

2.2 Methods

CNNs are a specific class of deep learning models that excel at processing visual data, particularly images [17]. Their architecture is designed to automatically extract hierarchical patterns and relationships within input data, enabling the identification of spatial structures. A key component of CNNs is the convolutional layer, which uses filters or kernels to scan the input data and capture basic features such as edges, textures and shapes. These filters traverse the input data, allowing the network to recognize different aspects of the image. Pooling layers are incorporated to reduce the spatial size of the feature maps and focus the network's attention on the most relevant features [18].

Common pooling strategies include average pooling and maximum pooling. To enhance the network's ability to model complex patterns, activation functions such as the rectified linear unit introduce non-linearity. In the final layers, fully connected layers connect neurons across successive layers, enabling the model to learn higher-level abstractions and make accurate predictions. CNNs have had a profound impact on several fields, particularly computer vision, due to their ability to automatically learn complex feature hierarchies, and have achieved significant success in tasks such as image classification, object detection, and face recognition [19].

The design of CNNs, in particular the convolutional and pooling layers, allows for the efficient capture of spatial dependencies in visual data, which contributes to their success in these applications [20]. The architecture used in this research is illustrated in Figure 3, which shows the overall methodology. In this design, the convolution and pooling layers form the initial stages of the network, while the final stages consist of a fully connected layer followed by a categorization layer. In essence, CNNs consist of successive trainable segments culminating in a discriminative classifier [21]. The layers work together to perform the training process, starting with the collection of input data. CNN architectures, ResNet, VggNet and DenseNet, were used to classify the four classes of weed images used in this study. All classification tasks were performed on a computer equipped with 16 GB of RAM and an Intel Core i7 processor running at 2.92 GHz.

2.2.1 ResNet

The ResNet architecture addresses the vanishing gradient problem common to very deep neural networks [22]. This problem, where gradients diminish as they move backwards through layers, can make it difficult for early layers to learn effectively. ResNet introduces residual learning, which allows the network to learn residual mappings rather than the full intended mappings directly.

The main feature of ResNet architectures is the inclusion of residual blocks, which contain shortcut connections that bypass certain layers. These shortcuts improve the flow of gradients during backpropagation, helping to mitigate problems such as the vanishing gradient problem and facilitating the training of very deep networks. In each residual block, the input is combined with the output of the convolutional layers, promoting a more efficient learning process.

ResNet models typically use global mean pooling, followed by fully connected layers and a softmax layer for classification tasks [23]. This design has proven highly effective in tasks such as image classification, object detection and semantic segmentation. ResNet has achieved cutting-edge performance on popular datasets such as ImageNet, and its robust results have made it a staple in many computer vision applications.

2.2.2 VggNet

VggNet is designed primarily as a series of convolutional layers, each followed by a max-pooling layer for downsampling, ensuring that the network remains simple and uniform in structure [24]. A notable feature of VggNet is the use of small filter sizes, such as 1×1 convolutional layers combined with 3×3 filters, which help to reduce dimensionality while capturing fine-grained features in the input images. This approach allows the network to effectively detect intricate details across images. In addition, the depth of the VggNet is another defining characteristic, with the network stacking multiple convolutional layers, enhancing its ability to learn increasingly discriminative and hierarchical representations of features [25]. The depth of the network contributes significantly to its performance, as each additional layer helps to refine the feature extraction process. In the later stages of the network, VggNet includes fully

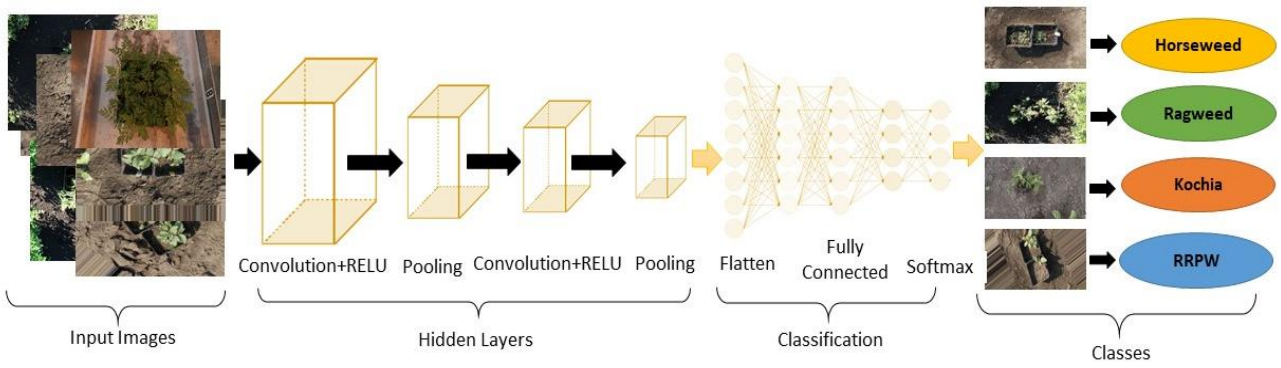


Figure 3. The general architecture of CNNs typically consists of convolutional and pooling layers

connected layers, followed by a softmax classifier to perform image classification. VggNet is also available in different configurations, such as Vgg16 and Vgg19, which differ mainly in the number of layers. Vgg16 consists of 16 weighting layers, while Vgg19 has 19 weighting layers [25]. These variations allow the network to be tailored for different levels of complexity, depending on the specific requirements of the task. VggNet has achieved remarkable success in image classification benchmarks, demonstrating its ability to learn robust and detailed feature representations. Its simple architecture and design principles have made it a popular choice for computer vision researchers and practitioners [26].

2.2.3 DenseNet

DenseNet is a neural network architecture introduced by Huang et al. [27]. The distinguishing feature of DenseNet is its dense connectivity, which differs from traditional architectures such as ResNet by connecting each layer to every other layer in a feed-forward fashion. In this configuration, each layer receives feature maps from all previous layers and forwards its own feature maps to all subsequent layers within a dense block, promoting extensive information sharing throughout the network. Dense blocks are the core elements of DenseNet. Each dense block consists of several layers, including activation, batch normalization, and convolutional layers, which work together in a tightly coupled fashion. Later layers in the block receive the accumulated feature maps from earlier layers, allowing efficient feature propagation and reuse throughout the network. DenseNet also incorporates a parameter called the growth rate, which determines the number of additional feature maps generated at each layer within a dense block, directly influencing the width of the network and its capacity for feature learning [28].

Between dense blocks, DenseNet uses transition layers that manage the spatial dimensions and regulate the number of feature mappings. These transition layers often consist of convolution, batch normalization, and pooling operations that effectively adjust the number of channels and downsample feature mappings. DenseNet also includes bottleneck layers, typically implemented as 1×1 convolutions, which streamline the computational requirements by reducing the input feature maps prior to the computationally intensive 3×3 convolutions. For classification tasks, DenseNet concludes with global average pooling, a fully connected layer and a softmax classifier, enabling the network to provide class predictions. DenseNet's densely connected architecture encourages feature reuse, facilitates effective gradient flow, and improves information transfer across layers, contributing to robust feature propagation. This structure effectively mitigates problems such as the vanishing gradient problem, supporting more efficient training and improved performance [29].

2.3 Classification metrics

Classification is a fundamental concept in machine learning and statistics, involving the categorization of data into distinct categories or classes based on specific features. The primary goal is to develop a model that can accurately predict the category of unseen data by learning patterns from a training data set. In this study, CA was used as a performance metric, calculated using the confusion matrix [30]. A multiclass confusion matrix is a tabular representation of the performance of a classification algorithm for multiple classes. It is structured as a square matrix, with rows representing the actual classes and columns representing the predicted classes. Each cell reflects the number of instances from a particular actual class that were predicted to belong to a particular class.

For example, Table 2 displays a confusion matrix for a four-class problem, A, B, C, and D. The rows correspond to the actual classes, while the columns represent the predicted classes [31]. The diagonal cells contain the number of correctly classified examples, such as the number of instances of class A that were correctly predicted as class A. Off-diagonal cells indicate misclassifications, such as the number of instances of class B that were incorrectly classified as class C. This matrix provides valuable insight into the performance of the model across all classes, allowing a detailed evaluation of its effectiveness.

Table 2. Four-class confusion matrix

Confusion Matrix	Predicted Classes			
	Predicted Class A	Predicted Class B	Predicted Class C	Predicted Class D
Actual Class A	AP1	BP1	BN1	BN1
Actual Class B	BN2	AP2	BP2	BN2
Actual Class C	BN3	BN3	AP3	BP3
Actual Class D	BN4	BN4	BP4	AP4

A multi-class classification problem with four categories, Class A, Class B, Class C and Class D, is represented by the confusion matrix shown in Table 2. Each cell in this matrix quantifies the number or frequency of instances belonging to a particular true class and its corresponding predicted class. For example, the correctly classified instances of class A are represented as AP1, while BP1 denotes instances that were misclassified as class A but actually belong to another class. Similarly, the matrix contains entries for true positives (AP), false positives (BP) and false negatives (BN) for each class. The overall CA for a four-class problem can be derived from this confusion matrix using Equation 1. The total number of trials is represented by the sum of all entries in the matrix, which includes correctly and incorrectly classified instances across all classes: AP1 + BP1 + BN1 + BN1 + BN2 + AP2 + BP2 + BN2 + BN3 + BN3 + AP3 + BP3 + BN4 + BN4 + BP4 + AP4. This comprehensive representation allows a detailed evaluation of the performance of the classification model.

Recall (RC) and Precision (PRC) are key performance metrics used to assess the effectiveness of classification models, particularly in multi-class classification scenarios [32]. PRC quantifies the accuracy of the model's positive predictions for each specific class. In a multi-class classification with four classes, the accuracy for each class is determined by the ratio of true positives to the total number

of predicted instances for that class [33]. On the other hand, RC measures the ability of the model to correctly identify all instances that belong to a particular class out of the total number of instances that truly belong to that class. For each class in a four-class classification task, recall is calculated as the ratio of true positives to the total number of true instances of that class. Equations 2 and 3 define the mathematical expressions for RC and PRC respectively. In Equation 2, the variable Ax is the sum of the true positives and false negatives for each class: $Ax = AP1 + BN1 + AP2 + BN2 + AP3 + BN3 + AP4 + BN4$. In Equation 3, Ab represents the sum of the true positives and false positives for each class: $AP1 + BP1 + AP2 + BP2 + AP3 + BP3 + AP4 + BP4$. These formulas allow the calculation of the recall and precision values, which are crucial for understanding the performance of the classification model.

$$CA = \frac{(AP1 + AP2 + AP3 + AP4)}{Total\ trials} x100 \quad (1)$$

$$RC = \frac{(AP1 + AP2 + AP3 + AP4)}{Ax} x100 \quad (2)$$

$$PRC = \frac{(AP1 + AP2 + AP3 + AP4)}{Ab} x100 \quad (3)$$

3 Results

In this study, transfer learning techniques were used for weed species classification. The starting point for training was the pre-trained parameters of three CNN models, ResNet, VggNet and DenseNet. These models, which are widely used for large datasets such as ImageNet, benefit from their ability to use extensive pre-training. The images within the ImageNet dataset are standardized to a resolution of 224×224 pixels, a size that balances computational efficiency with optimal performance for deep neural networks. Accordingly, all training and test images in this study were resized to 224×224 pixels prior to experimentation. A 5-fold cross-validation approach was used to evaluate model performance, and details of the CNN architectures used are shown in Table 3.

During training, the first five layers of the VggNet model were frozen to preserve pre-trained features, while batch sizes varied between 8 and 128 depending on the performance of each model. Of the architectures, DenseNet had the smallest number of layers and parameters, while VggNet had the simplest structure with 24 layers. All models were trained for 20 epochs at a learning rate of 0.001, and the best performing parameters based on validation set accuracy were selected for testing. To comprehensively assess the performance of the models, evaluation metrics such as accuracy, precision, recall and the confusion matrix were used, particularly due to the unbalanced distribution of the weed species dataset.

For model evaluation, 80% of the dataset was used for training and validation to optimize the CNN parameters, while the remaining 20% was reserved for testing. The dataset consisted of 3424 images for each weed species, randomly divided into training and test sets. Specifically, the training set contained 1032, 1150, 878, and 364 images for horseweed, kochia, ragweed, and redroot pigweed, respectively, while the test set contained 207, 230, 175, and 73 images for these classes. The results of the confusion matrix for the ResNet architecture during the 5-fold cross-validation are shown in Figure 4. As shown in Figure 4a, the CA for the first fold reached 99.00%, with 207 trials of class A, 223 trials of class B, 175 trials of class C and 73 trials of class D being correctly classified.

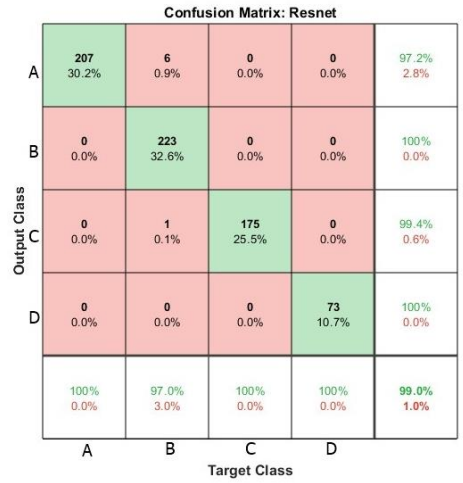
Table 3. Details of the CNN architectures used in the study

Model	Parameters				
	Batch size	Number of layers	Max Epochs	Optimizer	Base learning rate
ResNet	8	72	15	SGD	0.0001
VggNet	8	41	15	SGD	0.0001
DenseNet	8	709	15	SGD	0.0001

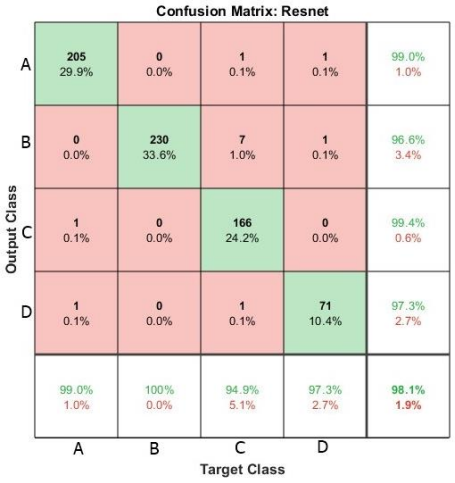
As depicted in Figure 4b, the classification accuracy (CA) for the second fold was 98.10%. In this fold, 205 trials of class A, 230 trials of class B, 166 trials of class C, and 71 trials of class D were accurately classified. Moving to the third fold, shown in Figure 4c, the CA increased to 99.10%, with correct classifications of 206 trials for class A, 229 for class B, 173 for class C, and 71 for class D. For the fourth fold, represented in Figure 4d, the CA was calculated as 98.80%. Specifically, 202 trials of class A, 228 of class B, 175 of class C, and 72 of class D were classified correctly. Similarly, the fifth fold achieved a CA of 99.00%, as shown in Figure 4e. In this case, 206 trials of class A, 227 of class B, 173 of class C, and 71 of class D were accurately identified. Finally, Figure 4f illustrates the combined confusion matrix for the 5-fold cross-validation. The average results show that 205.20 trials of class A, 227.40 of class B, 172.40 of class C, and 71.60 of class D were correctly classified across all folds.

Figure 5 showed the confusion matrix result for each fold of the 5-fold cross-validation using the VggNet architecture. The first fold had a CA of 99.70%, as shown in Figure 5a. Specifically, 206 trials were correctly classified as class A, 229 trials as class B, 175 trials as class C, and 73 trials as class D. Furthermore, a CA of 98.50% was calculated for the second fold, as shown in Figure 5b. Specifically, 204 trials were correctly classified as class A, 228 trials as class B, 173 trials as class C, and 70 trials as class D. In Figure 5c, the third fold had a CA of 99.10% and 205 class A trials, 226 class B trials, 175 class C trials and 73 class D trials were correctly classified. For the fourth fold, shown in Figure 5d, a CA of 99.00% was obtained and 200 trials of class A, 230 trials of class B, 175 trials of class C and 73 trials of class D were correctly classified. Furthermore, as shown in Figure 5e, the 5-fold had a CA of 99.70% and 206 trials of class A, 228 trials of class B, 176 trials of class C and 72 trials of class D were correctly classified. Figure 5f showed the 5-fold validation confusion matrix. In this case 204.20, 228.20, 172.80 and 72.20 trials of class A, B, C and D respectively were correctly classified.

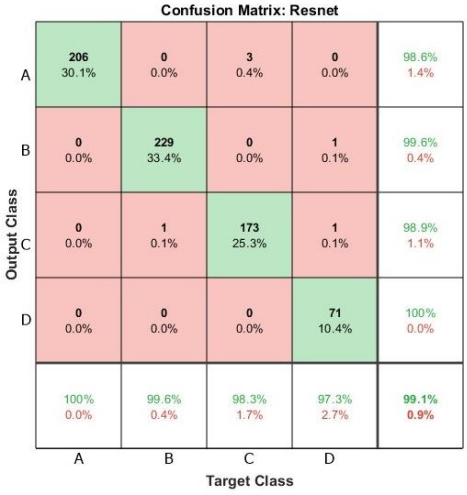
Figure 6 illustrated the confusion matrix results for each fold of the 5-fold cross-validation using the DenseNet architecture. In the first fold, as shown in Figure 6a, CA



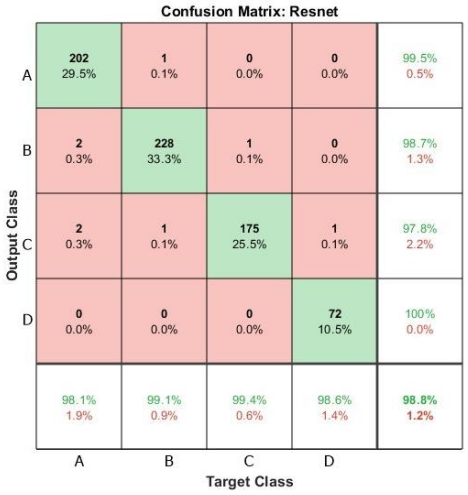
a)



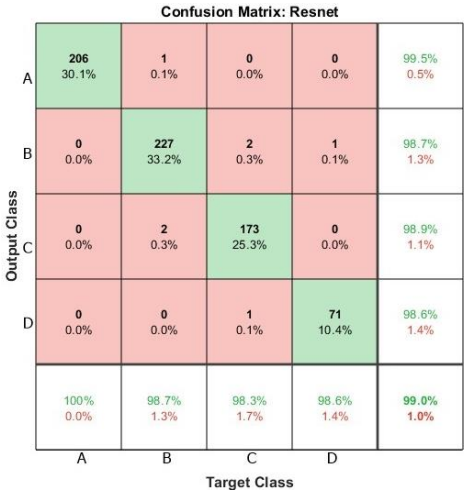
b)



c)



d)

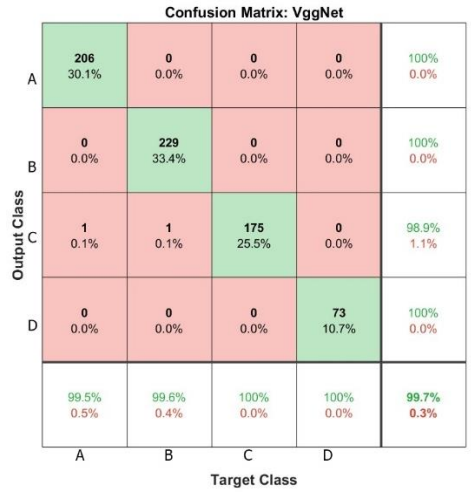


e)

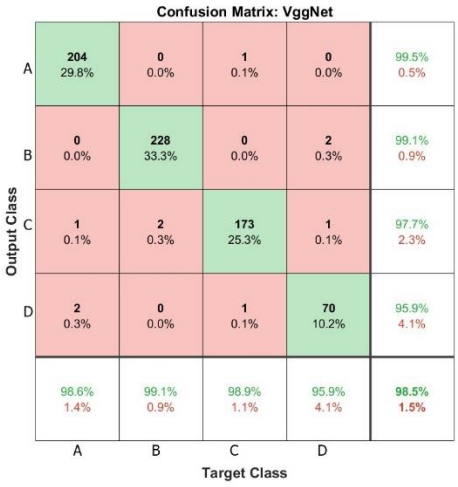
Confusion Matrix	Predicted Classes			
	Predicted Class A	Predicted Class B	Predicted Class C	Predicted Class D
Actual Class A	205.20	0.40	0.60	0.20
Actual Class B	1.60	227.40	1.00	0.00
Actual Class C	0.80	2.00	172.40	0.40
Actual Class D	0.20	0.60	0.40	71.60

f)

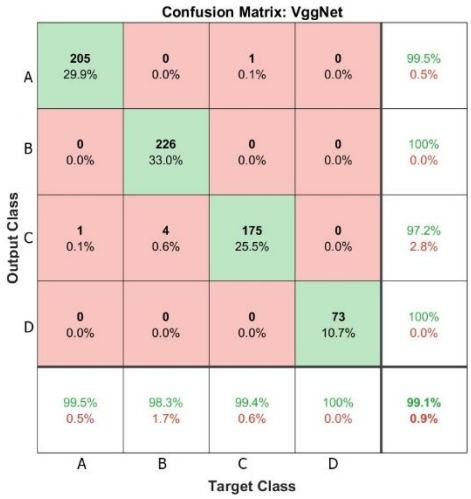
Figure 4. Confusion matrix result of a) 1st, b) 2nd, c) 3rd, d) 4th, e) 5th fold and f) mean 5-fold validation for ResNet



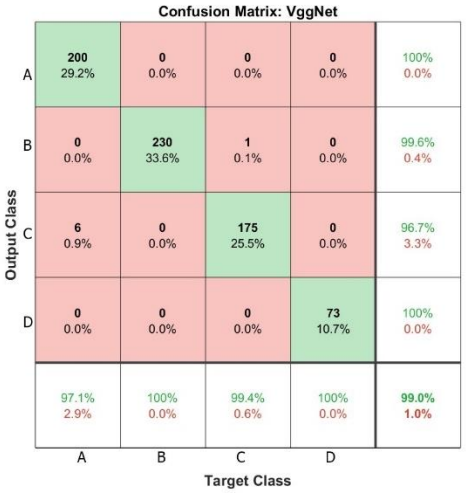
a)



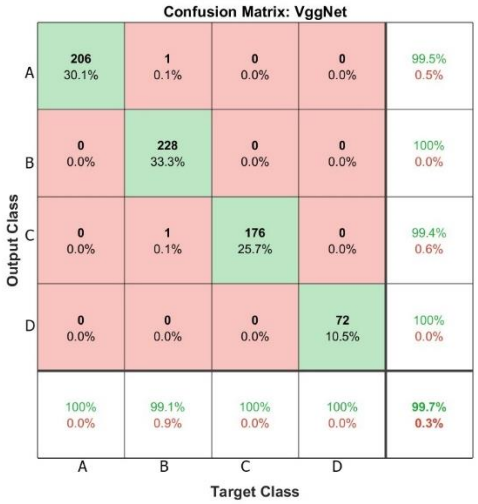
b)



c)



d)

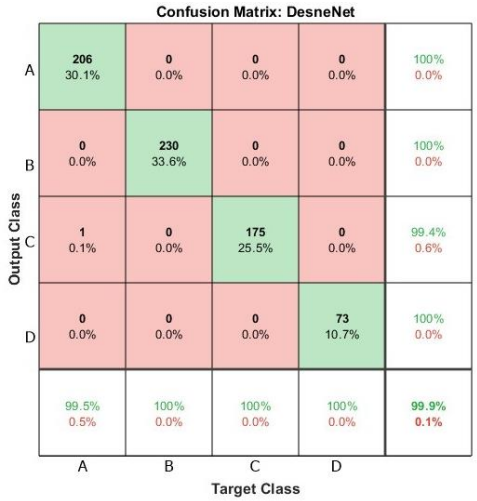


e)

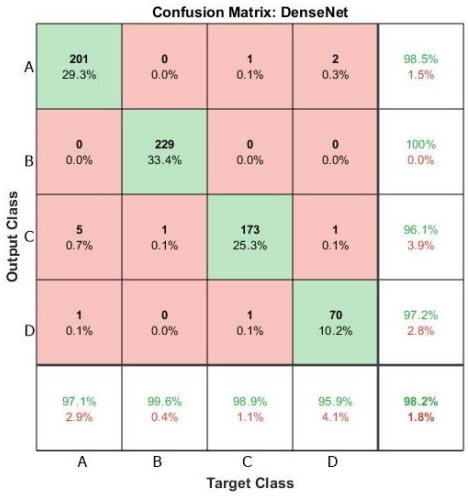
Confusion Matrix	Predicted Classes			
	Predicted Class A	Predicted Class B	Predicted Class C	Predicted Class D
Actual Class A	204.20	0.00	1.80	0.40
Actual Class B	0.20	228.20	1.60	0.00
Actual Class C	0.40	0.20	174.80	0.20
Actual Class D	0.00	0.40	0.20	72.20

f)

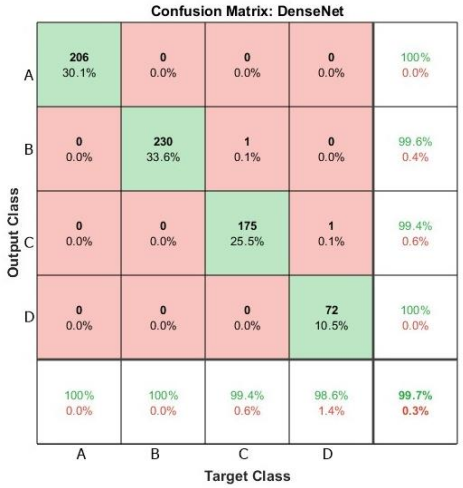
Figure 5. Confusion matrix result of a) 1st, b) 2nd, c) 3rd, d) 4th, e) 5th fold and f) mean 5-fold validation for VggNet



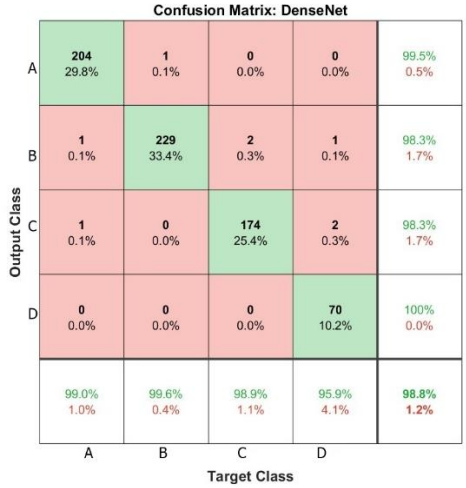
a)



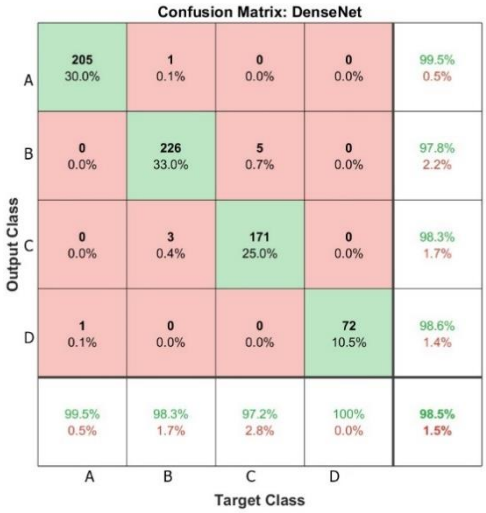
b)



c)



d)



e)

Confusion Matrix	Predicted Classes			
	Predicted Class A	Predicted Class B	Predicted Class C	Predicted Class D
Actual Class A	204.40	0.20	1.40	0.40
Actual Class B	0.40	228.80	1.80	0.00
Actual Class C	0.20	1.60	173.60	0.20
Actual Class D	0.40	0.20	0.80	71.40

f)

Figure 6. Confusion matrix result of a) 1st, b) 2nd, c) 3rd, d) 4th, e) 5th fold and f) mean 5-fold validation for DenseNet

reached 99.90%, with correct classifications of 206 trials for class A, 230 for class B, 175 for class C, and 73 for class D. For the second fold, depicted in Figure 6b, the CA was calculated as 98.20%, with 201 trials correctly classified as class A, 229 as class B, 173 as class C and 70 as class D. The third fold achieved a CA of 99.70%, with 206 trials correctly classified as class A, 230 as class B, 175 as class C and 72 as class D. The fourth fold, illustrated in Figure 6d, achieved a CA of 98.80%, with 204 trials correctly classified as class A, 229 as class B, 174 as class C, and 70 as class D. Lastly, as shown in Figure 6e, the fifth fold recorded a CA of 98.50%, with 205 trials correctly classified as class A, 226 as class B, 171 as class C and 72 as class D. The overall results for the 5-fold validation are summarized in Figure 6f, with the average number of correctly classified trials being 204.40 for class A, 228.80 for class B, 173.60 for class C and 71.40 for class D.

The classification results for CA, RC and PCR for all CNN architectures are shown in Table 4. As shown in the table, the highest CA is achieved by VggNet with 99.21%. The CA achieved with VggNet exceeds that of ResNet and DenseNet by 0.41% and 0.17% respectively. Furthermore, the RC values were calculated as 98.70%, 99.22% and 98.86% for ResNet, VggNet and DenseNet respectively, while the PCR values were calculated as 98.87%, 99.15% and 99.03%, respectively. These CA results highlight the effectiveness of the VggNet architecture in classifying these 4-class weed images. In addition to the results, an example image obtained from the training process for each weed images class is shown in Figure 7.

Table 4. Classification results for each CNN architecture

Model	Classification Results (%)		
	CA	RC	PRC
ResNet	98.80	98.70	98.87
VggNet	99.21	99.22	99.15
DenseNet	99.04	98.86	99.03

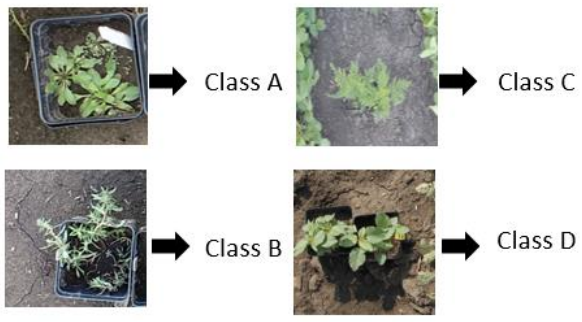


Figure 7. Example image for each weed image class

Table 5 showed the average central processor (CPU) times for training three deep learning models: VggNet, ResNet and DenseNet. Of these, VggNet had the shortest processing time, with an average of 350.03 minutes, due to its relatively simpler architecture. ResNet required more time, with an average of 425.15 minutes, reflecting the additional computational overhead introduced by its residual links. DenseNet, with its densely connected layers, had the highest CPU time of 530.20 minutes, highlighting the increased computational demands of its complex structure. These results demonstrate the correlation between model complexity and CPU time, with more complex architectures tending to require greater computational resources.

Table 5. Average central processor time results

Results in terms of minutes	Each CNN architecture		
	VggNet	ResNet	DenseNet
Average CPU time	350.03	425.15	530.20

4 Conclusion

This study represents a significant advance in the accurate and simultaneous identification of multiple weed species using deep learning techniques, in particular the VggNet model. The exceptional performance of VggNet, which achieved an impressive CA of 99.21%, highlights its effectiveness in accurately distinguishing between different weed species. The inclusion of high resolution images from different environmental conditions in the dataset further emphasizes the robustness and adaptability of the proposed models. A thorough evaluation using metrics such as accuracy, precision, recall and confusion matrix analysis highlights the effectiveness of the deep learning models in addressing the challenges of weed species discrimination. These results highlight the potential of integrating deep learning-based classification systems into agricultural practices. Such systems can provide scalable and efficient solutions for weed identification and management, thereby supporting precision agriculture and increasing crop productivity.

Future research can explore the integration of additional environmental and biological factors to further refine model performance, and expand the dataset to include more weed species and scenarios. The promising results of this study lay the foundation for the development of automated, non-invasive tools that can significantly contribute to sustainable agricultural practices.

Declaration of competing interest

The authors declare that none of the work reported in this study could have been influenced by any known competing financial interests or personal relationships.

Similarity rate (iThenticate): 19%

References

[1] F. Ahmed, H. A. Al-Mamun, A. H. Bari, E. Hossain, & P. Kwan, Classification of crops and weeds from digital images: A support vector machine approach. Crop Protection, 40, 98-104, 2012. https://doi.org/10.1016/j.cropro.2012.04.024.

[2] T. Luo, J. Zhao, Y. Gu, S. Zhang, X. Qiao, W. Tian & Y. Han, Classification of weed seeds based on visual images and deep learning. Information Processing in

Agriculture, 10(1), 40-51, 2023. https://doi.org/10.1016/j.inpa.2021.10.002.
[3] R. Bongiovanni, and J. Lowenberg-DeBoer, Precision agriculture and sustainability. Precision agriculture, 5, 359-387, 2004. https://doi.org/10.1023/B:PRAG.0000040806.39604.aa.
[4] A. S. Baghel, A. Bhardwaj and W. Ibrahim, Optimization of Pesticides Spray on Crops in Agriculture using Machine Learning. Computational Intelligence and Neuroscience, 2022. https://doi.org/10.1155/2022/9408535.
[5] R. Gerhards, D. Andujar Sanchez, P. Hamouz, G. G. Peteinatos, S. Christensen and C. Fernandez-Quintanilla, Advances in site-specific weed management in agriculture—A review. Weed Research, 62(2), 123-133, 2022. https://doi.org/10.1111/wre.12526.
[6] A. Monteiro, and S. Santos, Sustainable approach to weed management: The role of precision weed management. Agronomy, 12(1), 118, 2022. https://doi.org/10.3390/agronomy12010118.
[7] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin and V. K. Asari, The history began from alexnet: A comprehensive survey on deep learning approaches. Computer Vision and Pattern Recognition, 2018. https://doi.org/10.48550/arXiv.1803.01164.
[8] S. Khan, M. Tufail, M. T. Khan, Z. A. Khan and S. Anwar, Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. Precision Agriculture, 22(6), 1711-1727, 2021. https://doi.org/10.1007/s11119-021-09808-9.
[9] B. Espejo-Garcia, N. Mylonas, L. Athanasakos, S. Fountas and I. Vasilakoglou, Towards weeds identification assistance through transfer learning. Computers and Electronics in Agriculture, 171, 105306, 2020. https://doi.org/10.1016/j.compag.2020.105306.
[10] K. Hu, G. Coleman, S. Zeng, Z. Wang and M. Walsh, Graph weeds net: A graph-based deep learning method for weed recognition. Computers and electronics in agriculture, 174, 105520, 2020. https://doi.org/10.1016/j.compag.2020.105520.
[11] J. Tang, D. Wang, Z. Zhang, L. He, J. Xin and Y. Xu, Weed identification based on K-means feature learning combined with convolutional neural network. Computers and electronics in agriculture, 135, 63-70, 2017. https://doi.org/10.1016/j.compag.2017.01.001.
[12] V. H. Trong, Y. Gwang-hyun, D. T. Vu and K. Jin-young, Late fusion of multimodal deep neural networks for weeds classification. Computers and Electronics in Agriculture, 175, 105506, 2020. https://doi.org/10.1016/j.compag.2020.105506.
[13] A. Olsen, D. A. Konovalov, B. Philippa, P. Ridd, J. C. Wood, J. Johns and R. D. White, DeepWeeds: A multiclass weed species image dataset for deep learning. Scientific reports, 9(1), 2058, 2019. https://doi.org/10.1038/s41598-018-38343-3.
[14] X. Jin, M. Bagavathiannan, P. E. McCullough, Y. Chen and J. Yu, A deep learning-based method for classification, detection, and localization of weeds in turfgrass. Pest Management Science, 78(11), 4809-4821, 2022. https://doi.org/10.1002/ps.7102.
[15] N. Rai, M. V. Mahecha, A. Christensen, J. Quanbeck, Y. Zhang, K. Howatt and X. Sun, Multi-format open-source weed image dataset for real-time weed identification in precision agriculture. Data in Brief, 51, 109691, 2023. https://doi.org/10.1016/j.dib.2023.109691.
[16] TzutalinLabelImg v1.8.1 (Version 1.8.1), https://github.com/HumanSignal/labelImg.
[17] T. Tao, & X. Wei, A hybrid CNN–SVM classifier for weed recognition in winter rape field. Plant Methods, 18(1), 29, 2022. https://doi.org/10.1186/s13007-022-00869-z.
[18] M. Manikandakumar & P. Karthikeyan, Weed classification using particle swarm optimization and deep learning models. Comput. Syst. Sci. Eng., 44(1), 913-927, 2023. https://doi.org/10.32604/csse.2023.025434.
[19] W. Rawat and Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449, 2017. https://doi.org/10.1162/neco_a_00990.
[20] J. C. Chen, T. L. Chen, H. L. Wang and P. C. Chang, Underwater abnormal classification system based on deep learning: A case study on aquaculture fish farm in Taiwan. Aquacultural Engineering, 99, 102290, 2022. https://doi.org/10.1016/j.aquaeng.2022.102290.
[21] A. Paul, R. Machavaram, D. Kumar & H. Nagar, Smart solutions for capsicum Harvesting: Unleashing the power of YOLO for Detection, Segmentation, growth stage Classification, Counting, and real-time mobile identification. Computers and Electronics in Agriculture, 219, 108832, 2024. https://doi.org/10.1016/j.compag.2024.108832.
[22] A. Sagingalieva, M. Kordzanganeh, A. Kurkin, A. Melnikov, D. Kuhmistrov, M. Perelshtein and D. V. Dollen, Hybrid quantum ResNet for car classification and its hyperparameter optimization. Quantum Machine Intelligence, 5(2), 38, 2023. https://doi.org/10.1007/s42484-023-00123-2.
[23] A. Pandey and K Jain, An intelligent system for crop identification and classification from UAV images using conjugated dense convolutional neural network. Computers and Electronics in Agriculture, 192, 106543, 2022. https://doi.org/10.1016/j.compag.2021.106543.
[24] J. Wei, Y. Ibrahim, S. Qian, H. Wang, G. Liu, Q. Yu and J. Shi, Analyzing the impact of soft errors in VGG networks implemented on GPUs. Microelectronics Reliability, 110, 113648, 2020. https://doi.org/10.1016/j.microrel.2020.113648.
[25] A. S. Paymode and V. B. Malode, Transfer learning for multi-crop leaf disease image classification using convolutional neural network VGG. Artificial

Intelligence in Agriculture, 6, 23-33, 2022. https://doi.org/10.1016/j.aiia.2021.12.002.

[26] X. Zhang, Y. Qiao, Meng, F., Fan, C., & Zhang, M. (2018). Identification of maize leaf diseases using improved deep convolutional neural networks. IEEE Access, 6, https://doi.org/10.1109/ACCESS.2018.2844405.

[27] Lu, T., Han, B., L. Chen, F. Yu and C. Xue, A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. Scientific Reports, 11(1), 15824, 2021. https://doi.org/10.1038/s41598-021-95218-w.

[28] L. Shan and W. Wang, DenseNet-based land cover classification network with deep fusion. IEEE Geoscience and Remote Sensing Letters, 19, 1-5, 2021. https://doi.org/10.1109/LGRS.2020.3042199.

[29] K. Zhang, Y. Guo, X. Wang, J. Yuan and Q. Ding, Multiple feature reweight densenet for image classification. IEEE Access, 7, 9872-9880, 2019. https://doi.org/10.1109/ACCESS.2018.2890127.

[30] E. Ergün, Deep learning based multiclass classification for citrus anomaly detection in agriculture. Signal, Image and Video Processing, 18(11), 8077-8088, 2024. https://doi.org/10.1007/s11760-024-03452-2.

[31] E. Ergün, Artificial intelligence approaches for accurate assessment of insulator cleanliness in high-voltage electrical systems. Electrical Engineering, 1-14, 2024. https://doi.org/10.1007/s00202-024-02691-3.

[32] M. Heydarian, T. E. Doyle and R. Samavi, MLCM: Multi-label confusion matrix. IEEE Access, 10, 19083-19095, 2022. https://doi.org/10.1109/ACCESS.2022.3151048.

[33] M. L. Zhang and Z. H. Zhou, A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8), 1819-1837, 2013. https://doi.org/10.1109/TKDE.2013.39.