



High-resolution mapping of Blueberry scorch virus incidence using RGB and multispectral UAV images and deep learning[☆]



Ali Jamali ^{a,*}, Bing Lu ^{a,**}, Rishi R. Burlakoti ^b, Siva Sabaratnam ^c, Margaret Schmidt ^a, Carolyn Teasdale ^c, Eric M. Gerbrandt ^d, Lilian Yang ^a, Jonathon McIntyre ^e, David McCaffrey ^f

^a Department of Geography, Simon Fraser University, 8888 University Dr W, Burnaby, BC, V5A 1S6, Canada

^b Science and Technology Branch, Agassiz Research and Development Centre, Agriculture and Agri-Food Canada, Agassiz, BC, V0M 1A0, Canada

^c British Columbia Ministry of Agriculture and Food, Abbotsford Agriculture Centre, Abbotsford, BC, V3G 2M3, Canada

^d British Columbia Blueberry Council, #275-32160 South Fraser Way, Abbotsford, BC, V2T 1W5, Canada

^e i-Open Group, 206-3670 Townline Rd, Abbotsford, BC, V2T 0H2, Canada

^f Miraterra Inc., 199 W 6th Ave, Vancouver, BC, V5Y 1K3, Canada

ARTICLE INFO

Keywords:

Blueberry
Scorch virus
UAV mapping
Deep learning
Vision transformer
CNN

ABSTRACT

Blueberry scorch caused by *Blueberry scorch virus* (BlScV) is a destructive disease, which can result in substantial yield decline and pose a significant threat to the viability of well-established highbush blueberry fields in North America and other regions. Early detection of the disease in the field, removal of infected bushes, and control of its spread via aphids to other fields or regions are critical for managing this disease. Visual assessment of Blueberry scorch symptoms is the predominant method for identifying and estimating the disease, which, however, is labour-intensive, tedious, and inefficient. Unmanned Aerial Vehicle (UAV)-based imaging is a powerful remote sensing tool for crop monitoring with several advantages, such as flexibility to acquire images of different pixel sizes, short revisit time intervals, reduced susceptibility to cloud interference, and flexibility to equip with different sensors. This study aims to collect UAV images to detect and map BlScV-infected blueberry plants using a cutting-edge deep learning model. Images of different pixel sizes acquired by an RGB sensor, and a multispectral sensor were compared to evaluate their detection accuracies. To ensure comprehensive information dependency extraction at close-, mid-, and long-ranges, the deep learning techniques developed in this study incorporate various computer vision-based mechanisms, such as Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNNs), and Self-Attention (SA) modules. Through these innovations, the deep learning algorithm, called InceptionLSA, obtained the highest average accuracy of 76.33% and 70.00% at a 20 cm pixel size of the multispectral and RGB images, respectively.

* I agree with the above statements and declare that this submission follows the policies of Solid State Ionics as outlined in the Guide for Authors and in the Ethical Statement.

^{*} Corresponding author.

^{**} Corresponding author.

E-mail addresses: alij@sfu.ca (A. Jamali), b_lu@sfu.ca (B. Lu).

1. Introduction

In North America, highbush blueberry (*Vaccinium corymbosum* L.) is a high-value fruit crop, known for its nutritional benefits and economic importance. However, blueberry production can be impacted by numerous bacterial, fungal, and viral diseases. Among these, Blueberry scorch caused by *Blueberry scorch virus* (BlScV) is one of the most concerning diseases, particularly in British Columbia (BC), Canada, and other states in the Pacific Northwest (PNW) of the USA (Martin and Tzanetakis, 2018; Wegener et al., 2006; Xie et al., 2018). The BlScV was first found in BC in the year 2000, and now it has spread to all blueberry growing regions in BC (over 11500 ha) (Sabaratnam, 2021). Management of BlScV is very challenging, as several commonly grown blueberry cultivars in the PNW regions including BC are moderately to severely susceptible to BlScV, and the disease control methods (i.e., roguing of infected bushes and application of aphicides to control the vector) are expensive and time-consuming (James et al., 2017; Sweeney et al., 2009). Moreover, diagnosis of BlScV infection in individual plants relies on visual assessment in the field for symptoms (e.g., blighting of blossoms and leaves), and laboratory detection of the virus using immuno-based assays and molecular techniques (Lee et al., 2023; Wegener et al., 2006). Visual assessment of Blueberry scorch symptoms on individual blueberry plants on large commercial farms is labour-intensive, tedious, and inefficient, as well as requiring specific expertise in identifying the disease symptoms. Moreover, such field-based assessment is not practical for understanding the spatial distribution pattern and prevalence of the disease in large fields (e.g., calculating the percentage of disease in one field and identifying which section of a field has higher disease severity). It is also challenging to monitor the temporal variation in the prevalence of BlScV to determine the rate of virus spread within the field. Remote sensing could possibly address these challenges in visual BlScV monitoring, as this tool can precisely assess spatial patterns and temporal variation of targeted ground features. The increasing accessibility of high-resolution remote sensing technologies has led to a rise in their use for large-scale agricultural disease assessment in recent decades (Chen et al., 2022; Lee et al., 2020; Mothapo et al., 2022). However, to the best of our knowledge, there has been no research on the use of remote sensing for BlScV detection and mapping in blueberry fields.

Unmanned Aerial Vehicles (UAVs) have attracted significant attention in the remote sensing field for imaging due to their various benefits, including high image spatial resolution, data acquisition on demand, and the flexibility to attach various sensors as well as the possibility to be flown at different altitudes. UAVs have been frequently used in agriculture for supporting farming practices, including assessment of crop health and detection of insects and diseases (Bouguettaya et al., 2023; de Castro et al., 2015). Different sensors, such as an RGB camera (e.g., DJI P1) or a multispectral sensor (e.g., MicaSense Altum), can be mounted on a UAV to acquire images with different spectral resolutions, which can influence the accuracy of crop monitoring. For instance, Zheng et al. (2018) found that the multispectral camera performed best among three onboard UAV sensors (RGB, color-infrared (CIR), and multispectral cameras) used for estimating nitrogen concentrations in rice plants at different growth stages (Zheng et al., 2018). Therefore, it is critical to evaluate the impacts of spectral resolution of different UAV sensors on crop monitoring. In addition, UAVs enjoy the flexibility of flying at various altitudes and thus acquiring images at different pixel sizes (i.e., spatial resolution) which is a critical factor in image classification and detection of agricultural features of interest (Barreto et al., 2020; Zhao et al., 2023). For instance, to assess the importance of pixel size for crop disease mapping, Zhao et al. (2023) proposed to classify healthy and root rot-infested cotton fields on images taken by multiple platforms and with different pixel sizes, including a UAV (3 cm pixel size), a manned aerial vehicle (MAV) (20 cm pixel size), Sentinel-2 satellite (10 m pixel size), together with several other resampled images with different image pixel sizes (e.g., 0.05, 0.1, 0.2, 0.5, 1, 2, 4, 5 and 10m) (Zhao et al., 2023). They compared these images with various pixel sizes and found that 4 m was the ideal pixel size for identifying fields with cotton root rot. Overall, UAVs are effective tools for detecting crop disease and investigating crop health status. Therefore, it is valuable to explore the potential of UAVs for detecting BlScV and mapping its spatial distribution in blueberry fields.

Another vital factor for successful disease recognition on remote sensing images is the selection of an appropriate image classification algorithm. Deep learning algorithms have proven to be superior to the traditional classifiers (e.g., random forest) in the computer vision field for various applications, such as image classification (Sharma and Guleria, 2022), object detection (Kaur and Singh, 2023), and image segmentation (Jamali et al., 2023). The main reason for the superiority of deep learning algorithms can be attributed to their distinct ability to recognize spectral and spatial patterns in images (Wu et al., 2019; Shafi et al., 2020). Other strengths of deep learning algorithms lie in their automatic feature engineering, high generalization capability, and ability to extract data dependencies. Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown their excellent capability in crop health assessment, disease mapping, and crop growth stage monitoring (Bouguettaya et al., 2023; X. Zhang et al., 2019). However, recent research suggests that transformer-based models, such as the Vision Transformer (ViT) (Alexander et al., 2021), may perform better than CNNs for image classifications (Jamali et al., 2022). It has been shown that ViTs have better capabilities over CNNs in capturing long-range data dependencies from images which increases the capabilities of such algorithms for various applications, such as image classification (Yan et al., 2023; A. Jamali et al., 2022). The better image classification capability in vision transformers can be further explained by their use of self-attention mechanisms that can effectively mix disparate spatial information, resulting in an infinitely broad receptive field, as opposed to CNNs with a limited receptive field. Computer vision has been growing reliant on attention mechanisms, particularly self-attention (Jamali et al., 2023; Alexander et al., 2021; Hong et al., 2021). Nevertheless, self-attention overlooks possible correlations between various samples and has quadratic complexity. Moreover, considering the high dimensionality and complexity of remote sensing data, the current main concern in using a self-attention mechanism in the remote sensing field is their high computational cost and their need for a significant number of ground truth data (labeled data). To fully utilize the capability of vision transformers in image recognition and classification in the remote sensing field, one solution would be the use of more data-efficient architectures utilizing self-attention modules with linear computational cost over the well-established architectures (e.g., ViT) with a quadratic computational cost. Such a computationally efficient technique will be utilized and discussed for

BlScV mapping in this study.

In summary, the main objective of this study is to utilize and assess UAV images with different spatial (ranging from 2 cm to 50 cm) and spectral (RGB vs. multispectral) resolutions, along with advanced deep learning mechanisms (e.g., computationally efficient linear vision transformers) for BlScV mapping. In addition, a second objective is to evaluate the classification capability of the developed model in recognizing various scales of disease severity (e.g., low to severe). These analyses will help us to better understand the capability of UAV-based remote sensing and deep learning models for detecting different scales of BlScV infection. The third objective is to obtain the most accurate BlScV classification results (e.g., disease maps) by testing different modeling techniques, including adjusting training data ratios (i.e., the amount of ground data for model training) and optimizing input image patch sizes (ranging from 5 to 13). Overall, the contributions of this study can be described as follows.

1. This research represents a pioneering effort using remote sensing for BlScV mapping to support the blueberry industry in controlling this virus, which is critical for many growing countries and regions.
2. One RGB sensor (DJI Xennmuse P1) and one multispectral sensor (MicaSense Altum) are compared for BlScV mapping, and their classification results are evaluated. Spectral resolutions needed for the BlScV mapping are thus discussed.
3. Images of different pixel sizes (e.g., 2–50 cm) are evaluated for classifying BlScV and investigating the impact of pixel size on classification accuracy, aiming to identify an optimal image pixel size for BlScV mapping.
4. We introduce an efficient attention-based deep learning model, called Inception Linear Self-Attention (InceptionLSA), characterized by linear computational complexity. Leveraging CNN, attention, and Multi-Layer Perceptron (MLP) modules, our architecture facilitates multi-scale feature extraction, ensuring effective capture of close-, mid-, and long-range information for precise BlScV mapping.

2. Study area and data collection

Two blueberry fields (Field A and Field B) in Pitt Meadows, BC were selected in this study for the mapping of BlScV incidence (Fig. 1). Pitt Meadows is in the Lower Fraser Valley approximately 30 km east of Vancouver, and the area is one of the major blueberry production regions in BC.

2.1. Field data collection

In June 2022, field surveys were conducted to visually assess the presence and severity of blueberry scorch symptoms. Blueberry

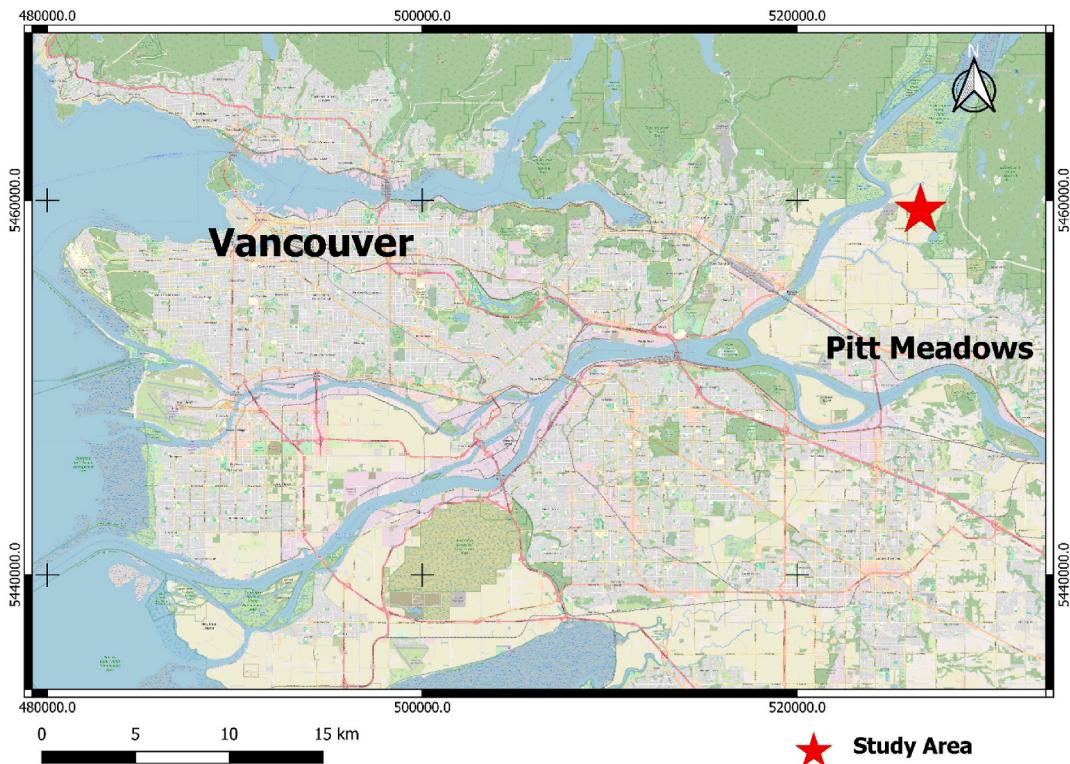


Fig. 1. Study area (indicated by star) located in the City of Pitt Meadows (approximately 30 km east of Vancouver), BC, Canada.

scorch severity was assessed using a rating scale of 0–5, where 0 indicates healthy plants and 5 indicates plants with the most severe symptoms (e.g., blighted leaves and flowers) affecting most of the canopy, as illustrated in Fig. 2. Plants showing blueberry scorch symptoms were identified and marked using flagging tape and precisely geo-located with centimeter-level accuracy via an Emlid Reach RS2 GPS unit (Emlid Tech Kft, Budapest, Hungary).

Healthy as well as symptomatic leaf samples were collected for laboratory testing to determine the presence of the virus, performed by Phyto Diagnostics Company Ltd. (North Saanich, BC). Healthy plants were also included in the survey to develop classification models for distinguishing healthy and diseased blueberry plants. Overall, 54 symptomatic and 77 healthy blueberry plants were surveyed in Field A, and 55 symptomatic and 75 healthy plants were surveyed in Field B. The surveyed plant data in both fields were combined to augment the reference dataset (i.e., a total of 109 symptomatic and 152 healthy samples in two fields) for training and validation of the classification model, which will be described in more detail in the modeling section. In Field A, blueberry plants exhibited disease severity ratings ranging from 0 to 5, comprising 49 samples on a scale of 1–3 and 5 samples on a scale of 4–5 (Fig. 3). Blueberry plants in Field B also had disease severity ratings from 0 to 5, with 50 samples on a scale of 1–3 and 5 samples on a scale of 4–5 (Fig. 4). Employing a disjoint data sampling technique, the reference data were divided into training and testing datasets (Audebert et al., 2019). To mitigate the potential reduction in accuracy and reliability resulting from spatial autocorrelation, we extracted the training and testing data from separate polygons (i.e., a 50 cm buffer was created around each blueberry plant).

2.2. UAV data collection and image processing

The UAV flights were conducted simultaneously on clear and sunny days between 10 a.m. and 2 p.m. in June 2022. A Matrice 300 RTK drone equipped with a DJI P1 RGB sensor (SZ DJI Technology Co Ltd., China) or a MicaSense Altum Multispectral sensor (AgEagle Aerial Systems Inc., USA), respectively, was utilized to collect high-resolution images in the study area (Fig. 4 and Table 1). The RGB and multispectral images collected were mosaiced and geo-corrected in Pix4D Fields software. A digital surface model (DSM) was also derived and used to exclude soil and grass backgrounds that are significantly lower in height than the blueberry plants. This allowed the classification model to focus solely on the blueberry plants. The processed RGB and multispectral images were resampled to image series with different pixel sizes, including, 2, 5, 10, 20, 30, and 50 cm, respectively, aiming to evaluate the impacts of pixel size on the classification accuracy of BiScV.



Fig. 2. Blueberry plants showing symptoms with a severity rating scale of 0–5. a) healthy (severity rating 0), b) severity rating 1, c) severity rating 2, d) severity rating 3, e) severity rating 4, and f) severity rating 5.

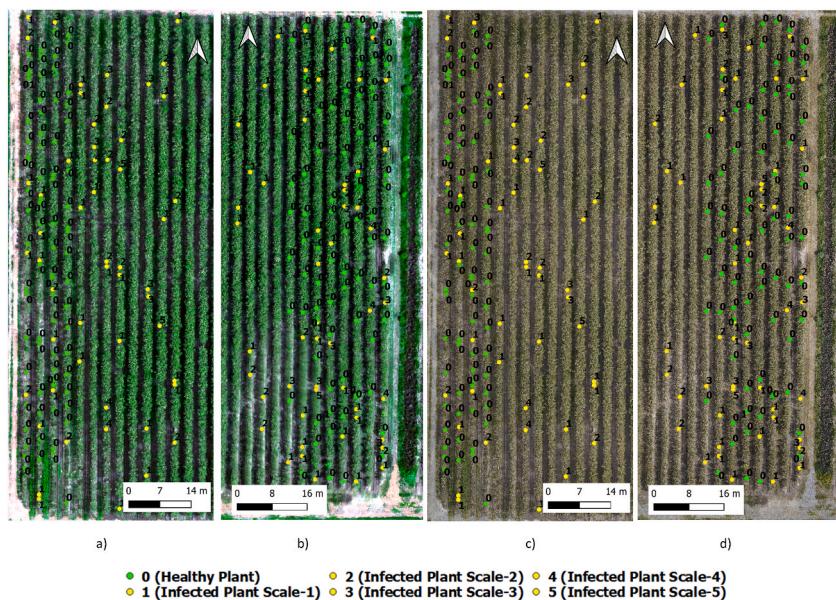


Fig. 3. Images of Fields A and B collected using Altum multispectral sensor (subfigure a and b, respectively) and P1 RGB sensor (subfigure c and d, respectively). Yellow and green dots indicate BlScV-infected and healthy blueberry plants, respectively, based on visual assessment and laboratory testing. Severity of disease symptoms are indicated by the scale 0–5 (i.e., 0 indicates a healthy plant without symptoms, 1 indicates lowest severity, and 5 indicates the highest severity). The spatial distribution of infected blueberry plants appeared to be random. Healthy plant samples were concentrated on one side of the field to allow for more efficient surveys, as their characteristics were consistent across the field. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

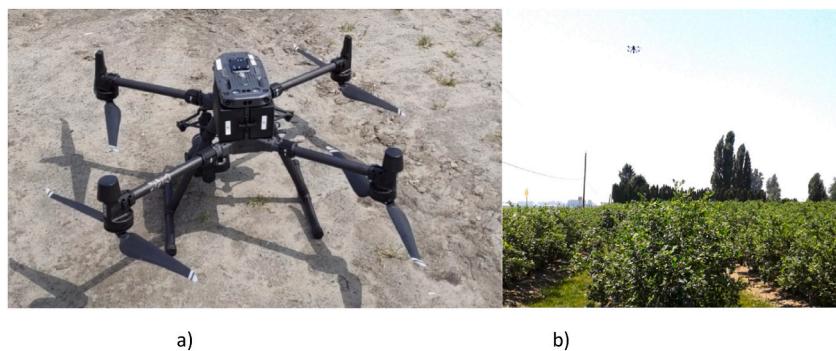


Fig. 4. UAV used for collecting images of blueberry fields, a) Matrice 300 RTK drone, and b) UAV imagery collection over blueberry plants.

Table 1

Specification of utilized RGB and multispectral UAV sensors in this research (Kersten et al., 2022; “Micasense Altum-PT” 2024).

| Sensor | ZENMUSE P1 | Altum-PT |
|-------------------|---|---|
| Producer | SZ DJI Technology Co Ltd., China | AgEagle Aerial Systems Inc., USA |
| Spectral Bands | Blue, Green, Red | Blue, Green, Red, Red Edge, Near-IR |
| Sensor Resolution | 8192 × 5460 (45 MP) 1.1 cm resolution achieved at 90 m altitude | 2064 × 1544 (3.2 MP per multispectral band) 1.8 cm resolution achieved at 90 m altitude |
| Dimensions | 198 × 166 × 129 mm | 110 × 80 × 69 mm |
| Weight | 800 g | 577 g |
| Power | 20W | 7.0W (average) |
| IP Rating | IP4X | IP4X |

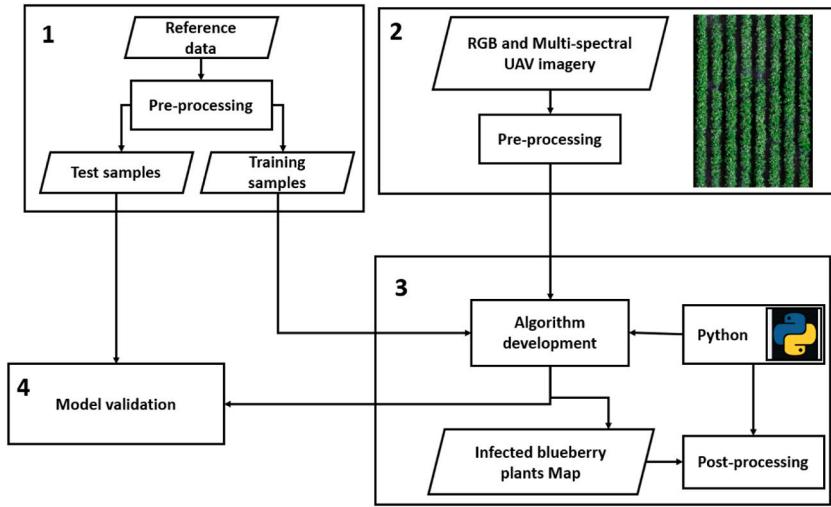


Fig. 5. The proposed workflow for precise BlScV mapping.

3. Classification methods and settings

Fig. 5 represents the proposed workflow diagram outlining the process for detecting infected blueberry plants using RGB and multi-spectral UAV imagery. Reference data is pre-processed and split into test and training data sets using a disjoint data sampling technique in Step 1. To address the low number of training data, we used a data augmentation technique in reference data pre-processing. In our methodology, an image size of 9x9 pixels is defined for the preprocessing pipeline, where *TensorFlow Keras* layers are used to perform various data augmentation techniques. The data augmentation model applies several transformations: it normalizes the data using *layers.Normalization()*, resizes the images to 9x9 pixels using *layers.Resizing()*, and introduces randomness with horizontal flipping, slight rotations (*RandomRotation* with a factor of 0.02), and zooming in both height and width by 20% (*RandomZoom*). The normalization layer is adapted to the training data by computing its mean and variance, ensuring that the input data is properly scaled during training for improved model convergence. This augmentation process helps increase the diversity of the training data, making the model more robust to variations and improving its generalization capabilities. In Step 2, RGB and multispectral UAV imagery undergoes pre-processing, including data normalization. Step 3 involves deep learning algorithm development using *Python* programming language, where the processed imagery is analyzed to identify infected blueberry plants. The output is BlScV detection maps, which undergo post-processing (i.e., removal of surrounding grass and mud). To eliminate the mud and grass areas surrounding the blueberry plants, a Digital Surface Model (DSM) was created using point cloud data in *QGIS 3.16.7* and the *LAStools* library. Since the blueberry plants in the study area were significantly taller than surrounding features like grass, this height difference allowed for the removal of irrelevant data in the final classification map. Finally, in Step 4, the developed model is validated against the test samples and several other deep learning models, including Efficient Net (Tan and Quoc, 2019), EdgeNeXt (Maaz et al., 2023), and DaViT (Ding et al., 2022).

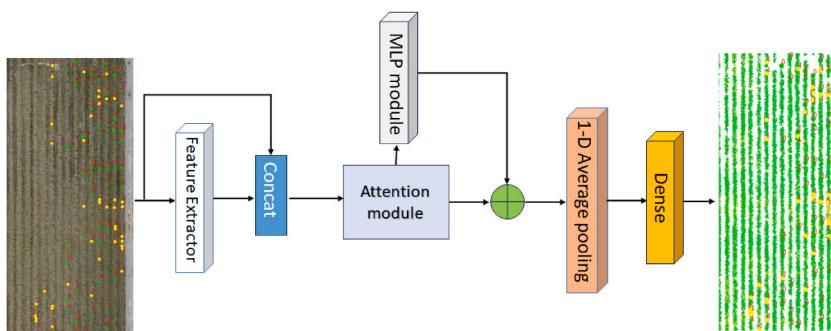


Fig. 6. The structure of the model developed for mapping blueberry scorch, where element-wise sum is illustrated by \oplus and layer concatenation by *Concat*. The image on the left indicates inputting image dataset with dots representing surveyed samples, and the image on the right indicates expected classification results with green representing healthy and yellow representing infected plants. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.1. The classification architecture

A deep learning algorithm was developed for mapping blueberry scorch on RGB and multispectral UAV images of different pixel sizes, respectively. The developed deep learning architecture uses the excellent feature extraction capability of CNN models in an inception-like module (i.e., feature extractor). Moreover, the developed model has the ability of vision transformers in global feature dependency extraction in an efficient external linear attention module (Fig. 6).

3.1.1. Attention module

Cutting-edge transformers can replace computationally extensive self-attention in popular architectures by utilizing external, small, and learnable shared memories (Guo et al., 2023). The complexity of such attention modules is linear (i.e., $O(n)$), and for better global dependency feature extraction, they implicitly consider the correlations among all the data samples. As seen in Fig. 7, the proposed attention module produces the attention map by calculating the similarities among the self-query vectors and an external learnable key memory. The resulting attention map is then multiplied by another external learnable value memory to create an improved feature map. Since the two memories are built using linear layers in execution, back-propagation can be utilized to optimize the system from the beginning to the end. The memories are shared throughout the whole data set and unaffected by specific samples, which strengthens the attention mechanism's regularization effect and increases its capacity for generalization.

Considering an input image $X \in R^{N \times d}$ (N and d present the number of image pixels and number of bands/channels), the self-attention mechanism linearly projects the input image into a value matrix $V \in R^{N \times d}$, a query matrix $Q \in R^{N \times d}$, and a key matrix $K \in R^{N \times d}$. The self-attention map can be defined as:

$$A = (a)_{ij} = \text{softmax}(QK^T)$$

$$O_{OUT} = AV$$

Where the attention matrix is presented by $A \in R^{N \times N}$ and the pair-wise similarity between the i_{th} and j_{th} elements are presented by $(a)_{ij}$. The output feature map is the enhanced feature representation of the input image, and the attention map is produced by calculating pixel-by-pixel similarity in the feature space with a high computational complexity of $O(dN^2)$. In a more simplified approach, from the input image of M , the attention map can be calculated as:

$$A = \text{softmax}(MM^T)$$

$$O_{OUT} = AM$$

However, we can calculate the attention input between the image pixels and an external memory unit $U \in R^{S \times d}$ as defined by:

$$A = (a)_{ij} = N(MU^T)$$

$$O_{OUT} = AU + M$$

where U serves as a learnable, input-independent parameter that retains the memory of the entire training dataset and N is a double-normalization function. To enhance the ability of the model, two different memory units of U_k and U_v can be used. Thus, the attention map can be defined as:

$$A = N(MU_k^T)$$

$$O_{OUT} = AU_v + M$$

The U_k and U_v are randomly initialized and learned through the procedure of model training. The double-normalization function as proposed by Guo et al. (2021) is calculated as:

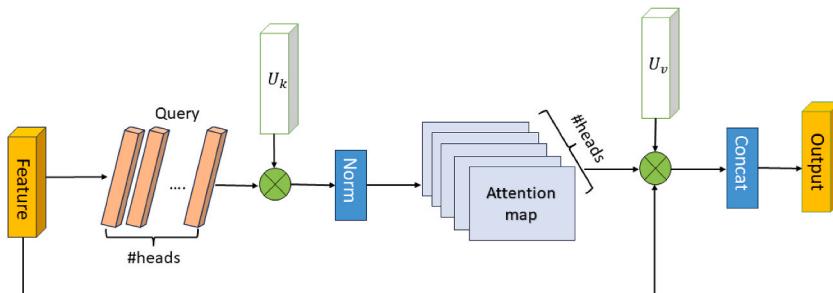


Fig. 7. The architecture of multi-head external attention module where matrix multiplication is illustrated by \otimes .

$$(\alpha')_{ij} = \mathbf{M}\mathbf{U}^T$$

$$\alpha''_{ij} = \exp(\alpha'_{ij}) / \sum_k \alpha'_{k,j}$$

$$\alpha_{ij} = \alpha''_{ij} / \sum_k \alpha''_{i,k}$$

Considering the H number of heads where h_i is the i_{th} head, the multi-head attention can be calculated as:

$$h_i = A(M_i, U_k, U_v)$$

$$O_{OUT} = Multihead(M, U_k, U_v) = concat(h_1, \dots, h_H)L_o$$

Where L_o is a linear transformation matrix used to fulfill the consistency between input and output feature map dimensions.

3.1.2. CNN and MLP modules

As depicted in Fig. 8a, to extract multi-scale feature representation, the CNN-based module consists of three Conv 2D layers with filter numbers of 128, 256, and 64 with kernel sizes of 1×1 , 3×3 , and 5×5 , respectively. A max-pooling layer (Mp) with a kernel size of 3×3 with a stride of size 1 is utilized as well. Considering the image input from the RGB and multispectral UAV sensors, the output feature map of the CNN module (X') can be defined as:

$$X' = \sum_{k=1,3,5} Conv2D_{k \times k}(X) + Mp(X)$$

The resulting feature map from the CNN module is then concatenated with the original feature map, as defined by:

$$X'' = X' + X,$$

The resulting output map (X'') is then fed to the attention module, and the obtained output map of the attention module (X''') is sent to the MLP module consisting of the element-wise sum of three dense layers of sizes 16, 32, and 16, respectively, as shown in Fig. 8b and defined by:

$$X_{MLP} = \prod_{k=16,32,16} D_k(X''')$$

3.1.3. Developed deep learning model (InceptionLSA)

Our deep learning model (InceptionLSA) incorporates an MLP module for linear feature extraction, employing dense layers of sizes 16 and 32. Additionally, a CNN module was specifically designed for multi-scale feature extraction, integrating point-wise convolutional layers of size 1×1 and depth-wise convolutions of sizes 3×3 and 5×5 . This design ensures that the model can effectively capture local data interactions and relationships critical for distinguishing between infected and healthy blueberry plants. To efficiently extract long-range data dependencies, we implemented a linear attention module, replacing computationally intensive self-attention in conventional architectures. This attention module utilizes two external, small, learnable shared memories with two cascaded linear layers and two normalization layers. The linear $O(n)$ complexity of the attention module implicitly considers

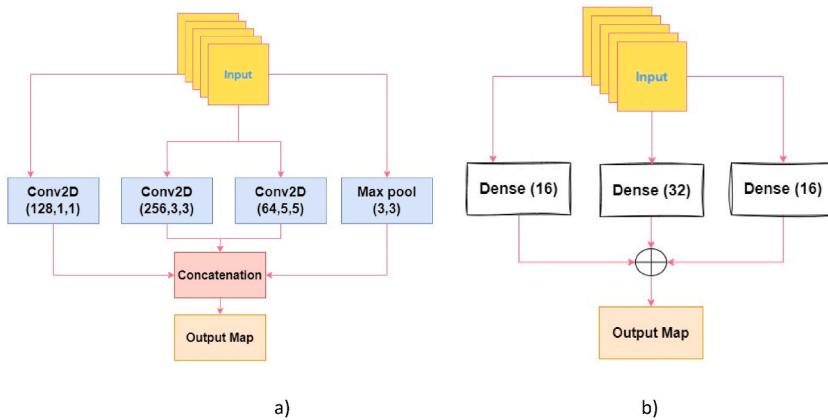


Fig. 8. The architecture of a) the feature extractor module and b) the MLP module designed in the developed algorithm where element-wise sum is illustrated by \oplus and layer concatenation by *Concatenation*.

correlations among all data samples, facilitating superior global dependency feature extraction. Considering the image input of $X^{9 \times 9 \times 8}$ for the multispectral Altum UAV sensor, the input image is sent to the CNN module, resulting in a feature map of $X^{9 \times 9 \times 456}$. Then the original input image is concatenated with the resulting feature map of the CNN module and fed to the attention module, resulting in an output map of $X^{16 \times 64}$. The output map of the external attention map is sent to the MLP module, resulting in an output map of $X^{16 \times 64}$. Afterwards, the resultant output map of the attention and MLP modules is added and sent to a global 1-D average pooling layer. The final layer is a dense layer of size 2 with a *SoftMax* activation function, as depicted in Fig. 6.

3.2. Evaluation metrics and models

Blueberry scorch disease identification data generated by Scorch Mapper were evaluated and compared to several other CNN- and ViT-based algorithms, including Efficient Net (Tan and Le 2019), EdgeNeXt (Maaz et al., 2023), and DaViT (Ding et al., 2022). Tan and Le (2019) introduced an innovative scaling method that systematically scaled the three dimensions of depth, width, and resolution using a simple yet highly effective compound coefficient. Employing neural architecture search, they first built a new baseline network. They afterwards scaled it up to produce a family of algorithms called Efficient Nets, which performed more accurately and efficiently than previous Conv Nets. Building resource-efficient multipurpose networks is highly desirable because of their many application domains. Maaz et al. (2023) proposed EdgeNeXt to successfully integrate the advantages of the CNN and Transformer models and presented a novel, efficient hybrid architecture. In particular, they introduced the split depth-wise transpose attention (STDA) encoder in EdgeNeXt, which implicitly increases the receptive field and encodes multi-scale features by splitting input tensors into various channel groups and using self-attention across channel dimensions in conjunction with depth-wise convolution. Dual Attention Vision Transformers (DaViT), an intuitive but powerful vision transformer architecture that preserves computational efficiency while capturing global context, was proposed by Ding et al. (2022). To take advantage of self-attention mechanisms, the model was developed using both "spatial tokens" and "channel tokens." The classification results of the developed deep learning models were evaluated in terms of average and overall accuracy, F1-score, and kappa index (Equations (1)–(4)).

$$\text{Overall Accuracy} = \frac{(TP + TN)}{\text{Total number of pixels}} \times 100 \quad (1)$$

$$\text{Average Accuracy} = \frac{\sum_{i=1}^n \text{Recall}_i}{n}, \text{Recall} = \frac{\text{True positive}}{(\text{True positive} + \text{False negative})} \quad (2)$$

$$\text{Kappa} = \frac{p_0 - p_e}{1 - p_e}, p_0 = \frac{\sum x_{ii}}{N}, p_e = \frac{\sum x_{ii} x_{ii}}{N^2} \quad (3)$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

3.3. Computing setting

We utilized a hardware setup featuring an Intel i7-10750H CPU with a base clock speed of 2.60 GHz, paired with an NVIDIA GeForce RTX 2070 GPU for accelerated computation, and 16 GB of RAM. This configuration ran on a 64-bit version of Windows 11, providing a stable and efficient environment for our computational needs. Our deep learning algorithms were implemented using Python TensorFlow, leveraging the framework's versatility for building and training models. We employed the Adam optimizer, known for its adaptive learning rate and computational efficiency, to optimize the parameters of our neural networks. To ensure thorough training and prevent underfitting or overfitting, we selected a maximum of 40 epochs as the stopping criteria, allowing sufficient time for the model to converge. Additionally, we set a batch size of 256 during training, balancing computational load and gradient updates for improved performance.

Table 2

Detection results of infected blueberry plants using *Altum* multispectral image with different pixel sizes in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Pixel size (cm) | 2 | 5 | 10 | 20 | 30 | 50 |
|----------------------------|-------|-------|-------|--------------|-------|-------|
| F1-score (Infected Plants) | 0.59 | 0.55 | 0.58 | 0.71 | 0.57 | 0.56 |
| F1-score (Healthy Plants) | 0.74 | 0.77 | 0.8 | 0.81 | 0.76 | 0.74 |
| AA ($\times 100$) | 66.46 | 65.48 | 68.03 | 76.33 | 66.49 | 65.25 |
| OA ($\times 100$) | 68.03 | 69.51 | 72.49 | 76.66 | 69.12 | 67.52 |
| KI | 0.33 | 0.32 | 0.38 | 0.52 | 0.33 | 0.31 |

4. Results

This section presents statistics of classification results from the developed deep learning model with RGB and multispectral images of different pixel sizes. Classification maps showing distributions of diseased plants were also generated.

4.1. Statistics of classification results

BlScV classification results of the deep learning model developed using Altum multispectral images with different pixel sizes are illustrated in [Table 2](#). Employing image pixel sizes ranging from 2 to 50 cm which provide plant details from high to low, we aimed to identify the optimal image pixel size for precise and efficient UAV-based BlScV mapping. The developed deep learning model attains its peak classification accuracy with a 20 cm pixel size, yields a kappa index (KI) of 0.52, has an average accuracy (AA) of 76.33%, and an overall accuracy (OA) of 76.66% ([Table 2](#)). Furthermore, the highest F1-score is observed for the 20 cm pixel size, with values of 0.71 and 0.81 for recognizing infected and healthy blueberry plants, respectively. Delving into specifics, the developed algorithm experiences a notable enhancement in AA, increasing by approximately 10 percentage points from image pixel sizes of 2–20 cm. However, there is a significant decline of about 11 percentage points from 20 to 50 cm pixel sizes. Notably, the classifier achieves its lowest AA of 65.25% at the pixel size of 50 cm ([Table 2](#)). Considering an input image patch size of 9 by 9 pixels, the lower classification accuracies observed for smaller image pixel sizes (2–10 cm) can be attributed to the limited object sizes, ranging from 18 × 18 cm to 90 × 90 cm. This limited object size hindered the extraction of necessary characteristics of blueberry plants. It further indicates that very small pixels providing excessive detail may hide the recognition of general patterns or characteristics needed for accurate BlScV detection. Pixels must generally be smaller than the plant, and thus 20 cm or less pixel sizes are needed to provide plant-scale details. Conversely, lower classification accuracies generated with larger pixel sizes (30–50 cm) were linked to challenges in isolating characteristics of infected plants from surrounding features, such as soil and grass between blueberry plants. Additionally, larger pixel sizes led to a significant loss of feature details, impeding the extraction of detailed attributes from blueberry plants. In other words, a significant percentage of pixels were mixed because they contained both background (e.g., soil and grass) and plant information, particularly around the edges of plant canopies. Influences of mixed pixels became evident when employing image pixel sizes larger than 30 cm, as seen in [Tables 2–3](#). This issue worsens when the image's pixel size becomes coarser (e.g., 50 cm) and approaches or exceeds the size of the plants.

Similar to the results of using multispectral imagery ([Table 2](#)), the developed classifier's AA of using RGB images witnessed a substantial increase of approximately 8 percentage points when transitioning from image pixel sizes of 2 (61.80%) to 20 cm (70.00%) ([Table 3](#)). Then, there was a significant decrease of about 19 percentage points in average accuracy when shifting from 20 (70.00%) to 50 cm (50.67%) of image pixel sizes. The architecture's lowest average accuracy, at 50.67%, was recorded when using an image pixel size of 50 cm, as outlined in [Table 3](#). Comparable to the outcomes derived from the multispectral UAV sensor, the deep learning algorithm demonstrated its best BlScV classification accuracy when employed with the RGB UAV sensor featuring an image pixel size of 20 cm. This resulted in a kappa index of 0.41, an average accuracy of 70.00%, and an overall accuracy of 73.10%, ([Table 3](#)). Furthermore, the RGB UAV sensor with a 20 cm image pixel size yielded the highest F1 score, boasting values of 0.62 and 0.79 for distinguishing infected and healthy blueberry plants, respectively.

The training and validation accuracy and loss curves in [Fig. 9](#) illustrate the performance of the developed deep learning architecture over 40 epochs, using UAV images with a pixel size of 20 cm. Graphs (a) and (b) show the accuracy for Altum multispectral and P1 RGB images, respectively, while (c) and (d) depict the corresponding loss curves. In both cases, the model shows better performance on the training set compared to the validation set. The lack of correlation between training and validation data contributes to the observed lower accuracy in validation performance compared to the higher accuracy of the train data.

4.2. Performance analysis of proposed architecture over other deep learning models

The use of both RGB and multispectral UAV images highlights the significant superiority of the InceptionLSA compared to other models, including EfficientNet, EdgeNeXt, and DaViT. As shown in [Table 4](#), the developed architecture achieves an average accuracy of 70% on P1 RGB images, outperforming Efficient Net, EdgeNeXt, and DaViT by approximately 14% each. In detecting healthy plants, the model demonstrated superior performance, achieving an F1-score of 0.79, significantly higher than Efficient Net (0.62), DaViT (0.59), and EdgeNeXt (0.58). This underscores the model's advanced capabilities over the other deep learning approaches.

Table 3

Detection results of infected blueberry plants using P1 RGB image with different pixel sizes in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Pixel size (cm) | 2 | 5 | 10 | 20 | 30 | 50 |
|----------------------------|-------|-------|-------|--------------|-------|-------|
| F1-score (Infected Plants) | 0.51 | 0.54 | 0.56 | 0.62 | 0.55 | 0.08 |
| F1-score (Healthy Plants) | 0.73 | 0.73 | 0.75 | 0.79 | 0.73 | 0.76 |
| AA (× 100) | 61.80 | 63.44 | 65.15 | 70.00 | 64.21 | 50.67 |
| OA (× 100) | 65.02 | 65.73 | 67.93 | 73.10 | 66.01 | 62.42 |
| KI | 0.24 | 0.27 | 0.31 | 0.41 | 0.28 | 0.16 |

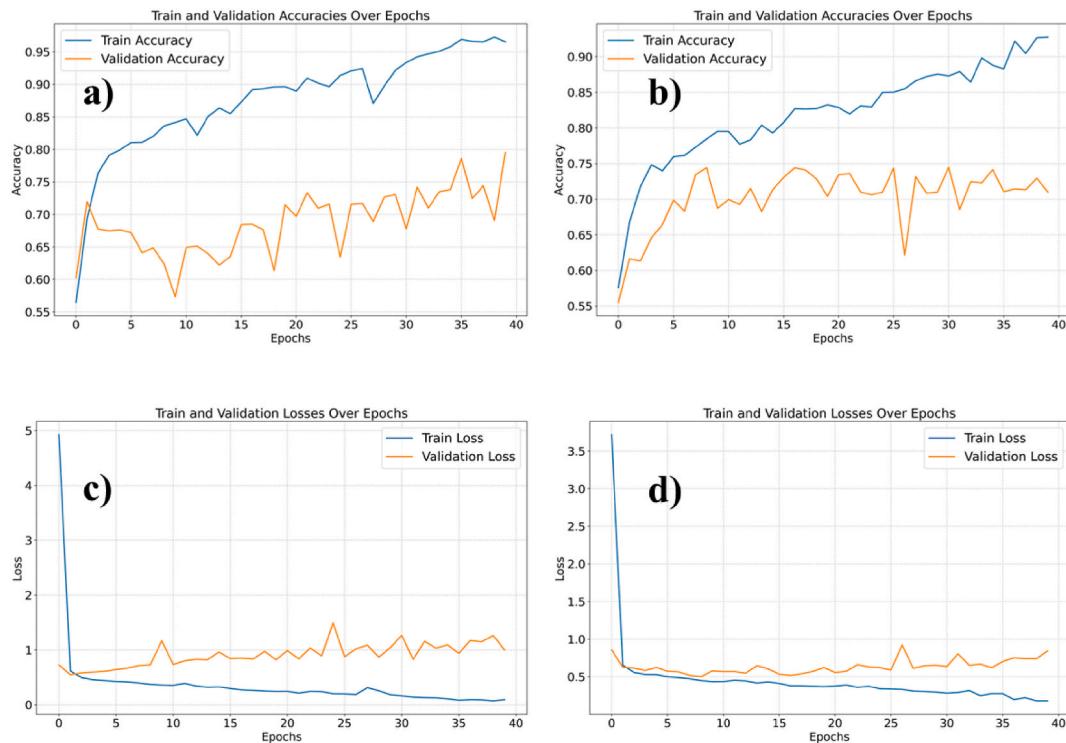


Fig. 9. The training and validation accuracy of the developed deep learning architecture over 40 epochs using 20 cm image pixel size of a) Altum multispectral b) P1 RGB images and train and validation loss of the model using c) Altum multispectral d) P1 RGB images (There is no correlation between train and validation data, so the accuracy of validation set is much lower than the train data set).

Table 4

Detection results of infected blueberry plants of developed deep learning models using P1 RGB image with 20 cm pixel size in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Model | Efficient Net (Tan and Le 2019) | EdgeNeXt (Maaz et al., 2023) | DaViT (Ding et al., 2022) | InceptionLSA (our model) |
|----------------------------|---------------------------------|------------------------------|---------------------------|--------------------------|
| F1-score (Infected Plants) | 0.48 | 0.51 | 0.51 | 0.62 |
| F1-score (Healthy Plants) | 0.62 | 0.58 | 0.59 | 0.79 |
| AA ($\times 100$) | 55.62 | 56.28 | 56.32 | 70.00 |
| OA ($\times 100$) | 56.16 | 54.92 | 55.38 | 73.10 |
| KI | 0.11 | 0.12 | 0.12 | 0.41 |

Table 5

Detection results of infected blueberry plants of developed deep learning models using Altum multispectral image with 20 cm pixel size in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Model | Efficient Net (Tan and Le 2019) | EdgeNeXt (Maaz et al., 2023) | DaViT (Ding et al., 2022) | InceptionLSA (our model) |
|----------------------------|---------------------------------|------------------------------|---------------------------|--------------------------|
| F1-score (Infected Plants) | 0.52 | 0.45 | 0.32 | 0.71 |
| F1-score (Healthy Plants) | 0.60 | 0.68 | 0.76 | 0.81 |
| AA ($\times 100$) | 57.98 | 56.46 | 56.04 | 76.33 |
| OA ($\times 100$) | 56.80 | 59.74 | 64.19 | 76.66 |
| KI | 0.15 | 0.13 | 0.14 | 0.52 |

Furthermore, when tested with Altum multispectral images (Table 5), the model achieved an average accuracy of 76.33%, surpassing Efficient Net by 18%, EdgeNeXt by 20%, and DaViT by 20%. The model also excelled in detecting infected plants, achieving an F1-score of 0.71, in contrast to Efficient Net (0.52), EdgeNeXt (0.45), and DaViT (0.32) (Table 5). These results demonstrate the model's superior classification ability, and its robust detection performance compared to other implemented deep learning models. The much lower classification accuracy of other deep learning models can be attributed to their need for a large amount of train data to reach their full potential in a classification task.

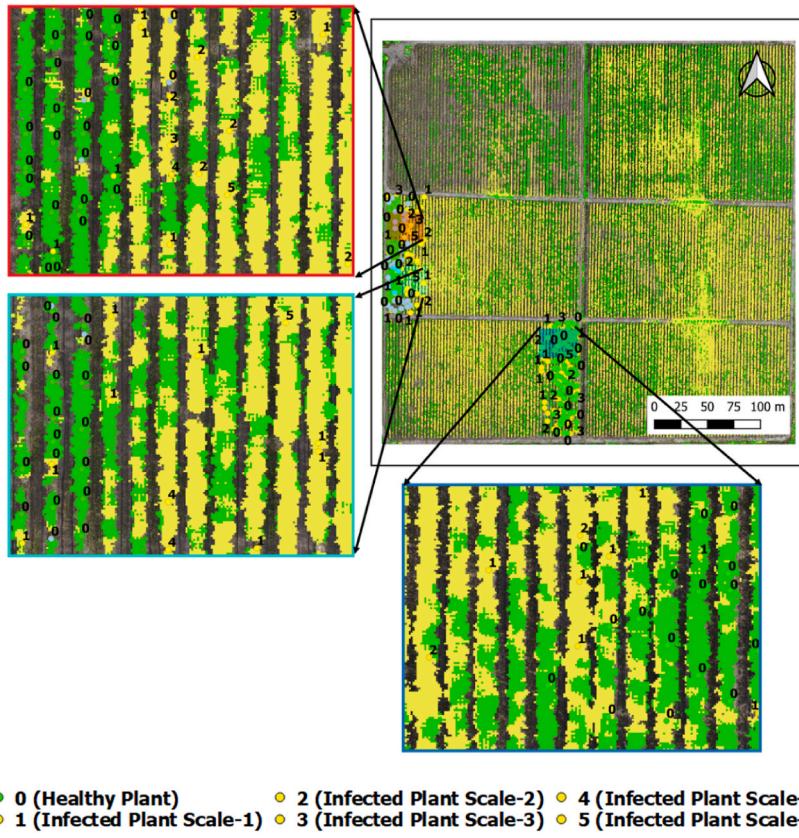


Fig. 10. Detection maps of infected blueberry plants produced by deep learning algorithm utilizing Altum multispectral image with a 20 cm pixel size. Two fields with surveyed plants are Field A (top left) and B (bottom right), respectively.

4.3. Classification maps

The BlScV classification maps generated by the developed deep learning algorithm with multispectral and RGB images for the image pixel size of 20 cm are depicted in Figs. 10 and 11, respectively. An intriguing observation emerges when comparing the classifications of infected blueberry plants between the two images. The model using the Altum multispectral UAV image (Fig. 10) exhibits a notably higher rate of detecting infected plants compared to the model utilizing the P1 RGB UAV imagery (Fig. 11). As discussed in the previous subsections regarding classification accuracies, the developed deep learning model using the multispectral UAV imagery resulted in slightly higher overall and per-class accuracy as compared to the RGB UAV imagery with an image pixel size of 20 cm. However, this heightened accuracy of multispectral UAV imagery is accompanied by a more assertive approach to detecting infected plants, making the model more proactive in identifying instances of infection. Field validation of the classification maps, together with the fact that the distribution of BlScV infection should be random (virus spread via aphids), indicate the infection map generated utilizing the RGB UAV imagery is more accurate than the multispectral UAV imagery, as seen in Fig. 11. The better representation of the blueberry scorch from the RGB UAV image can be attributed to various factors. One important factor is the enhanced generalization capability of the model developed with the RGB UAV data compared to multispectral UAV images. Essentially, although the overall accuracy of the deep learning model using RGB UAV imagery is slightly lower than that of multispectral data, it yields better classification maps. The better generalization ability may be because the model utilizing the RGB imagery has fewer inputting feature variables (than the multispectral), enabling the deep learning architecture to capture more generalized features or patterns. In contrast, the red edge band, NIR band, and NDVI with multispectral UAV imagery are sensitive to various vegetation stresses (e.g., water and nutrient shortages). Therefore, the deep learning model using multispectral UAV imagery detected blueberry scorch together with other vegetation stresses, which led to over-classifications of BlScV as shown in Fig. 10.

4.4. Ablation study

An ablation study comparing various model mechanisms for the detection of infected blueberry plants using P1 RGB and Altum multispectral images with a 20 cm pixel size is presented in Table 6. The results obtained exhibit the Kappa index (KI), average accuracy (AA), overall accuracy (OA), and F1-scores for the detection of infected and healthy plants. For P1 RGB images, the inclusion of both self-attention and CNN modules performed similarly to just using CNN model for infected plants (0.62 vs. 0.63), but it achieved a

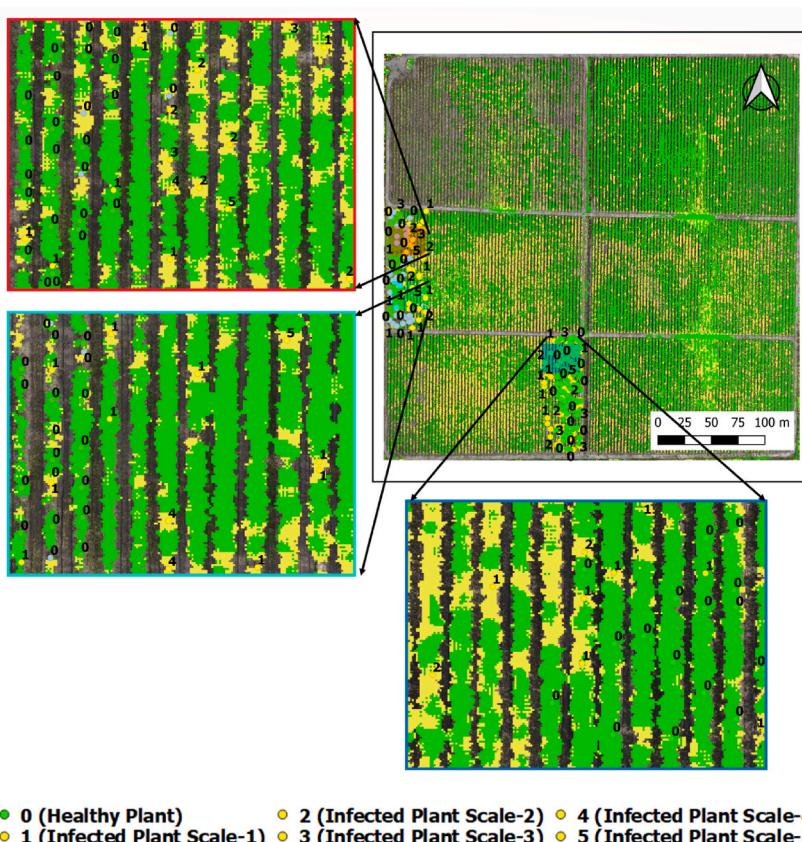


Fig. 11. Detection maps of infected blueberry plants produced by deep learning algorithm utilizing P1 RGB image with a 20 cm pixel size. Two fields with surveyed plants are Field A (top left) and B (bottom right) respectively.

Table 6

Detection results of infected blueberry plants of various mechanisms of the developed deep learning models using P1 RGB and *Altum* multispectral image with 20 cm pixel size in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Model | P1 RGB | | Altum Multispectral | |
|----------------------------|-------------|----------------------|---------------------|----------------------|
| | CNN | Self-attention + CNN | CNN | Self-attention + CNN |
| F1-score (Infected Plants) | 0.63 | 0.62 | 0.63 | 0.71 |
| F1-score (Healthy Plants) | 0.73 | 0.79 | 0.82 | 0.81 |
| AA ($\times 100$) | 69.45 | 70.00 | 71.22 | 76.33 |
| OA ($\times 100$) | 69.22 | 73.10 | 75.29 | 76.66 |
| KI | 0.37 | 0.41 | 0.45 | 0.52 |

Table 7

Detection results of infected blueberry plants using the *Altum* multispectral UAV image with different pixel sizes in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Resolution (cm) | 10 | 20 | 30 | 50 |
|---------------------------|-------|--------------|-------------|-------|
| F1-score (L1-2) | 0.42 | 0.47 | 0.46 | 0.42 |
| F1-score (L3-5) | 0.11 | 0.30 | 0.30 | 0.05 |
| F1-score (Healthy Plants) | 0.75 | 0.83 | 0.57 | 0.75 |
| | | | | |
| OA ($\times 100$) | 60.96 | 69.23 | 50.50 | 61.29 |
| AA ($\times 100$) | 43.18 | 51.98 | 15.33 | 41.82 |
| KI | 0.23 | 0.38 | 0.17 | 0.21 |

higher F1-score for healthy plants (0.79) compared to the CNN module (0.73). Compared to the CNN module, which produced an AA of 69.45% and an OA of 69.22%, the use of both self-attention and CNN modules was better overall, with an AA of 70.00% and an OA of 73.10%. For Altum multispectral images, the use of both self-attention and CNN modules accomplished more effectively than the CNN module alone, detecting infected plants with an F1-score of 0.71 (vs. 0.63), but slightly worse for healthy plants (0.81 vs. 0.82). Together with higher OA (76.66%) and AA (76.33%), the inclusion of self-attention with the CNN module additionally featured a higher Kappa index (0.52), indicating stronger model performance across evaluation metrics.

4.5. Distinguishing low and high BlScV disease severity scales

To assess the developed model's ability to identify and distinguish various disease severity scales, we categorized infected plants into two groups. The first group consisted of the infected blueberry plants with rating scales 1 and 2, while the second group consisted of the infected plants with rating scales 3 to 5. Results, depicted in [Tables 7–8](#), showcased the highest accuracy achieved when utilizing a 20 cm pixel size for both multispectral and RGB UAV imagery. Specifically, for the multispectral image, the deep learning architecture achieved a KI, AA, and OA of 0.38, 51.98%, and 69.23%, respectively. Meanwhile, with RGB imagery at a 20 cm pixel size, the model yielded a KI, AA, and OA of 0.27, 45.30%, and 64.77%, respectively.

4.6. Effect of training ratio

In this study, various proportions of training instances, 5%, 10%, 30%, 50%, and 90%, were employed to train the proposed model, with the remaining samples allocated for validation to assess the model's efficacy. It should be mentioned that deep learning models require a significant amount of training data to attain their best classification capability. However, the acquisition of training/ground truth data in the field is time-consuming and costly. Thus, we need to develop more data-efficient deep learning models. Therefore, testing the capability of the model using various data ratios can help us showcase the data efficiency of the developed architecture. The quantity of training samples emerges as a pivotal factor influencing the classification performance of an algorithm. The findings consistently reveal an enhanced classification accuracy for the developed model as the training ratios increase from 5% to 100% of training data (the ratio presents the percentage of utilized data out of all available training data), applied to both RGB and multispectral UAV sensors. Specifically, utilizing the multispectral UAV system led to a notable enhancement in the classification accuracy of the model—approximately 11%, 13%, and 37% for overall accuracy, average accuracy, and kappa index, respectively, across training ratios from 5 to 100%. Similarly, employing the RGB UAV system showcased a substantial improvement in classification accuracy, roughly 15%, 17%, and 54% for average accuracy, overall accuracy, and kappa index, respectively, as depicted in [Fig. 12](#). Overall, the developed model achieved high classification accuracies, even with a low number of training data, which highlights the data efficiency of the developed deep learning architectures. This result showcases the capability of the specially designed deep learning model capable of producing acceptable classification accuracy with a limited number of ground truth data.

4.7. Effect of input image patch size

To classify high-resolution RGB and Multispectral UAV imagery, spatial information is extremely important. The efficacy of a deep learning model can be strongly influenced by the spatial dimension of the input image patch. To get the most consistent results, the effectiveness of a model must be compared across a range of image patch sizes. Using various image patch sizes, the developed model's performance is assessed. Results indicate that the deep learning model's classification performance improved for both RGB and multispectral UAVs when the image input patch size increased from 5 to 9. However, a noticeable decline in classification accuracy occurred when the patch size further increased to 13×13 for both UAV sensors. Specifically, with the multispectral UAV sensor, the developed model showed enhancements of approximately 12% in OA, 14% in AA, and 36% in the KI when transitioning from a 5×5 to a 9×9 image patch size. Conversely, increasing the image patch size from 9 to 11 resulted in a decrease of about 4% in overall accuracy, 7% in AA, and 20% in the kappa index. For the RGB UAV sensor, the classification accuracy of the model increased by around 9%, 10%, and 33% for AA, OA, and KI, respectively, with the shift from a 5×5 to a 9×9 image patch size (see [Fig. 13](#)). However, a decrease of approximately 5%, 6%, and 22% in these metrics occurred when the image patch size increased from 9 to 11. For effectively extracting this vital information, image pixel size and input image patch size are significantly important. As experiments indicated, smaller feature sizes may not reveal global interactions, while much bigger object sizes may include irrelevant information leading to a decrease in the classification accuracy of deep learning models.

5. Discussion

Blueberry scorch, caused by the Blueberry scorch virus (BlScV), is a devastating disease that can lead to significant yield loss, threatening the sustainability of highbush blueberry fields in North America and beyond. Effective management relies on early detection, removal of infected plants, and controlling the spread via aphids. However, traditional visual assessment of symptoms is labor-intensive and inefficient. This study explored the use of UAV-based imaging as a modern, remote sensing tool to detect BlScV-infected plants. UAVs offer flexibility in image acquisition, with advantages like different pixel sizes, short revisit times, and the ability to carry various sensors. This research compared the performance of RGB and multispectral sensors in detecting BlScV using advanced deep learning models, aiming to improve disease detection and mapping accuracy and support real-world disease management.

The results obtained in this research illustrated the higher classification accuracy of the multispectral UAV images over the RGB

Table 8

Detection results of infected blueberry plants using the P1 RGB UAV image with different pixel sizes in terms of F1-score (OA = overall accuracy, AA = average accuracy, and KI= Kappa index).

| Resolution (cm) | 10 | 20 | 30 | 50 |
|---------------------------|-------|--------------|-------|-------|
| F1-score (L1-2) | 0.44 | 0.43 | 0.50 | 0.27 |
| F1-score (L3-5) | 0.19 | 0.17 | 0.00 | 0.00 |
| F1-score (Healthy Plants) | 0.73 | 0.78 | 0.66 | 0.74 |
| OA ($\times 100$) | 60.34 | 64.77 | 56.18 | 60.41 |
| AA ($\times 100$) | 44.70 | 45.30 | 42.73 | 36.20 |
| KI | 0.22 | 0.27 | 0.22 | 0.80 |

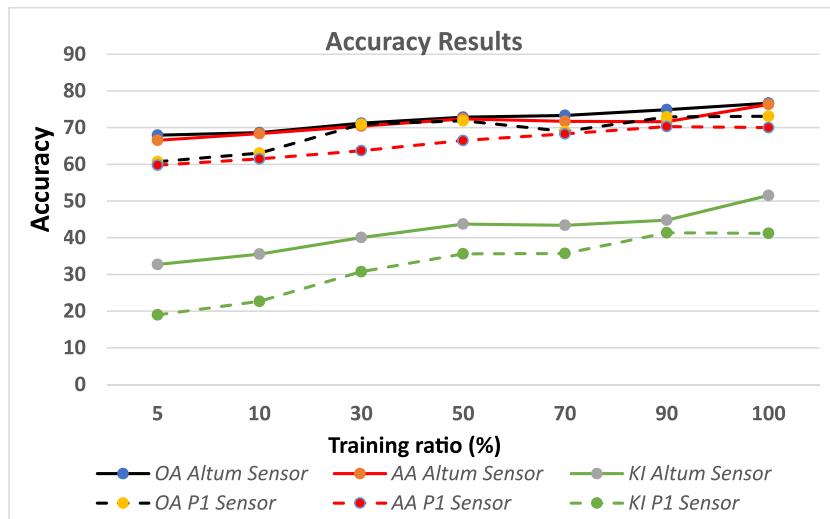


Fig. 12. The classification accuracies of the RGB and Multispectral UAV imaging systems in terms of average accuracy, overall accuracy, and kappa index obtained by the developed deep learning model with different percentages of training samples.

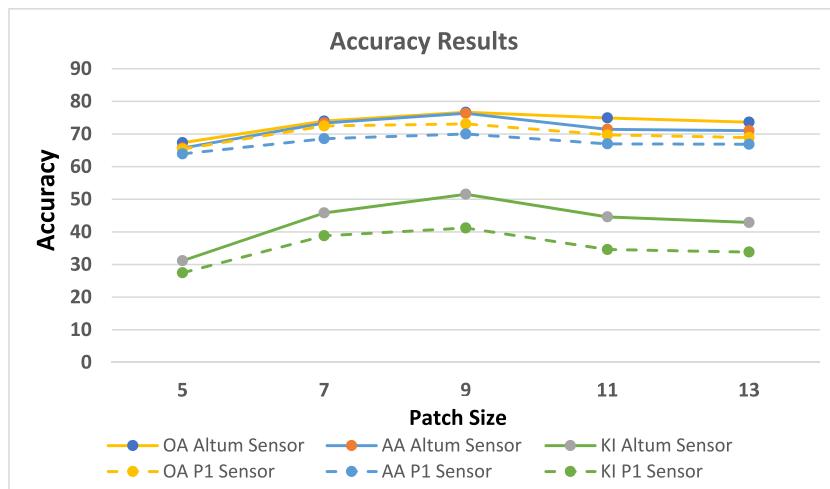


Fig. 13. The classification accuracies of the RGB and Multispectral UAV imaging systems in terms of average accuracy, overall accuracy, and kappa index obtained by the developed deep learning model with different input image patch sizes.

UAV images. This can be ascribed to the additional spectral bands, specifically the Red Edge and Near Infra-Red (NIR), incorporated in the Altum sensor. These supplementary bands enhance spectral information, causing the Altum sensor to be more sensitive to variations in vegetation stress and health. Furthermore, the deep learning model applied to the Altum multispectral images leverages the

Normalized Vegetation Difference Index (NDVI), a widely utilized spectral index for crop health assessment. The addition of these additional spectral bands and the NDVI in the classification model led to superior classification accuracy (AA 76.33%) of Altum multispectral images with 20 cm pixel sizes compared to the model with P1 RGB images with the same pixel sizes (AA 70.00%). Such incorporation of spectral indices for the assessment and mapping of crop stress and health is widely used (Stanton et al., 2017; L. Zhang et al., 2019). The NDVI and other structure- and chlorophyll-sensitive indices calculated using visible, red-edge, and near-infrared bands are commonly utilized for crop health mapping (L. Zhang et al., 2019). Therefore, crop management can significantly benefit from the rapid mapping of these spectral indices (Jiang et al., 2021).

Previous studies have also attempted to compare the performance of RGB and multispectral UAV imagery for different classification or regression tasks. For example, Marcial-Pablo et al. (2019) used a UAV equipped with both RGB and RGB + NIR sensors to acquire high-resolution imagery over a maize field throughout the growing and senescence stages to estimate the vegetation cover. They concluded that the accuracies of multispectral UAV images with 2.1 cm pixel size were superior to those of the RGB images which had a pixel size of 1.25 cm. Moreover, Fuentes-Peailillo et al. (2018) compared NDVI calculated with UAV multispectral images to four VIs calculated with RGB images for identifying soil and vegetation cover. The study findings suggested that RGB indices can be used to recognize the same spatial patterns compared to multispectral images. Even though the pixel count (for soil or vegetation cover) yielded comparable results between RGB and multispectral imagery, a visual examination of the data revealed that the RGB indices had higher misidentified vegetation rate.

Regarding the optimal UAV image pixel size or flight altitude for crop health/disease assessment, there has also been a diversity of findings. For instance, Borra-Serrano et al. (2015) reported that for identifying infestation of weeds in sunflower fields with UAV flights at altitudes ranging from 30 to 100 m, the best results were obtained for the 60 m UAV flight altitude. Their results also illustrated that higher flight altitudes (e.g., higher than 100 m) will result in low-quality UAV images unsuitable for weed detection in sunflower fields. Similarly, Zhao et al. (2023) compared images with pixel sizes between 3 cm and 10 m for the identification of cotton root rot and concluded that 4 m proved to be the ideal resolution. In our research, it was found that a 20 cm image pixel size is ideal for blueberry scorch virus mapping with accuracy higher than those using 2, 5, 10, 30, and 50 cm image pixel size. Overall, results in this study highlight the importance of selecting optimal UAV image pixel size and UAV sensor type (e.g., RGB vs multispectral) for different classification or regression tasks in the remote sensing field as the optimal selection may differ across various applications.

The developed deep learning algorithm demonstrated optimal performances for the identification of infected blueberry plants with an image pixel size of 20 cm. Our choice of a 9×9 image patch size in the developed model led to a significant increase in classification accuracy. Therefore, the optimal object size for accurate BlScV mapping was determined to be 1.8×1.8 m. Within this dimension, the model successfully extracted crucial data relationships and image characteristics from infected and healthy blueberry plants across close-, mid-, and long-range data relationships. Smaller feature sizes might not reveal general relationships, according to experiments, while larger object sizes might contain unnecessary information that lowers deep learning models' classification accuracy.

6. Conclusion

In this study, we assessed the use and applicability of UAV imaging systems for effective blueberry scorch detection and mapping, aimed to support real-world virus control and plant management. RGB and multispectral UAV images with different pixel sizes were acquired and compared to evaluate the impacts of image spectral and spatial resolution on detection accuracy. We developed an advanced deep learning architecture that leverages cutting-edge CNN and vision transformers mechanisms to enhance the precision of blueberry scorch virus mapping. Different training ratios and image patch sizes were tested in the model to further improve its performance. The obtained results show that the BlScV-infected plants can be identified with sufficient accuracy, which demonstrates the effectiveness and applicability of developed high-resolution UAV-based techniques for BlScV detecting and mapping. The optimal setting for accurate BlScV mapping was found to be a UAV image pixel size of 20 cm with an input image patch size of 9 by 9 pixels. The average accuracy reached 76.33 and 70.00% using the multispectral and RGB UAV images with this 20 cm pixel size, respectively. Using image pixel sizes lower or higher than 20 cm decreased accuracy. Overall, the results achieved in this study indicate that mapping BlScV infection using high-resolution UAV images and advanced deep learning models is a powerful tool for monitoring the distribution of BlScV infection in large blueberry fields. Future research can explore the potential integration and effectiveness of IoT (e.g., for detecting aphids that spread BlScV), artificial intelligence, and UAV imaging for more comprehensive BlScV detection and monitoring to facilitate more timely and effective virus control.

CRediT authorship contribution statement

Ali Jamali: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Bing Lu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Rishi R. Burlakoti:** Writing – review & editing, Investigation. **Siva Sabaratnam:** Writing – review & editing, Investigation. **Margaret Schmidt:** Writing – review & editing. **Carolyn Teasdale:** Writing – review & editing. **Eric M. Gerbrandt:** Writing – review & editing. **Lilian Yang:** Writing – review & editing. **Jonathon McIntyre:** Writing – review & editing, Resources, Data curation. **David McCaffrey:** Writing – review & editing.

Ethical Statement for Solid State Ionics

Hereby, I/Ali Jamali/consciously assure that for the manuscript/High-resolution mapping of Blueberry scorch virus incidence

using RGB and multispectral UAV images and deep learning/the following is fulfilled.

- 1) This material is the authors' own original work, which has not been previously published elsewhere.
 - 2) The paper is not currently being considered for publication elsewhere.
 - 3) The paper reflects the authors' own research and analysis in a truthful and complete manner.
 - 4) The paper properly credits the meaningful contributions of co-authors and co-researchers.
 - 5) The results are appropriately placed in the context of prior and existing research.
 - 6) All sources used are properly disclosed (correct citation). Literally copying of text must be indicated as such by using quotation marks and giving proper reference.
 - 7) All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.
- The violation of the Ethical Statement rules may result in severe consequences.

To verify originality, your article may be checked by the originality detection software iThenticate. See also <http://www.elsevier.com/editors/plagdetect>.

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.

Acknowledgement

This research was funded by Canada's Digital Technology Supercluster, Mitacs (IT27380), i-Open Technologies Inc., Terramera Inc., and the Natural Sciences and Engineering Research Council of Canada, Discovery Grant [RGPIN-2022-03679] to Dr. Bing Lu. Thanks to the owners of farm fields for permitting us to collect data there. We also thank a group of research assistants who helped with fieldwork.

Data availability

The data that has been used is confidential.

References

- Alexander, Kolesnikov, Alexey, Dosovitskiy, Dirk, Weissenborn, Georg, Heigold, Jakob, Uszkoreit, Lucas, Beyer, Matthias, Minderer, et al., 2021. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. Computer Vision and Pattern Recognition. <https://doi.org/10.48550/arXiv.2010.11929>.
- Audebert, Nicolas, Bertrand, Le Saux, Sébastien, Lefevre, 2019. Deep learning for classification of hyperspectral data: a comparative review. IEEE Geoscience and Remote Sensing Magazine 7 (2), 159–173. <https://doi.org/10.1109/MGRS.2019.2912563>.
- Barreto, Abel, Paulus, Stefan, Varrelmann, Mark, Mahlein, Anne-Katrin, 2020. Hyperspectral imaging of symptoms induced by rhizoctonia solani in sugar beet: comparison of input data and different machine learning algorithms. J. Plant Dis. Prot. 127 (4), 441–451. <https://doi.org/10.1007/s41348-020-00344-8>.
- Borra-Serrano, Irene, Peña, José M., Torres-Sánchez, Jorge, Mesas-Carrascosa, Francisco J., López-Granados, Francisca, 2015. Spatial Quality Evaluation of Resampled Unmanned Aerial Vehicle-Imagery for Weed Mapping. Sensors 15 (8), 19688–19708. <https://doi.org/10.3390/s150819688>.
- Bouguettaya, Abdelmalek, Zarzour, Hafed, Ahmed, Kechida, Taberkit, Amine Mohammed, 2023. A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images. Cluster Comput. 26 (2), 1297–1317. <https://doi.org/10.1007/s10586-022-03627-x>.
- de Castro, Ana I., Ehsani, Reza, Plotz, Randy C., Crane, Jonathan H., Buchanon, Sherrie, 2015. Detection of laurel wilt disease in avocado using low altitude aerial imaging. PLoS One 10 (4), e0124642. <https://doi.org/10.1371/journal.pone.0124642>. Public Library of Science.
- Chen, Fengnong, Zhang, Yao, Zhang, Jingcheng, Liu, Lianmeng, Wu, Kaihua, 2022. Rice false smut detection and prescription map generation in a complex planting environment, with mixed methods, based on near earth remote sensing. Rem. Sens. 14 (4). <https://doi.org/10.3390/rs14040945>.
- Ding, Mingyu, Xiao, Bin, Codella, Noel, Luo, Ping, Wang, Jingdong, Lu, Yuan, 2022. DaViT: dual attention vision transformers. In: Avidan, Shai, Brostow, Gabriel, Cissé, Moustapha, Farinella, Giovanni Maria, Hassner, Tal (Eds.), Computer Vision – ECCV 2022. Springer Nature Switzerland, Cham, pp. 74–92.
- Fuentes-Pealillido, F., Ortega-Farias, S., Rivera, M., Bardeen, M., Moreno, M., 2018. Comparison of vegetation indices acquired from RGB and multispectral sensors placed on UAV. In: 2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA), pp. 1–6. <https://doi.org/10.1109/ICA-ACCA.2018.8609861>.
- Guo, Meng-Hao, Cai, Jun-Xiong, Liu, Zheng-Ning, Tai-Jiang, Mu, Martin, Ralph R., Hu, Shi-Min, 2021. PCT: point cloud transformer. Computational Visual Media 7 (2), 187–199. <https://doi.org/10.1007/s41095-021-0229-5>.
- Guo, M.-H., Liu, Z.-N., Mu, T.-J., Hu, S.-M., 2023. Beyond self-attention: external attention using two linear layers for visual tasks. IEEE Trans. Pattern Anal. Mach. Intell. 45 (5), 5436–5447. <https://doi.org/10.1109/TPAMI.2022.3211006>.
- Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., Chanussot, J., 2021. SpectralFormer: rethinking hyperspectral image classification with transformers. IEEE Trans. Geosci. Rem. Sens. <https://doi.org/10.1109/TGRS.2021.3130716>, 1–1.
- Jamali, A., Mahdianpari, M., Mohammadimanesh, F., Homayouni, S., 2022. A deep learning framework based on generative adversarial networks and vision transformer for complex wetland classification using limited training samples. Int. J. Appl. Earth Obs. Geoinf. 115 (December), 103095. <https://doi.org/10.1016/j.jag.2022.103095>.
- Jamali, A., Roy, S.K., Bhattacharya, A., Ghamisi, P., 2023. Local window attention transformer for polarimetric SAR image classification. Geosci. Rem. Sens. Lett. IEEE 20, 1–5. <https://doi.org/10.1109/LGRS.2023.3239263>.
- James, J. Polashock, Lawrence Caruso, Frank, Caruso, Lawrence, Schilder, Annemiek C., 2017. Compendium of Blueberry, Cranberry, and Lingonberry Diseases and Pests, Second. American Phytopathological Society (APS Press, St. Paul, USA <https://www.cabdirect.org/cabdirect/abstract/20173305725>.
- Jiang, Rui, Sanchez-Azofeifa, Arturo, Laakso, Kati, Wang, Pei, Xu, Yan, Zhou, Zhiyan, Luo, Xiwen, Lan, Yubin, Zhao, Genping, Chen, Xin, 2021. UAV-based partially sampling system for rapid NDVI mapping in the evaluation of rice nitrogen use efficiency. J. Clean. Prod. 289 (March), 125705. <https://doi.org/10.1016/j.jclepro.2020.125705>.

- Kaur, Ravpreet, Singh, Sarbjeet, 2023. A comprehensive review of object detection with deep learning. *Digit. Signal Process.* 132 (January), 103812. <https://doi.org/10.1016/j.dsp.2022.103812>.
- Kersten, Thomas, Wolf, Joshua, Lindstaedt, Maren, 2022. Investigations into the accuracy of the uav system dji Matrice 300 rtk with the sensors zenuimuse P1 and L1 in the hamburg test field. In: XXIV ISPRS Congress "Imaging Today, Foreseeing Tomorrow", 6–11 June 2022. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Nice, France, pp. 339–346. <https://repos.hcu-hamburg.de/handle/hcu/877>.
- Lee, Sue Han, Goéau, Hervé, Bonnet, Pierre, Joly, Alexis, 2020. New perspectives on plant disease characterization based on deep learning. *Comput. Electron. Agric.* 170 (March), 105220. <https://doi.org/10.1016/j.compag.2020.105220>.
- Lee, Hae Min, Song, Eun Gyeong, Ryu, Ki Hyun, 2023. Development of a multiplex PCR for simultaneous detection of blueberry red ringspot virus and blueberry scorch virus including an internal control. *Res. Plant Dis* 29 (1), 94–99. <https://doi.org/10.5423/RPD.2023.29.1.94>. The Korean Society of Plant Pathology.
- Maaaz, Muhammad, Shaker, Abdelrahman, Cholakkal, Hisham, Khan, Salman, Zamir, Syed Waqas, Anwer, Rao Muhammad, Khan, Fahad Shahbaz, 2023. EdgeNeXt: efficiently amalgamated CNN-transformer architecture for mobile vision applications. In: Karlinsky, Leonid, Michaeli, Michael, Tomer, Nishino, Ko (Eds.), Computer Vision – ECCV 2022 Workshops. Springer Nature Switzerland, Cham, pp. 3–20.
- Marcial-Pablo, Mariana de Jesús, Gonzalez-Sanchez, Alberto, Jimenez-Jimenez, Sergio Iván, Ontiveros-Capurata, Ronald Ernesto, Ojeda-Bustamante, Waldo, 2019. Estimation of vegetation fraction using RGB and multispectral images from UAV. *Int. J. Rem. Sens.* 40 (2), 420–438. <https://doi.org/10.1080/01431161.2018.1528017>. Taylor & Francis.
- Martin, Robert R., Tzanetakis, Ioannis E., 2018. High risk blueberry viruses by region in North America; implications for certification, nurseries, and fruit production. *Viruses* 10 (7). <https://doi.org/10.3390/v10070342>.
- Mothapo, Mologadi Clodean, Dube, Timothy, Abdel-Rahman, Elfatih, Sibanda, Mbulisi, 2022. Progress in the use of geospatial and remote sensing technologies in the assessment and monitoring of tomato crop diseases. *Geocarto Int.* 37 (16), 4784–4804. <https://doi.org/10.1080/10106049.2021.1899303>. Taylor & Francis.
- Sabarathnam, Siva, 2021. *Blueberry scorch*. British Columbia: ministry of agriculture, food and fisheries. <https://www2.gov.bc.ca/assets/gov/farming-natural-resources-and-industry/agriculture-and-seafood/animal-and-crops/plant-health/phu-blueberryscorchviruss.pdf>.
- Shafi, U., Mumtaz, R., Iqbal, N., Zaidi, S.M.H., Zaidi, S.A.R., Hussain, I., Mahmood, Z., 2020. A multi-modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning. *IEEE Access* 8, 112708–112724. <https://doi.org/10.1109/ACCESS.2020.3002948>.
- Sharma, S., Guleria, K., 2022. Deep learning models for image classification: comparison and applications. In: 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 1733–1738. <https://doi.org/10.1109/ICACITE53722.2022.9823516>.
- Stanton, Carly, Starek, Michael J., Elliott, Norman, Brewer, Michael, Maeda, Murilo M., Chu, Tianxing, 2017. Unmanned aircraft system-derived crop height and normalized difference vegetation index metrics for sorghum yield and aphid stress assessment. *J. Appl. Remote Sens.* 11 (2), 026035. <https://doi.org/10.11117/JRS.11.026035>.
- Sweeney, M.E., Wegener, L.A., Ring, S., Raworth, D.A., 2009. Management of blueberry scorch virus on highbush blueberry in BRITISH columbia. In: *Acta Horticulturae*. International Society for Horticultural Science (ISHS), Leuven, Belgium, pp. 313–318. <https://doi.org/10.17660/ActaHortic.2009.810.40>.
- Tan, Mingxing, Quoc, Le, 2019. EfficientNet: rethinking model scaling for convolutional neural networks. In: Chaudhuri, Kamalika, Salakhutdinov, Ruslan (Eds.), Proceedings of the 36th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 97. PMLR, pp. 6105–6114. In: <https://proceedings.mlr.press/v97/tan19a.html>.
- Wegener, L.A., Punja, Z.K., Martin, R.R., Bernardy, M.G., MacDonald, L., 2006. Epidemiology and identification of strains of blueberry scorch virus on highbush blueberry in British Columbia, Canada. *J. Indian Dent. Assoc.* 28 (2), 250–262. <https://doi.org/10.1080/07060660609507294>. Taylor & Francis.
- Wu, Harvey, Wiesner-Hanks, Ty, Stewart, Ethan L., DeChant, Chad, Nicholas, Kaczmar, Gore, Michael A., Nelson, Rebecca J., Lipsom, Hod, 2019. Autonomous detection of plant disease symptoms directly from aerial imagery. *The Plant Phenome Journal* 2 (1), 190006. <https://doi.org/10.2135/tpj2019.03.0006>. John Wiley & Sons, Ltd.
- Xie, L.X., Zheng, S., Zhang, L.J., Zhang, X.Y., Li, T., 2018. Blueberry scorch virus detected on blueberry plants imported into China. *Plant Dis.* 102 (8), 1673. <https://doi.org/10.1094/PDIS-12-17-1867-PDN>. Scientific Societies.
- Yan, H., Zhang, E., Wang, J., Leng, C., Basu, A., Peng, J., 2023. Hybrid conv-ViT network for hyperspectral image classification. *Geosci. Rem. Sens. Lett. IEEE* 20, 1–5. <https://doi.org/10.1109/LGRS.2023.3287277>.
- Zhang, Liyuan, Zhang, Huihui, Niu, Yaxiao, Han, Wenting, 2019. Mapping maize water stress based on UAV multispectral remote sensing. *Rem. Sens.* 11 (6). <https://doi.org/10.3390/rs11060605>.
- Zhao, Hengqian, Yang, Yifeng, Yang, Chenghai, Song, Rui, Guo, Wei, 2023. Evaluation of spatial resolution on crop disease detection based on multiscale images and category variance ratio. *Comput. Electron. Agric.* 207, 107743. <https://doi.org/10.1016/j.compag.2023.107743>.
- Zheng, Hengbiao, Cheng, Tao, Dong, Li, Xiang, Zhou, Yao, Xia, Tian, Yongchao, Cao, Weixing, Zhu, Yan, 2018. Evaluation of RGB, color-infrared and multispectral images acquired from unmanned aerial systems for the estimation of nitrogen accumulation in rice. *Rem. Sens.* 10 (6). <https://doi.org/10.3390/rs10060824>.