Applying Colour-Based Feature Extraction and Transfer Learning to Develop a High Throughput Inference System for Potato (*Solanum tuberosum L.)* Stems with Images from Unmanned Aerial Vehicles after Canopy Consolidation.

Joseph K. Mhango[a], Ivan G. Grove[b], William Hartley[a], Edwin W. Harris[a] and James M. Monaghan[a]

[a]*Crops and Environment Research Centre, Harper Adams University, TF10 8NB, Edgmond, Shropshire, United Kingdom.*
[b]*Curious Raven Imagery, Market Drayton, Shropshire, United Kingdom.*

Corresponding Author:
James M Monaghan
Postal Address: *Fresh Produce Research Centre, Harper Adams University, TF10 8NB, Edgmond, Shropshire, United Kingdom.*
Email Address: jmonaghan@harper-adams.ac.uk

Acknowledgements

Acknowledgement to the Crops and Environment Research Centre of Harper Adams University (TF10 8NB, Edgmond, Shropshire, United Kingdom) and the Agriculture and Horticulture Development Board's Potatoes Division (Sutton Bridge, Spalding, Lincolnshire, PE12 9YD, United Kingdom) for providing the experimental sites and for this study and the National Centre for Precision Farming at Harper Adams University for the operation of UAVs during image acquisition.

Abstract

Potato (Solanum tuberosum) stem density variation in the field can be used to inform harvest timing to improve tuber size distribution. Current methods for quantifying stem density are manual with low throughput. This study examined the use of Unmanned Aerial Vehicle imagery as a high-throughput alternative. A colour-based feature extraction technique and a deep convolutional neural network (CNN) were compared for their effectiveness in enumerating apical meristems as a proxy to subtending stems. Two novel colour indices, named the Cumulative Blue Differences Index and Blue Difference Normalized Index, showed significant differences (P<0.001) between meristematic leaves and mature leaves in comparison to other indices. The two indices were used to generate 500 pseudo-labelled human-corrected images as training data for the CNN. Benchmarked against a human labelled test dataset, the CNN performed better with a normalized Root Mean Square Error (nRMSE) of 0.09 than the sole use of the image analysis algorithm (nRMSE = 0.3) in predicting the number of meristems in a canopy at 52 days after planting. Furthermore, the CNN had better precision (Intersection over Union [IOU]: 0.49 and 0.56, respectively) than the image analysis algorithm (IOU: 0.33 and 0.13, respectively). Meristem counts in both approaches showed a linear relationship with actual subtending stem counts (P<0.001). This study demonstrates the validity of using traditional image analysis and CNNs to generate meristem detectors with *acceptable nRMSE. Transfer learning with CNN is proposed for developing meristem detectors for evaluating stem density variation from UAV images in the field.*

Key words: Vegetation Indices, Deep Learning, Potato, Plant Population, Phenotyping, Machine Vision

1. Introduction

At emergence, Potato (*Solanum tuberosum L.)* seed tubers produce variable sprout numbers depending on the physiological status of the seed, which results in variable stem numbers per potato plant (Knowles & Knowles, 2006). Estimation of spatial variation in plant density is important in potato production, with several studies linking it to tuber size and total yield variations at harvest (Bleasdale, 1965; Gray, 1972; Knowles & Knowles, 2006; Love & Thompson-Johns, 1999; Wurr, 1974). Potato growers normally have a contracted range of optimum tuber size, outside of which the value of the produce declines. Therefore, it is in the interest of growers to determine the factors that cause tuber size variation in the field and tailor management practices to mitigate the effects.

At the stem level, a negative correlation between potato stem density and mean tuber size has been widely recognized (Goeser et al., 2012; Mangani et al., 2015), and predictive models have been produced to describe potato tuber size distribution using stem density as a covariate (Bussan et al., 2007). To counteract the negative effect of stem density on average tuber size, several studies propose delayed harvesting to prolonging tuber bulking period as a strategy to increase tuber size (O'Brien & Allen, 1992; Rębarz et al., 2015; Waterer, 2007).

With this background, there is interest in techniques for determining stem density within an actively growing crop to enable spatially and temporally variable downstream crop management like vine desiccation, to maximize yield within desired tuber size classes. Manual stem counting in randomly or systematically selected quadrants around the field give approximations of stem densities which can be geospatially interpolated to the whole field, however this is a laborious, sometimes destructive, and a low throughput method. The validity of data interpolation relies on assumption of a random distribution of stem numbers or the unpredictable chance of establishing enough spatial autocorrelation in stem numbers to model the variation with minimal error, which is not always possible.

Using the spectral reflectance of potato plants to determine stem numbers from canopies is a potential approach for estimating accurate stem density from RGB or multispectral sensors mounted on Unmanned Aerial Vehicles (UAVs). This approach offers a high-throughput solution to estimate variation in stem population across the entire field without interpolation. Potato stems terminate with leaf primordia forming the growing tip of the stem, which sometimes convert to floral primordia depending on the variety (Firman et al., 1995). The leaf primordia therefore represents a distinct unit which can be used to estimate the total number of stems in a closed canopy. Plant canopies exhibit unique, species-dependent, responses to incident radiation, generally showing high absorption in the ultraviolet and Blue (490–450 nm) spectra, high reflectance in the Green (560–520 nm) spectrum, high absorption in the Red (700–635 nm) spectrum and high reflectance in the near-infrared portion (800-2500 nm) (Gates et al., 1965). Variability in chlorophyll content, water content and cell-to-air space ratio in the leaves directly influences spectral reflectance of plants in the visible (400-700 nm) spectrum (Cochrane, 2000), which can enable the use of computer vision and image analysis techniques to decompose consolidated crop canopies and enumerate features of interest based on their spectral reflectance.

Multispectral sensors with the near-Infrared band operating around 750 - 850 nm, enable the use of well-defined vegetation indices like the Normalized Difference Vegetation Index (NDVI) to assess and classify crop canopy components. Sankaran et al. (2015) used NDVI to extract and count emerging potato plant clusters from images taken at 15 metres above ground using a UAV at 32 days after planting with R^2 values of up to 0.82 when regressed to manual plant counts. However, predictive power was lost as the canopy gradually consolidated at 43 days after planting. The most widely used colour index for individual green plant segmentation from canopy remote sensing data is the Excess Green Index (ExG), first proposed by Woebbecke et al. (1995). The index has been used, in combination with other indices, for enumerating plant stands in wheat (Jin et al., 2017), rapeseed (Zhao et al., 2017), and in potatoes (Li et al., 2019), where the index was used to detect potato plant clusters at emergence. These techniques provide sufficient accuracy for counting clusters of stems from the same mother tuber at emergence before canopy closure. However, individual stem enumeration after full crop establishment, which is the level of accuracy required in precision farming for variable desiccation management, has not yet been reported.

This study hypothesized that a colour index to extract clusters of leaf primordia and enumerate them as a proxy to actual stem counts would potentially offer a solution. Following the spectral properties of plants outlined in Gates et al. (1965), an ideal colour index would be one that is sensitive to the differences between Blue and Green reflectance since the meristematic tips have less chlorophyll, thereby exhibiting lower reflectance in the Blue range compared to older leaves. The performance of object detectors based on image colour calculations depends on the acquisition of high quality imagery with optimum light conditions which are not always possible in the field. A deep learning approach therefore potentially presents a more robust approach with respect to variation in image quality. Ground-truth labelled data is essential in deep learning training pipelines and forms the basis of model evaluation in so called supervised learning models. Labelling a large dataset of leaf primordia from closed canopies has a large time cost as it requires expertise in identifying irregular leaf primordia. Semi-supervised learning therefore becomes a potentially important solution. Pseudo-labelling is a widely used technique to train Convolutional Neural Networks from non-labelled data with high accuracy. It involves the generation of an accurate model from a limited labelled dataset then using the model on unlabelled data and selecting all predicted labels that have high confidence as new labels, which helps to expand the labelled dataset.

The objective of this study was to use the spectral properties of plants in the visible wavelengths to develop colour indices for enhancing primordial features in canopies and use them to infer variation in actual stem number. The study also tested the use of colour indices for developing an automatic labelling algorithm for generating a training dataset for transfer learning using Faster Regions with Convolutional Neural Network” (Faster R-CNN) to generate a robust potato meristem detector for inferring variations in stem number.

2. Materials and Methods

2.1 Feature Engineering: Development of Colour Indices

2.1.1 Data Acquisition

Development and evaluation of colour indices was conducted using potato canopy imagery collected from Harper Adams University, Shropshire, England (52°46'26.8"N, 2°25'48.9"W) on a dark brown stone-less sandy loam soil. The images were collected from Amora and Maris Piper varieties at 91 and 50 days after planting respectively as shown in Table 1. Four different cameras, with varying sensor sizes and resolutions were used to generate variable sensor sharpness and colour resolving power. This enabled evaluation of the ability of the colour indices to distinguish meristem and old leaf pixels at different sensor sharpness and colour resolving powers so as to select the indices with the most consistent performance across sensors.

Table 1: Specifications of Unmanned Aerial Vehicle cameras and crop stage used in the study.

Location	Camera Description (Alias)	Variety	Days after Planting
	Colour Index Development		
HAU, Shropshire, England	DJI™ Mavic Air UAV equipped with a 1/2.3-inch CMOS sensor producing 12 MP still images at 88° FOV (Mavic)	Amora	91
HAU, Shropshire, England	DJI™ Inspire equipped with a Zenmuse X3 Camera equipped with 1/2.3-inch CMOS sensor producing 12.4 MP still images at 90° FOV (Inspire)	Amora	91
HAU, Shropshire, England	3DR Solo UAV mounted with a Mapir™ Survey 3N Camera with a Sony ExmorTM R IMX117 sensor, f/3.0 Aperture and 41° FOV (Mapir)	Amora	91
HAU, Shropshire, England	3DR Solo UAV mounted with a GoPro™ Hero 3+ Black Edition camera equipped with a 1/2.3-inch sensor with 12 MP and a fisheye lens with a 94.4° FOV (GoPro)	Maris Piper	50
	Model Training Data		
HAU, Shropshire, England	Phantom 4 pro UAV equipped with a Hasselblad L1D-20c aerial camera with a 1 inch CMOS sensor producing 20 MP still images at 70° FOV	Maris Piper	48
HAU, Shropshire, England	Phantom 4 pro UAV equipped with a Hasselblad L1D-20c aerial camera with a 1 inch CMOS sensor producing 20 MP still images at 70° FOV	Pentland Dell	48

HAU, Shropshire, England	Phantom 4 pro UAV equipped with a Hasselblad L1D-20c aerial camera with a 1 inch CMOS sensor producing 20 MP still images at 70° FOV	Amora	82
	Model Testing Data		
Shawbury, Shropshire	Mavic Air UAV equipped with a 1/2.3-inch CMOS sensor producing 12 MP still images at 88° FOV at 20 m attitude.	45 different varieties (see Appendix A)	52

MP=Mega pixels. CMOS = complementary metal-oxide-semiconductor. FOV = Field of View, DJI = Dà-Jiāng Innovations™, Shenzhen, China. HAU = Harper Adams University.

The crops were grown using standard UK commercial practices for all inputs. For the Amora variety, the aerial images were captured with 4 different visible light (Red, Green and Blue) sensors on UAV systems as outlined in Table 1. The Maris Piper crop was added to increase variation in the dataset as well as capture soil background because the canopy had not fully consolidated to cover the soil. Additionally, the Maris Piper images were collected on a day where the field was partially irrigated, providing the option of sampling pixels from both dry and wet soil. The captured images were manually evaluated to exclude pictures with distortion or blur due to UAV speed and a total of 5 images were selected from each camera, resulting in 20 images from which colour indices were evaluated.

All image processing was done using Matlab™ R2020a. From each image, pixels from Meristems, Mature Leaves, wet Soil and Dry Soil features were manually selected, and reflectance values for Red, Green and Blue were extracted. For each canopy feature, grayscale values were calculated from all selected pixels and the first 25 pixels from either side of the mean were selected to create 50 pixels per feature per camera. The final dataset for the evaluation of colour indices therefore contained 500 labelled data points of features with their Red, Green and Blue values.

2.1.2 Visible Light Colour Index Selection

Several colour indices were evaluated, with inclusion based on a literature search of widely used visible light spectrum indices as shown in Table 2. The ExG is most widely used for segmenting vegetation against soil background, however ~~we are unaware of any~~ evaluation of its potential for differentiating leaf age, health or stress has not been reported. Other indices included the Excess Red (ExR) Index for automatic segmentation of vegetation from soils, the difference between the Excess Green and Excess Red (xGxR), the Colour Index of Vegetation Extraction (CIVE) and the Excess Blue index (ExB), which is analogous to the ExR index. The Normalized Difference Red/Green Redness index (NDGR), normally used for enhancing the contrast between red backgrounds and vegetation was also considered.

Following the spectral reflectance properties of meristematic leaves described by Gates et al. (1965), it was hypothesized that the difference between Blue and Red, and Blue and Green reflectance at the pixel level would have the highest potential for maximizing contrast between meristems and older leaves. It was expected that meristematic pixels would reflect more Red than Blue while mature leaves would reverse this order due to darkening of the leaf as chlorophyll accumulated. Due to the darker hue of soil, it was assumed that soil pixels would exhibit negligible differences in reflectance among the three light bands. Two novel indices were therefore derived based on these premises. The Red to Blue difference was plotted against the Green to Blue difference using all the selected pixels from section 2.1.1 as data points. Visual observations showed that the Manhattan Distance or Euclidian Norm of each point as a vector from the Cartesian origin would provide an index that maximizes the difference between meristems and non-meristems. These two distances were therefore simplified into linear expressions, with the Manhattan distance termed as the Cumulative Blue Difference Index (CBDI) and the Euclidian Norm termed as the Blue Difference Norm Index (BDNI).

Table 2: Descriptions and equations of popularly cited and custom colour indices based on standard Red, Green and Blue bands.

Index Name	Formula	Source
Excess Green	$ExG = 2G - R - B$	(Woebbecke et al., 1995)
Excess Red	$ExR = 1.4R - G$	(Meyer & Neto, 2008)
Excess Green minus Excess Red	$ExG - ExR = 3G - 2.4R - B$	(Zhao et al., 2017)
Colour Index for Vegetation Extraction	$CIVE = 0.441R - 0.811G + 0.385B + 18.78745$	(Kataoka et al., 2003)
Excess Blue	$ExB = 1.4B - G$	(Guijarro et al., 2011)
Normalized Difference Blue Index	$NDGR = \frac{R-G}{R+G}$	(Bannari et al., 1995)
Cumulative Blue Difference Index	$CBDI = R + G - 2B$	Generated in this study
Blue Difference Norm Index	$BDNI = \sqrt{(R-B)^2 + (G-B)^2}$	Generated in this study

All colour indices were calculated for all pixel features. The resulting dataset was comprised of pixel values as observations, the four features as factors and the 8 colour indices as continuous independent variates. Differences in means from sensors which only contained Meristem and Leaf features were analysed using the T-Test while data from the GoPro sensor, which contained dry and wet soil features apart from the canopy vegetation features, was analysed by ANOVA using R version 4.0.2 (R Core Team, 2020) adopting a Completely Randomized Design and means were compared using Fisher's unprotected Least Significant Difference. The colour indices with significant highest differences in index value between meristematic and other features were selected for further analysis.

Following selection of colour indices, an algorithm was designed to estimate the number of stems in an image using Matlab™ R2020a. Briefly, the model consisted of (1) K-means clustering for segmenting the image into foreground pixels of interest and background pixels and (2) establishing an objective process for consolidating fragmented pixels of leaflets into single meristem units. These two components were used to create and test a feature-extraction and object detection process as illustrated in Fig.1.

2.1.3 Image Segmentation using k-Means Clustering

Both Otsu-based and adaptive threshold methods have been extensively reported in literature for green vegetation segmentation, but they perform poorly in images where the frequency of target and non-target pixels does not result in a bi-modal distribution of intensities (Yang et al., 2012). Meristematic tips constitute a small percentage of the image area in comparison to the rest of the canopy and non-canopy features. Consequently, intensity histograms from indices that maximize the reflectance of meristems against all other features like the CBDI and BDNI are expected to follow an exponential decay, which renders Otsu-based threshold selection unreliable for effective segmentation.

Since the images predominantly contained 3 object classes (i.e. Soil, and the non-meristematic canopy component as background and the meristems as the foreground), k-means clustering with three means was applied as the most appropriate method for clustering the three classes and segmenting the foreground without computation of a threshold. Using k-means clustering for segmentation is effective to separate object classes by minimizing the intra-class squared distance (Hartigan & Wong, 1979), and is widely applied in foreground segmentation in canopy images collected from UAVs and remote sensing systems (Chen et al., 2019; Cinat et al., 2019; Sun et al., 2019). Accordingly, the pixel with the maximum greyscale intensity in each image was identified, then k-means clustering was performed on each image and the cluster containing the identified pixel was chosen to represent the meristems. The image was then binarized with the selected cluster as foreground and all other clusters as background.

2.1.4 Noise Reduction and Final Bounding Box Generation

Due to the compound nature of potato leaves, the k-means-based segmentation produced unconsolidated meristem binary objects. Therefore, it was necessary to consolidate unconnected meristematic pixels that belonged to the same stem to minimize the chances of double-counting, while ensuring that meristems belonging to adjacent stems were not wrongly attributed. Morphological operations like erosion or dilation have the risk of consolidating some independent but proximal objects in a binary image (Pesaresi & Benediktsson, 2001). To avoid this, a custom noise reduction technique was created by shrinking every binary object in the image to its centroid pixel, followed by a pixel-wise iterative range search to index all other pixels located within a Euclidian distance that corresponded to the average size of a stem in the image. The average size of a stem at the plot level was estimated by calculating the average number of pixels per foreground object in each binary image. As a result, all pixels located in close proximity to each other were indexed together and considered to originate from the same primordia, then joined together. Pixels that were separated by a distance more than the estimated average stem size were not connected and constituted separate instances of a meristem. The number of connected components was then used as an approximation of the number of stems in the image and a minimum bounding box was generated to approximate the location and size of each stem, signalling the end of the algorithm. The flow chart of the algorithm was as illustrated in Fig.1

2.2 Development of a Transfer Learning Model

An aerial survey with a UAV was conducted at Harper Adams University on 09th June 2020 to develop a model training dataset of images images collected at 15 m altitude using a Phantom 4 pro UAV as specified in Table 1. Varieties covered in the survey were Maris Piper (0.5 ha) and Pentland Dell (0.5 ha) at 48 days after planting and Amora (9.4 ha) at 82 days after planting. The images were then cropped into 500 images of 500 pixels wide and 1500 pixels long then processed using the developed algorithm to generate bounding boxes around proposed potato meristems. The generated bounding for each of the 500 images were inspected and corrected manually by deleting erroneous detections and adjusting the extent of each valid box to fit the extent to which a human would label the data. Wu et al. (2020) emphasize on the computational constraint of training faster R-CNN object detectors from large UAV images, in their case 5472 X 3648 pixels, which necessitated the cropping of their images to 1000 X 800 pixels for optimized computation. The sensor used in this study produced 5472 X 3648 pixels, which were cropped to 1500 X 500 pixels to approximate the size of each plot in a compiled test dataset. All generated bounding boxes were stored as pseudo-labels to create a training dataset for deep learning with a CNN.

Fuentes et al. (2017) tested the Visual Geometry Group's (VGG-16) CNN (Simonyan & Zisserman, 2015) against deeper residual networks in the similar task of deep feature extraction of disease-caused leaf colour changes in tomatoes and found that the VGG-16 performed better than the deeper networks with up to 83% mean average precision. To keep the number of network backbone layers minimal for producing the simplest model with faster training times, A Faster R-CNN model (Ren et al., 2017) with the VGG-16 network backbone and imagenet weights was chosen. Faster R-CNN is a unified framework that learns an object region proposal network from a CNN feature map, classifies each proposed region and localizes the class of the object within the region with the introduction of anchor boxes, from which object bounding boxes are learnt and refined by regression. To convert a VGG-16 CNN into a Faster R-CNN object detector, a region proposal network was trained on the final convolutional feature map and the last max pooling layer was replaced by an ROI-max-pooling layer after which Faster R-CNN's classification and box regression layers were added to achieve object detection and localization. The training was conducted on an Nvidia GeForce GTX 1070 GPU with 8 GB Video RAM for 11 hours. The model was trained with a fixed learning rate of 0.0001, a single image mini batch size and 48 anchor boxes. The anchor boxes used in this study were predetermined iteratively by estimating an increasing number of anchor boxes and their sizes with each iteration, then checking their IoU with the ground truth data using the *estimateAnchorBoxes* function in Matlab™ R2020a. The final number and size of anchor boxes was chosen by observing the asymptote of the scatter plot of the determined IoU against the number of anchor boxes. Loss was optimized using the stochastic gradient descent with 0.95 momentum. The model converged after 50000 iterations in 100 epochs. The flow chart of the training pipeline is as illustrated in Fig.2.

2.3 Model Testing

2.3.1 Data Acquisition

The traditional image analysis algorithm and the deep learning model were tested for performance accuracy using a training dataset of images collected from 45 potato varieties (see Appendix A for a list of the varieties) grown at Eaton Upon Tern Runway (Fig.3), Shawbury, Shropshire, England (52°48'19.3"N, 2°30'41.8"W) on a Clayey Loam soil. The potato varieties came from the four determinacy classifications (Group one to four) and some varieties were non-classified by determinacy. Twenty Seven of the 45 varieties were drawn from the top 50 varieties grown in the UK in 2019 in terms of area planted. The varieties acted as a source of variation in stem numbers per unit area and canopy colour intensity. Differential performance between varieties was not considered in order to generalize model accuracy across varieties. The 45 varieties were planted on 2019-04-29, uniformly managed throughout the season and harvested on 2019-09-10. The ground-truth number of above ground stems was manually determined on 2019-09-10 before harvest. The number of visible meristems on top of the canopy was also manually counted. To create the model testing image dataset, two adjacent rows of 5 metres each per variety were imaged on 2019-06-20 at 52 days after planting at 20 m altitude using a Mavic Air UAV as specified in Table 1.

2.3.2 Image Processing and Data analysis

The aerial images were cropped manually around each of the 45 varieties plots to create an image for each plot for analysis then bounding box labels were manually defined for all meristems in each cropped image using Matlab™ R2020a's "image labeller" application. For each image in the test set, meristem counts were generated using the image analysis algorithm and the Faster R-CNN model then compared to the manually counted number of meristems. Bounding boxes were generated from the two predictive models. For each image representing a plot, the bounding boxes for the ground truth, Faster R-CNN detections and image analysis detections were converted into binary masks then confusion matrices were computed for the two detection models against the ground truth. The rates of true positives (TP) and false positives (FP) were computed for each of the 45 plots from confusion matrices. From these metrics, classification precision, as a measure of model performance, was computed as follows:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Precision was chosen over Recall and F1-score metrics because the image analysis approach was based on the detection of minor colour aberrations at the apex of the plant. This meant that bounding boxes for the image analysis approach were expected to be smaller than the human-verified ground truth where shape features that distinguish a meristem from older leaves were identified and the bounding boxes expanded. This was projected to cause a high rate of false negatives which would penalize Recall and subsequent F1-scores and therefore, the precision metric was used. The most important output of the model for practical decision support is the detection of the presence of a meristem for calculating stem density, while its size and extent are secondary considerations. Furthermore, the Intersection over Union (IoU) was calculated as follows:

$$Intersection\ Over\ Union = \frac{B_1 \cap B_2}{B_1 \cup B_2}$$

Where B_1 represents the ground truth and B_2 represents the predicted bounding boxes from the two models.

The final dataset contained the variety, breeder, number of manually counted above ground stems, number of manually counted meristems, and the number of meristems predicted by the image analysis and Faster R-CNN approaches. Observed vs Predicted plots were plotted for each prediction against the ground truth data to examine the residuals then the Root Mean Square Error (RMSE) was calculated for each model as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(Predicted - Observed)^2}{n}}$$

The RMSE was divided by the mean of the observed stem or meristem counts to calculate the normalized RMSE (nRMSE).

3. Results

3.1 Feature Engineering and Selection of Appropriate Indices

When the difference between Green and Blue light was plotted against the difference between Red and Blue light, grouped by the pixel source, four distinct clusters were visually discernible in Cartesian space. The data points of the meristems clustered in the first quadrant, the mature leaf data points clustered in the second quadrant while the two soil sources clustered near the origin (Fig.4). Upon visual assessment (Fig.4), the meristem data points were clustered at the largest Euclidian distance from the origin, followed by mature leaves. From this assessment, the CBDI and BDNI were considered to have potential to represent the overall variation linearly and guaranteeing that the meristematic pixels would be at the maxima of this variation's range, therefore enabling a predictable threshold selection. The CBDI and BDNI were calculated and compared with established RGB-based colour indices.

All the colour indices exhibited significant statistical differences between pixels from meristems and older leaves in mean intensity values ($P<0.05$), a trend sustained across all sensor types except for the ExR in the Mapir camera (Table 2). This suggests that the features of interest occur in exclusive quantiles of the range of each index, and therefore a k-means clustering approach would adequately segment the image into meristems and background, eliminating the need for determination of a subjective threshold. To maximize the chances of accurate segmentation, it was necessary to select colour indices that maximize the value of the meristematic pixels while maintaining a large separation with all other features. To achieve this, the separation between the maximum and minimum quantile of mature leaves and meristems was evaluated and the percentage overlap was calculated, with the aim of selecting the colour index with the largest difference between the values of meristem and non-meristem features.

Of all the colour indices, the BDNI (Fig.5-a) and CBDI (Fig.5-b) had the most consistent maximization of the index values for meristematic pixels, with distinct visual separation between the targeted and background features in boxplots of the indices. Boxplots of ExG, (Fig.5-c) and CIVE (Fig.5-d) showed an overlap of index values between the mature leaves and meristems. The minimum quantiles of meristematic pixels were consistently higher than the maximum quantiles of the mature leaves in only the CBDI and BDNI (Table 3). The opposite trend was consistently observed in the ExB, where the meristematic pixels had negative values due to a Blue colour deficit while the soils and mature leaves had excess Blue, resulting in a positive index, except in the Mavic and Mapir images. Unlike the EXG-ExR (Fig.5-e), NDGR (Fig.5-f) ExR (Fig.5-g), indices, the ExB index (Fig.5-h) achieved a linear positioning of features that would enable clustering. However, there were overlaps in the index values of meristem and mature leaf features from images obtained with all cameras except the GoPro, with the Inspire having a 28% overlap (Table 3).

As illustrated in table 3, the performance of the indices was consistent across camera type used to collect the data. Index values of the features of interest overlapped in 6 out of 8 indices in images taken with the Mapir Camera. The Mavic, Inspire and GoPro cameras had overlaps in five, four and two indices respectively. The ExG index consistently showed overlapping in all the cameras while the CBDI and BDNI indices only overlapped in the Mapir camera images. Based on this analysis, the CBDI and BDNI were chosen for use in the k-means clustering. In all further images, the two indices where calculated at each pixel and the resulting images normalized to 8-bit range then the two indices were combined into a single channel by averaging. The resultant grayscale image was then used for k-means clustering and subsequent stem count generation and bounding box approximation for both the image analysis and Faster R-CNN approaches.

Table 3: Mean values of colour indices calculated from the pixels of Leaves, Meristems, Wet soil and Dry soil using four different cameras

Camera[a]	Feature	BDNI	CBDI	CIVE	ExB	ExG	xGxR	ExR	NDGR[c]
GoPro	Leaf	40.3 (2.7)	12.3 (11.7)	-15.6 (2.3)	19.8 (9.8)	87.8 (5.6)	97.5 (11.2)	-9.7 (7.6)	-20 (3.0)
GoPro	Meristem	82.1 (17.8)	115.6 (25.2)	-6.7 (5.6)	-32.1(13.0)	70.2 (14.6)	27.2 (15.8)	42.9 (8.9)	-3.0 (0.2)
GoPro	Wet Soil	4.9 (1.9)	-10.0 (3.2)	23.6 (1.0)	56.9 (5.9)	-8.3 (2.1)	41.1 (9.2)	33.6 (7.7)	1.0 (0.2)
GoPro	Dry Soil	7.4 (2.4)	5.2 (4.3)	21.8 (1.6)	39.1 (5.8)	-1.6 (3.8)	-64.3 (9.0)	62.7 (5.9)	1.0 (0.3)
GoPro	Gap[b]	0.3***	1.4***	-3.0***	0.1***	-3.0***	0.1***	2.5***	27***
Inspire	Leaf	33.5 (4.9)	-16.5 (4.9)	-3.1 (7.1)	19.3 (4.8)	54.0 (17.5)	75.8 (26.6)	-21.5 (10.3)	-3 .0(1.0)
Inspire	Meristem	87.8 (14.7)	115.4 (22.6)	-28.7 (5.2)	-28.2 (16.4)	125.0 (13.0)	103.9 (19.8)	21.2 (11.1)	-1.0 (0.4)
Inspire	Gap[b]	1.0***	1.5***	-1.0***	-2.8***	-0.4***	-7.0***	0.1***	0.1***
Mapir	Leaf	81.2 (17.9)	114.3 (25.6)	-6.6 (5.6)	-31.7 (13.2)	69.9 (14.7)	27.5 (15.6)	42.4 (8.4)[NS]	-3.0 (0.02)
Mapir	Meristem	150.1 (16.1)	211.20 (22.5)	-31.9 (8.3)	-91.4 (15.9)	135.6 (21.04)	93.2 (27.7)	42.41 (9.6)[NS]	-0.6 (0.02)
Mapir	Gap[b]	-1.0***	-1.0***	1.9***	2.4***	-1.6***	-2.1***	-0.1	-1.0***
Mavic	Leaf	50 (5.5)	43.20 (10.9)	-23.6 (3.9)	-14.9 (14.6)	107.2 (9.5)	131.8 (17.9)	-24.6 (12.0)	-3.0 (0.7)
Mavic	Meristem	127.4 (16.6)	170.40 (24.3)	-47.5 (7.7)	-77.5 (18.8)	171.80 (17.9)	170.5 (28.3)	1.4 (13.2)	-2.0 (0.3)
Mavic	Gap[b]	1.5***	2.1***	-0.1***	-0.8***	-0.2***	-4.5***	-5.1***	28***

a = Camera alias, unless otherwise stated, there was significant difference in mean index values between meristem and non-meristem features in each camera ($P<0.05$). Standard deviations from each mean are expressed in parentheses. b = the interval between the minimum value of meristems and the maximum value of a mature leaf, expressed as a proportion of the range x 10^{-1} (x 10^{-2} for NDGR), negative values indicate that the ranges of the two features overlap. c = NDGR x 10^{-2}. NS=No significant difference between meristem and non-meristem features. CBDI = Cumulative Blue Difference Index, ExG = Excess Green Index, NDGR = Normalized Difference Green Redness, ExR = Excess Red, ExB = Excess Blue, xGxR = Excess Blue to Excess Red difference, CIVE = Colour Index of Vegetation Extraction, BDNI = Blue Difference Normalized Index.

3.2 Model Testing

3.2.1 Mean Stem Counts

Table 4: Performance of image analysis and Convolutional Neural Network approaches in the enumeration of meristems in all the varieties. The varieties are grouped into determinancy types for presentation purposes and "Actual Stem Number" information is included to illustrate the difference between meristem and stem counts.

Variety Group	Actual Meristem Number	Actual Stem Number	Image Analysis Meristem Prediction	CNN Meristem Prediction
1	77.5 (2.1)	40.5 (3.5)	82.0 (2.8)	70.0 (2.8)
2	82.6 (12.2)	50.3 (13.2)	102.0 (32.1)	75.0 (8.5)
3	78.4 (7.9)	47.4 (6.7)	87.1 (22.6)	73.0 (8.0)
4	67.3 (3.1)	48.7 (5.1)	74.0 (16.1)	67.7 (1.5)
UG[a]	81.1 (9.1)	49.3 (11.5)	92.2 (28.2)	75.8 (10.0)
Grand Mean	**79.8**	**48.7**	**90.6**	**74.5**
RMSEm[b]			**24.1**	**7.3**
nRMSEm[b]	**-**	**-**	**0.3**	**0.1**
RMSEs[c]			**46.9**	**26.8**
nRMSEs[c]	**0.7**	**-**	**0.9**	**0.6**

RMSE = Root Mean Square Error. nRMSE = Normalized Root Mean Square Error. a = Unknown variety group. b = RMSE or nRMSE with meristem ground truth as the observed variable. c = RMSE or nRMSE with manual stem counts as the observed variable

Actual main stem counts from the field validation showed that the average number of above-ground stems per determinacy group had low variation ranging from 47 to 50 stems while there was more variation in the actual number of meristems counted, ranging from 67.3 in group 4 varieties to 82.6 in group 2 varieties (Table 4). Faster R-CNN had a better predictive accuracy for the total number of meristems (nRMSE=0.09) than the image analysis method (nRMSE=0.3). Both Faster R-CNN and image analysis algorithms had low accuracy in predicting the actual number of stems in the plot (nRMSE was 0.6 and 0.9 respectively) and the same observation was made when manually labelled meristem were compared to the manual stem counts (nRMSE = 0.7) as shown in table 4.

Least squares linear models of the predicted meristem counts against manual meristem counts showed an R^2 value of 0.57 (Fig.6) and 0.73 (Fig.7) for the image analysis method and Faster R-CNN learning method respectively. Additionally, there was a significant ($P<0.001$) relationship between manual counts of main stems and Faster R-CNN meristem detections (Fig.8) as well as counts from the image analysis approach (Fig.9).

3.2.2 Localization Accuracy

The image analysis method had a low mean IoU (0.3) and Precision (0.1) compared to the Faster R-CNN method (IoU = 0.5, Precision = 0.6) against the ground truth bounding boxes (Table 5). The Image Analysis algorithm had an average bounding box size that was closer to the average size of the ground truth boxes than observed in the faster R-CNN model (Table 5). The Inter-quartile Range (IQR) showed that there was more spread in the bounding box predictions of the image analysis method than the Faster R-CNN method, which predicted more equally sized bounding boxes (IQR=159.6).

Table 5: Means and standard deviations (in parentheses) of detection and localization performance metrics of the image analysis and Convolutional Neural Network against manually labelled meristem data

	Ground Truth	**Image Analysis**			**Faster R-CNN**		
Variety Group	**BB[a] Size**	**IoU[b]**	**Pr[c]**	**BB[a] Size**	**IoU[b]**	**Pr**	**BB[a] Size**
1	2991.1 (657.5)	0.4 (0.3)	0.2 (0.2)	1739.8 (584.2)	0.5 (0.5)	0.6 (0.2)	3161.1 (45.2)
2	2382.7 (745.9)	0.31 (0.5)	0.1 (0.7)	1253.8 (222.2)	0.4 (0.5)	0.4 (0.1)	3168.9 (130.6)
3	2985.4 (366.9)	0.32 (0.7)	0.1 (0.6)	2468.9 (304.2)	0.5 (0.7)	0.6 (0.2)	3159.0 (111.4)
4	2851.7 (	0.3 (0.5)	0.1 (0.7)	938.8 (343.7)	0.4 (0.3)	0.6 (0.1)	3015.9 (172.2)
UG[d]	2996.3 (553.5)	0.34 (0.9)	0.4 (0.9)	2132.6 (204.9)	0.5 (0.5)	0.6 (0.1)	3125.9 (110.6)
Mean	**2930.2**	**0.3**	**0.1**	**2009.8**	**0.5**	**0.6**	**3129.2**
IQR[e]	**814.4**			**609.9**			**159.6**

a = Bounding Box. b = Intersection over Union, standard deviation values are x 10-1. c = Precision, standard deviation values are x 10^{-1}. d = Unknown Variety Group. [e]Interquartile Range

4. Discussion

4.1 Feature Engineering and Development of Colour Indices

Both the CBDI and BDNI indices achieved better classification of the meristematic leaves than the other indices compared. The CBDI and BDNI indices were derived in such a way as to take advantage of the theory that plant leaves exhibit variable reflectance of the Blue wavelength based on the age of the leaves, and the maximization of index values in meristematic features in line with the projected spectral signature of Gates et al. (1965). The Excess Blue index equally agrees with the findings of Gates et al. (1965) as it shows sensitivity to the diminished level of Blue light reflectance in meristematic structures, leading to lower index values than older leaves and soil.

In agreement with findings from Woebbecke et al. (1995), The Excess Green index adequately separated soils from canopy features. However, the index showed insensitivity to the amount of reflected green light between the meristematic structures and leaves, though the matured leaves had a higher mean reflectance than the meristems. The range of the Excess Green index and all other indices (Fig.5) in meristematic leaves overlaps with the range of the matured leaves, reflecting different levels of chlorophyll in meristematic leaves as affected by the age of the leaf. This is expected as noted by Gates et al. (1965) that a sharp drop in Red reflectance accompanies the continued increase in green reflectance with leaf age as proto-chlorophyll is converted to chlorophyll.

Though potatoes generally contain a larger concentration of the lighter shaded chlorophyll-a than chlorophyll-b (Anžlovar et al., 1996), a noticeable difference in Blue reflectance can be expected in mature leaves compared to the meristems which still have proto-chlorophyll. This is confirmed by Gates et al. (1965) who illustrates a slightly higher reflectance in the Blue range from mature leaves than younger leaves in reflectance curves. The CBDI and BDNI achieve better classification of meristematic leaves because they take this Blue light reflectance into account in relation to green reflectance. The difference between these two wavelengths is responsible for the high Manhattan distance and Euclidian norm from the origin in the meristems (Fig.4). The results also show that the difference between Blue and green reflectance is minimal in soils, showing more reflectance in the green range than the Blue range in dry soils. This is in agreement with soil reflectance curves reported by Huete (2004) which show a linear increase in reflectance from Blue to Near Infrared. Baumgardner et al. (1986) reported similar curves which consistently show more Red than Blue reflectance in soil. The findings for dry soils in this study concur with Baumgardner et al. (1986), however, wet soils were found to reflect more Blue light than Red. Huete (2004) and Baumgardner et al. (1986) discussed a decrease in reflected energy which makes soils appear darker, consistent with the high reflectance of Blue wavelength observed. These findings make the CBDI and BDNI ideal as they minimize the index values of soils and mature leaves in comparison to meristems. Comparison of the boxplots of the two indices additionally shows that the BDNI can be used as a general colour index as it additionally separates vegetation from soils, while the overlap between mature leaves and soils in the CBDI would make it unsuitable as a general colour index.

When targeting sparse features that do not show a peak in the feature space's histogram, Otsu-method binarization of an image is known to produce non-satisfactory segmentation. K-means segmentation adopted in this study provides an alternative that formulates clusters of features based on the variation in the feature space (Yang et al., 2012) rather than a subjective segmentation threshold. Where the feature space is defined by the Manhattan distances using the CBDI or Euclidian distances in the BDNI, automatic selection of a cluster of interest as a basis for binarizing the image is made possible since the meristems cluster is bound to occur at the upper quantiles of the histogram.

4.2 Model Testing

Observed vs predicted plots of the number of meristems in the image analysis and Faster R-CNN methods had R^2 values of 0.57 and 0.73 respectively (Fig.6 and 7). Faster R-CNN has an advantage over image analysis with a low nRMSE of 0.09 compared to 0.3 nRMSE observed in the image analysis. With no previous studies on potato stem detection, these results can be benchmarked against models that detect variation in leaf colour and shape due to viral leaf yellow mottling and crinkling akin to the underdeveloped leaves of meristematic tips. In this respect, the Faster R-CNN performs comparably to findings by Duarte-Carvajalino et al. (2018) where convolutional neural networks achieved a maximum of 0.82 R^2 value for the detection of incidences of Late Blight (*Phytophthora infestans*) on potato leaves when compared to manually labelled ground truth data. Comparably, Sugiura et al. (2016) similarly developed an image analysis protocol for estimating the severity of late blight with R^2 of 0.77. The results presented here show that the Faster R-CNN approach is as efficient as other studies that aim to detect objects of interest in potato canopies that are humans identify based on colour and leaf shape. The difference between predicted counts and observed counts in the image analysis approach show the need to account for more variation within the image by improving the image segmentation and the algorithm's inclusion criteria of an independent stem. Improvements in the image segmentation can be achieved by further feature engineering to generate more robust colour indices. Furthermore, although K-means clustering and subsequent cluster segmentation overcomes the problems of Otsu-based segmentation in non-bimodal data, the hard-coding of cluster number introduces the possibility of misclassification of ambiguous pixels, a double-edged sword that caused both false positives and false negatives (Kanungo et al., 2002). More in-depth studies into possible adaptive threshold selection techniques at the image level are needed to generate robust clustering and threshold selection rules to improve accuracy. Differences between predictions and observations in the Faster R-CNN model can partially be attributed to the limited variation in the training dataset, generated from two potato varieties, against the testing dataset which contained 45 varieties with variable canopy characteristics.

The performance of region-based CNNs is influenced by the adequate determination of the number of anchor boxes and their sizes at the training phase (Zhao et al., 2019). The irregularity of potato meristems means there needs to be a representative compendium of anchor boxes to cover the high variation in ground truth bounding box sizes. In this study, the ground truth bounding boxes had a high IQR of 814.38 pixels compared to the predicted bounding boxes of the CNN (159.61) and image analysis (609.86) on the test dataset (table 5). The CNN model produced regular (equally-sized) but larger bounding boxes than the ground truth while the image analysis approach produced smaller bounding boxes than the ground truth but were more variably sized, more naturally representing the variation in sizes of meristematic tips. In subsequent studies with the CNN approach, a more exhaustive method of anchor box size estimation is warranted, but equally so is the development of the model from lower resolution imagery at higher UAV altitude to reduce the ground truth IQR of the test dataset and potentially improve the model accuracy, though this comes at a cost of more errors in labelling low resolution imagery. These observations signal potential improvements to the data collection and hyper-parameter settings which may improve model accuracy in future studies. The small bounding boxes in the image analysis approach were reflective of the results of k-means clustering on the novel colour indices which were highly optimized to maximize values of meristematic pixels against mature leaves. However, the high R^2 values observed in both models show that there is a significant correlation between the predicted and actual meristem counts, as well as actual main stem counts (Fig.8 and 9), which shows that both models can be used in mapping this variation at field scale, a key desire for farmers who seek to vary vine desiccation dates based on stem density to manage potato tuber sizes and their distribution at harvest.

The faster R-CNN model achieved higher precision (0.56) and mean IoU (0.49) across the variety groups compared to the image analysis method (0.13 and 0.33 precision and IoU respectively), showing better efficiency at learning the features that a human labeller would identify with meristems, as well as the effect of the human-verification and adjustment of training labels in section 2.2 on the final model. In the absence of potato meristem segmentation studies, precision scores were benchmarked against the Potato Virus Y (Polder et al., 2019), whose primary symptom is chlorotic foliage akin to the signal being detected by the image analysis approach to label stems. Polder et al. (2019) found precision scores between 0.23 and 0.54 when a fully convolutional network was used to achieve semantic segmentation of Potato Virus Y. This is comparable with the performance of the faster R-CNN approach but outperforms the image analysis method. While the image analysis approach also adequately identifies the presence or absence of a meristem, the size and centroid of its resultant bounding boxes is less consistent since the system is purely based identifying the extent of the colour aberration at the very tip of the youngest leaves and not learning any other advanced features as in the Faster R-CNN. As a result, the image analysis approach produces highly variable meristem sizes within an image as shown by the high IQR. However, its inclusion in the Faster R-CNN pipeline is justified as it speeds up the labelling of a large dataset, allowing a human-labeller to only correct the computer generated labels. Potato meristems are not difficult to annotate for domain non-experts. The image analysis method allows the generation of initial annotations to guide labellers and train non-expert labellers to identify canopy features of interest from which they can simply adjust bounding box extents and hence speed up the annotation process.

For the purposes of deriving a management or phenotyping tool for evaluating variations in stem sizes across different stem densities, the establishment of a significant linear relationship between predicted stem counts and actual counts is important despite the presence of residuals because the linear relationship can be used to model spatial variation in stem density at field scale. While Sankaran et al. (2015) reported a predictive model with R^2 values of 0.83 for modelling plant density variation at emergence using the NDVI, they observed that predictive accuracy was lost as the canopy consolidated and they were not able to successfully run the prediction after 43 days from emergence. Furthermore, the effective unit of plant density in the potatoes is the stem, which can only be evaluated when all potential stems have developed, after plant canopy consolidation (Wurr & Morris, 1979). The overall 0.73 R^2 value in this study's CNN method gives a level of accuracy that is comparable to Sankaran et al. (2015) while offering the desired ability to enumerate the preferred unit of plant density, which can be incorporated in vine desiccation decision support systems for manipulating tuber size distribution at harvest and in high throughput phenotyping. With 40 tubers planted per plot, the actual stem counts found in this study mean that the average number of above ground stems per plant (1.21) falls within the ranges (1-4.4) reported in literature (Wurr & Morris, 1979). Most plants had one or two primary stems due to physiologically young seed tubers, stored below induced dormancy-breaking temperature. While the meristems represent the termination of both primary and secondary stems as well as sympodial branches, it can be noted that the average number of meristems per plant (2.01) also falls within the range of the number of main stems per potato plant reported in literature (Wurr & Morris, 1979) and further suggests that most plants in this study produced one or two primary stem and one secondary stem. The potato main stem always terminates with a meristem in all varieties and a sympodial branch continues growth in indeterminate varieties (Almekinders & Struik, 1996). The average number of secondary branches per stem reported in literature is minimal ranging from 0.5 to 0.9 branches per main stem (Vos & Biemond, 1992; Wurr & Morris, 1979). Therefore, while the number of meristems does not directly correspond to the number of main stems, its density variation across the field is a predictable proxy for stem density variation, which is the main desired unit of potato plant density whose determination at field-scale had so far been elusive (Wurr & Morris, 1979).

The number of main stems formed by a potato is largely variable and contingent upon the physiological age, plant population density and other agronomic and management factors

(Knowles and Knowles, 2006). The number of secondary stems formed is also dependant on factors that affect apical dominance like inherent determinacy characteristics and frost events (Chang et al., 2014). Additionally, differences in growth rates between stems means some meristems are occluded from view at the top of the canopy by other leaves, hence cannot be captured by UAV. These factors all contribute to the residuals between the number of actual main stems and the number of meristems detected at the top of the canopy. The results of this study suggest that the number of meristems visible at the top of the canopy can be predicted using a CNN with low residuals (nRMSE = 0.09). Predicting the actual number of stems from the meristems proved to be less accurate due to the influence of secondary stems that also terminated in a meristem. However, this study established that the predicted number of meristems at the top of the canopy explains a large portion of the variation in the actual number of stems, providing a statistical route for generating 2D density maps of the variation in stem density from UAV, using meristem density as a proxy. Future studies must focus on generating methods for distinguishing a meristem originating from a main stem from those originating from branches and secondary stems. Unlike the physiologically young seed used in this study, temperature-primed physiologically old seed is mostly used in commercial production, with low apical dominance, forming multiple primary main stems at emergence and only branching late in the season after flowering (Knowles and Knowles, 2006). To partially solve the problem of secondary stems, it if therefore suggested that the meristem detection models should be used before significant branching occurs. Future studies must also focus on determining the optimum timing of imagery for minimizing the probability of detecting secondary meristems.

5. Conclusion

This study represents the first attempt to enumerate potato stem number after canopy consolidation using UAV based sensors. The prospect of accurately mapping variation in stem density across a field enables the possibility of using precision agriculture techniques to manipulate potato tuber size distribution through variable harvesting dates and other in-season management practices. This study provides evidence that deep learning and image analysis approaches can be used to accurately enumerate potato meristems and estimate stem density variation in 45 UK potato varieties. Based on the spectral properties of plants, the colour indices developed in this study should also have potential applicability in mapping physiological maturity and leaf discolouration due to biotic or abiotic stress. More studies to test the wider applicability of these indices are therefore recommended. The study has also demonstrated the validity of automated labelling for generating a large dataset of pseudo-labelled ground truth data which can be more rapidly quality-checked and adjusted by a human labeller then used to train deep learning models that learn the features of interest and achieve high IoU with manually labelled test data.

6. Declarations

6.1 Funding

This work was supported by the Agriculture and Horticulture Development Board (Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, United Kingdom) [Project Number 11140054].

6.2 Conflicts of interest/Competing interests

The authors declare that they have no conflict of interest.

6.3 Availability of data and material

The data for this study shall be made available upon request.

6.4 Code availability

The custom code for this study was written in Matlab™ R2020a. The code shall be made available upon request.

CRediT Authorship Contribution Statement

Conceptualization, Methodology, Software, Investigation, Writing – Original Draft, Formal Analysis, Visualization, **Joseph K. Mhango**; Investigation, Funding Acquisition, Writing – Review & Editing, Supervision, **Ivan G. Grove**; Conceptualization, Writing – Review & Editing, Supervision, **William Hartley**; Writing – Review & Editing, Supervision **Edwin Harris**; Writing – Review & Editing, Supervision, Funding Acquisition, Project Administration, **James M. Monaghan**.

7. References

Almekinders, C. J. M., & Struik, P. C. (1996). Shoot development and flowering in potato (Solanum tuberosum L.). In *Potato Research*. https://doi.org/10.1007/BF02358477

Anžlovar, S., Kovač, M., & Ravnikar, M. (1996). Photosynthetic pigments in healthy and virus-infected potato plantlets (Solanum tuberosum L.) grown in vitro. *Phyton - Annales Rei Botanicae*.

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*. https://doi.org/10.1080/02757259509532298

Baumgardner, M. F., Silva, L. R. F., Biehl, L. L., & Stoner, E. R. (1986). Reflectance properties of soils. *Advances in Agronomy*. https://doi.org/10.1016/S0065-2113(08)60672-0

Bleasdale, J. K. A. (1965). Relationships between set characters and yield in maincrop potatoes subject to the Cambridge Core terms of. *J. Agric. Sci.* https://doi.org/10.1017/S0021859600016683

Bussan, A. J., Mitchell, P. D., Copas, M. E., & Drilias, M. J. (2007). Evaluation of the effect of density on potato yield and tuber size distribution. *Crop Science*. https://doi.org/10.2135/cropsci2007.01.0026

Chang, D. C., Sohn, H. B., Cho, J. H., Im, J. S., Jin, Y. I., Do, G. R., Kim, S. J., Cho, H. M., & Lee, Y. B. (2014). Freezing and Frost Damage of Potato Plants: a Case Study on Growth Recovery, Yield Response, and Quality Changes. *Potato Research*, *57*(2), 99–110. https://doi.org/10.1007/s11540-014-9253-5

Chen, Y., Hou, C., Tang, Y., Zhuang, J., Lin, J., He, Y., Guo, Q., Zhong, Z., Lei, H., & Luo, S. (2019). Citrus tree segmentation from UAV images based on monocular machine vision in a natural orchard environment. *Sensors (Switzerland)*. https://doi.org/10.3390/s19245558

Cinat, P., Di Gennaro, S. F., Berton, A., & Matese, A. (2019). Comparison of unsupervised algorithms for Vineyard Canopy segmentation from UAV multispectral images. *Remote Sensing*. https://doi.org/10.3390/rs11091023

Cochrane, M. A. (2000). Using vegetation reflectance variability for species level classification of hyperspectral data. *International Journal of Remote Sensing*. https://doi.org/10.1080/014311600050021303

Duarte-Carvajalino, J. M., Alzate, D. F., Ramirez, A. A., Santa-Sepulveda, J. D., Fajardo-Rojas, A. E., & Soto-Suárez, M. (2018). Evaluating late blight severity in potato crops using unmanned aerial vehicles and machine learning algorithms. *Remote Sensing*. https://doi.org/10.3390/rs10101513

Firman, D. M., Obrien, P. J., & Allen, E. J. (1995). Appearance and growth of individual leaves in the canopies of several potato cultivars. *The Journal of Agricultural Science*. https://doi.org/10.1017/S0021859600084884

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. *Sensors (Switzerland)*. https://doi.org/10.3390/s17092022

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral Properties of Plants. *Applied Optics*. https://doi.org/10.1364/ao.4.000011

Goeser, N. J., Mitchell, P. D., Esker, P. D., Curwen, D., Weis, G., & Bussan, A. J. (2012). Modeling Long-Term Trends in Russet Burbank Potato Growth and Development in Wisconsin. *Agronomy*. https://doi.org/10.3390/agronomy2010014

Gray, D. (1972). Spacing and harvest date experiments with Maris Peer potatoes. *The Journal of Agricultural Science*. https://doi.org/10.1017/S0021859600032263

Guijarro, M., Pajares, G., Riomoros, I., Herrera, P. J., Burgos-Artizzu, X. P., & Ribeiro, A. (2011). Automatic segmentation of relevant textures in agricultural images.

Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2010.09.013

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*. https://doi.org/10.2307/2346830

Huete, A. R. (2004). Remote Sensing for Environmental Monitoring. In *Environmental Monitoring and Characterization*. https://doi.org/10.1016/B978-012064477-3/50013-8

Jin, X., Liu, S., Baret, F., Hemerlé, M., & Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2017.06.007

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithms: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2002.1017616

Kataoka, T., Kaneko, T., Okamoto, H., & Hata, S. (2003). Crop growth estimation system using machine vision. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM*. https://doi.org/10.1109/AIM.2003.1225492

Knowles, N. R., & Knowles, L. O. (2006). Manipulating stem number, tuber set, and yield relationships for northern- and southern-grown potato seed lots. *Crop Science*. https://doi.org/10.2135/cropsci2005.05-0078

Li, B., Xu, X., Han, J., Zhang, L., Bian, C., Jin, L., & Liu, J. (2019). The estimation of crop emergence in potatoes by UAV RGB imagery. *Plant Methods*. https://doi.org/10.1186/s13007-019-0399-7

Love, S. L., & Thompson-Johns, A. (1999). Seed piece spacing influences yield, tuber size distribution, stem and tuber density, and net returns of three processing potato cultivars. *HortScience*. https://doi.org/10.21273/hortsci.34.4.629

Meyer, G. E., & Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2008.03.009

O'Brien, P. J., & Allen, E. J. (1992). Effects of seed crop husbandry, seed source, seed tuber weight and seed rate on the growth of ware potato crops. *The Journal of Agricultural Science*. https://doi.org/10.1017/S0021859600012193

Pesaresi, M., & Benediktsson, J. A. (2001). A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/36.905239

Polder, G., Blok, P. M., de Villiers, H. A. C., van der Wolf, J. M., & Kamp, J. (2019). Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. *Frontiers in Plant Science*. https://doi.org/10.3389/fpls.2019.00209

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.

Rębarz, K., Borówczak, F., Gaj, R., & Frieske, T. (2015). Effects of Cover Type and Harvest Date on Yield, Quality and Cost-Effectiveness of Early Potato Cultivation. *American Journal of Potato Research*. https://doi.org/10.1007/s12230-015-9441-0

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2016.2577031

Mangani, R, Mazarura, U., Mtaita, T. A. & Shayanowako, A. (2015). Growth, yield and quality responses to plant spacing in potato (Solanum tuberosum) varieties. *African Journal of Agricultural Research*, 10(6), 571-578. https://doi.org/10.5897/ajar2014.8665

Sankaran, S., Khot, L. R., & Carter, A. H. (2015). Field-based crop phenotyping: Multispectral aerial imaging for evaluation of winter wheat emergence and spring stand. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2015.09.001

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Sugiura, R., Tsuda, S., Tamiya, S., Itoh, A., Nishiwaki, K., Murakami, N., Shibuya, Y., Hirafuji, M., & Nuske, S. (2016). Field phenotyping system for the assessment of potato late blight resistance using RGB imagery from an unmanned aerial vehicle. *Biosystems Engineering*. https://doi.org/10.1016/j.biosystemseng.2016.04.010

Sun, S., Song, H., He, D., & Long, Y. (2019). An adaptive segmentation method combining MSRCR and mean shift algorithm with K-means correction of green apples in natural environment. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2018.08.011

Vos, J., & Biemond, H. (1992). Effects of nitrogen on the development and growth of the potato plant. 1. leaf appearance, expansion growth, life spans of leaves and stem branching. *Annals of Botany*. https://doi.org/10.1093/oxfordjournals.aob.a088435

Waterer, D. (2007). Vine desiccation characteristics and influence of time and method of top kill on yields and quality of four cultivars of potato (Solanum tuberosum L.).

Canadian Journal of Plant Science. https://doi.org/10.4141/P06-074

Woebbecke, D. M., Meyer, G. E., Von Bargen, K., & Mortensen, D. A. (1995). Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the American Society of Agricultural Engineers*. https://doi.org/10.13031/2013.27838

Wu, J., Yang, G., Yang, H., Zhu, Y., Li, Z., Lei, L., & Zhao, C. (2020). Extracting apple tree crown information from remote imagery using deep learning. *Computers and Electronics in Agriculture*. https://doi.org/10.1016/j.compag.2020.105504

Wurr, D. C. E. (1974). Some effects of seed size and spacing on the yield and grading of two maincrop potato varieties: I. Final yield and its relationship to plant population. *The Journal of Agricultural Science*. https://doi.org/10.1017/S0021859600050206

Wurr, D. C. E., & Morris, G. E. L. (1979). Relationships between the number of stems produced by a potato seed tuber and its weight. *The Journal of Agricultural Science*. https://doi.org/10.1017/S0021859600038089

Yang, X., Shen, X., Long, J., & Chen, H. (2012). An Improved Median-based Otsu Image Thresholding Algorithm. *AASRI Procedia*. https://doi.org/10.1016/j.aasri.2012.11.074

Zhao, B., Ding, Y., Cai, X., Xie, J., Liao, Q., & Zhang, J. (2017). Seedlings number identification of rape planter based on low altitude unmanned aerial vehicles remote sensing technology. *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering*. https://doi.org/10.11975/j.issn.1002-6819.2017.19.015

Zhao, Z. Q., Zheng, P., Xu, S. T., & Wu, X. (2019). Object Detection with Deep Learning: A Review. In *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2018.2876865

Figure Titles

Fig.1 Flow chart of the image analysis algorithm for generating meristem objects and infering stem number

Fig.2 Flow chart of the Faster R-CNN algorithm for training a potato meristem object detector

Fig.3 Aerial image of the testing site for the image analysis and convolutional neural network algorithms

Fig.4 The difference between Green and Blue colour plotted against the difference between Red and Blue colour in pixels selected from four prevalent features in a potato canopy

Fig.5 Index values of Meristems, Leaves, Dry soil and Wet soil using eight colour indices, from images taken before canopy consolidation of partially irrigated Sandy Loam soil. a - Blue Difference Normalized Index, b - Cumulative Blue Difference Index, c - Excess Green Index, d - Colour Index of Vegetation Extraction, e - Excess Green minus Excess Red Index, f - Normalized Difference Green Redness Index, g - Excess Red, h - Excess Blue Index

Fig.6 Observed vs Predicted of the number of meristems in potato canopies when predictions were made using the traditional image analysis approach

Fig.7 Observed vs Predicted number of meristems when predictions were made using a Convolutional Neural Network-based object detector

Fig.8 Observed number of stems vs Predicted number of meristems when predictions were made using a Convolutional Neural Network

Fig.9 Observed number of stems vs Predicted number of meristems when predictions were made using the traditional image analysis approach

Appendices

Appendix A A list of the varieties used to test the object detection models

Purpose	**Breeder**	**Variety**
Chipping	Agrico	Agria
Crisping	HZPC	Alcander
Chipping	HZPC	Althea
Chipping	HZPC	Alverstone Russet
Crisping	Agrico	Arsenal
Chipping	HZPC	Asterix
Unknown	Unknown	Babylon
Crisping	PepsiCo	Brooke
Chipping	HZPC	Challenger
Crisping	Agrico	Corsica
Prepack	Agrico	Desiree

Prepack	Agrico	Estima
Crisping	HZPC	Heraclea
Chipping	HZPC	Innovator
Chipping	HZPC	Ivory Russet
Prepack	Greenvale	Jelly
Prepack	Unknown	King Edward
Crisping	Meijer	Lady Clair
Prepack	Branston	Lanorma
Prepack	Branston	Laura
Salad	Agrico	Maris Peer
Chipping	Agrico	Maris Piper
Crisping	Agrico	Markies
Prepack	Meijer	Melody
Prepack	HZPC	Mozart
Prepack	IPM	Nectar
Prepack	HZPC	Panther
Chipping	SCRI	Pentland Dell
Chipping	Agrico	Performer
Chipping	Norika	Pirol
Chipping	Higgind Group	Ramos
Chipping	IPM	Rooster
Chipping	McCains	Royal
Chipping	Unknown	Russett Burbank
Chipping	HZPC	Sagitta
Crisping	Stet	SHC1010
Crisping	PepsiCo	Shelford
Unknown	Unknown	Sorentina

Prepack	HZPC	Sunita
Crisping	HZPC	Taurus
Unknown	Unknown	Thalassa
Crisping	Unknown	Titan
Crisping	HZPC	Triple 7
Unknown	Unknown	VDW 07-197
Crisping	Stet	VR808