Weed Science

www.cambridge.org/wsc

Deep learning–based detection and quantification of weed seed mixtures

Shahbaz Ahmed[1], Samuel R. Revolinski[2], P. Weston Maughan[3], Marija Savic[4], Jessica Kalin[4] and Ian C. Burke[5]

[1]Postdoctoral Researcher, Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA; [2]Assistant Professor, University of Kentucky, Department of Plant & Soil Science, Lexington, KY, 40503; [3]Graduate Research Assistant, Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA; [4]Research Associate, Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA and [5]Professor, Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA

Rapid Communication

Cite this article: Ahmed S, Revolinski SR, Maughan PW, Savic M, Kalin J, Burke IC (2024) Deep learning–based detection and quantification of weed seed mixtures. Weed Sci. **72**: 655–663. doi: 10.1017/wsc.2024.60

Received: 12 April 2024
Revised: 14 June 2024
Accepted: 20 August 2024
First published online: 21 November 2024

Associate Editor:
Muthukumar V. Bagavathiannan, Texas A&M University

Keywords:
Machine learning; mean average precision; object detection; seedbank quantification; YOLOv8

Corresponding author:
Shahbaz Ahmed;
Email: shahbaz.ahmed@wsu.edu

Abstract

Knowledge of weed seeds present in the soil seedbank is important for understanding population dynamics and forecasting future weed infestations. Quantification of the weed seedbank has historically been laborious, and few studies have attempted to quantify seedbanks on the scale required to make management decisions. An accurate, efficient, and ideally automated method to identify weed seeds in field samples is needed. To achieve sufficient precision, we leveraged YOLOv8, a machine learning object detection to accurately identify and count weed seeds obtained from the soil seedbank and weed seed collection. The YOLOv8 model, trained and evaluated using high-quality images captured with a digital microscope, achieved an overall accuracy and precision exceeding 80% confidence in distinguishing various weed seed species in both images and real-time videos. Despite the challenges associated with species having similar seed morphology, the application of YOLOv8 will facilitate rapid and accurate identification of weed seeds for the assessment of future weed management strategies.

Introduction

The soil weed seedbank is a repository of viable weed seeds within the soil profile that usually encompasses a diversity of weed species at various depths and densities (Teo-Sherrel et al. 1996). The weed seedbank is composed of recently shed seeds, as well as older seeds dating back several years (Buhler et al. 1997; Mahé et al. 2020). Common practices such as tilling can effectively shuffle the soil seedbank, depositing recently added seeds deeper into the soil profile, while at the same time lifting to the surface older, dormant seeds, that may have been there for many years (Feledyn-Szewcyzk et al. 2020). The presence of viable seeds in the soil can vary greatly; it is difficult and time-consuming to quantify and remains a significant challenge for both researchers and growers (Buhler et al. 1997; Mahé et al. 2020). Understanding the composition of the weed seedbank could be a critical component in management decisions if such information could be timely and accurately collected (Ambrosio et al. 2004; Creech et al. 2008; Luo et al. 2021; Mahé et al. 2020).

Identification of weed seeds at the species level presents challenges due to seed size and the often-inconspicuous appearance of seeds. Some species may require the germination of seeds or the use of molecular markers for accurate identification (Hussain et al. 2017). The challenge is further complicated by the presence of large volumes of diverse seeds that may be present in the soil seedbank – therefore, a single sample often consumes considerable processing time and requires specialized experience (Karlík and Poschlod 2014; Morgensen et al. 2005).

Developing an accurate and efficient identification method is essential for overcoming the challenges associated with weed seed identification and seedbank quantification. Automated solutions that can handle diverse weed species with complex morphology and structures across different regions and environmental conditions would substantially increase the throughput, facilitating the processing of a far greater number of samples. The incorporation of advanced technologies such as the use of artificial intelligence for object detection in agriculture has opened avenues for addressing challenges more efficiently (Eli-Chukwu 2019; Xu et al. 2021). Object detection models offer a valuable alternative to traditional methods such as germination assays or manual seed counting for the task of quantifying seedbanks.

Object detection with machine learning is increasingly used in crop research due to its ability to learn distinct features for precise identification and localization of objects in an image or video frame (Wosner et al. 2021; Wu et al. 2019; Zhang et al. 2022). Such models have been successfully deployed to detect and identify diseases, pests, and weeds and for crop monitoring

© The Author(s), 2024. 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.

and growth analysis (Alfonso et al. 2020; Khalid et al. 2023; Pérez-Porras et al. 2023; Verma et al. 2021; Yao et al. 2024). Object detection models use convolutional neural networks (CNNs), a class of deep-learning algorithms, to detect various objects in images. CNN architecture, inspired by the human brain, works by employing convolutional layers to learn the hierarchical features of input images (Alzubaidi et al. 2021). The series of layers in a CNN use kernels or filters to scan the input data to detect features of an image (Alzubaidi et al. 2021). Image classification models, such as GoogLeNet and AlexNet, have been deployed to identify seeds with high accuracy; however, these image classification methods were based on either a single seed or single species at a time (Gulzar et al. 2020; Luo et al. 2021).

You Only Look Once (YOLO), a popular object detection model, incorporates CNNs to predict bounding boxes and class probabilities all in one forward pass, facilitating rapid and efficient detection with high accuracy (Jocher et al. 2023). Several versions of YOLO have been released, with each model building upon addressing the limitations of its predecessor (Jiang et al. 2022). Different object detection models can be evaluated for their accuracy of prediction using mean average precision (mAP). mAP facilitates an evaluation of the accuracy and recall, offering the average precision (AP) of the method (Zhao et al. 2019). The most recent version, YOLOv8, demonstrates an improved mAP of 53.9% on the Microsoft Common Objects in Context dataset (Lin et al. 2014), a widely used benchmark dataset in computer vision research, surpassing the AP of YOLOv5, which achieved 50.7% (Lee and You 2024). Recently, YOLO models have been highly successful in identifying weed plants within cornfields (Hasan et al. 2024).

In this study, we employ a YOLOv8 model to streamline and accelerate the identification process of weed seeds. The objective of this study is to assess the efficacy of a deep-learning approach in accurately detecting weed seeds, thereby establishing a robust identification workflow. As proof of concept, we conducted comprehensive training and testing of the model using seed data from 19 distinct weed species, ensuring thorough validation of its performance across diverse samples. Through our research, we aim to develop a solution that allows for efficient assessments of large weed seedbank samples, significantly reducing the time required compared with conventional methods.

Material and Methods

Dataset Preparation

The red-green-blue (RGB) images for the weed seedbank dataset were obtained using an IVESTA3 digital microscope (Leica Microsystems, 10 Parkway N, Suite 300, Deerfield, IL 60015) with an external light source. The microscope was stationed on a stable platform with a constant working distance for all seed types. The image processing was optimized by using 0.61× object lens for precise focus. A total of 485 images were captured at a resolution of 4,000 by 3,000 with a pixel size of 1.55 µm by 1.55 µm using a high-resolution camera integrated in the microscope body. The image dataset included 19 weed species, that are common in eastern Washington, elutriated from soil samples from field experiments conducted in the region (Table 1). All the seeds analyzed in these experiments were obtained and cleaned from elutriated samples and weed seed collection; thus, no seeds were added to the soil. By imaging real samples collected in-field, this project demonstrates the direct utility of the methods presented here for finding and identifying seeds from weeds in realistic datasets. The dataset included weed seeds of different sizes, shapes, and colors, accounting for the wide spectrum of weed species encountered in eastern Washington dryland cropping systems. Images present in the dataset were separated into two categories. The first category consisted of images containing seeds of single weed species to carry out a focused training for individual species, while the images in the second category featured the coexistence of multiple weed species within single frame to train the model to identify and differentiate various weed species together.

Data Labeling

The images in this project were labeled manually using labelImg (Tzuta 2015) to create ground-truth data. The approach facilitated identification of weed seed in images by drawing bounding boxes around the individual seeds in the image. The output of the annotation was saved a text format. The output generated from this exercise represents annotated information for a specific bounding box. The format includes class ID indicating the numeric identifier for the specific weed species and bounding box coordinates (x_center, y_center, width, height), providing the position and dimensions of the bounding box relative to the image dimensions (Table 2; Figure 1). The annotated dataset generated using labelImg was crucial in teaching the model to effectively identify and classify different weed species.

Model Training and Validation

The annotated dataset generated through the labeling process in the previous step was utilized to train the YOLOv8 model. This dataset, consisting of images of various weed instances and their corresponding bounding box annotations, served as the training ground truth. To facilitate the training and cross-validation process, the annotated dataset was split into two subsets: a training and validation set, with 80% of data allocated to the training set. The remaining 20% of data was allocated to the validation set to assess model performance in a 20% hold-out cross-validation scheme. The validation set of the hold-out data scheme is data set aside and not used for training the model, so the model can be assessed on images that were not used to train it, ensuring that the reported accuracy of the model is not due to overfitting. The training process was set to 100 epochs with an image size of 640 pixels (width and height). YOLOv8 utilizes an image size of 640 by 640 pixels to optimize computational efficiency and speed during object detection. The fixed input size enables faster processing and inference times compared with larger image sizes. However, the reduction in resolution may affect the model's ability to accurately detect and localize objects, particularly smaller ones or those requiring finer details. During the training process, YOLOv8 leveraged a transfer learning strategy by initializing its weights with pretrained values from Common Objects in Context (Lin et al. 2014), which contains 200,000 images for object detection. Transfer learning, also known as weight transfer, is a technique commonly used in deep learning, where knowledge gained from training on one task or dataset is transferred to another related task or dataset. Abstract data features learned by CNNs for classification problems often can be applied to other classification problems, but learning these abstract features can initially require hundreds of thousands of training images. Thus, using pretrained networks can lead to CNNs with high accuracy without requiring hundreds of thousands of training images. Transfer learning will be essential for weed scientists to use for deploying deep-learning

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press

Table 1. List of species with instances used in a study to develop a machine learning program for weed seed identification.

Species			
Common name	Scientific name	EPPO code	Instances
Italian ryegrass	*Lolium perenne* L. ssp. *multiflorum* (Lam.) Husnot	LOLMU	80
Wild mustard	*Sinapis arvensis* (L.) Andrz. Ex Besser	SINAR	23
Catchweed bedstraw	*Galium aparine* L.	GALAP	15
Flixweed–tumble mustard	*Descurainia sophia* (L.) Webb ex Prantl/ *Sisymbrium altissimum* L.	DESSO-SSYAL	33
Common lambsquarters	*Chenopodium album* L.	CHEAL	58
Mayweed chamomile	*Anthemis cotula* L.	ANTCO	467
Kochia	*Bassia scoparia* (L.) A.J. Scott	KCHSC	55
Prickly lettuce	*Lactuca serriola* L.	LACSE	14
Prickly lettuce–annual sowthistle	*Lactuca serriola* L./ *Sonchus oleraceus* L.	LACSE-SONOL	48
Pigweed	*Amaranthus* spp.	AMARE	415
Rattail fescue	*Vulpia myuros* (L.) C.C. Gmel.	VLPMY	61
Cheatgrass	*Bromus tectorum* L.	BROTE	47
Russian thistle	*Salsola tragus* (L.) Scop.	SASKT	25
Brassicas	*Brassica* spp.		296
Wheat	*Triticum aestivum* L.	TRZAX	5
Henbit	*Lamium* spp.	LAMAM	77
Nightshade species	*Solanum* spp.	SOLDU	30
Bull thistle	*Cirsium vulgare* (Savi) Ten.	CIRVU	4

Table 2. Example of bounding box data from.txt output generated by labelImg during labeling process.

		Bounding box center			
Species code	Class index	*x* coordinate	*y* coordinate	Width	Height
LAMAM[a]	15	0.874625	0.323167	0.06525	0.108333
AMARE[b]	9	0.6725	0.023667	0.056	0.044
AMARE	9	0.51925	0.248	0.0515	0.063333
AMARE	9	0.3765	0.572167	0.053	0.061667
AMARE	9	0.29775	0.600167	0.0385	0.062333
AMARE	9	0.600125	0.782667	0.04075	0.053333
AMARE	9	0.788875	0.595833	0.03775	0.059667
AMARE	9	0.808125	0.911333	0.04425	0.044

[a]Henbit (*Lamium* spp.).
[b]Pigweed (*Amaranthus* spp.).

models, because weed scientists will seldom have hundreds of thousands of images for training models, but using previous model weights classification models can be developed using much smaller image collections. The training process used here involved further fine-tuning of these weights using the weed seed dataset. The model was trained to minimize focal loss function for the bounding boxes, and the stochastic gradient descent loss function was used for the classification, thus refining the model's ability to differentiate between weed species.

Performance Metrics

The model performance was assessed by using precision, recall, F1 score, and the mAP. Precision is calculated as the ratio of true positive predictions to the total number of actual positives (Equation 1):

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad [1]$$

Recall represents the ability of model to capture all positive instances. It is calculated as the ratio of true positive predictions to the total number of actual positives (Equation 2:

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad [2]$$

F1 score provides a measure of model performance on both precision and recall by calculating the harmonic mean of precision and recall (Equation 3):

$$\text{F1 score} = 2{*}\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad [3]$$

mAP compares detected box with the ground-truth bounding box to return a score. A higher mAP value denotes the accuracy of the model detection. The first step in the mAP calculation requires the estimation of AP of each class. The mean of these APs for all classes is used to produce mAP. The equation to calculate mAP is (Equation 4):

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \quad [4]$$

These performance metrics play a crucial role in evaluating the effectiveness of the model in detecting and classifying objects, in this case, weed seeds. Precision measures the proportion of correctly identified positive cases out of all cases identified as

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press

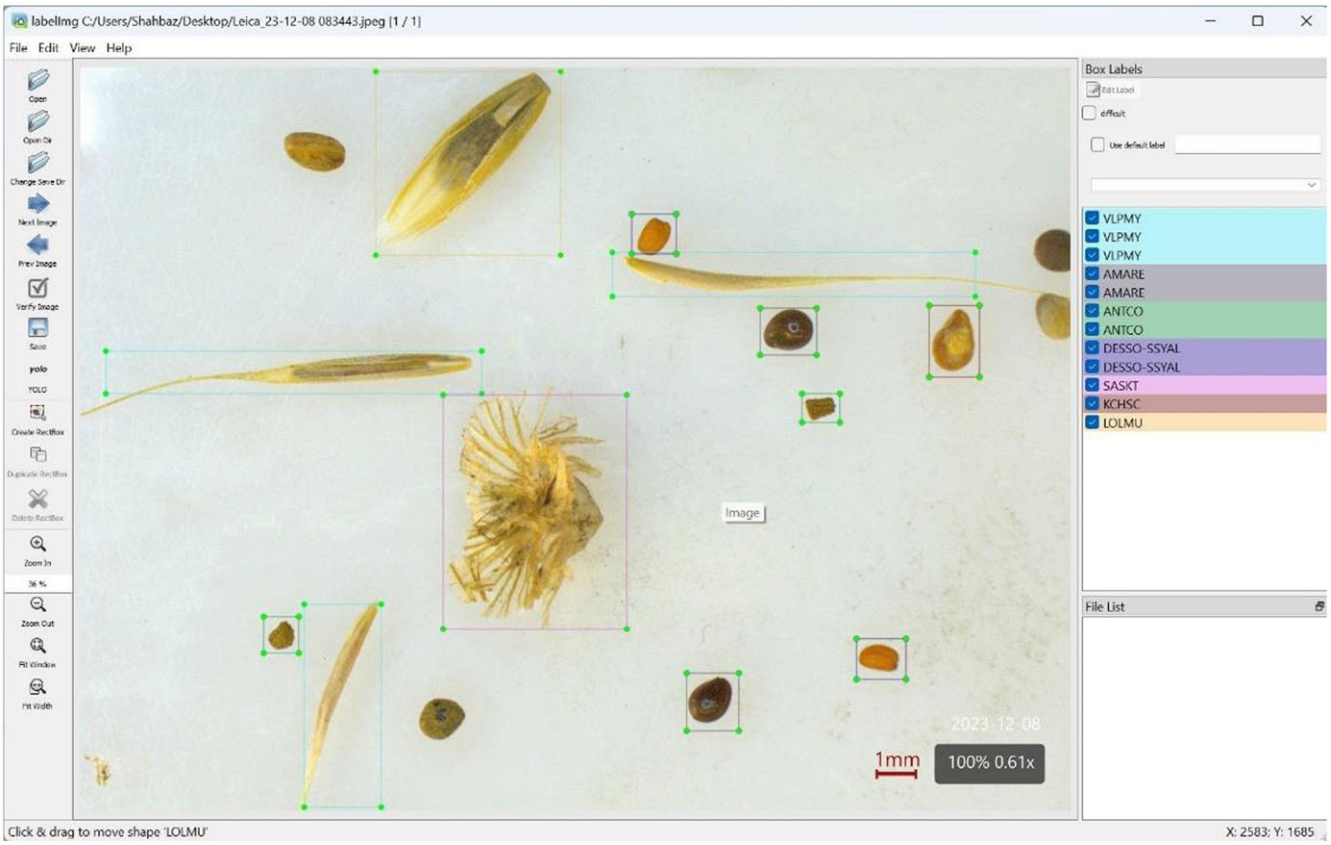


Figure 1. labelImg-generated bounding boxes drawn around weed seeds belonging to different classes. Each colored box and number corresponds to a distinct class of weeds. Bounding boxes aid in identifying and localizing weed seeds within images, assisting in the accurate classification and detection of various weed species. Understanding the representation of bounding boxes is fundamental for interpreting object detection results.

positive, providing insight into the ability of model to avoid false positives. Recall, on the other hand, assesses the model capability to capture all positive instances, indicating its sensitivity to true positives. The F1 score balances precision and recall, providing a single metric to gauge overall model performance.

AP and mAP extend this assessment to object detection tasks by evaluating the accuracy of bounding box predictions. AP calculates the precision–recall curve for each class, while mAP aggregates AP values across all classes to provide a comprehensive measure of detection accuracy. Together, these metrics offer a detailed understanding of the model's strengths and weaknesses, enabling researchers to fine-tune algorithms for improved performance.

Additionally, a normalized confusion matrix was used to examine the levels of misclassification between species. A confusion matrix is a matrix with output (predicted) classes along the *y* axis and the target (true) classes along the *x* axis, with each cell containing the number of misclassifications between classes. The confusion matrix was then normalized by dividing the value in each column by the total number of incidences of the class for which the column represents the true values. In our study, the normalized confusion matrix is used to examine patterns of misclassification between the seeds of weed species.

Results and Discussion

Image Acquisition

Images obtained from the microscope played a pivotal role in enhancing the accuracy of seed labeling, particularly for small seeds. The microscopic images provided clear and discernible details, enabling the differentiation of even small and dark-colored seeds. Importantly, the use of a microscope ensured that images were sufficiently large during labeling, eliminating the need to zoom in to locate specific seed types and draw labels around them. The aspect is critical, as zooming in during manual labeling may result in inaccuracies that can affect the performance of the detection model during inference. Therefore, the availability of high-quality images from the microscope not only facilitated straightforward labeling but also contributed to ensuring the robustness and accuracy of the subsequent detection process. It is also important to recognize that YOLOv8 typically resizes input images to a smaller size, usually 640 by 640 pixels. Resizing can lead to a loss of information, especially for small objects in images. In some cases, the downsized images may make small objects undetectable to the model, presenting a challenge for accurate detection. Images from a microscope for small weed seeds ensure that they appear larger, so that when the model resizes them, they remain sufficiently detailed for detection.

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press

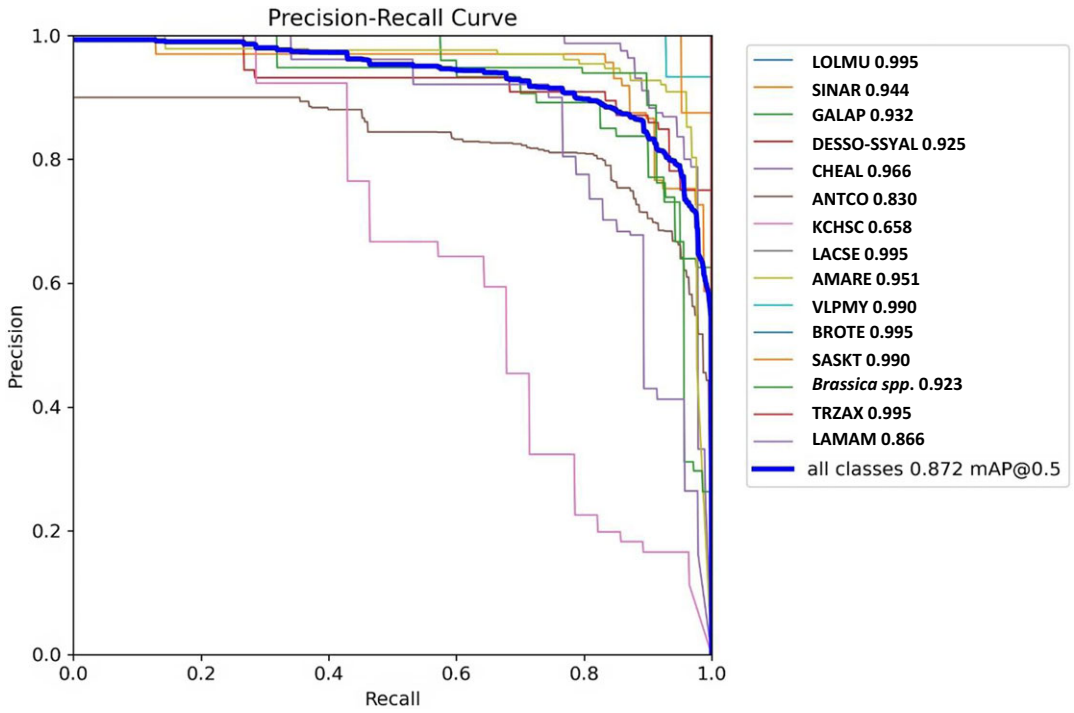


Figure 2. The precision–recall curve generated during evaluation of weed species detection model. The *x* axis represents the recall, which measures the proportion of true positive detections out of all actual positive instances. The *y* axis represents precision, indicating the proportion of true positive detections out of all predicted positive instances. Each point on the curve corresponds to a different threshold used for classification, with the curve representing the trade-off between precision and recall. A higher area under the curve (AUC) indicates improving model performance in distinguishing between weed species. Abbreviations: LOLMU, *Lolium perenne* L. ssp. *multiflorum* (Lam.) Husnot; SINAR, *Sinapis arvensis* (L.) Andrz. ex Besser; GALAP, *Galium aparine* L.; DESSO-SSYAL, *Descurainia sophia* (L.) Webb ex Prantl/*Sisymbrium altissimum* L.; CHEAL, *Chenopodium album* L.; ANTCO, *Anthemis cotula* L.; KCHSC, *Bassia scoparia* (L.) A.J. Scott; LACSE, *Lactuca serriola* L.; AMARE, *Amaranthus* spp.; VLPMY, *Vulpia myuros* (L.) C.C. Gmel.; BROTE, *Bromus tectorum* L.; SASKT, *Salsola tragus* (L.) Scop.; *Brassica* spp.; TRZAX, *Triticum aestivum* L.; LAMAM, *Lamium* spp.

Species Classification Metrics

The performance of the YOLOv8 model was evaluated through classification matrices, precision recall curve, precision, recall, and mAP. The precision recall curve and area under the precision recall curve (AUC) for all classes of weed species in the dataset is used to assess weed seed identification performance, where a higher AUC indicates greater performance (Figure 2). We observed variability in the AUC values across weed species, indicating varying levels of classification performance. For instance, cheatgrass (*Bromus tectorum* L.) exhibited a higher AUC (0.995) compared with kochia [*Bassia scoparia* (L.) A.J. Scott] (0.658), suggesting that the model achieved better discrimination for *B. tectorum*. The lower AUC for *B. scoparia* compared with *B. tectorum* indicates a comparatively lower level of discriminative ability of the model, suggesting that it had reduced capability in distinguishing *B. scoparia* from other weed species in the dataset compared with *B. tectorum*.

Despite having a larger number of training images for *B. scoparia* compared with *B. tectorum*, the lower ability of the model to accurately identify and discriminate *B. scoparia* seeds (as indicated by an AUC of 0.66) can be attributed to several factors, including the inherent characteristics of the weed species and the challenges they pose for object detection models. *Bassia scoparia* seed, being smaller in size and lacking distinct shape features compared with *B. tectorum*, presents a greater difficulty for the model in accurately recognizing and delineating its boundaries. The limited visual cues and variations in appearance may hinder the model's ability to differentiate *B. scoparia* from other weed species present in the dataset. Additionally, the smaller size of *B. scoparia* seeds may result in less prominent features, making them harder to detect amid cluttered backgrounds or similar-looking objects.

In contrast, *B. tectorum*, with its larger size and more pronounced shape characteristics, provides clearer visual cues for the model to identify and distinguish (AUC of 0.995). The distinctiveness of *B. tectorum* seeds facilitates better recognition and classification, despite the smaller number of training images available. The larger size and more-defined shape of *B. tectorum* seeds likely contribute to higher accuracy in detection compared with *B. scoparia*. It is essential to consider the complexity of the training process and the representation of different classes within the dataset. While the number of training images is important, the quality and diversity of these images also play a crucial role in model performance. The presence of diverse examples and variations within the *B. tectorum* training set may have enabled the model to learn more robust features for accurate detection, despite the smaller quantity of data.

Precision measures the proportion of correctly identified positive cases out of all cases identified as positive, providing insight into the ability of model to avoid false positives (Figure 3). The precision of the model for bounding box prediction improved considerably from 0.43 at epoch 1 to 0.82 at epoch 100. The increased precision value indicates the model's ability to make accurate predictions with minimum false positives. Recall, unlike precision, gauges the model's ability to identify all positive instances, reflecting its sensitivity to true positives. For instance, a recall value of 0.76 signifies the model's effectiveness in capturing true positives. The mean mAP at 50% intersection over union (IoU) (mAP50), a crucial metric for object detection, increased consistently with each training iteration, reaching a value of 0.82 by epoch 100. mAP50 evaluates the AP of the model across different object classes when there is at least a 50% IoU between predicted

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press

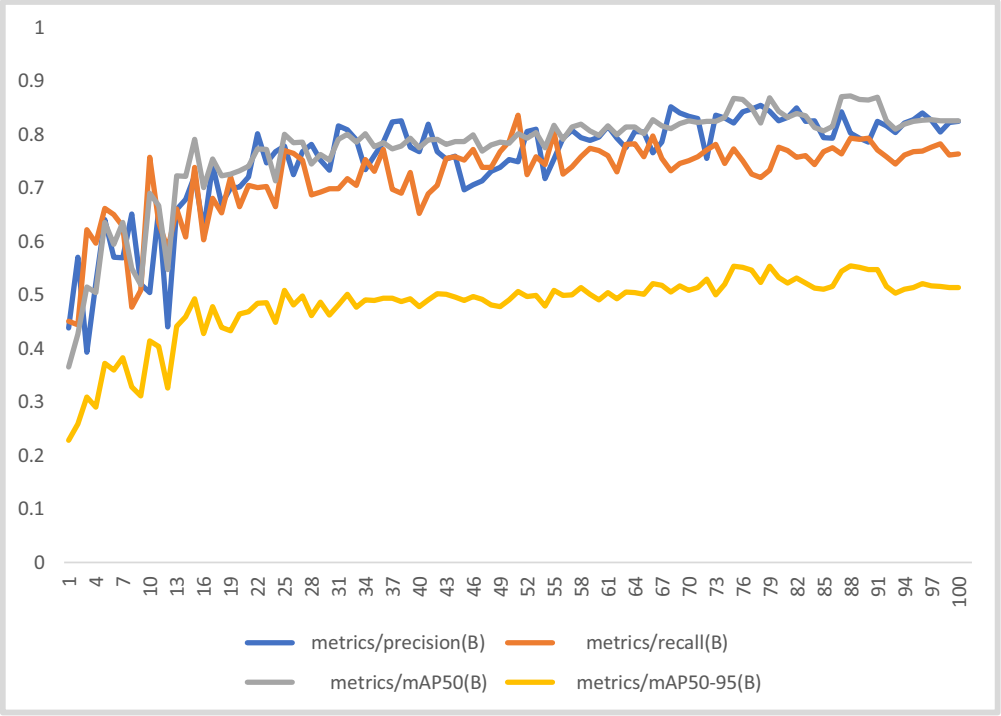


Figure 3. Evolution of performance metric precision, recall, mean average precision calculated at an intersection over union threshold of 0.50 (mAP50), and the average of the mean average precision calculated at varying intersection over union thresholds, ranging from 0.50 to 0.95 (mAP50-95) over the course of training epochs from 0 to 100. The x axis represents each epoch, indicating the progress of the model during training. The y axis showcases the improvement in performance metrics with each epoch, providing insights into how the model's accuracy, recall rate, and mean average precision vary throughout the training process. Understanding these metrics is crucial for assessing the effectiveness and progress of the object detection model.

bounding boxes and ground-truth boxes (precise location and class of object), indicating how well the model detects objects with a moderate level of overlap with the ground truth. Additionally, mAP50-95, which considers mAP at various IoU thresholds from 50% to 95%, improved from 0.22 to 0.51 over the course of training, highlighting the learning capacity of the model over different rounds of training (Figure 3). The improvement in these values is important to handle variation present between weed species. In addition to the accuracy, the YOLOv8 model also displayed an impressive speed in processing both images and real-time videos, making it suitable for large-scale applications and for extension functions where real-time identification would be useful for training and demonstration.

The progressive increase in the precision signifies the ability of the model to distinguish between different species, contributing to minimizing false positives. Adaptability along with processing speed is crucial for real-world field application in handling large datasets. However, despite these advancements, it is important to acknowledge potential challenges associated with the model's performance. For instance, different weed species exhibit varying precision scores based on factors such as size and color, with lower precision for small seeds.

Detection Performance

The YOLOv8 model performed remarkably well in detecting, counting, differentiating, and classifying seeds from cleaned field samples using both still images and real-time videos for the majority of weed species. Challenges did arise in cases of increased similarity between seed morphology and color, such as between flixweed [*Descurainia sophia* (L.) Webb ex Prantl] and tumble mustard (*Sisymbrium altissimum* L.) (Figure 4). The constraint, however, was limited to seed pairs that were within the same taxonomic family, such as *S. altissimum* and *D. sophia* or annual sowthistle (*Sonchus oleraceus* L.) and prickly lettuce (*Lactuca serriola* L.) We anticipate that similar constraints will be identified with *Amaranthus* spp. seed.

The normalized confusion matrix (Figure 5) depicts a clear pattern where species with similar seeds in size are frequently confused with one another. While common lambsquarters (*Chenopodium album* L.), a small-seeded Chenopodiaceae species, was never identified as another species, other small-seeded species were frequently misidentified as *C. album*. In the hold-out validation, 32% of *B. scoparia*, 11% of background material, 5% of *S. oleraceus*, 3% of mayweed chamomile (*Anthemis cotula* L.), and 3% of *Amaranthus* spp. were misclassified as *C. album* (Figure 5). Interestingly, *B. scoparia* represented the direct opposite of *C. album*. Misclassification in *B. scoparia* was high, with 50% of samples being misidentified in the hold-out validation, as 32% was identified as *C. album*, 4% as *A. cotula*, and 14% as background material (Figure 5). Additionally, there was a high rate of background material misclassified as weed seeds. None of the background material was classified as background material, with 54% classified as *A. cotula*, 11% as *C. album*, 9% as *Lamium* spp., 8% as *Amaranthus* spp., 5% as Italian ryegrass (*Lolium multiflorum* Lam.), 4% as both catchweed bedstraw (*Galium aparine* L.) and *S. altissimum*, 3% as *Brassica* spp., 2% as *B. scoparia*, and 1% as both rattail fescue [*Vulpia myuros* (L.) C.C. Gmel.] and *Solanum* spp., indicating that the model is overly sensitive to false positives when encountering background material (Figure 5). However, the model was able to identify large-seeded weeds with nearly 100%

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press

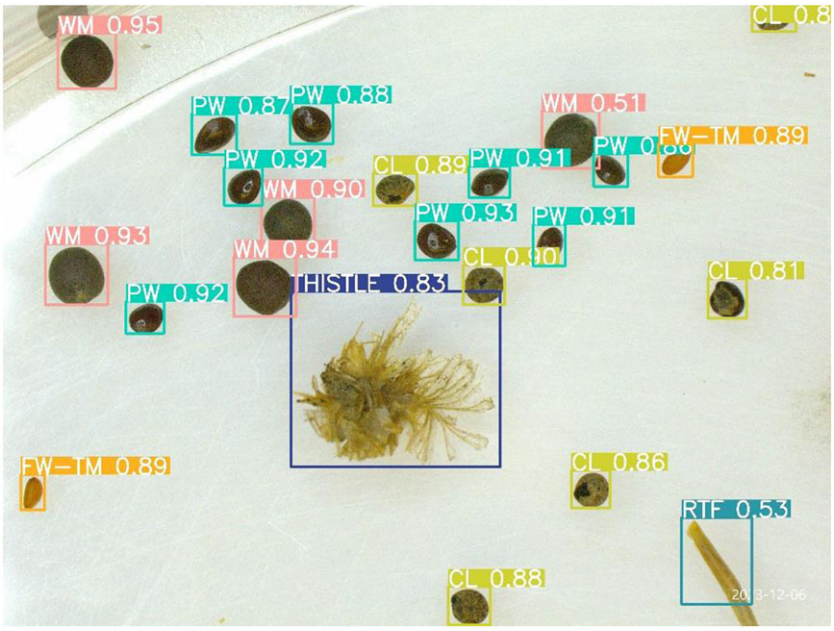


Figure 4. Example detection results generated by the model, with bounding boxes drawn around different weed seeds. Each bounding box indicates the location and size of a detected seed with a confidence score. The confidence score is a measure of the model's certainty in its prediction, ranging from 0 to 1, with 1 indicating complete certainty. In this case, for example, a score of 0.95 suggests a very high level of confidence in the accuracy of the detection and identification.

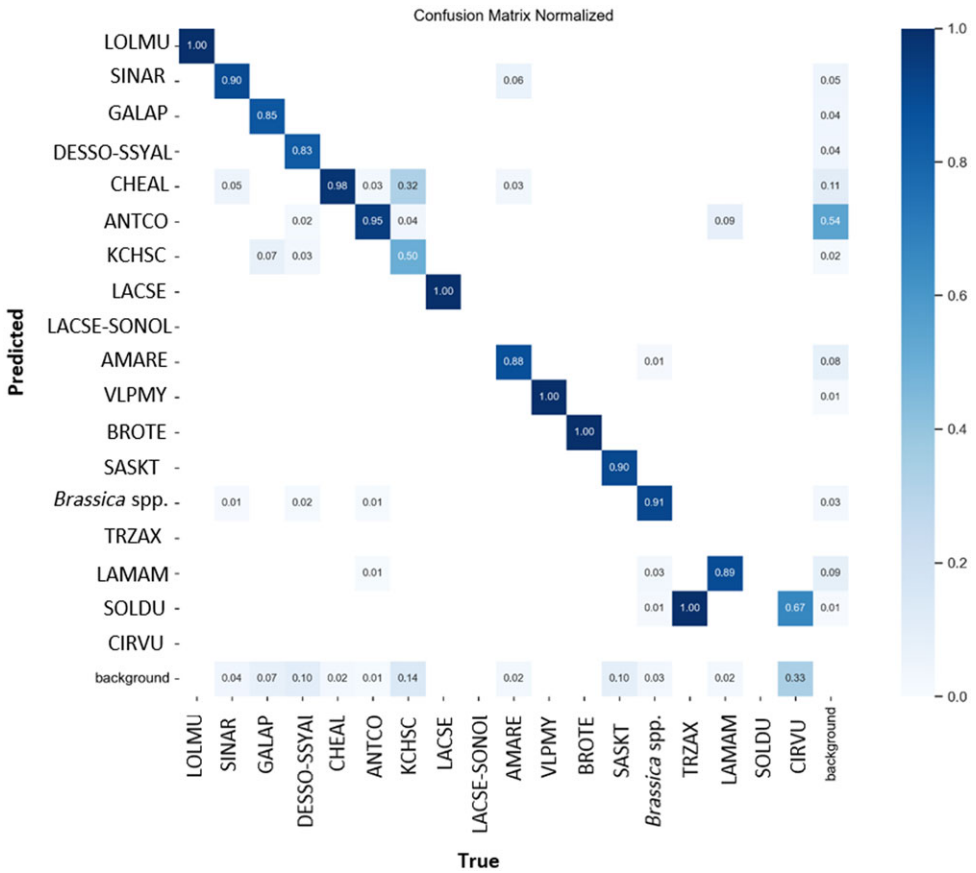


Figure 5. Normalized confusion matrix of seed classification. Rows represent predictions of species while columns represent the true species. Total incidences of miss classification are divided by the number of incidences of each class. Abbreviations: LOLMU, *Lolium perenne* L. ssp. *multiflorum* (Lam.) Husnot; SINAR, *Sinapis arvensis* (L.) Andrz. ex Besser; GALAP, *Galium aparine* L.; DESSO-SSYAL, *Descurainia sophia* (L.) Webb ex Prantl/*Sisymbrium altissimum* L.; CHEAL, *Chenopodium album* L.; ANTCO, *Anthemis cotula* L.; KCHSC, *Bassia scoparia* (L.) A.J. Scott; LACSE, *Lactuca serriola* L.; LACSE-SONOL, *Lactuca serriola* L./*Sonchus oleraceus* L.; AMARE, *Amaranthus* spp.; VLPMY, *Vulpia myuros* (L.) C.C. Gmel.; BROTE, *Bromus tectorum* L.; SASKT, *Salsola tragus* (L.) Scop.; *Brassica* spp.; TRZAX, *Triticum aestivum* L.; LAMAM, *Lamium* spp.; SOLDU, *Solanum* spp.; CIRVU *Cirsium vulgare* (Savi) Ten.

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press

accuracy (Figure 5). While the model is useful for classifying and counting seeds from larger-seeded grass species, caution should be taken when using the model for smaller-seeded weeds when there is non-seed material present.

The ability of the model to precisely identify different species differed based on the size and the color of the seed. For example, the model performed with increased precision on larger seeds with distinguishing features than on smaller seeds with similar color spectra. Because we only used 19 weed species to evaluate performance, the risk of misclassification escalates as the number of weed species to be identified increases, especially if the model encounters challenges in distinguishing between morphologically similar species, or when morphology varies significantly within species. Addressing these challenges requires continued research and refinement of the model, potentially incorporating additional features or training strategies to improve performance across a broader range of weed species. Furthermore, ongoing efforts to expand the diversity of the training dataset can enhance the model's ability to generalize to novel weed species, mitigating the risk of misclassification in practical agricultural scenarios. While the YOLOv8 model demonstrates considerable promise in weed seed detection, ongoing optimization and adaptation are necessary to address the inherent complexities of real-world weed management applications.

Application

Rapid and accurate quantification of the weed seedbank is a critical capability gap in weed science. Utilizing high-throughput systems, including automated weed seed identification, renders many of the research needs on weed seedbank dynamics that have been proposed for decades potentially feasible (Buhler et al. 1997; Khan et al. 2021). Spatially explicit quantification of weed seedbanks, obtained by collecting soil samples across a landscape and using YOLOv8 to classify and count weed seeds in images, could be utilized to understand the relationship between the seedbank and final weed populations, the emergence dynamics of weed species, the effects of management practices on weed seed dynamics, and the empirical effect of management inputs on weed seedbank and weed densities. The use of such systems could also be grower targeted—with just a few samples of soil at the time of planting or tilling, growers could make critical management decisions earlier, enhancing their integrated weed management programs or increasing the impact of decision support systems. Because YOLOv8 is compatible with any three-band images such as RGB images, it will be applicable to identifying seeds with any RGB camera that has adequate resolution, but likely will require a microscope.

Although the YOLOv8 model is adequate for many tasks in weed seedbank assessment, it has critical limitations. First, YOLOv8 cannot identify bounding boxes of overlapping objects. If many seeds are overlapping in an image, the YOLOv8 model likely would not be able to accurately count the number of seeds of each species. This limitation can be overcome by splitting dense samples into multiple images and lightly shaking the paper to spread the seeds adequately, so they are not overlapping.

Second, many of the smaller weed seeds appear as specks in images, making it difficult for the classifier to distinguish them from background material. Another challenge with samples from elutriators is the presence of soil coating on the objects, giving them a similar color and making it difficult for the model to make accurate predictions. While it is unlikely that elutriators will ever be able to remove all background material from samples, a second washing along with higher-resolution images and a specific classifier that separates small seeds from background material could alleviate the misclassification between small seeds and background material.

The integration of high-resolution microscopy images was pivotal in carrying out the object detection analysis for small weed seeds and enhancing the overall robustness of the detection process. While handheld devices such as phones or scanners offer a faster alternative for image acquisition, the small size of the seeds led to inaccuracies in labeling and detection. Images from a microscope provided consistent and precise magnification, eliminating the need for additional programming to adjust for scale variations. This ensures accurate and reliable measurements, which are often a challenge with smartphone images due to variable camera settings and distances. Evaluation of metrics further highlighted the inherent challenges associated with small and similarly colored seeds, emphasizing the importance of dataset quality and diversity in model training. While the YOLOv8 model demonstrated high accuracy in detection and classification tasks, challenges persist in accurately distinguishing between morphologically similar species and addressing variations in seed size and color, particularly in seedbank samples with substantial soil particles obtained directly after the elutriation step. Incorporating images of seedbank samples taken post-elutriation into the dataset would likely improve detection accuracy. This limitation could not be addressed in the current study due to time constraints and the unavailability of sufficiently diverse elutriated seedbank images for training. Addressing these challenges will require continued research, refinement of the model, and expansion of the diversity of the training dataset. Despite these limitations, this work presents proof of concept that the integration of automated weed seed identification systems holds great promise for driving innovation and improving integrated weed management systems.

Acknowledgments. We thank Cody Willmore and Vishal Sonawane for their assistance.

Funding statement. The study was supported by the Washington Grain Commission, the R. J. Cook Endowment for Wheat Research, and Pacific Northwest Herbicide Resistance Initiative.

Competing interests. The authors declare no competing interests.

References

Alfonso M, Fonteijn H, Fiorentin FS, Lensink D, Mooij M, Faber N, Polder G, Wehrens R (2020) Tomato fruit detection and counting in greenhouses using deep learning. Front Plant Sci 11:571299

Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, Santamaría J, Fadhel MA, Al-Amidie M, Laith Farhan L (2021) Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8:53

Ambrosio L, Iglesias L, Marin C, Del Monte JP (2004) Evaluations of sampling methods and assessments of the sample size to estimate the weed seedbank in the soil taking into account spatial variability. Weed Res 44:224–236

Buhler D, Hartzler R, Forcella F (1997) Implications of weed seedbank dynamics to weed management. Weed Sci 45:329–336

Creech JE, Westphal A, Ferris VR, Faghihi J, Vyn TJ, Santini JB, Johnson WG (2008) Influence of winter annual weed management and crop rotation on soybean cyst nematode (*Heterodera glycines*) and winter annual weeds. Weed Sci 56:103–111. https://doi.org/10.1614/WS-07-084.1

Eli-Chukwu NS (2019) Applications of artificial intelligence in agriculture: a review. Eng Technol Appl Sci Res 9:4377–4383

Feledyn-Szewcyzk B, Smagacz J, Kwiatkowski CA, Harasim E, Woźniak A (2020) Weed flora and soil seedbank composition as affected by tillage system in three-year crop rotation. Agriculture 10:186

Gulzar Y, Hamid Y, Soomro AB, Alwan AA, Journaux L (2020) A convulsion neural network-based seed classification system. Symmetry 12:2018

Hasan ASMM, Diepeveen D, Laga H, Jones MGK, Sohel F (2024) Object-level benchmark for deep learning-based detection and classification of weed species. Crop Prot 177:106561

Hussain M, Ali S, Tahir MN, Shah GA, Ahmed I, Sarwar MA, Latif S (2017) A comparative study of soil weed seedbank determination in Pothwar region by using different methodologies. Pak J Agric Res 30:310–315

Jiang P, Ergu D, Liu Fangyao, Cai Y, Ma B (2022) A review of YOLO algorithm developments. Procedia Comput Sci 199:1066–1073

Jocher G, Chaurasia A, Qiu J (2023) Ultralytics YOLO. Version 8.0.0. https://docs.ultralytics.com/models/yolov8/. Accessed: October 11, 2023

Karlík P, Poschlod P (2014) Soil seed-bank composition reveals the land-use history of calcareous grasslands. Aca Oecol 58:22–34

Khalid S, Oqaibi HM, Aqib A, Hafeez Y (2023) Small pests detection in field crops using deep learning object detection. Sustainability 15:6815

Khan S, Tufail M, Khan MT, Khan ZA, Anwar S (2021) Deep learning-based identification system of weeds and crops in strawberry and pea fields for precision agriculture sprayer. Precis Agric 22:1711–1727

Lee JH, You SJ (2024) Balancing privacy and accuracy: exploring the impact of data anonymization on deep learning models in computer vision. IEEE Access 12:8346–8358

Lin T, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár Zitnick CL (2014) Microsoft COCO: common objects in context. Pages 740–754 *in* Fleet D, Pajdla T, Schiele B, Tuytelaars, eds. Computer Vision—ECCV 2014. Lecture Notes in Computer Science 8693. Cham, Switzerland: Springer

Luo T, Zhao J, Gu Y, Zhang S, Qiao X, Tain W, Han Y (2021) Classification of weed seeds based on visual images and deep learning. Inf Process Agric 10:40–51

Mahé I, Cordeau S, Bohan DA, Derrouch D, Dessaint F, Millot D, Chauvel B (2020) Soil seedbank: old methods for new challenges in agroecology? Ann Appl Bio 178:23–28

Morgensen SHAC, Allen PS, Jeffery LS (2005) Retrieval and identification of weed seeds from landscape soils. Seed Technol 27:37–48

Pérez-Porras FJ, Torres-Sánchez J, López-Granados F, Mesas-Carrascosa FJ (2023) Early and on-ground image-based detection of poppy (*Papaver rhoeas*) in wheat using YOLO architectures. Weed Sci 71:50–58

Teo-Sherrel CPA, Mortensen DA, Keaton ME (1996) Fates of weed seeds in soil: a seeded core method of study. J Appl Ecol 33:1107–1113

Wosner O, Farjon G, Bar-Hillel A (2021) Object detection in agricultural contexts: a multiple resolution benchmark and comparison to human. Comput Electron Agric 189:106404

Wu N, Zhang Y, Na R, Mi C, Zhu S, He Y, Zhang C (2019) Variety identification of oat seeds using hyperspectral imaging: investigating the representation ability of deep convolutional neural network. RSC Adv 9:12635

Tzuta L (2015) LabelImg (Free Software: MIT License). https://github.com/tzutalin/labelImg.

Verma S, Tripathi S, Singh A, Ojha M, Saxena R (2021) Insect detection and identification using YOLO algorithms on soybean crop. Pages 272–277 *in* TENCON 2021–2021 IEEE Region 10 Conference (TENCON), Auckland, New Zealand, December 7–10, 2021

Xu P, Yang R, Zeng T, Zhang J, Zhang Y, Tan Q (2021) Varietal classification of maize seeds using computer vision and machine learning techniques. J Food Process Eng 44:e13846

Yao Q, Zheng X, Zhou G, Zhang J (2024) SGR-YOLO: a method for detecting seed germination rate in wild rice. Front Plant Sci 14:1305081

Zhang C, Li T, Zhang W (2022) The detection of impurity content in machine-picked cotton based on image processing and improved YOLO V4. Agronomy 12:66

Zhao, ZQ, Zheng P, Xu ST, Wu X (2019) Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst 30:3212–3232

https://doi.org/10.1017/wsc.2024.60 Published online by Cambridge University Press