www.cambridge.org/wsc

Research Article

Cite this article: Kong X, Han K, Liu T, Maity A, Li A, Jin X, Yu J (2025). Development of a robust deep learning model for weed classification across diverse bermudagrass turfgrass regimes in China and the United States. Weed Sci. **73**(e50), 1–10. doi: 10.1017/wsc.2025.10021

Received: 6 November 2024 Revised: 14 April 2025 Accepted: 28 April 2025

Associate Editor:

Prashant Jha, Louisiana State University

Keywords:

Computer vision; precision herbicide application; turfgrass weed management; neural networks

Corresponding authors: Aimin Li; Email: lam@qlu.edu.cn; Xiaojun Jin; Email: xiaojunjin@njfu.edu.cn; Jialin Yu; Email: jialin.yu@pku-iaas.edu.cn

© The Author(s), 2025. Published by Cambridge University Press on behalf of Weed Science Society of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https:// creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



Development of a robust deep learning model for weed classification across diverse bermudagrass turfgrass regimes in China and the United States

Xiaotong Kong¹, Kang Han², Teng Liu², Aniruddha Maity³, Aimin Li⁴, Xiaojun Jin⁵ and Jialin Yu⁶

¹Intern, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, Shandong, China; and Graduate Student, Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China; ²Research Assistant, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China; ³Assistant Professor, Department of Crop, Soil and Environmental Sciences, Auburn University, Alabama, USA; ⁴Associate Professor, Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China; ⁵Associate Professor, National Engineering Research Center of Biomaterials, Nanjing Forestry University, Nanjing, China and ⁶Professor and Principal Investigator, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Shandong, China

Abstract

Automatic precision herbicide application offers significant potential for reducing herbicide use in turfgrass weed management. However, developing accurate and reliable neural network models is crucial for achieving optimal precision weed control. The reported neural network models in previous research have been limited by specific geographic regions, weed species, and turfgrass management practices, restricting their broader applicability. The objective of this research was to evaluate the feasibility of deploying a single, robust model for weed classification across a diverse range of weed species, considering variations in species, ecotypes, densities, and growth stages in bermudagrass turfgrass systems across different regions in both China and the United States. Among the models tested, ResNeXt152 emerged as the top performer, demonstrating strong weed detection capabilities across 24 geographic locations and effectively identifying 14 weed species under varied conditions. Notably, the ResNeXt152 model achieved an F1 score and recall exceeding 0.99 across multiple testing scenarios, with a Matthews correlation coefficient (MCC) value surpassing 0.98, indicating its high effectiveness and reliability. These findings suggest that a single neural network model can reliably detect a wide range of weed species in diverse turf regimes, significantly reducing the costs associated with model training and confirming the feasibility of using one model for precision weed control across different turf settings and broad geographic regions.

Introduction

Turfgrass is ubiquitously grown in various landscapes, including home lawns, golf courses, parks, school playgrounds, and sports fields (Stier et al. 2013). Weed control is a constant task for turf management, as weeds compete with turfgrasses for essential nutrients, sunlight, and water, potentially compromising both aesthetics and functionality of the turf. Implementing cultural practices, such as mowing and irrigation, can reduce weed infestation but rarely achieves complete weed control (Busey 2003; Neal 2020). Currently, the most effective way to control weeds is the application of various preemergence and postemergence herbicides (Kraehmer et al. 2014). However, weeds in natural environments are often unevenly distributed, leading to the broadcast application of herbicides across entire areas, including those without weed presence. Moreover, many commonly used herbicides, such as atrazine (photosystem II inhibitor) and monosodium methanearsonate (MSMA, arsenical herbicide), are classified as restricted-use pesticides due to their potential environmental impact (Kudsk and Streibig 2003; McElroy and Martins 2013; USEPA 2023a, 2023b; WSSA 2023). While manual spot spraying can significantly reduce herbicide inputs compared with broadcast applications by targeting only weed-infested areas, it is labor-intensive, time-consuming, and impractical for large-scale applications. Broadcast applications, on the other hand, are more efficient for large areas but often result in excessive herbicide use and environmental risks (Kudsk and Streibig 2003; Yu and McCullough 2016).



Computer vision-based automated weed detection and precision spraying technology is a promising solution for significantly reducing herbicide input and weed control costs (Bhakta et al. 2019; Gerhards et al. 2022; Jin et al. 2023b). Deep learning, a subfield of machine learning, employs multi-layered neural networks to simulate the way the human brain connects and transmits information between neurons (LeCun et al. 2015). Neural networks are a core component of deep learning and consist of a mathematical model composed of many artificial neurons, also known as nodes or units. Each neuron receives inputs from other neurons, weights and processes these inputs through an activation function, and then passes the result to the neurons in the next layer (Yang and Wang 2020). Deep convolutional neural networks (DCNNs) have achieved remarkable success in many applications, such as facial recognition (Singh et al. 2020), natural language processing (Chowdhary 2020), self-driving cars (Maqueda et al. 2018), and automated detection of structural flaws (Luo et al. 2019).

In recent years, the application of DCNNs in agriculture has grown exponentially. For example, Zhao et al. (2023) developed a neural network capable of detecting the germination status and estimating the total number of germinated rice (Oryza sativa L.) seeds. Ahmad Loti et al. (2021) documented a neural network that effectively identifies and differentiates various diseases in pepper (Piper aduncum L.). Additionally, previous studies have demonstrated the efficacy of DCNNs for detecting weeds in a variety of cropping systems (Dang et al. 2023; Sharpe et al. 2020; Yu et al. 2020). For example, researchers developed a neural network to detect weeds growing in soybean [Glycine max (L.) Merr.] (dos Santos Ferreira et al. 2017). Andrea et al. (2017) developed a neural network that accurately and reliably classifies weeds in corn (Zea mays L.) stands. Osorio et al. (2020) successfully employed a neural network to detect weeds in lettuce (Lactuca sativa L.). Moreover, You et al. (2020) proposed a neural network-based semantic segmentation method to distinguish weeds from crops in complex agricultural settings.

DCNNs have also demonstrated great performance in detecting weeds growing in bermudagrass [Cynodon dactylon (L.) Pers.] turf (Jin et al. 2022a; Xie et al. 2021; Yu et al. 2019a, 2019b). The use of DCNNs to detect weeds in both dormant and actively growing bermudagrass was first reported by Yu et al. (2019a), who compared three image classification neural networks, including DetectNet (NVIDIA, 2016), VGGNet (Simonyan and Zisserman 2014), and GoogLeNet (Szegedy et al. 2015). These neutral networks were evaluated for their ability to detect and classify several broadleaf weed species, including dollar weed (Hydrocotyle spp.), Florida pusley (Richardia scabra L.), and old world diamond-flower (Oldenlandia corymbosa L.), in actively growing bermudagrass turf. It was found that DetectNet achieved an excellent F1 score of 0.99, outperforming the other two neural networks. In a subsequent study, Yu et al. (2020) documented that VGGNet achieved excellent performance, surpassing AlexNet and GoogLeNet in detecting dallisgrass (Paspalum dilatatum Poir.), doveweed [Murdannia nudiflora (L.) Brenan], smooth crabgrass [Digitaria ischaemum (Schreb.) Schreb. ex Muhl.], and tropical signalgrass [Urochloa adspersa (Trin.) R. Webster] in actively growing bermudagrass.

To develop a commercially viable smart sprayer employing a neural network model, the spray system must detect various weed species across different turfgrass regimes, irrespective of species, ecotypes, densities, growth stages, and geographic locations. However, detecting and differentiating weeds from turfgrass can be challenging, particularly when dealing with grass weeds that share similar morphological characteristics with turfgrass species. In contrast, broadleaf weeds often exhibit distinct features that facilitate their identification. Previous neural networks were typically designed to identify a single or a limited number of weed species in specific geographic areas (Jin et al. 2022a, 2022b, 2023a). In addition, due to phenotypic plasticity, significant morphological variations may occur among the weed ecotypes from different turfgrass management regimes or geographic areas (Kerr et al. 2019). For example, a dwarf ecotype of goosegrass [Eleusine indica (L.) Gaertn.], with an average internode length of only 0.2 cm, has been found in a golf course in Florida (Kerr et al. 2019), in contrast to wild ecotypes with an average internode length of 7 cm (Saidi et al. 2016). Consequently, these networks may struggle to detect different weed ecotypes, species, or those growing in mixed stands at varying growth stages and densities across diverse turf regimes and geographic regions. Moreover, bermudagrass is widely utilized in various turf sites, including home lawns, golf courses, school playgrounds, and sport fields.

Research has shown that training image size and quantity significantly influence the performance of neural networks in weed detection. Zhuang et al. (2022) evaluated multiple neural networks, including AlexNet and VGGNet, for the detection and classification of broadleaf weed seedlings in wheat (Triticum aestivum L.) and found that, for a small training dataset (5,500 negative and 5,500 positive images), increasing the size of the training images from 200 \times 200 pixels to 300 \times 300 or 400 \times 400 pixels resulted in a decrease in the F₁ scores of both networks. However, for larger training datasets (11,000 negative and 11,000 positive images), increasing the image size improved the performance of all studied networks, regardless of the image sizes. Therefore, the objectives of this research were (1) to assess the ability of image classification neural networks trained on datasets from limited geographic regions to generalize weed detection performance across diverse bermudagrass turf regimes and locations and (2) to examine how varying the size of training datasets impacts the performance of eight different neural network models.

Materials and Methods

Neural Network Models

This study evaluated eight neural network models for weed detection in turfgrass systems. These included AlexNet (Krizhevsky et al. 2012), a pioneering DCNN with five convolutional layers; GoogleNet (Szegedy et al. 2015), which employs an Inception architecture for efficient feature extraction; and VGGNet (Simonyan and Zisserman 2014), known for its deeper structure with multiple 3×3 convolution kernels. Additionally, ResNet101 and ResNet152 (He et al. 2016), which utilize residual learning, were assessed to enhance deep network training. To improve multi-scale feature extraction, Res2Net and ResNeXt (Gao et al. 2019) were also included. Finally, PoolFormer, a transformer-based model with a simplified MetaFormer architecture (Yu et al. 2022), was tested for its ability to generalize weed detection across diverse conditions. A summary of each model's architecture and advantages is provided in Table 1.

Image Acquisition

The training, validation, and testing dataset images were mainly captured in four cities in Florida and Georgia in the United States,

Table 1. Eight neural networks evaluated in the study.

Model name	Key features	Notable advantages ^a
AlexNet	5 convolutional layers, 3 fully connected layers	First deep CNN using GPUs, ILSVRC 2012 winner
GoogleNet	Inception architecture, 22 layers	Efficient computation, ILSVRC 2014 winner
VGGNet	16 layers, 3×3 convolution kernels	Deeper architecture, improved feature extraction
ResNet101/152	Residual learning framework, 101/152 layers	Mitigates vanishing gradient, ILSVRC 2015 winner
Res2Net	Multi-scale feature representation	Enhanced hierarchical feature extraction
ResNeXt	Aggregated residual transformations	Improved multi-scale feature extraction
PoolFormer	MetaFormer with Transformer structure	Simplified design, strong generalization

^aAbbreviations: CNN, convolutional neural network; GPU, graphics processing unit; ILSVRC, ImageNet Large Scale Visual Recognition Challenge.

Table 2. Details of training, validation, and testing dataset images.^a

Location	GPS coordinates	Turfgrass sites	Photo acquisition dates	Broadleaf weeds	Grasses	Sedges
Bradenton, FL, USA	(27.4963,—82.5745)	Golf courses	OctNov. 2018	None	Eleusine indica, Digitaria ischaemum, Urochloa adspersa	None
Tampa, FL, USA	(27.9473, -82.4584)	Golf courses, roadsides, and sport fields	JulDec. 2018	Murdannia nudiflora, Richardia scabra	E. indica, D. ischaemum	None
Riverview, FL, USA	(27.8139,—82.4167)	Golf courses	Jul. 2018–Feb. 2019	Hydrocotyle spp., Oldenlandia corymbosa	D. ischaemum, E. indica, U. adspersa	Cyperus rotundus
Griffin, GA, USA	(33.2608,—84.2521)	Turfgrass research facility on the University of Georgia Griffin campus	Jul. 2018, Oct. 2018	Paspalum dilatatum, Diodia virginiana, Chamaesyce maculata, Trifolium repens	D. ischaemum, E. indica	Cyperus compressus, Kyllinga odorata

^aTraining, validation, and testing dataset images collected from four locations in Florida and Georgia, USA.

covering various turf regimes infested with a variety of broadleaf, grass, and sedge weeds, as detailed in Table 2 and Figure 1.

In Bradenton, FL, USA (27.4963, -82.5745), images were captured multiple times at a golf course between October and November 2018, predominantly featuring annual grass weeds such as *E. indica*, *D. ischaemum*, and *U. adspersa*, with visually estimated turf cover exceeding 90% and bare soil cover below 10%.

In Tampa, FL, USA (27.9473, -82.4584), images were captured multiple times between July and December 2018 from golf courses, roadsides, and sports fields, primarily featuring *M. nudiflora*, *R. scabra*, *E. indica*, and *D. ischaemum*.

In Riverview, FL, USA (27.8139, -82.4167), images were collected from the rough of a golf course between July 2018 and February 2019. These images included broadleaf weeds such as *Hydrocotyle* spp. and *O. corymbosa*, along with grasses like *E. indica*, *D. ischaemum*, and *U. adspersa*. Additionally, low-density pre-flowering purple nutsedge (*Cyperus rotundus* L.) was present. Turfgrass cover ranged from 70% to 80%, while bare ground covered 20% to 30%.

In Georgia, USA, images were captured in July and October 2018 at the turfgrass research facility at the University of Georgia Griffin Campus (33.2608, -84.2521). These images featured low-density, pre-flowering annual sedges (*Cyperus compressus* L.) exhibiting a clump growth habit, as well as broadleaf weeds, grasses, and fragrant kyllinga (*Kyllinga odorata* Vahl). The broadleaf weeds included spotted spurge [*Chamaesyce maculata* (L.) Small; syn.: *Chamaesyce maculata* L.], Virginia buttonweed (*Diodia virginiana* L.), and white clover (*Trifolium repens* L.), while the grasses included *P. dilatatum, E. indica*, and *D. ischaemum*.

Regarding model robustness testing, this study constructed an additional robustness testing dataset. These data were collected from 24 different locations across the United States and China, as shown in Figures 2 and 3. Each test scenario in the dataset consists of 100 turfgrass images without weeds and 100 images containing weeds.

All images used in this study were captured using a Sony[®] Cyber-Shot camera (Sony, Minato, Tokyo, Japan) with a resolution of $1,920 \times 1,080$ pixels. They were captured between 0900 and 1700 hours under various weather and outdoor lighting conditions, including clear, cloudy, and partly cloudy conditions.

Training and Testing

To align with the input requirements and maintain compatibility across the deep learning model architectures evaluated in this study, the original images were resized to 480×480 pixels. These resized images were then divided into two datasets. The small training dataset contained 10,000 positive images (with weeds) and 10,000 negative images (without weeds). To construct the large training dataset, an additional 40,000 images were added—20,000 positive and 20,000 negative—resulting in 30,000 positive and 30,000 negative images. Both the validation and the testing datasets contained 1,000 positive and 1,000 negative images each.

In this study, eight DCNNs were trained on the small and large training datasets for 100 epochs each. After the optimal weights for each model were obtained, model performance was evaluated on the testing dataset. The best-performing model was then selected to undergo final robustness testing on the testing dataset. Jin et al. (2022b) employed a classification-based approach for weed detection, wherein images were segmented into grid cells and classified based on the presence of weeds. This method enables both detection and localization, with classification evaluation metrics effectively reflecting the model's performance in weed detection.

Training dataset



Figure 1. Image examples from the training dataset. The dataset includes Digitaria ischaemum and Paspalum dilatatum at the 3- to 5-tiller stage and Murdannia nudiflora and Oldenlandia corymbosa at full maturity before flowering.



Figure 2. Images of weed species and their respective turf sites in the United States used for neural network robustness testing.

A confusion matrix was employed to evaluate the performance of each model. This matrix compares the predicted results of a classifier with the true labels, categorizing them into four categories: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

In the confusion matrix:

- TP represents the number of instances correctly predicted as positive.
- FN represents the number of instances incorrectly predicted as negative when they are actually positive.
- FP represents the number of instances incorrectly predicted as positive.
- TN represents the number of instances correctly predicted as negative.

In the present research, TP represents the model correctly identified the target weed; TN represents the model correctly



Figure 3. Images of weed species and their respective turf sites in China used for neural network robustness testing.

identified images without the target weed; FP represents the model incorrectly predicted the target weed; and FN refers to instances where the model failed to predict the presence of the actual target weed.

The confusion matrix was used to calculate various performance metrics of a classification algorithm. In this study, precision (Equation 1), recall (Equation 2), F_1 score (Equation 3), and Matthews correlation coefficient (MCC) (Equation 4) were calculated using the results from the confusion matrix.

Precision measures the ability of the developed neural network to correctly identify the targets and was calculated using the following formula:

$$Precision = \frac{TP}{TP + FP}$$
[1]

Recall provides an estimation of the developed neural network's ability to identify its targets and was calculated using the following formula:

$$\operatorname{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$
[2]

The F_1 score, defined as the harmonic mean of precision and recall, was calculated using the following formula:

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$
[3]

MCC is a universal evaluation metric used to assess the performance of classification models, measuring the correlation between predicted and actual labels. It applies to both binary and multiclass classification tasks. In this study, we applied MCC to evaluate the performance of binary classification models, specifically distinguishing between weed (presence) and non-weed (absence) in turfgrass images (Sokolova and Lapalme 2009). It was calculated using the following equation:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$
[4]

Experimental Configuration

The experiments were conducted using the PyTorch (Meta) deep learning framework (v. 1.13.0) with CUDA 11.6 (NVIDIA). To ensure fairness of the experiments, none of the models were initialized with pretrained weights. All training and testing procedures were executed on a workstation equipped with an Intel[®] CoreTM i9-10920X CPU @ 3.50 GHz, an NVIDIA RTX 3080 Ti GPU, and 128 GB of memory. The operating system used was Ubuntu 20.04.1.

The hyperparameter settings were as follows: the image size for the training process was set to 480×480 pixels, with stochastic gradient descent as the optimizer. The base learning rate was set to 0.1, and weight decay was applied with a value of 0.0001. The batch size was 16, and the learning rate policy was set to "step." Momentum was set to 0.9, and the model was trained for 100 epochs. The output layer was configured with two nodes, corresponding to the binary classes (weed vs. non-weed), using a softmax activation function. This configuration ensured a controlled and consistent setup for evaluating the model's performance.

Table 3. Testing results of neural networks for classification of weeds while growing in turfgrasses.

Training dataset ^a	Neural network	Precision	Recall	F ₁ score	MCC ^b
Small dataset	AlexNet GoogleNet VGGNet16 ResNet101 ResNet152 Res2Net Res2Net	0.8801 0.7897 0.9990 0.9740 0.9813 0.9950 0.9970	0.9840 0.9800 0.9970 0.9799 0.9980 0.9970 0.9990	0.9292 0.8746 0.9980 0.9769 0.9896 0.9960 0.9980	0.8560 0.7408 0.9960 0.9540 0.9791 0.9920 0.9960
Large dataset	Poolformer AlexNet GoogleNet VGGNet16 ResNet101 ResNet152 Res2Net ResNeXt152 Poolformer	0.9549 0.9980 0.9950 0.9872 0.9921 0.9921 0.9794 0.9842 0.9940	0.9950 0.9960 1.0000 0.9990 1.0000 1.0000 1.0000 1.0000	0.9745 0.9970 0.9975 0.9935 0.9955 0.9960 0.9896 0.9921 0.9970	0.9488 0.9940 0.9950 0.9871 0.9910 0.9920 0.9792 0.9841 0.9940

^aThe small training dataset contained 10,000 positive images (with weeds) and 10,000 negative images (without weeds), while the large training dataset contained 30,000 positive

and 30,000 negative images.

^bMCC, Matthews correlation coefficient.

Result and Discussion

Model Evaluation

Among the evaluated neural networks, VGGNet16, trained with 20,000 images, achieved the highest precision in the testing dataset, reaching 0.9990 (Table 3). It also exhibited the highest recall and F_1 score and an MCC value exceeding 0.9960. Res2Net and ResNeXt152 also performed well, with accuracies of 0.9950 and 0.9970, respectively. Notably, ResNeXt152 had a recall of 0.9990, slightly surpassing VGGNet16, with similar F_1 and MCC scores.

Initially, AlexNet and GoogleNet showed lower performance, with precision of 0.8801 and 0.7897 and MCC values below 0.9. However, their performance improved significantly when trained with 60,000 images. AlexNet's precision increased to 0.9980, surpassing VGGNet16, and its MCC reached 0.9940. GoogleNet achieved a perfect recall (1.0), a precision of 0.9950, and the highest MCC among all models.

ResNet152 and Poolformer achieved a recall of 1.0 with a precision of 0.9921 and 0.9940, respectively. While ResNet101 and ResNet152 showed further improvements, VGGNet16, Res2Net, and ResNeXt152 experienced overall performance declines, possibly due to their simpler architectures and limited learning capacities (Chollet 2021; Hastie et al. 2009). Models with complex structures excelled on smaller datasets but faced overfitting risks on larger datasets (Szegedy et al. 2015).

VGGNet16 and ResNeXt152 consistently demonstrated excellent performance across all metrics, with MCC values >0.98 and F_1 scores >0.99. ResNeXt152 outperformed all models in recall and showed the best performance in confusion matrices (Figures 4 and 5). Therefore, this study selected the top two weights of ResNeXt152 for weed classification in various turfgrass regimes.

Single Weed Species Classification

Weed classification in turfgrass systems has been identified as a critical component of precision agriculture, with recent advancements in deep learning offering promising solutions for reducing herbicide use and improving management efficiency (Beckie et al. 2019; Bhakta et al. 2019). Previous studies have often been confined

to collecting data from specific regions for training neural networks and subsequently evaluating them within the same geographic locations (Jin et al. 2022c; Xie et al. 2017; Yu et al. 2019a, 2019b). Although these studies have demonstrated effective weed detection, the testing datasets in these works were limited to specific turfgrass management regimes and geographic locations. In this study, we tested model performance across 24 diverse scenarios in China and the United States, encompassing various weed species, ecotypes, densities, and growth stages, as detailed in Table 4.

The models trained with large datasets demonstrated consistent precision exceeding 0.97 across all classifications for *D. ischaemum*. They achieved F_1 scores ranging from 0.988 to 0.998, with perfect recall of 1.0 for the first five scenarios. Even small datasets yielded strong performance, with MCC values above 0.81 across the same five scenarios. However, the classification of *D. ischaemum* in commercial landscapes was less effective, likely due to lower recall (below 0.58). These results confirm that ResNeXt152, when trained on large datasets, can accurately classify *D. ischaemum* at different growth stages in various turfgrass settings.

For *P. dilatatum* classification in Auburn, AL, USA, the model trained on a small dataset showed acceptable recall but lower precision (0.7333). In contrast, the model trained on the large dataset achieved a precision of 0.9686, emphasizing the importance of dataset size in minimizing false detections. Both models trained with large datasets performed well in other scenarios, with MCC exceeding 0.95, and the neural network maintained excellent performance even with low-density *P. dilatatum*, indicating that ResNeXt152 can accurately classify *P. dilatatum* in various turf conditions.

For *M. nudiflora* classification, tests were conducted at locations in Miami, FL, and Tifton, GA, USA. Both small and large datasets yielded robust results, with precision consistently above 0.97 and F_1 scores not falling below 0.96. Recall rates were optimal in three out of four scenarios, further affirming the efficacy of ResNeXt152 in *M. nudiflora* classification.

Testing on single weed species at additional locations revealed that the model trained with a large dataset achieved near-perfect classification (precision and recall exceeding 0.99, MCC > 0.98) for species including *Hydrocotyle* spp., *R. scabra*, green kyllinga *[Kyllinga brevifolia* Rottb.], *O. corymbosa*, *D. virginiana*, and *T. repens*.

Finally, experiments with *U. adspersa* at two Florida golf courses demonstrated that the model trained on a small dataset performed well in low-density scenarios (precision: 0.9745, MCC: 0.9696), while the model trained on a large dataset achieved consistent excellence (all metrics > 0.99). These results highlight ResNeXt152's ability to classify weeds across varying densities in bermudagrass regimes.

Mixed Weed Species Classification

After confirming the accuracy of ResNeXt152 in single weed species classification, the study extended its evaluation to complex scenarios involving the coexistence of multiple weed species. As previously reported, multiple classifier neural network models, including DenseNet, EfficientNetV2, ResNet, RegNet, and VGGNet, are capable of effective weed detection and classification (Jin et al. 2022b). These models achieved high accuracy, with F_1 scores of at least 0.946 on the test dataset. However, the training and test images used in this study were collected from geographically proximate areas, which may limit generalizability.



Figure 4. Confusion matrices of models trained on the small dataset and tested on the robustness testing dataset.



Figure 5. Confusion matrices of models trained on the large dataset and tested on the robustness testing dataset.

In the present study, the test locations for evaluating the model's performance in classifying different weed species growing in turfgrass across multiple states in the United States and provinces in China were selected. The model performed exceptionally well in scenarios featuring multiple weed species. It achieved a precision and recall of 1.0000 and 0.9732 for *D. ischaemum* and *O. corymbosa* on a golf course and roadside when trained on the small dataset, and 0.9946 and 1.0000 when trained on the large dataset. In a sod farm with *K. brevifolia, C. rotundus*, and *D. ischaemum*, the large dataset model outperformed the smaller dataset model. In a city park with *C. maculata* and *D. ischaemum*, both models demonstrated outstanding performance, with evaluation metrics exceeding 0.99. In conclusion, ResNeXt152 proves effective in accurately classifying multiple weed species across turf management regimes. The

experimental results demonstrate that data collected from a limited geographic area can be used to train a neural network capable of effectively classifying a wide variety of weed species in bermudagrass across different locations. While this study focused on bermudagrass, the methodology is potentially applicable to other cool- and warm-season turfgrass species. Future work will investigate its effectiveness for weed classification in these turfgrass types.

In summary, this study assessed the performance of eight classification neural networks trained with varying numbers of images. Through a comparative analysis, ResNeXt152 emerged as the most effective model among those evaluated. Additionally, this research highlights the practicality of utilizing a single neural network model for weed classification in turfgrass regimes with diverse uses across diverse geographic

					Testing res	ults ^c						
					Small datas	set				Large d	ataset	
Weed species ^b	Turfgrass site	Location	GPS coordinates	Photo acquisition dates	Precision	Recall	F ₁ score	MCC	Precision	Recall	F ₁ score	мсс
Digitaria ischaemum	Roadsides	Hefei, Anhui, CHN	(31.8611, 117.2808)	Jul. 2019	0.9926	1.0000	0.9963	0.9925	0.9824	1.0000	0.9911	0.9822
	Parks	Hefei, Anhui, CHN	(31.8611, 117.2808)	Aug. 2019	0.8343	1.0000	0.9097	0.8190	0.9932	1.0000	0.9966	0.9932
	Sod farms	Shuyang, Jiangsu, CHN	(34.1297, 118.7969)	Jun. 2019	0.9968	1.0000	0.9984	0.9968	0.9968	1.0000	0.9984	0.9968
	Sod farms	Shuyang, Jiangsu, CHN	(34.1297, 118.7969)	May 2021	0.8852	0.9585	0.9204	0.8380	1.0000	1.0000	1.0000	1.0000
	Roadsides	Nanjing, Jiangsu, CHN	(32.0603, 118.7969)	May 2021	0.8737	1.0000	0.9326	0.8693	0.9774	1.0000	0.9886	0.9771
	Commercial landscapes	Nanjing, Jiangsu, CHN	(32.0603, 118.7969)	May 2021	0.4167	0.3509	0.3810	0.3136	1.0000	0.5800	0.7342	0.6391
Paspalum dilatatum	Commercial landscapes	Auburn, AL, USA	(32.6097, —85.4808)	Oct. 2018	0.7333	0.9894	0.8423	0.6751	0.9686	0.9893	0.9788	0.9574
	Golf courses	Atlanta, GA, USA	(33.7490, —84.3880)	Oct. 2018	0.9520	1.0000	0.9754	0.9510	0.9958	1.0000	0.9979	0.9958
P. dilatatum (low weed density)	Commercial landscapes and roadsides	Tifton, GA, USA	(31.2304, —83.5237)	Aug. 2018	0.9867	1.0000	0.9933	0.9867	0.9733	0.9865	0.9799	0.9595
Murdannia nudiflora	Golf courses	Miami, FL, USA	(25.7617, —80.1918)	Oct. 2018	0.9901	1.0000	0.9950	0.9900	0.9756	1.0000	0.9877	0.9753
	Roadsides	Tifton, GA, USA	(31.2304, - 83.5237)	Oct. 2018	1.0000	0.9304	0.9639	0.9338	0.9936	1.0000	0.9968	0.9936
Oldenlandia corymbosa	Golf courses and roadsides	Orlando, FL, USA	(28.5384, —81.3789)	Oct. 2018	1.0000	0.9039	0.9495	0.9091	0.9913	0.9913	0.9913	0.9825
Kyllinga brevifolia	City parks and roadsides	Nanjing, Jiangsu, CHN	(32.0603, 118.7969)	Jun. 2019	0.8272	0.9437	0.8816	0.7578	1.0000	1.0000	1.0000	1.0000
Diodia virginiana	Golf course	Griffin, GA, USA	(33.2608,	Oct. 2018	1.0000	0.9531	0.9760	0.9543	1.0000	0.9950	0.9975	0.9950
Trifolium repens	Roadsides and parks	Shanghai, CHN	(31.2304, 121.4737)	May 2019	0.9751	0.9800	0.9776	0.9550	0.9950	1.0000	0.9975	0.9950
Hydrocotyle spp.	Golf courses	Lavaca County, TX, USA	(29.7536, —96.7143)	Sept. 2018	0.9005	0.8805	0.8904	0.7874	0.9955	0.9933	0.9944	0.9889
Richardia scabra	Golf courses	Miami, FL, USA	(25.7617, —80.1918)	Oct. 2018	1.0000	0.9898	0.9949	0.9897	1.0000	1.0000	1.0000	1.0000
Urochloa adspersa (high weed density)	Golf courses	Alafaya, FL, USA	(28.6039, —81.2074)	Sept. 2018	1.0000	1.0000	1.0000	1.0000	0.9949	1.0000	0.9975	0.9949
U. adspersa (low weed density)	Golf course	Alafaya, FL, USA	(28.6039, —81.2074)	Sept. 2018	0.9745	0.9948	0.9845	0.9696	1.0000	0.9948	0.9974	0.9948
U. adspersa (high weed density)	Golf course	Lakeland, FL, USA	(28.0394, —81.9516)	Sept. 2018	1.0000	0.9901	0.9950	0.9900	1.0000	1.0000	1.0000	1.0000

U. adspersa (low weed density)	Golf course	Lakeland, FL, USA	(28.0394, —81.9516)	Sept. 2018	0.7969	1.0000	0.8870	0.7661	0.9933	1.0000	0.9966	0.9933
O. corymbosa, D. ischaemum	Golf courses and roadsides	Homestead, FL, USA	(25.4672, —80.4533)	Oct. 2018	1.0000	0.9732	0.9864	0.9734	0.9946	1.0000	0.9973	0.9946
K. brevifolia, Cyperus rotundus, D. ischaemum	Sod farms	Shuyang, Jiangsu, CHN	(34.1297, 118.7969)	May 2019	0.9770	0.8563	0.9127	0.8425	0.9910	0.9565	0.9735	0.9484
Chamaesyce maculata, D. ischaemum	City parks	Nanjing, Jiangsu, CHN	(32.0603, 118.7969)	Aug. 2019	1.0000	0.9924	0.9962	0.9937	0.9924	1.0000	0.9962	0.9924
^a Neural network was trained using sma ^b Low density indicates weed coverage i fMCC, Matthews correlation coefficient.	ill and large dataset. n the images is visually less than 10% of	the total image pixels; hig	h density indicates t	he weed coverage in the in	nages is visua	lly more tha	ו 80% of th	ie total imag	ge pixels.			

regions in both China and the United States. The study demonstrated that ResNeXt152 achieved robust classification performance across 24 different locations, covering 6 turf sites with distinct uses and 14 weed species with varying densities and growth stages—all using a single training session. Future research will focus on expanding testing and validation to additional global locations and integrating the developed neural models into the machine vision subsystem of a smart sprayer prototype.

Funding statement. This work was supported by the Key R&D Program of Shandong Province, China (grant no. 202211070163), the National Natural Science Foundation of China (grant no. 32072498), the Taishan Scholars Program, China, and the Weifang Science and Technology Development Plan Project (grant no. 2024ZJ1097).

Competing interests. The authors declare no conflicts of interest.

References

- [USEPA] U.S. Environmental Protection Agency (2023a) Ingredients Used in Pesticide Products—Atrazine. http://www.epa.gov/ingredients-used-pestici de-products/atrazine. Accessed: February 27, 2025
- [USEPA] U.S. Environmental Protection Agency (2023b) Monosodium Methanearsonate (MSMA), an Organic Arsenical. https://www.epa.gov/ ingredients-used-pesticide-products/monosodium-methanearsonate-msmaorganic-arsenical. Accessed: February 27, 2025
- [WSSA] Weed Science Society of America (2023) Herbicide Mechanism of Action Classification. https://wssa.net/herbicides/herbicide-mechanism-ofaction. Accessed: February 27, 2025
- Ahmad Loti NN, Mohd Noor MR, Chang SW (2021) Integrated analysis of machine learning and deep learning in chili pest and disease identification. J Sci Food Agric 101:3582–3594
- Andrea CC, Daniel BBM, Misael, JBJ (2017) Precise weed and maize classification through convolutional neuronal networks. Pages 1–6 in Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting. Salinas, Ecuador: Institute of Electrical and Electronics Engineers
- Beckie HJ, Ashworth MB, Flower KC (2019) Herbicide resistance management: recent developments and trends. Plants 8:161
- Bhakta I, Phadikar S, Majumder K (2019) State-of-the-art technologies in precision agriculture: a systematic review. J Sci Food Agric 99:4878–4888
- Busey P (2003) Cultural management of weeds in turfgrass: a review. Crop Sci 43:1899–1911
- Chollet F (2021) Deep Learning with Python. 2nd ed. New York: Simon & Schuster. Pp 121–138
- Chowdhary KR (2020) Natural language processing. Pages 603–649 *in* Chowdhary KR, ed. Fundamentals of Artificial Intelligence. New Delhi: Springer.
- Dang F, Chen D, Lu Y, Li Z (2023) YOLOWeeds: a novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. Comput Electron Agric 205:107655
- dos Santos Ferreira A, Freitas DM, da Silva GG, Pistori H, Folhes MT (2017) Weed detection in soybean crops using ConvNets. Comput Electron Agric 143:314–324
- Gao SH, Cheng MM, Zhao K, Zhang XY, Yang MH, Torr P (2019) Res2net: a new multi-scale backbone architecture. IEEE Trans Pattern Anal Mach Intell 43:652–662
- Gerhards R, Andujar Sanchez D, Hamouz P, Peteinatos GG, Christensen S, Fernandez-Quintanilla C (2022) Advances in site-specific weed management in agriculture—a review. Weed Res 62:123–133
- Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York: Springer. Pp 9–41
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Pages 770–778 in Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV: Institute of Electrical and Electronics Engineers

Fable 4. (Continued)

- Jin X, Bagavathiannan M, Maity A, Chen Y, Yu J (2022a) Deep learning for detecting herbicide weed control spectrum in turfgrass. Plant Methods 18:94
- Jin X, Bagavathiannan M, McCullough PE, Chen Y, Yu J (2022b) A deep learning-based method for classification, detection, and localization of weeds in turfgrass. Pest Manag Sci 78:4809–4821
- Jin X, Liu T, Chen, Y, and Yu J (2022c) Deep learning-based weed detection in turf: a review. Agronomy 12:3051
- Jin X, Liu T, McCullough PE, Chen Y, Yu J (2023a) Evaluation of convolutional neural networks for herbicide susceptibility-based weed detection in turf. Front Plant Sci 14:1096802
- Jin X, Liu T, Yang Z, Xie J, Bagavathiannan M, Hong X, Chen Y (2023b) Precision weed control using a smart sprayer in dormant bermudagrass turf. Crop Prot 172:106302
- Kerr RA, Zhebentyayeva T, Saski C, McCarty LB (2019) Comprehensive phenotypic characterization and genetic distinction of distinct goosegrass (*Eleusine indica* L. Gaertn.) ecotypes. J Plant Sci Phytopathol 3:95–100
- Kraehmer H, Laber B, Rosinger C, Schulz A (2014) Herbicides as weed control agents: state of the art: I. Weed control research and safener technology: the path to modern agriculture. Plant Physiol 166:1119–1131
- Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Commun ACM 60:84–90
- Kudsk P, Streibig J (2003) Herbicides-a two-edged sword. Weed Res 43:90-102
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436-444
- Luo Q, Gao B, Woo W L, Yang Y (2019) Temporal and spatial deep learning network for infrared thermal defect detection. NDT E Int 108:102164
- Maqueda AI, Loquercio A, Gallego G, García N, Scaramuzza D (2018) Eventbased vision meets deep learning on steering prediction for self-driving cars. Pages 5419–5427 *in* Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT: Institute of Electrical and Electronics Engineers
- McElroy J, Martins D (2013) Use of herbicides on turfgrass. Planta Daninha 31:455-467
- Neal JC (2020) Turfgrass Weed Management—An IPM Approach Handbook of Integrated Pest Management for Turf and Ornamentals. Boca Raton, FL: CRC Press. Pp 275–292
- NVIDIA (2016) DetectNet: Deep Neural Network for Object Detection in DIGITS. NVIDIA Developer Blog. https://developer.nvidia.com/blog/dete ctnet-deep-neural-network-object-detection-digits/. Accessed: June 6, 2025
- Osorio K, Puerto A, Pedraza C, Jamaica D, Rodríguez L (2020) A deep learning approach for weed detection in lettuce crops using multispectral images. Agric Eng 2:471–488
- Saidi N, Kadir J, Hong LW (2016) Genetic diversity and morphological variations of goosegrass [*Eleusine indica* (L.) Gaertn] ecotypes in Malaysia. Weed Turfgrass Sci 5:144–154
- Sharpe SM, Schumann AW, Yu J, Boyd NS (2020) Vegetation detection and discrimination within vegetable plasticulture row-middles using a convolutional neural network. Precis Agric 21:264–277
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for largescale image recognition. arXiv:1409.1556

- Singh NS, Hariharan S, Gupta M (2020) Facial recognition using deep learning. Pages 375–382 in Proceedings of Advances in Data Sciences, Security and Applications: Proceedings of ICDSSA 2019. New Delhi, India: Bharati Vidyapeeth's College of Engineering
- Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inform Process Manag 45:427–437
- Stier JC, Steinke K, Ervin EH, Higginson FR, McMaugh PE (2013) Turfgrass benefits and issues. Pages 105–145 *in* Stier JC, Horgan BP, Bonos SA, eds. Turfgrass: Biology, Use, and Management. Madison, WI: American Society of Agronomy, Soil Science Society of America, Crop Science Society of America
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A (2015) Going deeper with convolutions. Pages 1–9 *in* Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA: Institute of Electrical and Electronics Engineers
- Xie S, Girshick R, Dollár P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. Pages 5987–5995 *in* Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognitio, Honolulu, HI: Institute of Electrical and Electronics Engineers
- Xie S, Hu C, Bagavathiannan M, Song D (2021) Toward robotic weed control: detection of nutsedge weed in bermudagrass turf using inaccurate and insufficient training data. IEEE Robot Autom Lett 6:7365–7372
- Yang GR, Wang XJ (2020) Artificial neural networks for neuroscientists: a primer. Neuron 107:1048–1070
- You J, Liu W, Lee J (2020) A DNN-based semantic segmentation for detecting weed and crop. Comput Electron Agric 178:105750
- Yu J, McCullough PE (2016) Efficacy and fate of atrazine and simazine in doveweed (*Murdannia nudiflora* L.). Weed Sci 64:379–388
- Yu J, Schumann AW, Cao Z, Sharpe SM, Boyd NS (2019a) Weed detection in perennial ryegrass with deep learning convolutional neural network. Front Plant Sci 10:1422
- Yu J, Schumann AW, Sharpe SM, Li X, Boyd NS (2020) Detection of grassy weeds in bermudagrass with deep convolutional neural networks. Weed Sci 68:545–552
- Yu J, Sharpe SM, Schumann AW, Boyd NS (2019b) Deep learning for imagebased weed detection in turfgrass. Eur J Agron 104:78–84
- Yu W, Luo M, Zhou P, Si C, Zhou Y, Wang X, Yan S (2022) Metaformer is actually what you need for vision. Pages 10809–10819 *in* Proceedings of the 2022 IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA: Institute of Electrical and Electronics Engineers
- Zhao J, Ma Y, Yong K, Zhu M, Wang Y, Luo Z, Huang X (2023) Deep-learningbased automatic evaluation of rice seed germination rate. J Sci Food Agric 103:1912–1924
- Zhuang J, Li X, Bagavathiannan M, Jin X, Yang J, Meng W, Chen Y (2022) Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. Pest Manag Sci 78: 521–529