Computers and Electronics in Agriculture 232 (2025) 110115

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Alfalfa detection and stem count from proximal images using a combination of deep neural networks and machine learning

Hazhir Bahrami [a,*], Karem Chokmani [a], Saeid Homayouni [a], Viacheslav I. Adamchuk [b], Md Saifuzzaman [b], Maxime Leduc [c]

[a] Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Québec, Canada
[b] Bioresource Engineering Department, McGill University, Macdonald Campus, Ste-Anne-de-Bellevue, QC, Canada
[c] Founder and Project Manager, Jasons Forage Systems, Canada

ARTICLE INFO

Keywords:
Alfalfa
Stem Count
Proximal Images
Deep learning
Deep Convolutional Neural Network
Synthetic data

ABSTRACT

Among various types of forages, Alfalfa (Medicago sativa) is a crucial forage crop that plays a vital role in livestock nutrition and sustainable agriculture. As a result of its ability to adapt to different weather conditions and its high nitrogen fixation capability, this crop produces high-quality forage that contains between 15 and 22 % protein. It is fortunately possible to improve the overall prediction of forage biomass and quality prior to harvest through remote sensing technologies. The recent advent of deep Convolution Neural Networks (deep CNNs) enables researchers to utilize these incredible algorithms. This study aims to build a model to count the number of alfalfa stems from proximal images. To this end, we first utilized a deep CNN encoder-decoder to segment alfalfa and other background objects in a field, such as soil and grass. Subsequently, we employed the alfalfa cover fractions derived from the proximal images to develop and train machine learning regression models for estimating the stem count in the images. This study uses many proximal images taken from significant number of fields in four provinces of Canada over three consecutive years. A combination of real and synthetic images has been utilized to feed the deep neural network encoder-decoder. This study gathered roughly 3447 alfalfa images, 5332 grass images, and 9241 background images for training the encoder-decoder model. With data augmentation, we prepared about 60,000 annotated images of alfalfa fields containing alfalfa, grass, and background utilizing a pre-trained model in less than an hour. Several convolutional neural network encoder-decoder models have also been utilized in this study. Simple U-Net, Attention U-Net (Att U-Net), and ResU-Net with attention gates have been trained to detect alfalfa and differentiate it from other objects. The best Intersections over Union (IoU) for simple U-Net classes were 0.98, 0.93, and 0.80 for background, alfalfa and grass, respectively. Simple U-Net with synthetic data provides a promising result over unseen real images and requires an RGB iPad image for field-specific alfalfa detection. It was also observed that simple U-Net has slightly better accuracy than attention U-Net and attention ResU-Net. Finally, we built regression models between the alfalfa cover fraction in the original images taken by iPad, and the mean alfalfa stems per square foot. Random forest (RF), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGB) methods have been utilized to estimate the number of stems in the images. RF was the best model for estimating the number of alfalfa stems relative to other machine learning algorithms, with a coefficient of determination (R^2) of 0.82, root-mean-square error of 13.00, and mean absolute error of 10.07.

1. Introduction

One of the most widespread biomes on Earth is grassland (31–69 % of global land surface). Grasslands are essential for global food security and provide several ecological services, such as erosion control, water harvesting, wildlife habitat support, and carbon sequestration (Squires et al., 2018). Some of the most significant components of the human diet or animal feed are grain and forage legumes (Vance et al., 2000). In recent years, precision agriculture has been developed and refined, increasing the use of advanced technology and previously untested systems to manage livestock more effectively (Wachendorf et al., 2018). Forage is vital to the livestock industry as a primary source of animal

* Corresponding author.
E-mail address: hazhir.bahrami@inrs.ca (H. Bahrami).

https://doi.org/10.1016/j.compag.2025.110115
Received 2 July 2024; Received in revised form 21 November 2024; Accepted 7 February 2025

nutrition. Monitoring the quality and availability of forage resources, particularly alfalfa, is crucial for ensuring optimal animal health and productivity. Agricultural producers can make informed decisions regarding grazing management, feed supplementation, and overall herd health by regularly assessing forage quantity and nutritional content.

Compared to other forage crops, alfalfa has an exceptionally high yield potential (Aldakheel et al., 2004). It is recognized that alfalfa or lucerne (*Medicago sativa*) is one of the most significant forage legumes (*Fabaceae*) and the most widely cultivated crop (over 80 countries) among forage legumes, covering an area of 35 million hectares (Kayad et al., 2016). Alfalfa is a temperate perennial crop typically grown in arid and semi-arid areas. Its well-developed root system can extend deep into the soil to obtain water (Quan et al., 2016). Due to its ability to fix atmospheric nitrogen, improve soil structure, and control weeds, alfalfa is an essential component of many crop rotations (Aldakheel et al., 2004). A study of crop parameters and yields conducted by Bahrami et al. (2022) has shown that alfalfa ranks seventh among the ten most frequently studied crops regarding biophysical parameters such as biomass and leaf area index (LAI). There is no doubt that one of the most considerable challenges to alfalfa production is weeds, given that these species compete with alfalfa for nutrients, space, sunlight, and water (Yang et al., 2022). They also reduce forage yield and nutritional value (Yang et al., 2022).

Alfalfa can be cut several times yearly and regrow quickly after each harvest (Gao and Zhang, 2021). As a result of climate conditions directly affecting the severity of environmental stresses and indirectly modulating plant hardiness, perennial crops such as alfalfa are more likely to survive the winter (McKenzie and McLean, 1980a, b; Suzuki, 1972). Various factors can lead to winter damage, including subfreezing and fluctuating temperatures, excessive soil moisture, ice encasement, and soil heaving (Andrews, 1987; Bélanger et al., 2006). In many forage-growing regions of Canada and other cold temperate countries, harsh winter climates cause frequent losses of stands and yield reductions (Bélanger et al., 2006). If this injury is evaluated early in the growing season, crop replacement decisions can be made more efficiently. Sparse or elderly stands may lack the yield capacity to produce sufficient yields in future cuts. In order to evaluate annual stem mortality, it becomes essential to assess the extent of damage caused by these environmental stresses using monitoring techniques. Stem counts offer a more precise assessment of yield potential compared to plant counts.

Stem counts also provides valuable insights into stand density, plant vigor, and potential yield, enabling farmers to make informed decisions regarding irrigation, fertilization, and harvest timing (David et al., 2021). Accurate estimation of stem counts plays a vital role in better crop yield projections and in monitoring the growth and development of crops (Dias et al., 2018). Yield estimation can be carried out automatically by counting the number of plants or tracking their growth (Hunt et al., 2010). Moreover, stem count is a crucial indicator of crop productivity, helping farmers and decision-makers optimize forage utilization and improve overall field performance (Dias et al., 2018). Current breeding programs typically assess plant density manually. Since this process is tedious, time-consuming, and expensive, operators count plants in the field over a limited area (David et al., 2021; Fernandez-Gallego et al., 2020). Several uncertainties and possible errors usually accompany the operators' sub-sampling methods (David et al., 2021; Fernandez-Gallego et al., 2020; Liu et al., 2016).

Nowadays, monitoring crop parameters such as yield, biomass, fractional vegetation cover, and other parameters is achievable at the field scale thanks to advances in remote sensing and ground- and mobile-based sensors (Hancock and Dougherty, 2007). Spectral-based remote sensing techniques mainly rely upon a crop's unique absorption and scattering characteristics, i.e., its spectral signature. In the visible range of the electromagnetic spectrum, absorption occurs at red and blue wavelengths (around 670 nm and 450 nm, respectively) and scatters in the green band (530–590 nm), resulting from the presence of chlorophyll and accessory photosynthetic pigments in plant leaves (Marshall and Thenkabail, 2015). There is wide variation in the accuracy of remote sensing methods, which is determined primarily by the modelers' ability to collect and scale up ground measurements for model calibration and validation (Lu, 2006). In-field estimation of the number of stems and other crop parameters, such as yield and biomass, let decision-makers better manage and organize forage production (Noland et al., 2018). Although the cost of in-situ spectral measurement technologies is comparatively high with higher spectral resolution and better resolution of contemporary systems, one of the best methods that can aid farmers and decision-makers in achieving rapid and direct assessments of a crop at field scales is remote sensing of canopy reflectance (Noland et al., 2018).

Several studies have focused on the use of various remote-sensing platforms, including ground-based sensors (Garriga et al., 2020; Hancock and Dougherty, 2007; Marshall and Thenkabail, 2015), aerial vehicles (Cazenave et al., 2019; Feng et al., 2020), and satellites (Azadbakht et al., 2022; Kayad et al., 2016), to assess alfalfa production. Li et al. (2023) examined the ability of several remote sensing images' reflectance and the vegetation indices derived from the bands to estimate the biomass and yield of alfalfa. In order to achieve this objective, they employed the MODIS surface reflectance product and Sentinel-2 Level2-A satellite imagery. The analyzed vegetation indices included normalized difference vegetation index (NDVI), normalized difference water index (NDWI), soil adjusted vegetation index, and normalized difference infrared index (NDII), to name a few. The inversion model utilized various models, including artificial neural network (ANN), Random Forest (RF), Support Vector Regression (SVR), and linear regression. According to their findings, moisture-based vegetation indices such as NDWI and NDII showed higher correlations with biomass and yield. Marshall and Thenkabail (2015) used a hyperspectral ground-based sensor and several other parameters, such as crop height, LAI, and vegetation cover fraction, to predict the wet biomass of several crops, including alfalfa. They reported a coefficient of determination (R^2) of 0.86 for alfalfa. Kayad et al. (2016) utilized Landsat satellite images to examine variability in alfalfa yield and to develop a model to evaluate how yield can be estimated using VIs (Vegetation Indices) extracted from Landsat data. Their yield maps showed spatial variability in alfalfa yield within the alfalfa field, and the correlation varied between 0.75 and 0.97 for four monitored harvests. Valente et al. (2020) utilized UAV-RGB (Unmanned Aerial Vehicle Red-Green-Blue) images (8 to 16 mm resolution), together with transfer learning (Kaya et al., 2019), excess green vegetation index and the Otsu algorithm (Otsu, 1979) to count spinach (*Spinacia oleracea*; Amaranthaceae) within a field. They reported an accuracy of 95 % over an area of 172 m^2. So far, no research has focused on estimating the number of alfalfa stems using remote sensing data.

Researchers are increasingly focusing on applying machine learning (ML) algorithms to agricultural production problems, as they can model complex non-linear relationships (Ranjbar et al., 2021). Artificial Neural Networks (ANNs) consist of simple linked processors called neurons, which generate sequences of activations with a real value (Akhavan et al., 2021a; Schmidhuber, 2015). Deep neural networks, particularly convolutional neural networks (CNNs), have revolutionized computer vision applications by learning complex patterns and characteristics from images. LeCun et al. (1989) first proposed CNN models, which utilize a stack of feature layers to automatically extract the efficient discriminating features for a given problem. The concentration of the first layers of CNN models is the image's low-level features (Bangare et al., 2022). CNN models provide advanced features as the model deepens and the number of layers increases. Substantial ML and Deep Learning (DL) algorithms have been proposed and developed for different types of remote sensing images, from high resolution to coarse resolution, thanks to the growth of ML in recent years.

Semantic segmentation is one of the critical tasks of machine learning for achieving pixel-level classification (Luo et al., 2023). Semantic segmentation produces pixel-level descriptions of objects

embedded in their spatial locations instead of making predictions about the whole image. Semantic segmentation models have become widely applied to problems in various areas, specifically in agriculture, such as weed segmentation (Zou et al., 2021), identification of pests and diseases (Singh et al., 2021), and crop cover and crop type analysis (Jadhav and Singh, 2018). In the case of severe occlusions in segmentation challenges, the U-Net network can slice low-level and high-level features while preserving edge information with a low computational workload.

Although the research on counting the number of stems is sparse, some research papers focus on detecting weeds, grass, and background objects in an image over a field. Yang et al. (2022) utilized several CNN models, such as AlexU-Net and ResU-Net, and photographic images in diverse sizes to detect weeds among alfalfa. They reported that images of 200 by 200 pixels delivered a more accurate model than other image sizes. They demonstrated that AlexU-Net was the most accurate model, with a precision and recall of more than 0.99. In contrast, ResU-Net was the least accurate model compared to other CNN models. Echeverría et al. (2021) investigated the potential for Sentinel-2 multispectral imagery to assess the fractional vegetation cover in rain-fed alfalfa in Spain. They utilized a maximum likelihood model to classify their ground images into five classes: alfalfa in the sun, alfalfa in the shade, soil in the sun, soil in the shade, and unclassified. Their results showed that this simple classification method achieved an accuracy of about 95 %.

While there is extensive research on biomass assessment and crop detection at the field scale, there is a notable deficiency in thorough studies regarding the quantification of alfalfa stems under varying weather conditions and growth stages across diverse locations. The main objective of this study is to develop a framework for counting stem numbers from proximal images using a combination of deep neural networks and machine learning algorithms. The specific objectives are 1) to generate a dataset comprising numerous synthetic and real images using a pre-trained algorithm and evaluate whether the semantic segmentation models can be trained well or not, 2) To evaluate the capabilities of multiple U-net models on the dataset and develop a U-net model that can accurately identify alfalfa and differentiate it from other objects including soil, grass, weeds, and crop residue, and 3) to model and predict the number of alfalfa stems using the fractional vegetation cover of class alfalfa calculated in objective 2.

2. Materials and methods

2.1. In-situ dataset

Over two years, ground measurements including soil sampling, stem counts, and crop heights along with the image dataset were collected in 461 alfalfa fields in four provinces of Canada: Nova Scotia, Quebec, Ontario, and Manitoba (Fig. 1). The details for each province are given in Table 1. The ground measurements consisted of 192 producers (farmers) and 33 advisors. The ground measurements are gathered when the crop reaches a height of 2–5 in.. A randomized design was implemented in each field. Spring and autumn stem counts were collected at each site (composed of three data points, often referred to as landmarks). In each quadrat, each stem (not the number of plants) of alfalfa above about 5 cm was counted. A ruler was utilized to measure the mean height of 5 stems.

For each landmark position, stem counts were measured using a rectangular quadrat (Fig. 2). An RGB iPad mini 5th generation camera was utilized to capture the images. All images have been captured about 1 m above the ground surface over various alfalfa fields selected for the tests. The brightness and contrast of images vary due to their acquisition on different dates, in different weather conditions, at various times of day, and various growth stages. Over 10000 images were captured throughout the field campaign. A total of 2222 images were utilized in this research. Only images captured in near-perpendicular position to the Earth's surface were utilized. The resolution of the images is 1632 by 1224 pixels. Patches of 256 by 256 pixels have been extracted from the original images. In total, 3447 alfalfa images, 5332 grass images, and 9241 background images were available. Also, 1011 alfalfa images, 269 grass images, and 380 background images were selected for the test dataset. The best and most accurate efforts have been made to include every possible situation in the training data: alfalfa, alfalfa in sun, alfalfa in shadow, and alfalfa in different growth stages. For the background,

Table 1
The details of the number of fields for each province.

Provinces	#N of fields
Manitoba	34
Nova Scotia	5
Ontario	13
Quebec	409

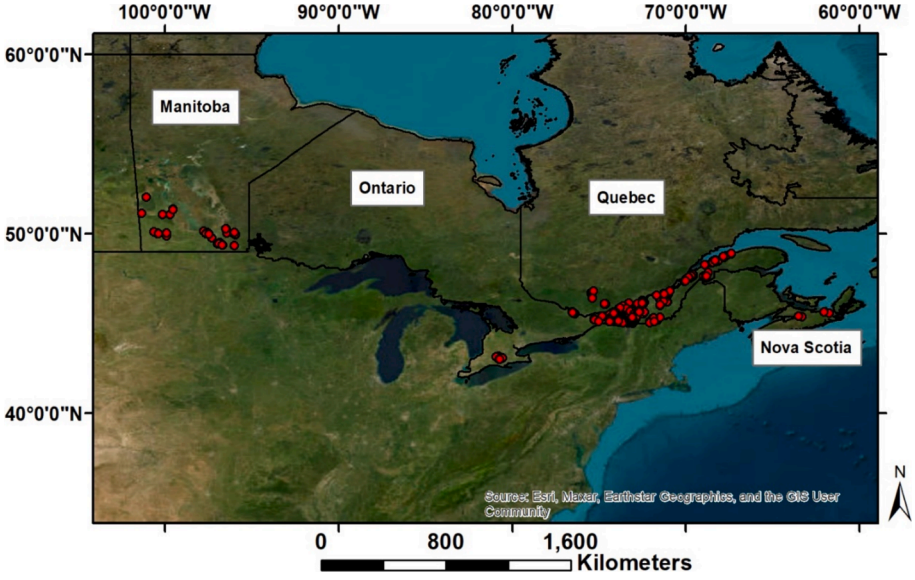

Fig. 1. Study area and data collection sites (red points) in 2021 and 2022.

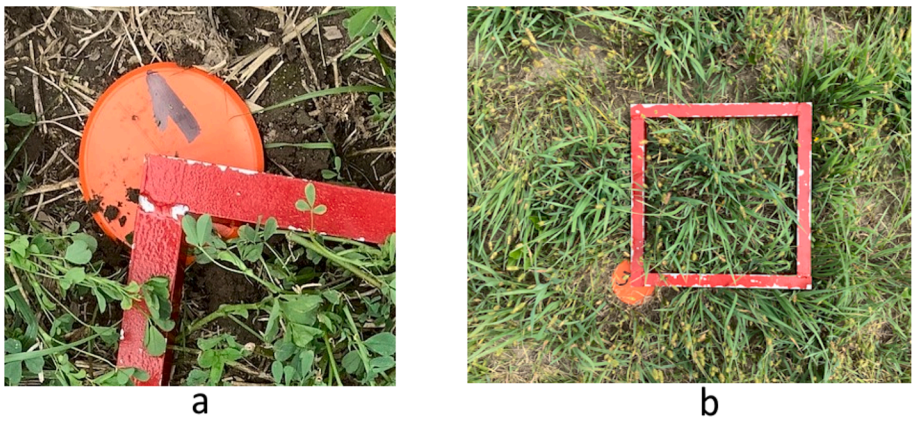

Fig. 2. Landmark (no. 1 of 3 with a distance of 1 m) placed in the sampling site (a), measurements were taken within the red rectangle (1 ft. x 1 ft.) placed on the corner of each landmark (b).

soil in sun and shadow, dead grass, and various quadrat types are included in the training data. The dataset used in this research is currently not publicly accessible.

A sample of patches for each class is displayed in Fig. 3.

2.2. Data Preparation and augmentation

Several challenges are associated with using supervised learning methods in agricultural image segmentation. One is the lack of sufficiently labeled datasets, which may negatively affect the training process (Kamilaris and Prenafeta-Boldú, 2018). Data augmentation is essential when using deep learning methods to solve minor sample problems since the diversity and quantity of training data affect the robustness and generalization of deep learning models (Kamilaris and Prenafeta-Boldú, 2018).

Data augmentation can be conducted by applying rotation, flip, image transformation, and enhancement methods or by creating synthetic images. One of the main advantages of synthetic images is the ability to generate diverse and precisely annotated datasets. Creating real-world datasets with pixel-level annotations can be time-consuming, labor-intensive, and expensive (Ge et al., 2023). Synthetic images enable researchers to generate ground truth labels automatically, accelerating model development and testing. Moreover, synthetic datasets can encompass a wide range of environmental conditions, lighting variations, and object configurations, aiding in the generalization of models to different scenarios. Therefore, in addition to popular augmentation methods, such as rotation (90°, 180°, 270°) and horizontal and vertical flip, synthetic images is used to augment the dataset and to reduce the number of samples that need to be manually labeled. A pre-trained ResU-Net model with attention gates trained in our lab by Maryam Rahimzad has been utilized to annotate training and testing data. This pre-trained model classifies three classes: vegetation, crop residue, and and background, including soil and shadow, utilizing RGB images. In that case, the model could only predict green pixels as vegetation. Since grass, weeds, and alfalfa have almost the same spectral reflectance (both are green), we must classify grass and alfalfa separately.

Since the classes are separated in our database (especially grass and alfalfa), synthetic images and masks for the model and conventional data augmentation methods have been utilized to increase the volume of data so the model can learn the problem better. Conventional augmentation methods utilized in this study consist of rotation 90, rotation 180, rotation 270, horizontal flip, and vertical flip, together with the original

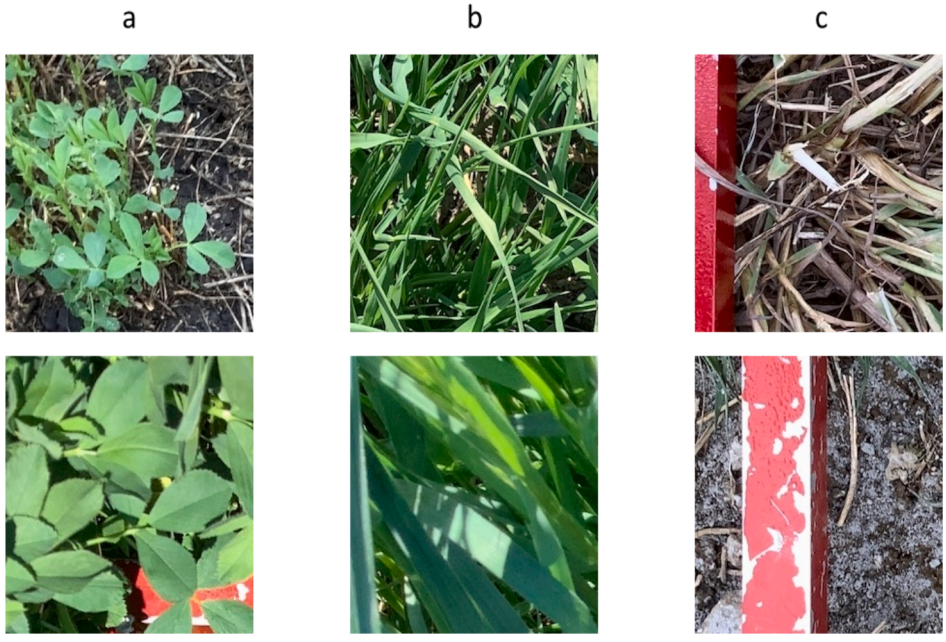

Fig. 3. Sample of patches for each class of a) alfalfa, b) grass, and c) background (combination of soil, quadrat, dead grass, and shadow, among others).

image without augmentation. Due to the large number of background and grass images compared to alfalfa images, only three augmentation methods for grass and one augmentation for background have been utilized.

Fig. 4 illustrates the fusion process of an alfalfa patch with a background class patch to create a synthetic patch for training. Based on the pre-trained model, the mask of class vegetation was extracted. By considering the mask of green areas (here is alfalfa) extracted from the pre-trained model, the alfalfa is extracted from the real patch. As a result, the vegetation part has been masked out precisely. For the remaining part of the image required to create a synthetic picture, a randomly selected image of background or grass has been considered. Finally, a synthetic image has been generated by combining the alfalfa patch and the background patch. As can be seen in Fig. 4, the synthetic image resembles a real image. As previously discussed, this tool can also simulate various conditions. Also, The original image was added to the training data as a real image.

An example of creating a synthetic image with a combination of alfalfa and grass can be demonstrated in Fig. 5.

It should be noted that we have a combination of grass and background pixels for grass images. As can be seen in Fig. 5, black is for the background (class 0), grey is for alfalfa (class 1), and white is for grass (class 2).

Using the pre-trained model, we annotated around 60,000 images in roughly an hour in this research. As stated in section 2.1 of the in-situ dataset, the quantity of patches across various classes is unequal. Data augmentation was employed to mitigate data imbalance to the greatest extent possible. The total number of pixels for each class is summarized in Table 2.

Several examples of the generated synthetic images and the corresponding mask can be seen in Fig. 6.

2.3. Proposed methodology

The primary goal of this study was to count the number of alfalfa stems in proximal images. Nonetheless, direct stem counting proved impossible in the proximal images utilizing existing and advanced models. Consequently, we proposed detecting alfalfa in the images in the first step and then utilizing the detected alfalfa fractional cover to quantify the number of stems. This research proposes a practical framework that combines deep CNN encoder-decoder models with machine learning regression models to meet the objective of stem counting. This study utilizes and assesses a few advanced encoder-decoder models to detect alfalfa in proximal images and distinguish it from other objects in fields, such as soil, grass, and weeds, using a combination of synthetic and real images. An evaluation was conducted to examine the impact of attention gates and residuals on semantic segmentation accuracy with a U-Net-based model. Following training the deep encoder-decoder models to detect alfalfa in the images, the subsequent task of estimating and counting the number of alfalfa stems was carried out by utilizing the alfalfa fractional cover computed for each image. Then, various machine learning regression algorithms, including random forest (RF), support vector regression, and Extreme Gradient Boosting (XGB), were trained to quantify the number of alfalfa stems measured within the fields. The regression models stated above were fed with fractional vegetation cover and aimed to predict the number of stems per square foot in the image (Fig. 7).

2.4. Network architecture

Fig. 8 illustrates an overview of the proposed architecture. Various parts of the model have been described below.

2.4.1. Encoder path

The encoder path consists of four layers, each containing a Unet block and a 2×2 Max-pooling layer. Our model takes an input of $256 \times 256 \times 3$, where the image size is 256×256, and we have three channels corresponding to the RGB band of an image. The unet blocks comprise two convoluted layers (Fig. 9-a). Each layer starts by a convolution layer followed by a Batch Normalization (BN) layer (i.e., input re-centering and re-scaling (Ioffe and Szegedy, 2015)) and a Rectified Linear Unit (ReLU) activation function (negative inputs = 0, positive values are returned as outputs). The first unet block has 16 feature maps of size 256×256. During downsampling (i.e., random removal of majority class observations), which reduces the dataset, especially class imbalances, the number of feature maps doubles, and each layer's size is halved. Therefore, the unet block at the fourth layer has 128 feature maps of size 32×32. The convolutional blocks in all encoder layers have a dilation rate of 2.

2.4.2. Bottleneck

A bottleneck layer connects an encoder and a decoder. It contains a unet block with 256 feature maps of size 16×16. In the bottleneck layer, the output moves in two directions. The output is fed into the convolutional transpose layer.

2.4.3. Decoder path

Each of the four layers of the decoding path begins with a U-Net block followed by a 2×2 up-convolution layer. A skip connection connects each decoder layer to the corresponding encoder layer. Upsampling (increasing the dimensions or cases of the dataset, cf. downsampling) is performed using a 2×2 up-convolution layer. The

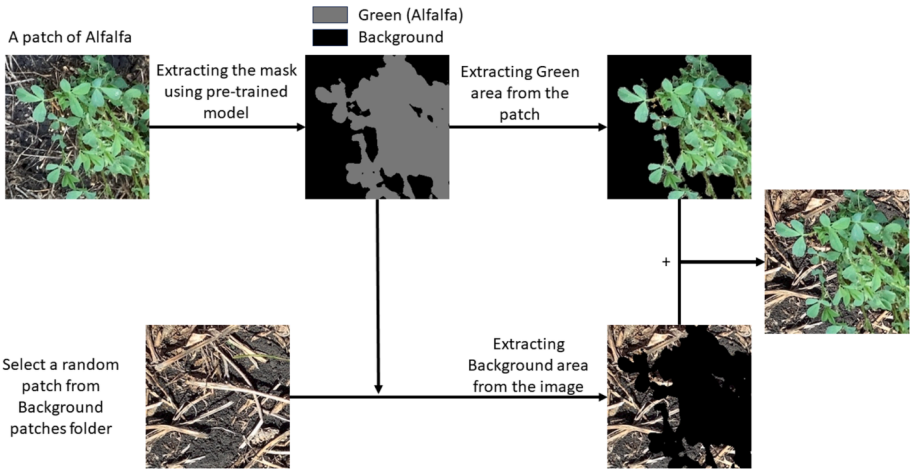

Fig. 4. An example is generating synthetic images by combining alfalfa and background objects (soil, crop residue, etc).

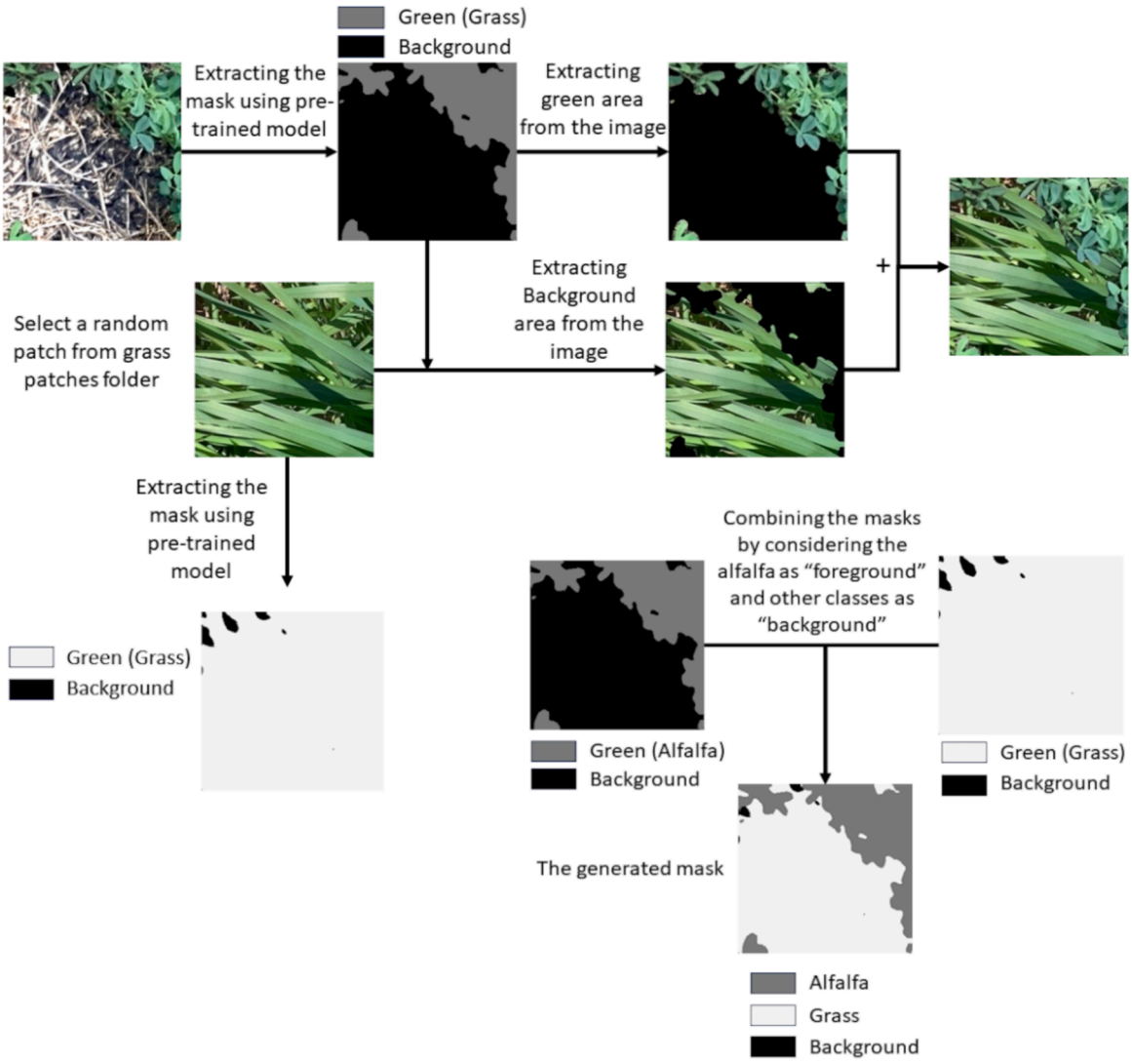

Fig. 5. An example of generating a synthetic image by combining alfalfa and grass.

Table 2
Details regarding the number of pixels for each class.

Class	Number of training pixels	Percentage (%)
Background	2,066,690,212	52
Alfalfa	1,030,889,364	26
Grass	866,824,136	22
Total	3,964,403,712	100

skip connection is concatenated with the upsampled decoder output from the previous layer. Unet blocks receive the output of concatenation. The same unet blocks are used as those in the encoder path. Each decoder layer reduces the number of feature maps by half while doubling their size. The output of the unet block at each layer is upsampled to the size of 256 × 256 and passed through a classification layer and softmax activation function (normalized exponential function).

The classification layer reduces the number of feature maps to three channels. Feature maps are converted into probabilities using softmax activation. Thus, each channel in the final output is transformed into a probability map corresponding to a particular class. Based on these outputs, the loss is calculated by comparing them to the ground truth.

Fig. 9 illustrates how our proposed model uses the residual blocks and the attention gate, respectively.

2.4.4. Loss function

The present study utilizes combined categorical focal loss (CFL) and weighted Jaccard loss (WJL) to segment Alfalfa images. The weights of the Jaccard loss were assigned according to the pixel count in the training data to address data imbalance. Consequently, the grass and alfalfa classes were given greater weights, while a smaller weight was assigned to the background class. A combined loss (CL) function is obtained by using the following equation to train the proposed model.

$$CL = WJL + CFL \tag{1}$$

2.4.5. Hyperparameters Optimization

In the final step of our investigation, we performed a comprehensive grid search to optimize multiple parameters and hyperparameters, evaluating various CNN architectures (Simple U-Net, Att U-Net, and Att ResU-Net), the initial number of features (8, 16, 32), batch sizes (8, 16, 24, 36), learning rates (1e-3, 1e-4, 1e-5), and optimizers (Adam, SGD). We evaluated each parameter and ultimately identified the optimal parameters for the models. We trained the models, computed the accuracy, and determined that the optimal number of initial features was 16, the ideal batch size was 24, the most effective learning rate was 1e-4, and the best optimizer was Adam (Kingma and Ba, 2014). We utilized these parameters to train the models.

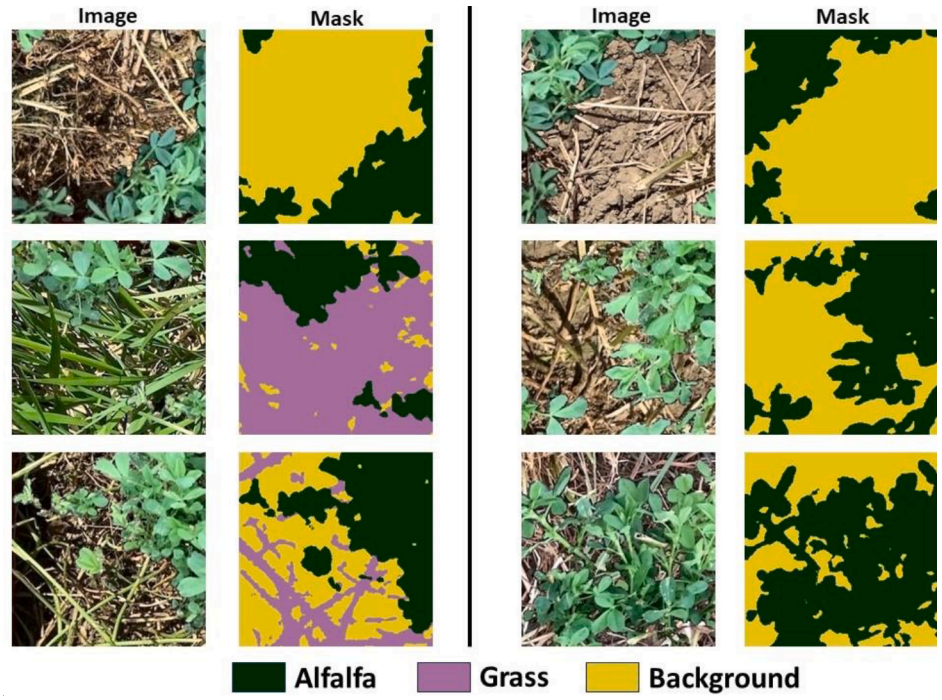

Fig. 6. Several data examples generated by the combination of real and synthetic images.

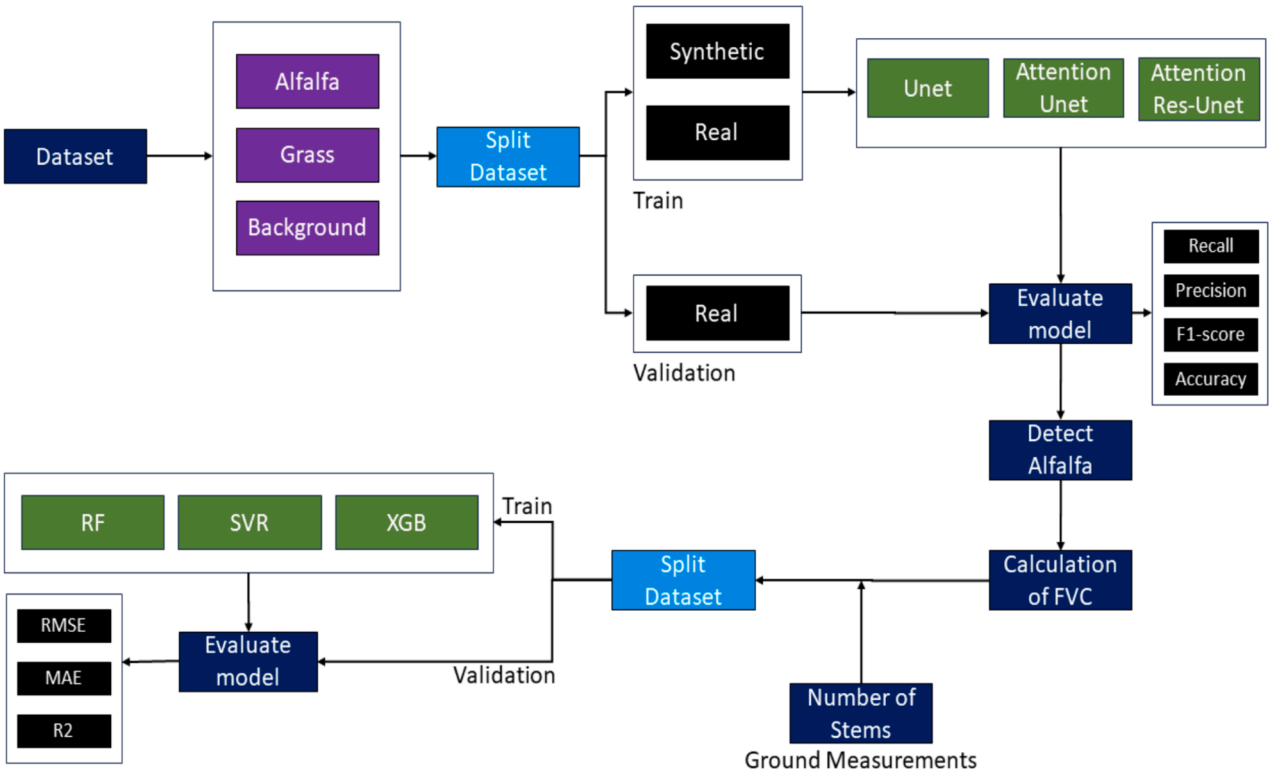

Fig. 7. The flowchart of the proposed methodology in this study.

2.4.6. Evaluation criteria

Confusion matrices are widely used to evaluate the performance of classification algorithms and form the basis of receiver operating characteristics (ROC) assessments and the Kullbeck-Leibler (relative entropy) divergence (Fawcett, 2006).

The confusion matrix is a table with two rows and two columns that reports the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Recall measures the model's ability to identify the positive cases and It becomes essential to identify and address missing positive cases. Recall is calculated as Equation 2. High recall indicates that the model has a strong ability to correctly identify positive cases. A model's precision is determined by its positive detection accuracy and is computed as Equation 3. High precision implies that the model has a minimal

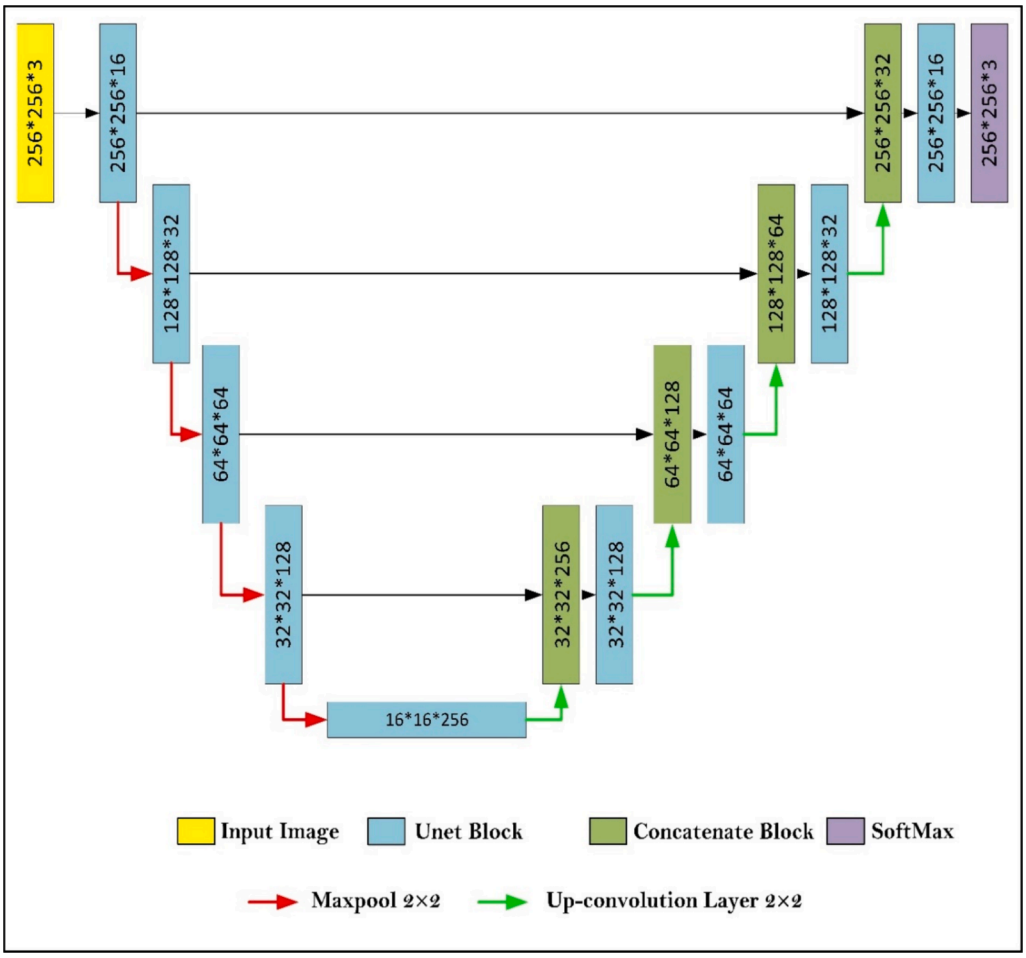

Fig. 8. The adapted architecture of Attention Residual U-Net in this study.

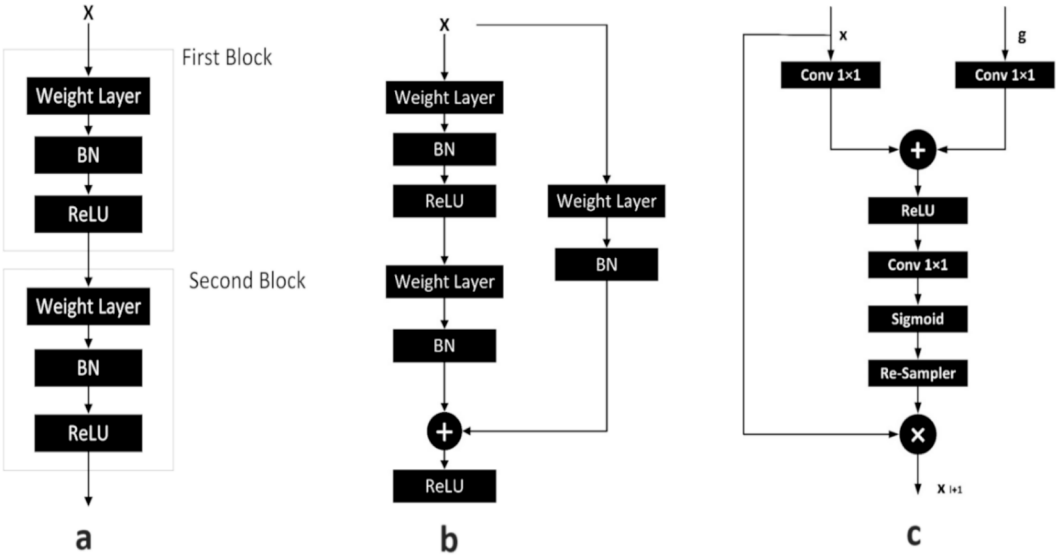

Fig. 9. The details of U-Net blocks (a), residual blocks (b), and the details of attention blocks (c).

occurrence of false positives. Also, the F1-score is a harmonic average of recall and precision, and it is useful when you need to take both false positives and false negatives into account. (Equation 4). Intersection over Union (IoU) is a commonly utilized evaluation metric for image segmentation models (Equation 5). It measures the overlap between the predicted segmentation mask and the ground truth mask. Finally, accuracy is a criterion that measures the percentage of correctly classified cases in the whole dataset and is calculated according to Equation 6.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$Percision = \frac{TP}{TP + FP} \tag{3}$$

$$F1 - score = 2^{*}\frac{Precision^{*}Recall}{Precision + Recall} = \frac{2^{*}TP}{2^{*}TP + FP + FN} \tag{4}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

2.5. Estimating alfalfa stems

When the model was trained, the original images of size 1632 by 1224 were given to the model. The number of pixels of alfalfa, divided by the total number of pixels, was then computed. If this value is calculated for an image as 0.5, it means that 50 percent of the image is alfalfa. Machine-learning algorithms were utilized to create a regression model between the alfalfa cover fraction, and the number of alfalfa stems in an image. The numbers of alfalfa stems were measured in a one-foot square quadrat (~30 × 30 cm). Many areas have recently used machine-learning algorithms in classification and regression problems. This study used regression models, i.e., RF (Random forest), XGB (Extreme Gradient Boosting), and SVR (Support Vector Regression), to estimate the number of alfalfa stems in an image. ML algorithms were implemented using the open-source Python Scikit-learn package. We had 2222 images, and the Alfalfa cover fraction has been calculated for these images. The data were divided into training and test datasets. Eighty percent of the data was selected to train the models, and the remaining data (i.e., 20 %) were used for testing. Grid Search Cross-Validation (GridSearchCV, a function in the Scikit-learn package) with a cross-validation value of 5 was used to tune the hyper-parameters of all machine-learning algorithms.

2.5.1. Random forest regression

As a robust ensemble learning method, RF is extensively used in classification and regression problems (Akhavan et al., 2021b). Ensemble learning refers to the process of producing multiple models and combining them in order to solve a particular problem. Two types of ensemble learning are boosting and bagging.The RF approach is a practical bagging approach based on many individual decision trees. The model then combines all predictions to achieve a better performance by combining every tree's prediction (Dangeti, 2017). The GridSearchCV parameters that were used in the RF are shown in Table 3.

2.5.2. Support Vector regression

One of the most widely used kernel-based machine learning algorithms is the support vector machines (SVMs) algorithm developed by Vapnik and his colleagues (Sheykhmousa et al., 2020). This algorithm can be used for various problems, particularly classification problems. While maintaining all algorithm characteristics, such as the maximal margin, SVM can also be applied to regression problems. In SVR, we can define the acceptable error level in our model and find a line (or hyperplane in higher dimensions) that fits the data well. In this manner, the points outside the tube receive penalization; however, the prediction function receives no penalization for the points inside the tube, either above or below the centerline. The Grid Search parameters used for the SVR model are shown in Table 4.

Table 3
GridSearchCV parameters that were used in the RF model.

Parameters	Description	Grid Search Values
n_estimators	No. of trees in the forest	25, 50, 100, 500
max_depth	Maximum depth of the trees	3, 4, 5

Table 4
GridSearchCV parameters set for the SVR.

Parameters	Description	Grid Search Values
Kernel	Specifies the kernel type to be used in the algorithm.	'linear', 'rbf'
Gamma	Kernel coefficient for 'rbf', 'poly' and 'sigmoid'.	0.001, 0.01, 0.1, 0.5
C	Penalty parameter	5, 10, 50, 100

2.5.3. Extreme gradient boosting

XGB (Brownlee, 2016) is one of the most popular gradient-boosting implementations, and it was first developed by Tianqi Chen in 2001 as a research project. The algorithm is based on a gradient-boosting framework and is an ensemble machine-learning algorithm. XGB enhanced a machine learning model's performance, speed, flexibility, and efficiency. The Grid Search parameters used for the XGB algorithms are shown in Table 5.

2.5.4. Evaluation criteria

Several criteria were used to evaluate prediction performance, including RMSE, Mean Absolute Error (MAE), and the coefficient of determination (R^2). The formula of the RMSE is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{N}} \tag{7}$$

where N is the number of observations.

RMSE provides a quantifiable indication of the extent to which these residuals are distributed. A smaller RMSE reflects a superior alignment between the model and the data. RMSE is highly sensitive to large errors due to its utilization of the squared residuals. MAE is calculated as the following equation:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N}|\hat{y}_i - y_i|}{N} \tag{8}$$

MAE is a simple and precise metric that quantifies the average size of errors in a given set of predictions, regardless of their direction. MAE exhibits greater resistance to large errors in comparison to RMSE, giving it valuable when seeking a statistic that is less susceptible to outliers. However, while RMSE squares each error prior to averaging, it disproportionately penalizes larger deviations. This is advantageous in cases where large mistakes are highly costly or undesirable. The last evaluation criterion used in this study is the R^2, calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2} \tag{9}$$

R^2 ranges from 0 to 1. A value of 1 for R^2 indicates that the regression model accurately predicts the dependent variable, whereas a value of 0 shows that the model fails to account for any of the variations in the dependent variable. Higher R^2 values generally indicate a better fit of the model.

Table 5
GridSearchCV parameters set for the XGB.

Parameters	Description	Grid Search Values
learning_rate	Shrinks the contribution of each tree	0.001,0.005,0.01,0.05,0.1,0.15,0.3
n_estimators	The number of boosting stages to conduct.	10, 30, 50, 100
max_depth	Limits the number of nodes in the tree.	4, 5, 6, 7

3. Results

After preparing the dataset, the U-Net models were trained on the dataset. Fig. 10-a and −b show the loss function plots for training and validation data. As can be seen, there is no sign of over-fitting. Nonetheless, Fig. 10 illustrates the unstable validation accuracy and loss throughout the training of a U-Net model. This could originate from a small batch size, which might produce noisy gradient estimates, consequently impacting loss stability. Due to GPU limitations, we could not increase the batch size beyond 32. Furthermore, as illustrated in section 2.2, the validation set is disproportionately small relative to the training data, even before data augmentation. This may be another cause of the instability of validation loss and accuracy compared to training metrics.

Table 6 shows the results of the model for validation data. All three classes have recall values greater than 0.9, as shown in the table. This model has demonstrated a very high accuracy with an overall accuracy of 0.97 for simple U-Net. The results showed that alfalfa and background have been detected with considerably high accuracy (precision of 0.99 for both classes and recall of 0.99 and 0.94 for background and alfalfa, respectively). The precision for the background is the highest among the classes, as 0.99 were reported using attention ResU-Net and simple U-Net for f1-score, recall, and precision, respectively. The results also showed that grass had been detected with the lowest accuracy compared to the other classes in terms of precision value, with the highest value of 0.80, resulting in precision.

Furthermore, the comparison results showed that simple U-Net has slightly better accuracy than other models. The overall accuracy for simple U-Net is 0.97, while the accuracy for attention U-Net and Attention ResU-Net is 0.96. The accuracy criterion for grass in simple U-Net is better than for attention U-Net and ResU-Net. Results of Intersection over Union (IoU) for training and validation data are summarized in Table 7. Both background and alfalfa have been well detected. The IoU of grass was low in all U-Net models. The IoU of grass in simple U-Net was better than attention U-Net and ResU-Net.

3.1. Prediction of the test data

Fig. 11 and Fig. 12 depict several examples of validation images of alfalfa and grass and the corresponding mask, along with the prediction of the model by attention ResU-Net. These results show that the model learned the problem relatively well. Fig. 11 illustrates that alfalfa can be effectively detected under various lighting and weather circumstances. The initial four rows of Fig. 11 depict alfalfa in bright environments, but the final row illustrates a dark scenario. Furthermore, nearly all images in Fig. 11 indicate that alfalfa shading can be accurately detected as the class background.

Fig. 13 shows a few examples of the original image taken with an iPad. The images were captured in the fields, and the model has not seen them during either training or validation. Model predictions indicated that it can accurately detect alfalfa in an image. The images were captured at various growth stages under varying weather and lighting conditions. The predictions indicate that the model can to detect alfalfa under dark (first two rows) and sunny (last four rows) circumstances. Yellow, green, and purple colors show the background, alfalfa, and grass.

3.2. Automated stem counting

In total, the alfalfa stem count data were available for 2222 images, together with the good condition of the imagery. Since the model input must be a size of 256 by 256 pixels, the images were first resized to the nearest multiple of 256 for both height and width. Patches of 256 by 256 were then extracted from the image and predicted by the model. The total coverage of alfalfa in each image has been calculated. The relationship between the alfalfa cover fraction and the number of alfalfa stems is shown in Fig. 14-a. The 2D histogram of alfalfa cover fraction and number of stems is depicted in Fig. 14-b. We utilized several ML algorithms to predict the average number of alfalfa stems in a square foot. It should be noted that, for the regression, the number of stems greater than 120 was truncated, having been replaced by 120.

The results of various ML algorithms can be seen in Fig. 15 and Table 8. The results showed a considerable correlation ($R^2 = 0.79$) between the alfalfa cover fraction in an image, and the number of alfalfa stems per square foot. We added average crop height (in centimeters) to the alfalfa cover fraction as an auxiliary feature, slightly improving prediction accuracy. The results showed that RF (Random forest), with the addition of height, slightly improved the accuracy of ML algorithms. The value of R^2 for predicting the number of stems using RF was 0.83. By adding height, RF showed a slight improvement compared to RF only using the alfalfa cover fraction, in which the value of R^2 was 0.82. Among the models that utilized only alfalfa cover fraction as input, RF was the best model in terms of mean absolute error (MAE); 10.07 was reported for MAE. The extreme gradient boosting (XGB) had the same R^2 value as the RF. Yet, the MAE of the XGB was lower than those of the RF (RMSE = 13.00; MAE = 10.09). SVR yielded the worst accuracy among ML algorithms with R^2 of 0.81, RMSE of 13.27, and MAE of 10.32. Among the models created using alfalfa cover fraction and average crop height, RF outperformed other ML algorithms with an R^2 of 0.83, RMSE of 12.59, and MAE of 9.64. XGB was second in terms of highest accuracy with R^2 of 0.83, RMSE of 12.7, and MAE of 9.68. In all models, saturation can be observed when the number of stems exceeds 100. One of the reasons that could be responsible for these results is that we do not have sufficient data for the number of images, which is more than 100.

Alfalfa yield may decrease with plant densities ranging from 40 to 55 stems per square foot. Research has shown that stands should be replaced if the plant density drops below 40 stems per square foot. Stem density beyond 55 does not restrict the production. According to the above rationale, the observation and estimation values of stems utilizing the RF regression model have been categorized into three classes. We

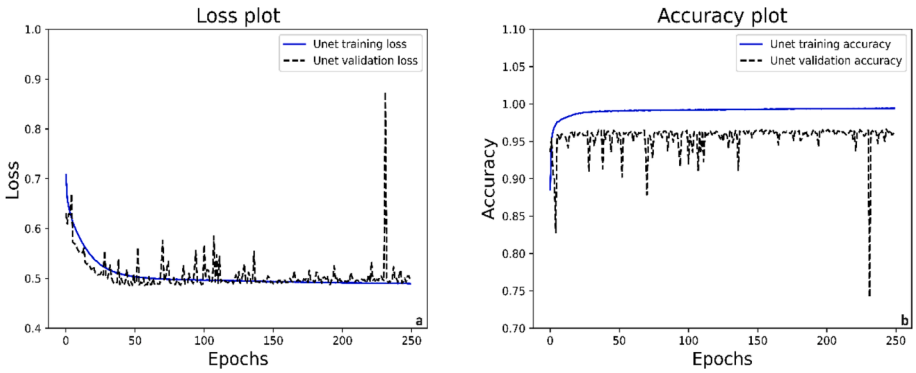

Fig. 10. Training and validation loss (a), and accuracy (b) of the model.

Table 6
The detaild results of the model.

	Attention ResU-Net			Attention U-Net			Simple U-Net		
	precision	**recall**	**f1-score**	**precision**	**recall**	**f1-score**	**precision**	**recall**	**f1-score**
Background	0.990	0.987	**0.989**	**0.994**	0.984	**0.989**	0.989	**0.989**	**0.989**
Alfalfa	0.990	0.934	0.961	0.989	0.936	0.962	**0.994**	**0.937**	**0.965**
Grass	0.792	0.978	0.874	0.784	0.979	0.872	**0.801**	**0.986**	**0.884**
Overall Accuracy	0.962			0.961			**0.965**		

Table 7
Results of Intersection over Union (IoU) for training and validation data using various U-Net models.

	IoU					
	Attention ResU-Net		**Attention U-Net**		**Simple U-Net**	
	Training	**Validation**	**Training**	**Validation**	**Training**	**Validation**
Background	0.983	0.977	0. 984	0.978	**0.985**	**0.979**
Alfalfa	**0.988**	0.929	**0.988**	0.929	**0.988**	**0.934**
Grass	0.972	0.780	0.973	0.779	**0.976**	**0.798**

identified five stems as errors in the prediction model. We categorized the classes as follows: 1) less than 35 stems as class 1, 2) between 35 and 60 as class 2, and 3) more than 60 as class 3. We computed the confusion matrix for the observed and predicted classes. It should be noted that the confusion matrix is not a straight regression product; rather, it is a post-processing output associated with the interpretation of findings. Fig. 16 shows the results. The graphic clearly shows that allowing a five-stem mistake allows the model to anticipate needing to re-seed the field with 86 % accuracy. The model can accurately predict class three with a 91 % success rate. Therefore, the model can effectively identify high-density alfalfa. Class 2 has the lowest accuracy compared to the other classes. The problem can arise at the boundary between classes since the number of stems is a continuous variable that we cut and divided into three classes. We believe that the border region (40 and 60 stems) is the source of this error. This prediction can accurately distinguish between classes 1 and 3 without any misclassified values.

3.3. Assessing the models

Upon ensuring the accuracy and proper training of the deep learning and machine learning models, we integrate the models into a sequential architecture to accomplish the ultimate goal of stem counting and make the models practical. We developed a Graphical User Interface (GUI) using Python. In this GUI framework, the user only needs to select the image captured of their fields. The alfalfa detection will be displayed within seconds. Furthermore, utilizing the alfalfa fractional vegetation cover computed in the previous step, the number of stems will be estimated and presented to the user. Fig. 17 illustrates three alfalfa images subjected to varying density conditions fed into the framework. The model did not detected any alfalfa stems in Fig. 17-a, indicating alfalfa absence in the image. The quantity of stems has been determined to be zero. Fig. 17-b depicts an area of relatively dense alfalfa. The alfalfa has been accurately identified, with an estimated density of 59 stems per square foot. Fig. 17-c presents a densely populated view of alfalfa. Most of the image has been covered by alfalfa, leading to a significant alfalfa fractional cover. The elevated value of fractional cover results in a high prediction of stems by the machine learning regression models (87 stems/foot^2).

4. Discussion

We applied pixel-based segmentation models to detect alfalfa and discriminate it from other objects over the alfalfa fields in this study. Pixel-based segmentation models have advantages and limitations compared to non-pixel segmentation techniques, such as the Segment Anything Model (SAM) or GroundedSAM. A pixel-based semantic segmentation model offers precise segmentation at the pixel level, making it suitable for tasks that demand great precision, such as precision agriculture. SAM and GroundedSAM are general-purpose models trained for broad application, which may not always correspond exactly with specific domain needs. U-Net designs can be readily modified for particular datasets by adjusting their depth, filter sizes, and loss functions, making them exceptionally adaptable for domain-specific applications. One drawback of pixel-based segmentation models is that training such models requires labeled pixel-wise ground truth, which is expensive and time-consuming to construct. SAM and GroundedSAM can work with limited input (e.g., bounding boxes or clicks) and do not consistently necessitate pixel-level annotations. Nonetheless, as previously stated, the dataset utilized in this study was generated using a pre-trained model, and the entire labeling procedure was not much more than one hour. Consequently, we were not faced with the issue of the laborious data labeling process. However, it should be noted that using more advanced U-Net models and transformers may improve the accuracy of alfalfa detection.

By interpreting the MAE calculated in almost all regression models, this value was ~10. This value tells us that the typical difference between our model's predictions and the actual alfalfa stems is ~10. Since the range of alfalfa stems was between 0 to ~120 stems, an MAE of ~10 indicates that the average error rate of the models is ~8.3 %, which is an acceptable value. Also, an RMSE of ~13 shows that we can expect 68 % of the stem values to be within 1 RMSE, given that the data is normally distributed. As shown in Fig. 18, most points are in the mean of ± RMSE. For the stem values of >80, the residual gets worse. However, in alfalfa fields, more than 55 stem values are considered highly dense alfalfa, and no re-seeding is required. Also, alfalfa stems under 40 are considered low density, and re-seeding may be considered for farmers. Therefore, based on the output of these models, we can expect that if the model prediction is <30 (40 – MAE), re-seeding may be necessary for farmers.

The results of this paper have shown that the number of alfalfa stems can be predicted from proximal imagery (iPad like) with relatively high accuracy, even though it's a complex process. Counting the number of stems depends upon various parameters, such as the height and tilt of the device taking the image, the growth stage and height of crops, whether the device is properly focused, and the weather and illumination conditions. If the device's height is very close to the crop's, the alfalfa cover fraction resulting from the U-Net model may be misleading. In this paper, we attempted to select the images with the imaging height at the standard level, meaning that the images were neither too close nor too far from the crop. Further, the device's tilt may result in the incorrect recognition of alfalfa in the images. Fugrthermore, we attempted to

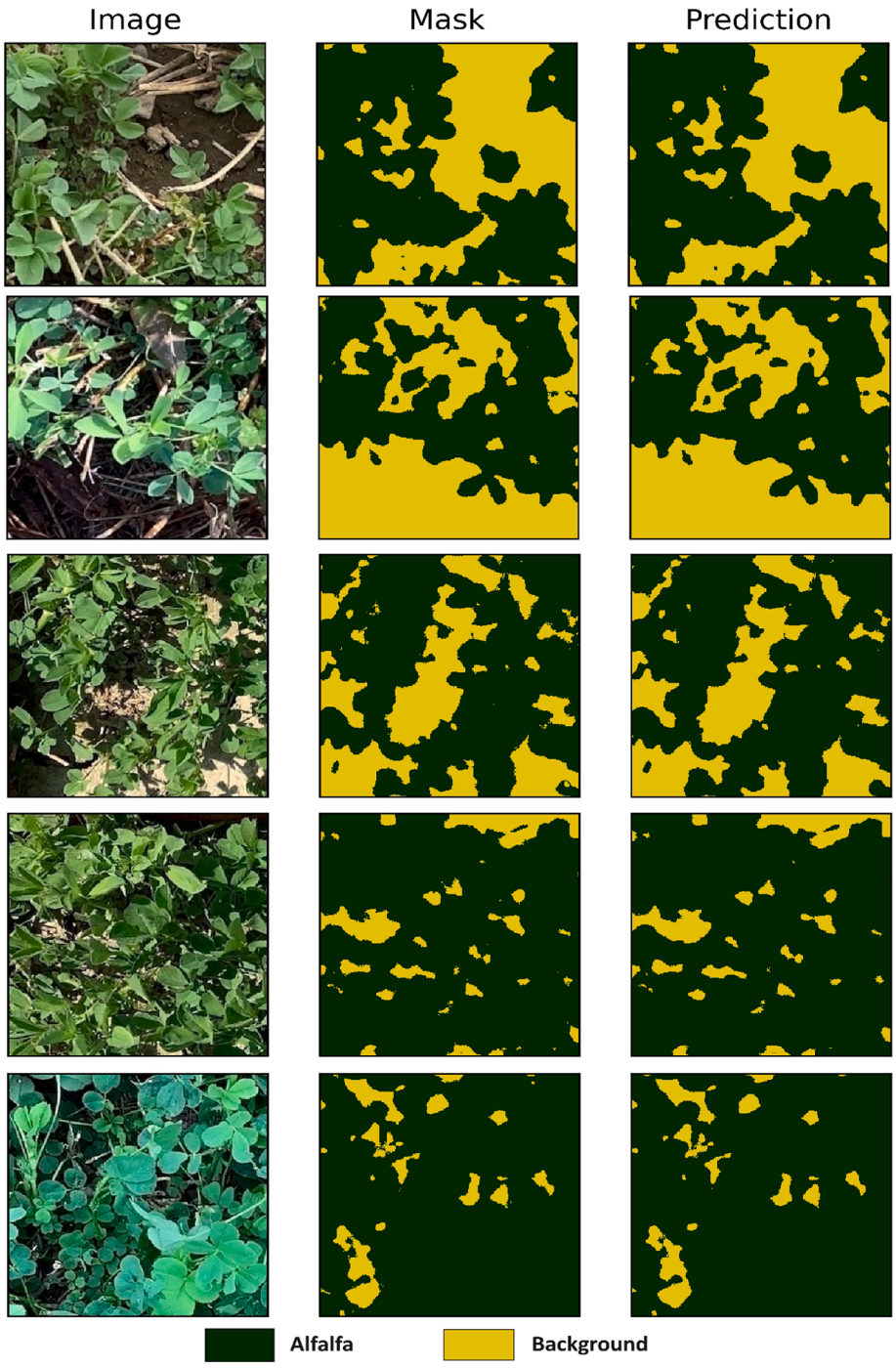

Fig. 11. Prediction of the model over images of the combination of alfalfa and background images.

include all growth stages in the training data. For the segmentation model if the crop is in the early growth stage, the alfalfa cover fraction results derived from the model may not correctly predict the number of stems in an image. Besides, color modifications and transformations can also be utilized for image correction. Non-linear filtering algorithms can also correct the blurry condition of images (Darwin et al., 2021). Although we tried to add crop height as an additional auxiliary parameter to address the problem of being at the early growth stage, the results showed that adding height does not considerably improve accuracy: the computation time and error rate increase when background complexity is present (Darwin et al., 2021). Furthermore, as the training data for this study was obtained from an RGB iPad Mini 5th generation, we remain unsure whether our model is compatible with RGB images captured by other sensors and devices. We will do testing in the near future, and if the model fails to perform with RGB images from other sensors, we intend to include more training data from these sensors and retrain the models accordingly. This ensures that the models are functional regardless of the device type.

While the model performs exceptionally well on numerous unseen data, there are cases where it struggles to differentiate between classes. Fig. 19 depicts several situations in which the model could not differentiate the classes appropriately. The model faces challenges with detecting grass branches in areas covered with alfalfa. We believe that this issue arose because the model incorrectly classified these branches as alfalfa stems. Fig. 19-a, b, and c illustrate examples where the model failed to distinguish the tiny grass branches in the images. The model

Fig. 12. Prediction of the model over combinations of grass and background images.

also encountered trouble differentiating certain weeds from the alfalfa class (Fig. 19-d). Although we included images of weeds in class grass and trained the model with these images, we observed that the model still needs more data to be fully able to discriminate various kinds of weeds from alfalfa. We believe this issue arose from an insufficient variety of weed data used in the training dataset. Owing to this lack of data, the models could not distinguish all kinds of weeds from other classes in the newly captured images. We intend to collect more images of various weed species in the alfalfa fields, incorporate them into the training dataset, and retrain the models. Another challenge that the model faced in identifying the classes was the cases that the color of the alfalfa leaves and stems was not quite green (Fig. 19-e and f). The number of images exhibiting this was limited. We believe this may be due to the alfalfa leaves and stems not being entirely fresh and vibrant.

The model demonstrated excellent performance on over 500 newly collected images, as we incorporated a comprehensive range of image types in the training dataset, including various weather conditions, lighting scenarios, and growth stages of alfalfa. The model correctly classified alfalfa and distinguished it from other objects. We intend to increase the training data with extra images to enhance the model's robustness and enable it to work in new and unseen scenarios. Besides, we believe including additional bands, such as near-infrared bands,

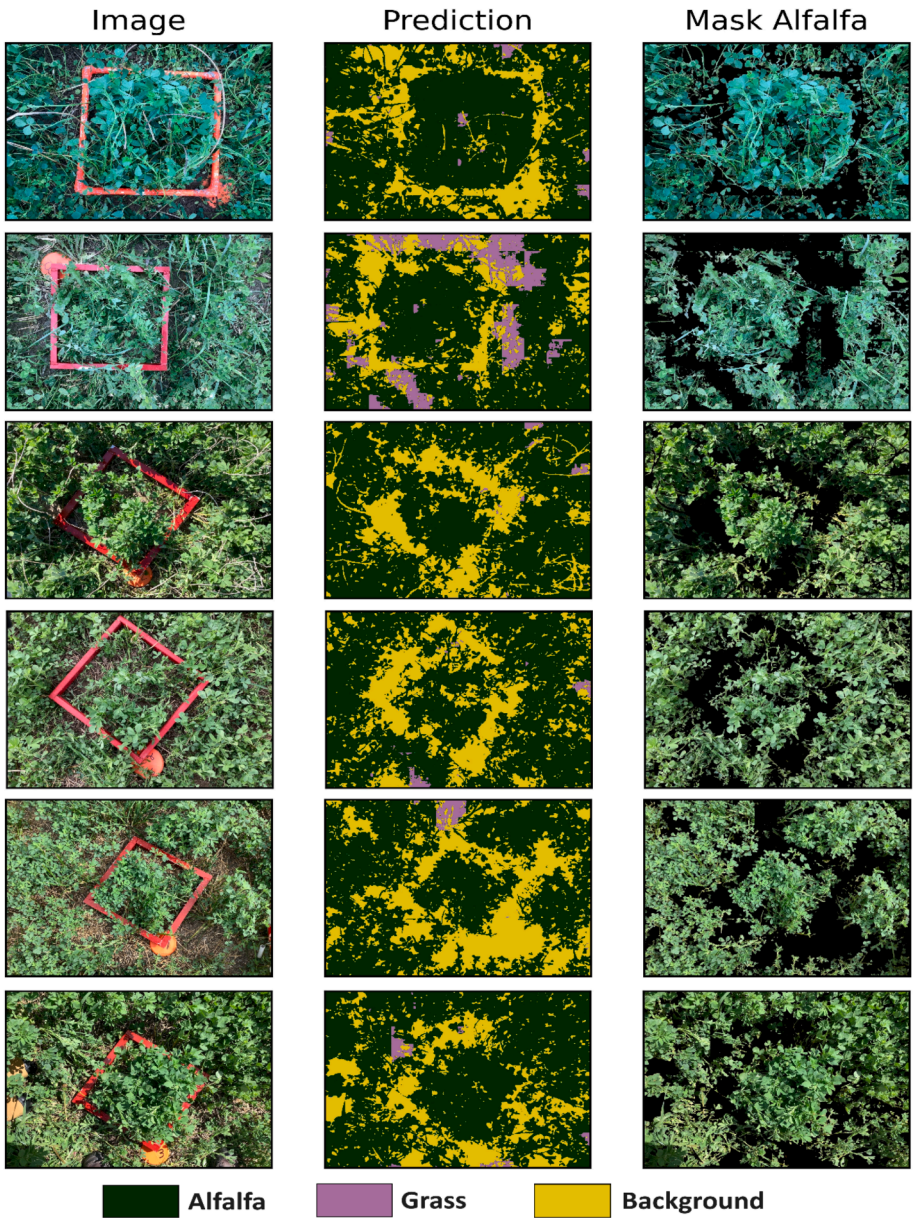

Fig. 13. Several examples of original iPad images, correspondingmask predicted by the model, and alfalfa mask class on the original image.

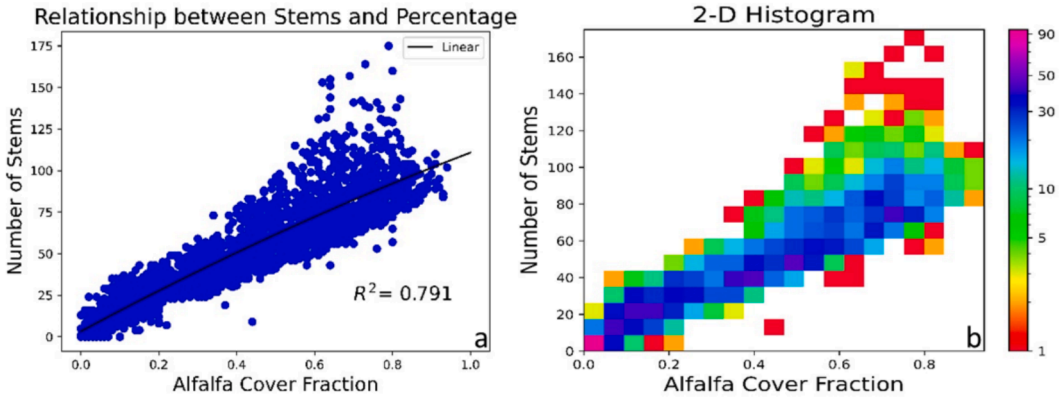

Fig. 14. Relationship between the number of alfalfa stems and alfalfa cover fraction in iPad images, presented as a) a scatter plot, and b) in a 2D histogram.

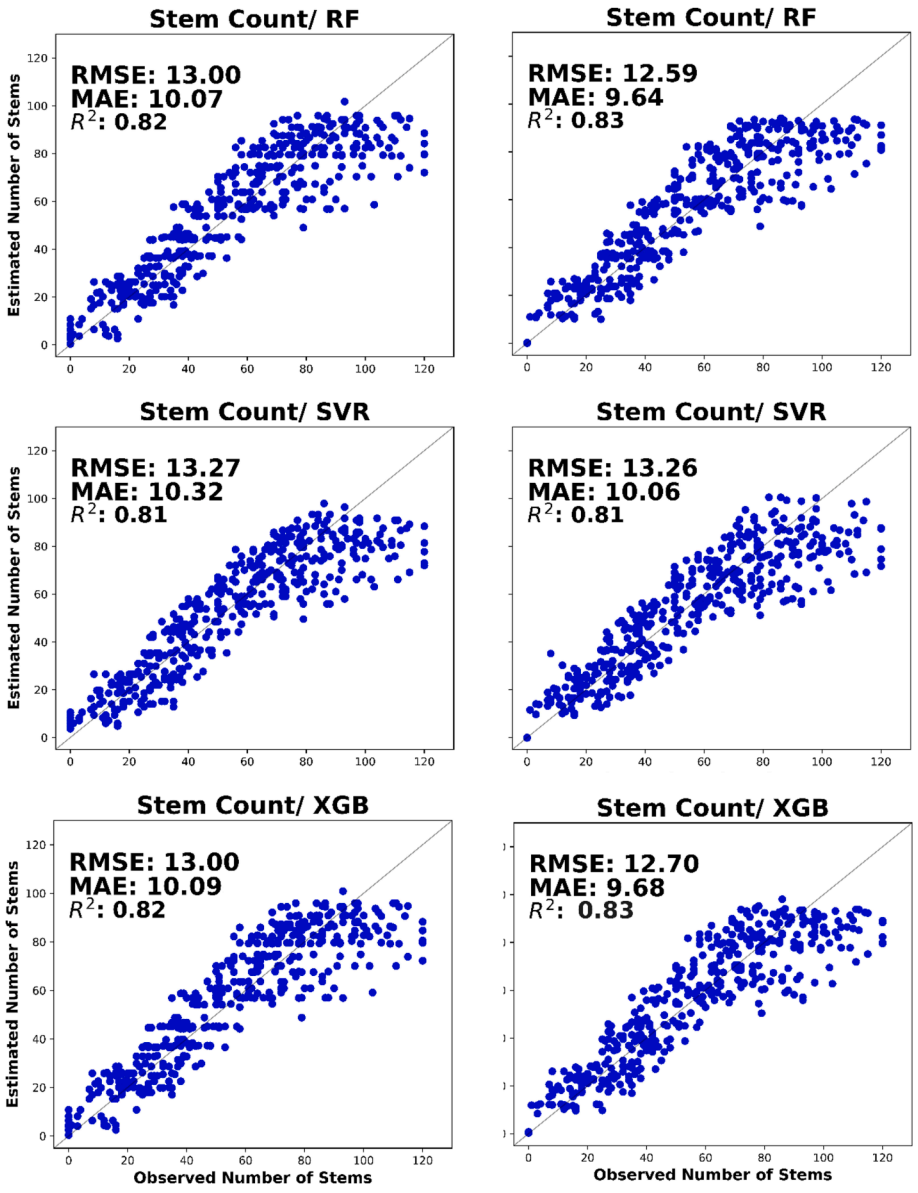

Fig. 15. Scatter plots of observed and predicted stem values in a) RF, b) RF including average crop height, c) SVR, d) SVR including average crop height, e) XGB, and f) XGB including average crop height.

Table 8
The results of machine learning regression models in estimating the number of alfalfa stems.

	RF		SVR		XGB	
	without height	with height	with height	without height	with height	without height
RMSE	**13.00**	**12.59**	13.27	13.26	**13.00**	12.70
MAE	**10.07**	**9.64**	10.32	10.06	10.09	9.68
R2	**0.82**	**0.83**	0.81	0.81	**0.82**	**0.83**

could enhance detection accuracy. Despite the potential high cost of utilizing equipment with near-infrared bands, we believe that incorporating data beyond RGB bands can considerably enhance the accuracy of alfalfa detection.

The methodology stated in this study aims to be applied through the construction of a mobile-based application and a website, both of which will be accessible to farmers throughout Canada as an initial step. Furthermore, we developed a graphical user interface program in Python. Users can easily upload an image or pass a directory, and all images will be processed, and results will be calculated and stored immediately. We evaluated the interface and found that an image requires 1.98 s to be fully processed. One drawback of the Python interface is that some individuals might not comprehend Python. Consequently, they will find it challenging to figure out the interface. To this end, the mobile-based application will be beneficial for the users. The application requires no prior knowledge of Python or any other programming language, allowing users to capture images and get the results immediately. Moreover, mobile-based applications are advantageous over web-based applications because they do not require an internet connection. This capacity makes mobile applications beneficial in remote areas and locations where the internet is unavailable.

It should be noted that this framework only works for alfalfa. However, the proposed encoder-decoder model used in this study can be applied in various domains of precision agriculture. This model could be particularly effective at differentiating crops from other background elements, including soil, agricultural residue, and dead crops or grass. By utilizing this model, we can detect the fresh and green crop sections

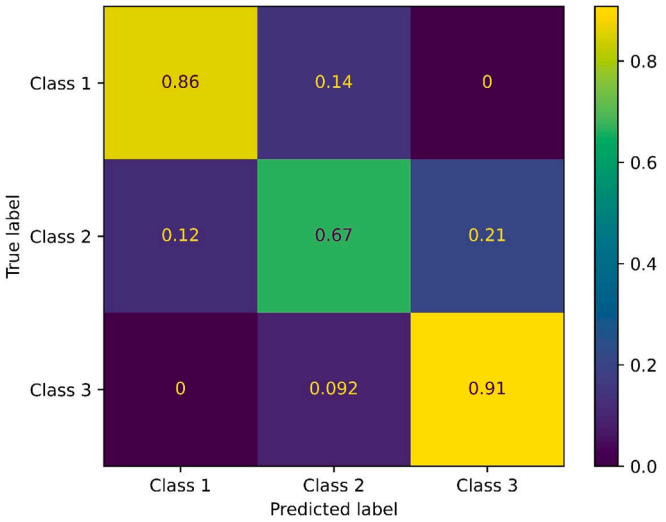

Fig. 16. The confusion matrix between the stems' observation and prediction values rcalculated using the RF regression model. Class 1 is defined as having between 0 and 35 stems, Class 2 as having between 35 and 60 stems, and Class 3 as having more than 60 stems.

and use this product for other post-processing sections, such as biomass and yield calculation. This model is highly effective for detecting grass and weeds in fields. Additional data is required to enhance the training set for improved detection of various weed types in agricultural fields.

5. Conclusion

This paper investigated the feasibility of various U-Net models to classify and detect Alfalfa using iPad images. After detecting alfalfa in the images, several machine learning regression models have been utilized to count the average number of alfalfa stems in a square foot. All U-Net models displayed strong capability in detecting alfalfa within the images. Testing the final models over real images showed that the model could detect alfalfa and easily distinguish it from grass and weeds across a crop field. It was further observed that utilizing synthetic images to simulate different conditions over a field was an excellent choice for annotating many images and training the model. The results showed that the simple U-Net was the best model with the highest accuracy among all U-Net models. The results also showed that by adding attention gates to the ResU-Net model, the model's accuracy did not considerably improve. A comparison between various regression models used by this study showed that RF was better at predicting the number of alfalfa stems than SVR and XGB. A real-time crop yield estimate can be obtained using the methods proposed in this paper. Although the results of this paper showed the capability of machine learning and deep learning models in detecting and estimating alfalfa over a field, some conditions must be met before taking images to have more accurate

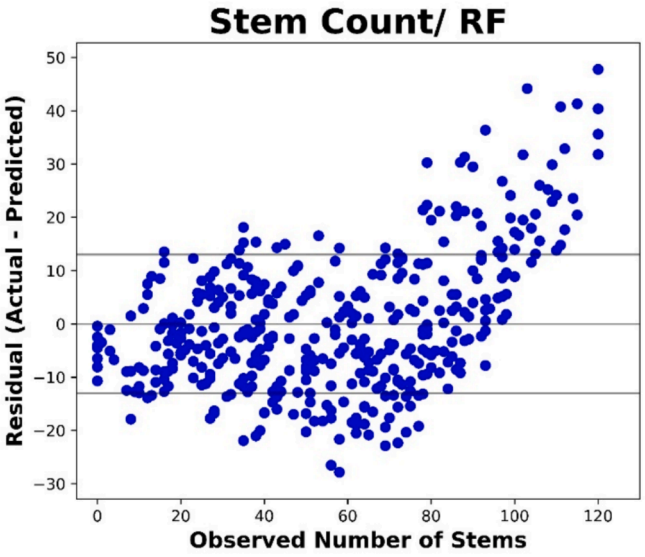

Fig. 18. Scatter plot of observed and residual values for RF model.

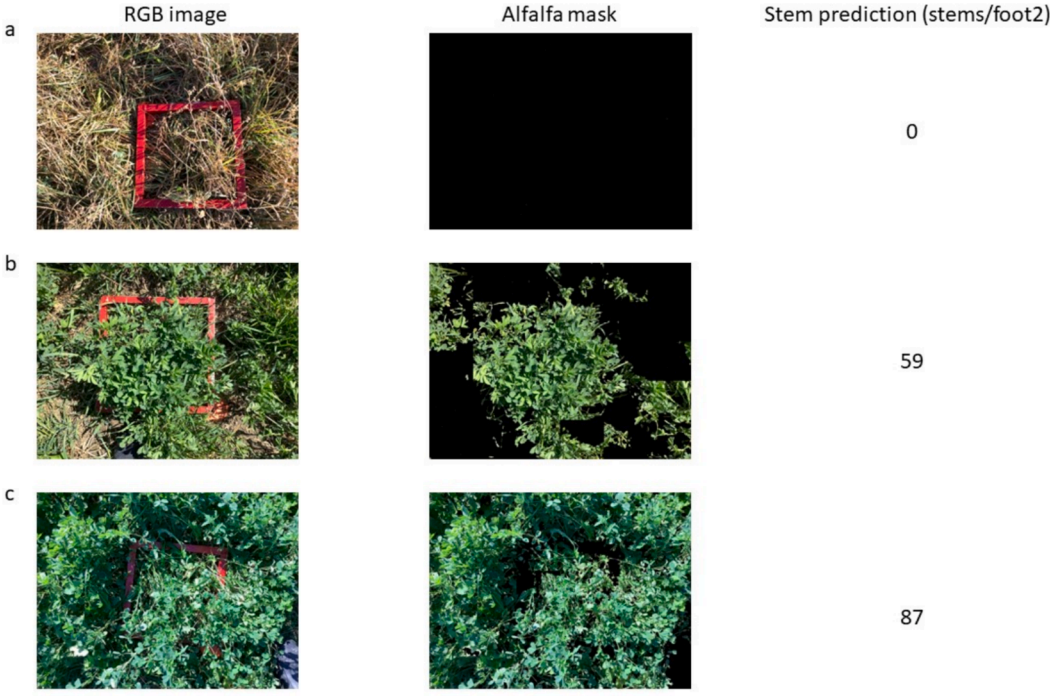

Fig. 17. Some examples of feeding the framework proposed in this study with different images of varying alfalfa densities.

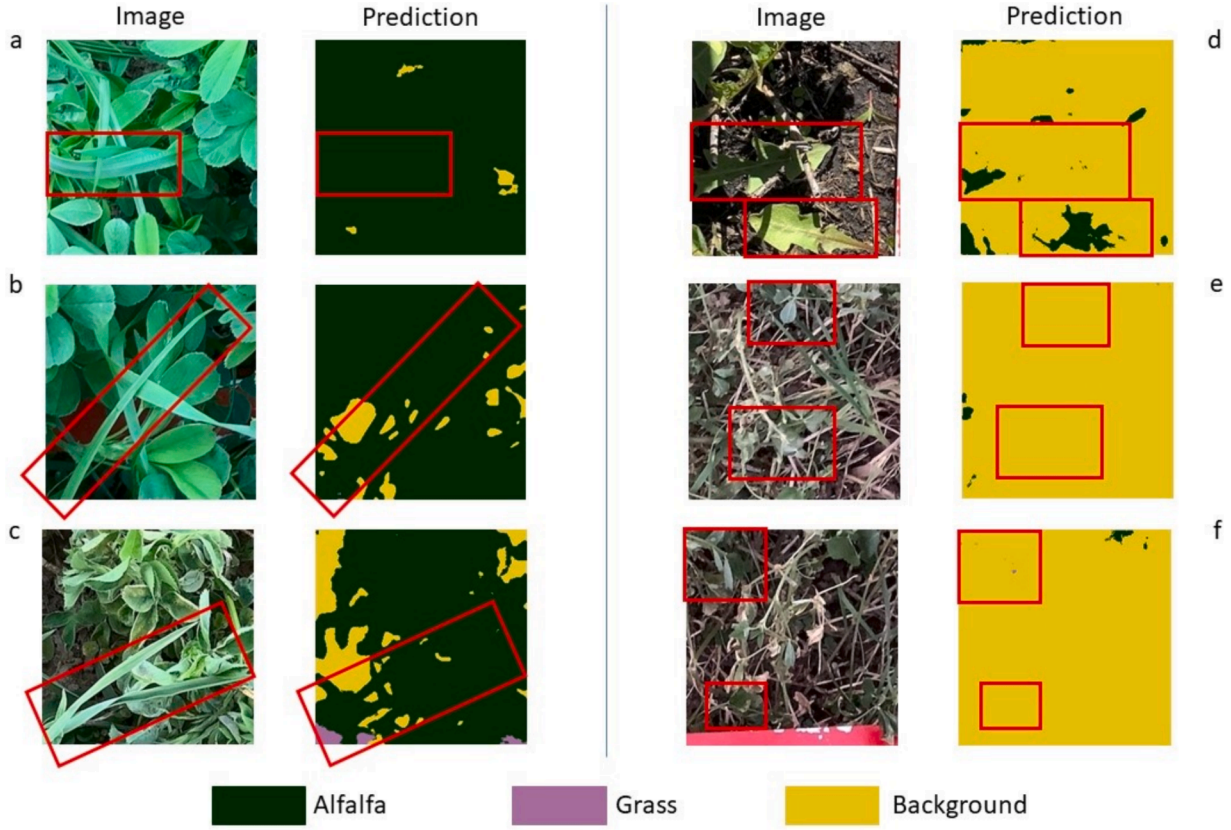

Fig. 19. Several examples of the model's challenges in differentiating the classes in the images.

results.

The methodology outlined in this study aims to be implemented through the development of an application and a website, both of which will be accessible to farmers throughout Canada as an initial step. Both the application and the website require only an image as input acquired by a mobile device (smartphone, tablette or similar device with RGB imaging capabilities). Through the utilization of the program, farmers may conveniently survey their fields, take an image, and swiftly identify the presence of alfalfa while accurately estimating the quantity of stems in a matter of seconds. This program resolves the time-consuming and laborious process of traditional counting methods. In our future work, we intend to utilize drone technology and various spatial–temporal satellite imagery to identify and estimate the fraction of alfalfa cover, as well as estimate the number of stems. All of the aforementioned models are intended to be implemented in both the application and the website.

CRediT authorship contribution statement

Hazhir Bahrami: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Conceptualization. **Karem Chokmani:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Saeid Homayouni:** Writing – review & editing, Writing – original draft, Supervision, Methodology. **Viacheslav I. Adamchuk:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Md Saifuzzaman:** Writing – review & editing, Methodology, Investigation. **Maxime Leduc:** Writing – review & editing, Resources, Project administration, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We gratefully acknowledge the Agri-Risk program of Agriculture and Agri-Food Canada (grant number CASPP-040), the Canadian Forage and Grassland Association (Edmonton, AB), Mitacs (grant number IT23205), Fonds de recherche du Québec - Nature et technologies (grant number 346438), De l'observation de la terre aux services d'information décisionnelle (DOTS), and the invaluable contributions of farmers and field advisors for their generous provision of data and financial support, without which this research would not have been possible. We also thank NVIDIA's Academic Grant program for providing two powerful NVIDIA Quadro RTX A6000 GPUs to our INRS Environmental and Northern Remote Sensing Laboratory. Finally, we would like to thank Maryam Rahimzad, a member of our Lab, for providing us with her great pre-trained model.

Data availability

The authors do not have permission to share data.

References

Akhavan, Z., Hasanlou, M., Hosseini, M., Becker-Reshef, I., 2021a. Soil moisture retrieval improvement over agricultural fields by adding entropy–alpha dual-polarimetric decomposition features. J. Appl. Remote Sens. 15, 034516.

Akhavan, Z., Hasanlou, M., Hosseini, M., McNairn, H., 2021b. Decomposition-based soil moisture estimation using UAVSAR fully polarimetric images. Agronomy 11, 145.

Aldakheel, Y., Assaedi, A., Al-Abdussalam, M., 2004. Spectral reflectance of alfalfa grown under different water table depths. In: International Conference on Water Resources and Arid Environment.

Andrews, C., 1987. Low-temperature stress in field and forage crop production–an overview. Can. J. Plant Sci. 67, 1121–1133.

Azadbakht, M., Ashourloo, D., Aghighi, H., Homayouni, S., Shahrabi, H.S., Matkan, A., Radiom, S., 2022. Alfalfa yield estimation based on time series of Landsat 8 and PROBA-V images: An investigation of machine learning techniques and spectral-temporal features. Remote Sens. Appl.: Soc. Environ. 25, 100657.
Bahrami, H., McNairn, H., Mahdianpari, M., Homayouni, S., 2022. A meta-analysis of remote sensing technologies and methodologies for crop characterization. Remote Sens. (Basel) 14, 5633.
Bangare, S., Rajankar, H., Patil, P., Nakum, K., Paraskar, G., 2022. Pneumonia detection and classification using CNN and VGG16. Int. J. Adv. Res. Sci., Commun. Technol. 12, 771–779.
Bélanger, G., Castonguay, Y., Bertrand, A., Dhont, C., Rochette, P., Couture, L., Drapeau, R., Mongrain, D., Chalifour, F.-P., Michaud, R., 2006. Winter damage to perennial forage crops in eastern Canada: Causes, mitigation, and prediction. Can. J. Plant Sci. 86, 33–47.
Brownlee, J., 2016. XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn. Machine Learning Mastery.
Cazenave, A.B., Shah, K., Trammell, T., Komp, M., Hoffman, J., Motes, C.M., Monteros, M.J., 2019. High-throughput approaches for phenotyping alfalfa germplasm under abiotic stress in the field. The Plant Phenome J. 2, 1–13.
Dangeti, P., 2017. Statistics for machine learning. Packt Publishing Ltd.
Darwin, B., Dharmaraj, P., Prince, S., Popescu, D.E., Hemanth, D.J., 2021. Recognition of bloom/yield in crop images using deep learning models for smart agriculture: A review. Agronomy 11, 646.
David, E., Daubige, G., Joudelat, F., Burger, P., Comar, A., de Solan, B., Baret, F., 2021. Plant detection and counting from high-resolution RGB images acquired from UAVs: comparison between deep-learning and handcrafted methods with application to maize, sugar beet, and sunflower. bioRxiv, 2021.2004. 2027.441631.
Dias, P.A., Tabb, A., Medeiros, H., 2018. Multispecies fruit flower detection using a refined semantic segmentation network. IEEE Rob. Autom. Lett. 3, 3003–3010.
Echeverría, A., Urmeneta, A., González-Audícana, M., González, E.M., 2021. Monitoring rainfed alfalfa growth in semiarid agrosystems using Sentinel-2 imagery. Remote Sens. (Basel) 13, 4719.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874.
Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., Luck, B., 2020. Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning. Remote Sensing 12 (12), 2028.
Fernandez-Gallego, J.A., Lootens, P., Borra-Serrano, I., Derycke, V., Haesaert, G., Roldán-Ruiz, I., Araus, J.L., Kefauver, S.C., 2020. Automatic wheat ear counting using machine learning based on RGB UAV imagery. Plant J. 103, 1603–1613.
Gao, F., Zhang, X., 2021. Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. Journal of Remote Sensing.
Garriga, M., Ovalle, C., Espinoza, S., Lobos, G., del Pozo, A., 2020. Use of Vis–NIR reflectance data and regression models to estimate physiological and productivity traits in lucerne (Medicago sativa). Crop Pasture Sci. 71, 90–100.
Ge, Y., Xu, J., Zhao, B.N., Joshi, N., Itti, L., Vineet, V., 2023. Beyond generation: Harnessing text to image models for object detection and segmentation. arXiv preprint arXiv:.05956.
Hancock, D.W., Dougherty, C.T., 2007. Relationships between blue-and red-based vegetation indices and leaf area and yield of alfalfa. Crop Sci. 47, 2547–2556.
Hunt Jr, E.R., Hively, W.D., Fujikawa, S.J., Linden, D.S., Daughtry, C.S., McCarty, G.W., 2010. Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. Remote Sens. (Basel) 2, 290–305.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. pmlr, pp. 448–456.
Jadhav, J.K., Singh, R., 2018. Automatic semantic segmentation and classification of remote sensing data for agriculture. Math. Models Eng. 4, 112–137.
Kamilaris, A., Prenafeta-Boldú, F.X., 2018. A review of the use of convolutional neural networks in agriculture. J. Agric. Sci. 156, 312–322.
Kaya, A., Keceli, A.S., Catal, C., Yalic, H.Y., Temucin, H., Tekinerdogan, B., 2019. Analysis of transfer learning for deep neural network based plant classification models. Comput. Electron. Agric. 158, 20–29.
Kayad, A.G., Al-Gaadi, K.A., Tola, E., Madugundu, R., Zeyada, A.M., Kalaitzidis, C., 2016. Assessing the spatial variability of alfalfa yield using satellite imagery and ground-based data. PLoS One 11, e0157166.
LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L. D., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1, 541–551.
Li, J., Wang, R., Zhang, M., Wang, X., Yan, Y., Sun, X., Xu, D., 2023. A method for estimating alfalfa (Medicago Sativa L.) forage yield based on remote sensing data. Agronomy 13, 2597.
Liu, T., Wu, W., Chen, W., Sun, C., Zhu, X., Guo, W., 2016. Automated image-processing for counting seedlings in a wheat field. Precis. Agric. 17, 392–406.
Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27, 1297–1328.
Luo, Z., Yang, W., Yuan, Y., Gou, R., Li, X., 2023. Semantic segmentation of agricultural images: a survey. Inform. Proc. Agri.
Marshall, M., Thenkabail, P., 2015. Developing in situ non-destructive estimates of crop biomass to address issues of scale in remote sensing. Remote Sens. (Basel) 7, 808–835.
McKenzie, J., McLean, G., 1980a. Changes in the cold hardiness of alfalfa during five consecutive winters at Beaverlodge, Alberta. Can. J. Plant Sci. 60, 703–712.
McKenzie, J., McLean, G., 1980b. Some factors associated with injury to alfalfa during the 1977–78 winter at Beaverlodge, Alberta. Can. J. Plant Sci. 60, 103–112.
Noland, R.L., Wells, M.S., Coulter, J.A., Tiede, T., Baker, J.M., Martinson, K.L., Sheaffer, C.C., 2018. Estimating alfalfa yield and nutritive value using remote sensing and air temperature. Field Crops Research 222, 189–196.
Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. 9, 62–66.
Quan, W., Liu, X., Wang, H., Chan, Z., 2016. Comparative physiological and transcriptional analyses of two contrasting drought tolerant alfalfa varieties. Frontiers in plant science 6, 1256.
Ranjbar, S., Zarei, A., Hasanlou, M., Akhoondzadeh, M., Amini, J., Amani, M., 2021. Machine learning inversion approach for soil parameters estimation over vegetated agricultural areas using a combination of water cloud model and calibrated integral equation model. J. Appl. Remote Sens. 15, 018503.
Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117.
Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., Homayouni, S., 2020. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, 6308–6325.
Singh, P., Verma, A., Alex, J.S.R., 2021. Disease and pest infection detection in coconut tree through deep learning techniques. Comput. Electron. Agric. 182, 105986.
Squires, V.R., Dengler, J., Hua, L., Feng, H., 2018. Grasslands of the world: diversity, management and conservation. CRC Press.
Suzuki, M., 1972. Winterkill patterns of forage crops and winter wheat in PEI in 1972. Can. Plant Dis. Surv. 52, 156–159.
Valente, J., Sari, B., Kooistra, L., Kramer, H., Mücher, S., 2020. Automated crop plant counting from very high-resolution aerial imagery. Precis. Agric. 21, 1366–1384.
Vance, C.P., Graham, P.H., Allan, D.L., 2000. Biological nitrogen fixation: phosphorus-a critical future need? Nitrogen fixation: From molecules to crop productivity 509–514.
Wachendorf, M., Fricke, T., Möckel, T., 2018. Remote sensing as a tool to assess botanical composition, structure, quantity and quality of temperate grasslands. Grass and forage science 73 (1), 1–14.
Yang, J., Bagavathiannan, M., Wang, Y., Chen, Y., Yu, J., 2022. A comparative evaluation of convolutional neural networks, training image sizes, and deep learning optimizers for weed detection in Alfalfa. Weed Technol. 36, 512–522.
Zou, K., Chen, X., Wang, Y., Zhang, C., Zhang, F., 2021. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. Comput. Electron. Agric. 187, 106242.